Digital watermarking for audio content protection

Lei, Baiying

2012

Lei, B. (2012). Digital watermarking for audio content protection. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/50758>

<https://doi.org/10.32657/10356/50758>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

**DIGITAL
WATERMARKING FOR
AUDIO CONTENT
PROTECTION**

LEI BAIYING

School of Electrical and Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2012

Acknowledgements

First and foremost, I would like to take this opportunity to express my sincerest gratitude and appreciation to my supervisor, Associate Professor Soon Ing Yann, for his valuable advice, patient guidance, supervision and support which made this study possible. I greatly appreciate his patience and encouragement that carry me through difficult times and for his insights and suggestions that help shape my research skills. His research intuition and experience guided me through the four years of PhD study. I am extremely grateful to him for his numerous as well as insightful suggestions to my thesis work.

I gratefully acknowledge the Nanyang Technological University for its financial support during the entire period of my candidature. I would also like to thank all the members and staff of the Media Technology Laboratory for providing me with a comfortable and pleasant working environment for doing my research.

Last, but not least, I must thank my family and my friends for their love and care over all these years. A PhD does not end within offices and laboratories. It is a life experience, and it has been the most significant one in my life so far.

Table of Contents

Acknowledgements.....	i
Table of Contents	ii
Summary	vii
List of Figures	x
List of Tables	xiii
List of Abbreviations.....	vi
Chapter 1.....	1
Introduction	1
1.1 Motivation.....	1
1.2 Scopes	7
1.3 Major Contributions of the Thesis	9
1.4 Organization of the Thesis	11
Chapter 2.....	12
Research Background	12
2.1 General Watermarking Scheme	12
2.2 Properties of Watermarking Systems	15
2.2.1 Imperceptibility	15
2.2.2 Robustness.....	16
2.2.3 Security.....	17
2.2.4 Capacity.....	17
2.2.5 Complexity	19
2.3 Classification of Watermarking Techniques.....	19

2.3.1 Blind, Semi-blind and Non-blind	20
2.3.2 Audible and Inaudiable.....	21
2.3.3 Robust, Semi-fragile, Fragile	21
2.4 Watermarking Applications	22
2.4.1 Watermarking for Copyright Protection.....	22
2.4.2 Watermarking for Authentication.....	23
2.4.3 Watermarking for Copy Protection or Usage control (Copy Control).....	24
2.4.4 Fingerprinting for Transaction Tracing	24
2.4.5 Owner Identification.....	25
2.5 Watermarking Methods	25
2.5.1 Time Domain Watermarking.....	25
2.5.2 Transform Domain Watermarking	27
2.5.3 Compressed Domain Watermarking	28
2.6 Watermarking Attacks and Evaluations	28
2.7 Chaotic Sequences in the Watermarking Field.....	30
2.8 SVD Principles and Properties	37
2.9 Literature Review	40
Chapter 3.....	46
A Perception-based Audio Watermarking Scheme in the Compressed	
Bitstream.....	46
3.1 Introduction.....	46
3.2 Related Work	48
3.3 MPEG-4 SLS Structure	51
3.4 Psychoacoustic Model and Significant State	55

3.4.1 Psychoacoustics	55
3.4.2 Significant State	57
3.5 Chaotic Watermarking	59
3.6 Watermark Embedding	60
3.7 Watermark Extraction	65
3.8 Performance Analysis	66
3.8.1 Error Analysis	66
3.8.2 Security Analysis	69
3.9 Experimental Tests and Results	72
3.9.1 Time and Spectrum Waves	72
3.9.2 Watermark Robustness	74
3.9.3 Objective Evaluation Test	75
3.9.4 Subjective Listening Test	76
3.9.5 Watermark Payload	77
3.10 Chapter Summary	77
Chapter 4.....	79
Blind and Robust Audio Watermarking Scheme Based on SVD-DCT	79
4.1 Introduction.....	79
4.2 Related Work	81
4.3 Watermark Embedding	83
4.3.1 Synchronization Code	83
4.3.2 Embedding Method	85
4.4 Watermark Extraction.....	91
4.5 Error Analysis	92

4.6 Experimental Results and Discussions	94
4.6.1 Perceptual Audio Quality Assessment	95
4.6.2 Robustness	102
4.6.3 Robustness to Stirmark Attacks	106
4.6.4 Embedding Capacity	109
4.6.5 Discussions	110
4.7 Chapter Summary	111
Chapter 5.....	113
A Robust Audio Watermarking Scheme Based on LWT and SVD	113
5.1 Introduction.....	113
5.2 Wavelet Lifting.....	116
5.3 Embedding Method.....	117
5.3.1 Watermark Preprocessing.....	117
5.3.2 Synchronization Code	118
5.3.3 Watermark Embedding.....	122
5.4 Watermark Extraction.....	125
5.5 Experimental Results	126
5.5.1 Imperceptibility Test	126
5.5.2 Subjective Listening Test	128
5.5.3 Robustness to Common Signal Processing Attacks	130
5.5.4 Robustness to Stirmark Attacks.....	132
5.5.5 Security Analysis.....	134
5.5.6 Algorithm Comparisons and Discussions	136
5.6 Chapter Summary	140

Chapter 6.....	141
An Optimal Audio Watermarking Scheme Based on Particle Swarm	
Optimization	141
6.1 Introduction.....	141
6.2 Methodology	144
6.2.1 PSO	144
6.2.2 Watermark Preprocessing.....	147
6.2.3 Watermark Embedding.....	148
6.2.4 Watermark Extraction	149
6.2.5 Performance Improvement Using PSO	150
6.2.6 False Positive Error Analysis	155
6.3 Experimental Results	156
6.3.1 Results of PSO Training.....	157
6.3.2 Imperceptibility	162
6.3.3 Robustness.....	164
6.3.4 Algorithm Comparison.....	168
6.4 Chapter Summary	168
Chapter 7.....	170
Conclusions and Recommendations	170
7.1 Conclusions.....	170
7.2 Recommendations for Further Research.....	172
List of Author's Publications	174
References	177

Summary

Rapid development of the digital technology and internet has made copyright protection and ownership verification of digital audio data hot research topics that have attracted a lot of interest from different parties. Since the traditional audio protection algorithms cannot solve all the problems of audio copyright protection and digital right management (DRM) due to the unique audio data features and algorithm limitations, this thesis is proposed to address the issue of audio copyright protection using watermarking techniques.

Firstly, a new perception-based watermarking scheme for MPEG-4 scalable to lossless (SLS) audio is proposed. Under the control of the psychoacoustic model, the significant part of integer modified discrete cosine transform (IntMDCT) coefficients are adaptively modified during MPEG-4 SLS audio compression taking robustness, imperceptibility and security into consideration. The chaotic watermark generated by chaos is then embedded to ensure security. Moreover, the adaptive spread spectrum (SS) method is exploited to further optimize the tradeoff between robustness and transparency of this scheme.

Secondly, a blind and robust audio watermarking scheme based on singular value decomposition (SVD) – followed by discrete cosine transform

(DCT) with the synchronization code technique is proposed. A binary watermark is embedded into the high frequency band of the SVD-DCT block. Chaotic sequence is adopted as the synchronization code and inserted into the host signal.

Thirdly, a new and robust audio watermarking scheme based on lifting wavelet transform (LWT) and SVD is proposed. Specifically, the watermark data is efficiently inserted in the low coefficient of the frequency subband of LWT audio taking advantage of both SVD and quantization index modulation (QIM). Furthermore, the synchronization code technique is also integrated in the hybrid LWT-SVD audio watermarking method.

Finally, a particle swarm optimization (PSO) and SVD based audio watermarking scheme is proposed. The watermark is embedded by modifying singular values (SVs) of the host audio signal with the SS method. The designed optimization technique employs PSO to search for the optimal watermark strength by continuously adjusting the fitness values so that the conflicting requirements of robustness and imperceptibility of the watermark are satisfied optimally. Moreover, the proposed scheme has the flexibility to include or remove various watermarking attacks during training for optimization.

The experimental results validate the practicality of the proposed methods and show that the proposed algorithms are very robust to typical signal manipulations. Furthermore, the results of comparison with other similar state-of-the-art audio watermarking techniques indicate that the proposed algorithms are more robust against most common signal processing attacks while preserving good imperceptibility.

List of Figures

Figure 1.1: The conflicting watermarking requirements.....	9
Figure 2.1: Typical watermarking system.....	13
Figure 2.2: The chaotic sequences with slight change and auto-correlation.....	31
Figure 2.3: Bifurcation diagram of logistic map.....	32
Figure 2.4: Time series, auto- and cross- correlation of different chaotic maps...	33
Figure 2.5: Comparison of auto- and cross-correlations of chaotic and PN sequence.....	34
Figure 3.1: Block diagram of SLS encoder.....	53
Figure 3.2: Signal, mask energy and allowed distortion in one frame.....	57
Figure 3.3: Proposed watermark embedding diagram.	61
Figure 3.4: Watermark embedding with perceptual model.....	64
Figure 3.5: Diagram of watermark message extraction.	66
Figure 3.6: Key space under different initial value difference.....	71
Figure 3.7: Original, watermarked and residue audio waveform in time domain.	73
Figure 3.8: Spectrogram of original, watermarked and residue audio.....	73
Figure 3.9: Subjective listening test results.....	77

Figure 4.1: Diagram of watermark embedding process.....	86
Figure 4.2: Sketch of SVD-DCT macroblock.....	87
Figure 4.3: Probability of false positive error under various M	92
Figure 4.4: Probability of false negative error under various M	93
Figure 4.5: Watermark signal; (a) original watermark; (b) extracted watermark without attack.....	95
Figure 4.6: Waveform and spectrogram of the original audio signal and the watermarked host audio signal; (a) time domain waveforms; (b) frequency domain spectrograms (SNR=36.98, SSNR=41.05, MOS=4.86, ODG=-0.53, SDG=-0.51, LLR=0.012, CD=0.44).	102
Figure 5.1: Decomposition and reconstruction of lifting wavelet.....	117
Figure 5.2: Data structure of hidden bitstream.....	119
Figure 5.3: Diagram of watermark embedding process.....	123
Figure 5.4: Waveform of original, watermarked and residual signals.....	128
Figure 5.5: Spectrogram of original and watermarked signal.....	128
Figure 5.6: SNR and SSNR results versus quantization step.....	128
Figure 5.7: MOS scores of the 15 test sequences.....	129
Figure 5.8: Key space under different initial value difference.....	136

Figure 6.1: The PSO flowchart.	146
Figure 6.2: Diagram of the watermark embedding process.	148
Figure 6.3: Diagram of the watermark extraction process.	150
Figure 6.4: Watermarking method with PSO training.	152
Figure 6.5: PSO training results of avemaria audio signal; (a) Optimal SF results (optimal SF =0.171); (b) Optimal SNR and SSNR results (optimal SNR=34.0847 and optimal SSNR=34.5148); (c) Optimal NC _m results (optimal NC _m =0.9982); (d) Optimal LLR results (optimal LLR=0.0402).	161
Figure 6.6: PSO training results of cymbal audio signal; (a) Optimal SF results (optimal SF =0.4590);(b) Optimal SNR and SSNR results (optimal SNR=31.2512) and optimal SSNR=37.6295); (c) Optimal NC _m results (optimal NC _m =0.9961); (d) Optimal LLR results (optimal LLR=0.0123).	161
Figure 6.7: Waveforms and spectrograms of the original and watermarked audio; (a) avemaria audio signal (SNR=34.0847dB, SSNR=34.5148dB, LLR=0.0267);(b)cymbal audio signal (SNR=31.2512dB, SSNR=37.6295dB, LLR=0.0123); (c)spectrogram of avemaria audio signal; (d) spectrogram of cymbal audio signal.	164

List of Tables

Table 3.1 MPEG-4 audio test sequences (16bits stereo).....	72
Table 3.2 BER of watermark extraction under attacks.	74
Table 3.3 Objective Evaluation Test Results.	75
Table 3.4 Watermark Payload for standard test sequences.....	77
Table 4.1 Test sequences of SQAM database (16 bits).	94
Table 4.2 MOS, SDG and ODG descriptions and results from the proposed scheme.....	98
Table 4.3 Objective and subjective assessment results of the proposed scheme.	102
Table 4.4 Results of robustness against common signal processing attacks achieved by the proposed scheme vs. that in [58].....	105
Table 4.5 Results of robustness against Stirmark attacks for audio achieved by the proposed scheme vs. SVD based [56] and DCT based [12] schemes.....	107
Table 4.6 Comparison results of capacity performance achieved by our proposed scheme vs. state-of-the-art audio watermarking algorithms.....	110
Table 4.7 Comparison results of SNR and MOS values achieved by the proposed scheme vs. traditional and SVD based methods.....	111
Table 5.1 Robustness to common signal processing attacks and comparison. ...	131

Table 5.2 Robustness to the Stirmark attacks.	133
Table 5.3 Algorithm comparison with DWT-SVD method regarding the robustness.	137
Table 5.4 Algorithm comparisons.	139
Table 6.1 The effect of the cognitive coefficients and inertia weights of PSO algorithm on the avemaria audio signal.	159
Table 6. 2 The effect of the cognitive coefficients and inertia weights of PSO algorithm on the cymbal audio signal.	159
Table 6.3 Optimal values under different particles (avemaria.wav).	159
Table 6.4 Optimal values under different particles (cymbal.wav).	160
Table 6.5 Optimal parameters after PSO of 12 test audios.	162
Table 6.6 NC results of 12 test audios.	162
Table 6.7 Perceptual quality of 12 test audio signals.	164
Table 6.8 Robustness test and algorithm comparison.	166
Table 6.9 Robustness results with and without PSO.	168
Table 6.10 Comparison results with ATS method and GA method.	169

List of Abbreviations

AAC	advanced audio coding
ATH	absolute threshold of hearing
ATS	adaptive tabu search
AAZ	advance audio zip
bps	bits per second
BER	bit error ratio
BPGC	bitplane Golomb coding
CBAC	context-based arithmetic coding
CBR	critical band rate
CD	cepstral distance
DCT	discrete cosine transform
DFT	discrete Fourier transform
DRM	digital right management
DST	discrete sine transform
DWT	discrete wavelet transform
GA	genetic algorithm
HAS	human auditory system
IFPI	International Federation of the Phonographic Industry

LEMC	Low energy mode coding
LLE	Lossless Enhanced
LLR	log-likelihood ratio
IntMDCT	integer modified discrete cosine transform
LPF	low pass filtering
LSB	least significant bit
LWT	lifting wavelet transform
MCLT	modulated complex lapped transform
MOS	mean opinion score
MPEG	Moving Picture Experts Group
NC	normalized cross-correlation coefficient
NMR	noise to mask ratio
ODG	objective difference grade
PAQM	perceptual audio quality measure
PCM	pulse code modulation
PEAQ	perceptual evaluation of audio quality
pdf	probability distribution function
QIM	quantization index modulation
PN	pseudo noise

PSM	pitch scale modification
PSO	particle swarm optimization
RSVD	reduced singular value decomposition
SDMI	Secure Digital Music Initiative
SDG	subjective difference grade
SF	scale factor
sfb	scale factor band
SLS	scalable to lossless
SMBA	StirMark for audio
SNR	signal to noise ratio
SPL	sound pressure level
SQAM	sound quality assessment material
SS	spread spectrum
SSNR	segmental signal to noise ratio
STFT	short time Fourier transform
SV	singular value
SVD	singular value decomposition
SVM	support vector machine
XOR	exclusive OR

Chapter 1

Introduction

1.1 Motivation

With the widespread availability of digital storage devices and the rapid growth of new technologies, a huge amount of digital information content has been produced and distributed quickly without much protection during transfer and communication. With the development and popularity of the internet, duplication, modification and forgery of digital multimedia have become more difficult to detect and prevent. Publishing content covered by intellectual property or copyright is then faced by the challenge of countering illegal access to the digital media and goods, while still enabling customers to enjoy their usage rights on the purchased content and products. Meanwhile, the implication of piracy of digital media could be terrible since the media data plays a major role in a wide range of products and services. Apart from this, the popularization of digital devices such as digital voice recorder allows more manipulation and tampering of the content than before, thus content providers are seeking new and effective techniques to protect their copyrights. To make things worse, improvements in media data reproduction, distribution and

storage have also greatly increased many chances for potential infringement. Concerns have been raised about the protection and enforcement of intellectual property rights for digital media. Although legal institutions exist for protecting intellectual property (trademarks and patents) owned by content creators and the traditional methods are very popular, they have clear limitations for audio applications and complementary technical measures are needed to sustain financial returns and ensure incentives for new creations. Consequently, the issues of intellectual property rights protection and digital right management (DRM) have attracted a lot of attention for developing effective ways of authenticating multimedia in these years using information hiding and watermarking techniques [1-5].

Obviously, it is desirable to provide a secure and efficient way to distribute and deliver multimedia data. Generally speaking, signature-based and digital watermarking approaches are two methodologies of multimedia authentication and protection. The signature-based idea is to sign off every multimedia signal by means of imprint characteristics of the source's owner [4]. This signature should be hidden in the signal and should not change its quality. Since the late 1990s, tremendous digital watermarking algorithms have been exploited and published, which is most likely due to the increase concern

over copyright protection of content and intensive efforts in securing multimedia data [6]. Moreover, digital watermarking [2, 4, 7-8] has been widely developed and investigated by researchers and gaining more and more popularity in the information security industry. The main advantage of using the digital watermarking techniques is that close relationship of the watermarked information and the media content requires no auxiliary information to be appended to the data. It is known that both encryption and watermarking techniques can protect the copyright information in their own ways. By hiding copyright information into the “essence” of the multimedia object to protect the media during transmission and after encryption, digital watermarking [9] complements encryption and provides protection for the media at all times even after decryption. As to the difference between these two techniques, the hidden information, i.e. the watermark, should be imperceptible and robust against malicious attacks and cannot be removed without damaging the original data. Moreover, the watermark persists within the work after decryption and may even survive after some malicious attacks. The role of watermarking is seen as a technique to complement encryption but does not replace encryption in preventing piracy.

Since the source to be protected in my work is audio, the embedding procedure

can exploit the imperfections of the human auditory system (HAS), which is also the basic idea of audio watermarking [7, 10]. In recent decades, it is known to all that a lot of basic concepts of watermarking are mostly applied to image or video signals. Compared to embedding watermarks into still image and video, hiding data in audio is much more challenging due to the reason that HAS is more sensitive than the human visual system [10]. Therefore, understanding of the human perception process and its inclusion in effective psychoacoustic models are the keys to a successful audio watermarking algorithm. With the proliferation of digital audio media and popularity of more audio standards, design and development of new and robust digital audio watermarking techniques for copyright protection and authentication are urgently required. In view of this, detailed research and development of this important topic and discussion on the important issues such as robustness and imperceptibility of the audio watermarking techniques will be meaningful and significant. Since the distribution of most audio content is carried out in the form of compressed streams, integrating the structure of the two processes, audio compression and watermarking, is very desirable for achieving codec protection purpose and potential commercial application. Furthermore, it is obvious that the digital audio watermarking technology is recognized as a

helpful and useful way in dealing with the audio copyright protection problem.

It is found that the most widely used way to increase watermarking security is through the use of keys to generate random sequences during the embedding process [2, 7, 11-12]. Chaotic sequence [13-19] is another way to enhance the watermarking confidentiality as it has several attractive features such as non-invertibility, asymmetry, ease of generation and the availability of a large number of sequences as well as sensitivity to initial conditions. As a matter of fact, a great amount of uncorrelated, random-like, deterministic and reproducible chaotic sequences can be obtained with small perturbations of the parameters. Therefore, the watermarking performance can be concentrated on the discrete-time dynamic system operation in chaotic states including Bernoulli, Chebyshev and logistic maps. Chaotic sequences can be explored to modulate and modify the information bits into the white noise-like wideband watermark signals.

Actually, how to adaptively modify the features of the watermarking scheme and thus survive the common signal processing related attacks in order to accommodate different robustness and inaudibility requirements is a challenging problem. In most of the applications, digital audio signals have to be compressed to facilitate storage and transmission, but the lossy compression

may result in a false alarm in the fragile watermark detection. Therefore, in order to provide content protection for audio, some recent works explore the watermarking schemes to achieve copyright protection and digital right management (DRM) goal [20-21]. Given such popularity, the interest in developing watermarking techniques for audio that are aligned with the audio standards is very high. DRM is a collection of techniques and technologies that enable technically enforced licensing of digital information. Specifically, DRM for audio enables the secure exchange of audio intellectual property [22].

This work is mainly based on the consideration of the relationship between HAS, chaotic characteristics, singular value decomposition (SVD) properties and watermarking requirements. The main goal of this research is to come up with novel audio watermarking schemes with improved and enhanced performance based on the understanding of the state-of-the-art audio watermarking methods in the surveyed papers in this field. With the exploration of fundamental issues of audio watermarking schemes, some hybrid and novel audio watermarking methods will be developed and designed to provide solutions for copyright protection and authentication for audio data. Therefore, the significance of this study is that the state-of-the-art research and development of the new watermarking schemes will boost DRM technologies,

popularize and enhance the development of audio copyright protection and authentication [23-26]. The potential copyright protection market will further promote the digital watermarking development and its related business. The outcome of this proposed research will serve the rapidly increasing demands of multimedia security and copyright protection in today's digital world. These applications [10] typically include a variety of domains such as professional audio, internet music, consumer electronics and broadcasting, pay-TV services, secret video telephony, video conferencing, multimedia short messages via mobile phones and personal digital assistants, distribution of commercial, medical, governmental and military images and/or videos, etc.

1.2 Scopes

There are some metrics to measure the effectiveness of a watermarking algorithm. The most important and critical requirements are stated in the International Federation of the Phonographic Industry (IFPI) and Secure Digital Music Initiative (SDMI), which specify that audio watermarking should have the following specifications [27]:

- 1) Imperceptibility (transparency or inaudibility): Imperceptibility is usually defined as the perceptual similarity between the original and the watermarked audios. The audio quality should not be degraded after the addition of the

watermark. It can be measured and evaluated using both objective and subjective performance tests. Signal to noise ratio (SNR) is usually used to evaluate the transparency and it should be more than 20 dB;

2) Capacity (payload): Capacity can be defined as the amount of data that can be embedded into the host audio signal without losing its imperceptibility. An appropriate capacity should usually be more than 20 bits per second (bps);

3) Robustness: Robustness is defined as detection reliability in the presence of malicious attacks, that is, watermarked signal should be able to not only withstand attacks such as unauthorized detection, removal and embedding, but also common audio signal processing operations and attacks even when the audio quality becomes very poor.

4) Security: In general, watermarked signals need to be secure and should not be accessible by unauthorized users. The watermarking algorithm security should rely on secret keys rather than the secrecy of the watermarking algorithm.

As shown in Figure 1.1, robustness, imperceptibility and payload are three main requirements that are mutually conflicting and contradictory. Therefore, tradeoffs are needed when designing and developing a new and successful watermarking scheme. When designing audio watermarking schemes for the

audio content protection, not only the tradeoff between efficiency and watermarking performance should be taken into consideration, but also some special features of audio watermarking, such as robustness, perceptibility, watermark payload, error tolerance, scalability, format-compliance and complexity.

Figure 1.1: The conflicting watermarking requirements.

In this thesis, robust digital audio watermarking algorithms for copyright protection are researched, which aims to develop and design new audio watermarking algorithms that improve the performance of the existing algorithms presented in the open literature and to verify their performance in the face of standard watermarking attacks.

1.3 Major Contributions of the Thesis

Four original audio watermarking methods are proposed in the different domains to carry out audio copyright protection. The main contributions of this

thesis are as follows:

- 1) Demonstration of a compressed domain audio watermarking technique for MPEG-4 SLS based on HAS and significant state. The adaptive spread spectrum (SS) watermarking scheme is introduced with the control of the SNR.
- 2) Development of a new synchronization technique in the proposed SVD-discrete cosine transform (DCT) based method. SVD algorithm is adopted because of its desirable and superior features. The binary watermark is inserted in selected frequency bands of SVD-DCT blocks by imposing a particular relationship on the selected pairs of DCT coefficients rather than directly embedding the watermark into SVs.
- 3) A classical blind additive audio watermarking algorithm in the lifting wavelet transform (LWT)-SVD domain has been proposed. In order to increase the robustness on desynchronization distortions, the proposed watermarking algorithm can be successfully combined with the introduced synchronization technique. The chaotic encryption is also used in the scheme to enhance security.
- 4) A particle swarm optimization (PSO) based audio watermarking has been developed too. PSO technique is an effective method for improving the watermarking performance and optimizing the tradeoff between the conflicting

requirements such as robustness and imperceptibility.

1.4 Organization of the Thesis

This thesis is organized as follows. Following the introduction, the background and literature review on different watermarking schemes for audio content protection are given in Chapter 2. The background material is presented to provide essential materials related to this work. Chapter 3 introduces a perception-based audio watermarking scheme in the compressed bitstream, which exploits the HAS, SS and chaotic encryption. Apart from the compressed domain watermarking scheme, a blind and robust audio watermarking scheme based on SVD-DCT is presented in Chapter 4. The SVD-DCT algorithm integrates both SVD and DCT to achieve good performance. Besides, synchronization code is introduced to withstand desynchronization attacks. Chapter 5 presents a novel LWT-SVD watermarking scheme with the exploration of the chaotic signals to improve the security. To optimize the watermarking parameters, an intelligent and optimal watermarking scheme based on PSO is described in Chapter 6. The watermarking scheme achieves an optimal tradeoff between robustness and imperceptibility by PSO training. Chapter 7 concludes the thesis and provides directions for future research.

Chapter 2

Research Background

2.1 General Watermarking Scheme

The watermarking process involves the imperceptible insertion of the watermark at the embedder and its perfect recovery at the detector. Therefore, the host signal can be viewed as a carrier or channel used to convey the watermarked information. The watermarked signal may be subjected to a variety of processes which affect its properties, such as storing in different media, cropping, noise addition, compression, sampling rate changes, or even malignant attacks aiming at the destruction of the inserted watermarks. As such, a watermarking system can be modeled from a secure data communication perspective. A typical watermarking system is illustrated in Figure 2.1. If the secret message, b (watermark), is embedded into a host signal, S , with a secret key, k , then the watermarked signal, W , is represented by

$$W = f(b, k, S). \quad (2.1)$$

In the digital watermarking field, a low-energy signal called watermark which depicts some data such as the copyright information about the host signal is imperceptibly hidden in a host signal. It can be a set of characters, a

trademark image, or even a short period of video. The host signal where the watermark is embedded is also referred to as a cover object since it covers the watermark. The cover signal can be any signal in the digital format such as a still image, an audio clip, a video sequence and a text document.

Figure 2.1: Typical watermarking system.

The secret message usually refers to the watermark that the system needs to imperceptibly transmit over the transmission channel. The secret key, which is in both the watermark embedder and detector, is used to modulate the watermark by the watermark modulator.

In the watermark embedder, the watermark embedding is usually carried out by adding a watermark signal to the host signal without distorting the imperceptibility of the host signal. Blind embedding treats the host signal as noise while for informed watermarking, the host signal is known to the

embedder during the watermark insertion which is more effective than blind embedding. In this way, informed watermark embedding can be denoted as:

$$W = S + a \times w \times b, \quad (2.2)$$

where a is a strength control factor used to ensure inaudibility of watermark embedding.

The watermark embedder inserts the watermark into the host audio in an imperceptible way and sends the watermarked signal (stego-signal). The watermarked signal is then transmitted over a noisy channel to the receiver.

The received signal, W^* , may get distorted due to possible channel noise, N .

The distorted signal at the receiver side is denoted as

$$W^* = W + N. \quad (2.3)$$

It should note that the distorted signal at the receiver side can be caused by convolution too.

The watermark detector needs to recover the secret message from the received signal with the help of the secret key. The watermark detector performs two functions: watermark decoding and watermark detection. The watermark function decodes the received watermark, W^* , using the secret key, k , and it determines whether the host signal is watermarked by the secret key, k , or not based on the recovered watermark. The communication channel is

usually hostile and noisy, and hence digital watermarking techniques should be resistant to both noise and security attacks. Most of the time, the received signal is a distorted version of the watermarked signal.

2.2 Properties of Watermarking Systems

In general, good watermarking techniques require some characteristics or properties [9, 28]. These characteristics apply to watermarking schemes for all kinds of data that can be watermarked, such as audio, image, video, formatted text and 3D models. A watermarking scheme is characterized by a number of defining features. The relative significance of each feature depends on the requirement of the application and the role of the watermark. The suitability of a given watermarking system for a given application may be judged in terms of the following features [1, 29].

2.2.1 Imperceptibility

Imperceptibility (also known as fidelity, inaudibility or transparency) is the perceptual similarity between the original and watermarked versions of the media object. One of the main requirements for watermarking is imperceptibility. The embedded watermark must be perceptually transparent. The data embedding process should not introduce any noticeable artifacts into the cover data. If annoying distortions are introduced by a watermarking

process, the commercial value of the watermarked data will be destroyed. For audio watermarking, it is therefore important to design a watermark which exploits HAS characteristics so as to maximize the watermark energy for high robustness under the constraint of the human audible threshold. In general, many factors affect the human perception of distortion, including the signal characteristics, distortion characteristics and the listening environment.

2.2.2 Robustness

Robustness of the watermarked data is another key requirement in watermarking [30-32]. Robustness refers to the ability to detect or extract the watermark after general signal processing modifications and/or malicious attacks. When in possession of a watermarked document, an unauthorized party should not be able to destroy the watermark without rendering the document useless at the same time. To ensure robustness, the watermark information is usually redundantly distributed over many samples (or features) of the cover media, thus providing a global robustness. Obviously, the watermark extraction is more robust if more of the watermarked data is available to the extraction process.

Researchers currently do not appear to have found a watermarking system which is 100% robust. When a watermark method is extremely robust, it tends to be perceptible. The more robust a watermark is, the more difficult it is to

remove from the watermarked signal. Many watermarking papers report robustness by measuring bit or detection error rate of the embedded watermark after performing specific attacks [30-32]. Actually, not all watermarking applications require robustness to all possible signal processing operations. For example, there are some watermarking applications such as tamper detection [33] and content authentication [34-36] which do not require such degree of robustness.

2.2.3 Security

In most copyright protection applications, the secrecy of the embedded data is of paramount importance [11, 37-38]. If security is a requirement, a secret key has to be adopted in both the embedding and extraction processes. A secure watermarking procedure cannot be broken unless the unauthorized user has access to the secret key. Generally, a watermarking scheme adopts one or more cryptographically secure keys to guarantee the security against attacks and erasure of the watermark. As a result, knowing the algorithms for embedding and extracting does not help unauthorized parties in detecting or removing the watermark.

2.2.4 Capacity

Capacity is often defined as the data payload or the number of information bits that can be embedded into the host media [1]. A watermark should include a

large number of identifying information. If it is not differentiable, it is often useless. It must be distinguishable from other watermarks. Capacity is a significant requirement as it directly impacts the watermark robustness [39]. The watermark robustness becomes lower if there is more information embedded. Generally, the capacity should be as high as possible. In these cases, the algorithms must be able to embed a large amount of data while maintaining the robustness and integrity of the cover signal. The data payload is dependent on the application and does not contain additional bits for error correction codes. For applications like copy control, the presence or absence of one-bit watermark is generally enough but for other applications such as fingerprinting or copyright protection, much more data needs to be added.

In fact, the data embedding payload is often evaluated in bits per second (bps). Several measures have been suggested for payload [1]. Here, the retrieval payload relative to the size of the marked signal is used and is denoted as

$$Payload = \frac{N_w}{Time}, \quad (2.4)$$

where N_w is the length of the watermark, $Time$ is the duration of the host audio.

From this equation, it can be seen that the embedding capacity increases as N_w decreases. However, a larger N_w results in a higher distortion. In principle, this method is independent of the watermarked signal in the sense that two

watermarked signals with the same sampling time and the same watermarking parameters would have exactly the same capacity.

2.2.5 Complexity

Computational complexity of the watermarking processes is another consideration [40]. The economics of deploying watermark embedders and detectors can be extremely complicated and their adoption depends on the business models involved. The concerned issues are speed and memory usage of the embedder and detector. In the case of audio watermarking, it is important that the algorithms must operate efficiently, that is, the watermark detector must be fast enough for realtime detection.

A watermarking scheme usually should satisfy some requirements such as transparency, robustness, blind detection and so on [1]. However, all the current watermarking methods only satisfy a portion of the requirements.

2.3 Classification of Watermarking Techniques

Digital watermarking techniques can generally be categorized in a number of ways dependent on different parameters and classification methods [1]. Various types of watermarking techniques exist and are listed together with their respective applications [5]. A watermarking system may be classified according to whether or not the original media and the secret key information are needed

in the watermark extraction algorithm for non-blind, semi-blind and blind watermarking systems. According to the visibility of the watermark, watermarking schemes may be classified into two types: visible techniques and invisible techniques. According to their robustness against attacks, watermarking technologies may also be classified as fragile, semi-fragile and robust.

2.3.1 Blind, Semi-blind and Non-blind

A blind (or public) watermark is often “blindly” extracted without any prior knowledge of the original host signal or any information related to the original data for watermark detection. The non-blind (also informed or private) in which the watermark is embedded in the original host signal and requires access to the original or some information of the original data for watermark extraction. The blind watermark has many more applications than the informed watermark. The only requirements for extracting the watermark are the watermarked cover object and the key employed to embed the watermark in the cover object. Some watermarking mechanisms require the original cover object which restricts their usefulness. Semi-blind watermarking system is also called semi-private watermarking. This type of watermarking technique does not use the original cover object, but it requires the secret keys and a copy of the embedded watermark data. Whether a watermarking system employs a blind or an

informed detection can be critical in determining whether it can be used for a given application. Generally, informed watermarking schemes are more robust while the blind watermarking approaches are more desirable in practice, as informed schemes require huge storage for the original data.

2.3.2 Audible and Inaudible

Audible (non-transparent) watermarks are embedded in audio content in such a way that they are only audible when the content is played. In inaudible (transparent) watermarking techniques, the watermark data is embedded in the cover object in a way that the watermark is intentionally perceptible to a human observer. The primary purpose of audible watermarks is to prevent unauthorized commercial use of the media. At the same time, the observer knows the institution that owns the material. Transparent watermarks are inaudible and they cannot be detected by just playing the digital content. In contrast to the audible watermark, the inaudible watermark is embedded into the host contents such that the watermark is not perceptible to an observer but may be extracted or detected by a computer program. The inaudible watermark is used to make an assertion about the ownership of an object.

2.3.3 Robust, Semi-fragile, Fragile

Robust watermarking is a method in which manipulation of the watermarked content will not influence the watermark and is widely used for the application

of copyright protection. On contrast, fragile watermarking is a method in which the watermark gets manipulated when the watermarked content is modified or tampered. Semi-fragile watermarking techniques are able to tolerate some degree of change to the watermarked content. Fragile and semi-fragile techniques are suitable for the authentication application. The robust watermarking methods needs to address copyright issue and takes the appropriate tradeoff between imperceptivity and robustness against manipulations [41-42].

2.4 Watermarking Applications

There are several application areas that range from copy protection to steganography [43]. Film and music makers, TV stations and courts are very much interested in using digital watermarking and cryptography as two complementary technologies. Watermarking applications can be found in different areas and may be categorized as follows:

2.4.1 Watermarking for Copyright Protection

One of the typical applications of watermarking is to protect copyright, which is probably the most prominent application of digital watermarking currently. Copyright protection is to embed some identification information so that the copyright owner can prevent other parties from claiming copyright on the work.

The product might have been manipulated by an adversary in an effort to remove or destroy the watermark or replace it with his own watermark. Detecting the invisible watermark provides the content owner with additional evidence of ownership. Copyright protection requires a high level of robustness. The driving force behind this application is the World Wide Web, which contains millions of freely available media that the rightful owners want to protect.

2.4.2 Watermarking for Authentication

When a user receives a multimedia file, it may be impossible to determine whether it is original or not. The problem with digital signature is that they can be lost during normal usage. By using watermarking, the signature can stay with the media and checks can be made to determine whether the data or the meaning of the content has been altered. It is even possible to provide localization information about where the content has been tampered. If the content has been modified, the copyright owner may hope to know how and to what extent, the content has been impaired. In authentication applications, the objective is not to protect the contents from being copied or stolen, but rather to authenticate the cover media and ensure its integrity. This can be done through fragile or semi-fragile watermarks that have a low robustness against certain manipulations, like compression. If the watermark is detected perfectly, the

data is genuine; otherwise, the data may have been corrupted. Among all the possible watermarking applications, authentication watermarks require the lowest level of robustness.

2.4.3 Watermarking for Copy Protection or Usage control (Copy Control)

The content provider is sometimes interested in limiting the level of control that the users may have on the purchased product. It is possible to embed copy information into a product and use this information to indicate the copy status of the data. Depending on the information of the copy status, the devices can allow or prohibit certain operations on the content. Some copy protection schemes allow a user to create a single copy but restrict the user from making additional copies from the copy.

2.4.4 Fingerprinting for Transaction Tracing

Watermarks such as user IDs, also known as fingerprints, are widely used for distinguishing an object from other object with similar characteristics. It can be used to track the source of pirated copies and enable legal action to be taken against the party who breached the licensing agreement. Tracing illegal copies is also one of the important issues that can be solved by fingerprinting. Fingerprinting is useful in monitoring or tracing illegally produced copies of the data that is being circulated. Watermarking for fingerprinting applications requires high robustness against standard data processing attacks.

2.4.5 Owner Identification

If the owner of a copyrighted material does not put a copyright notice on the distributed material, the award for the copyright holder can be limited in the event that the material has not been used as agreed upon. The problem with copyright notices is that they can be easily removed and may only cover a portion of the material. As watermarks can be made both imperceptible and inseparable from the material that contains them, they are a better alternative to copyright notices.

2.5 Watermarking Methods

There are three major watermarking schemes in multimedia. The first is time domain watermarking. The second is in various transform domains [44], such as Discrete Cosine Transform (DCT), Singular Value Decomposition (SVD), Discrete Fourier Transform (DFT) etc. The third is only for compressed image, audio or video files, that is, compressed domain watermarking. In general, for audio watermarking, it is fast, efficient to implement watermarking in time domain method [30], while the watermarking has an advantages of high robustness in transform domain.

2.5.1 Time Domain Watermarking

Spatial domain or time domain based watermarking [30, 45] is one of the

fundamental techniques of digital watermarking. The most straightforward way to add a watermark in the time domain is the addition of a pseudo random noise sequence to the host signal. A pseudo random noise sequence consists of floating point numbers, which are random values generated by a seed. Time domain embedding techniques are very simple and effective but they are not robust against all kinds of attacks, especially the cropping attack. Embedding watermarks into the time domain components of the cover objects is a straightforward method. It is one of the fundamental schemes used since the beginning of digital watermarking in 1993 [4]. A well-known classic time domain based watermarking method is the least significant bits (LSB) modification scheme. In summary, the time domain watermarking techniques can be categorized into three main groups: SS based coding, echo hiding [46-47] and phase coding.

Generally speaking, time domain based watermarking schemes possess advantages such as easy implementation and low complexity, as compared to other domain based watermarking schemes. However, they also possess shortcomings such as weak robustness when compared to other domain based watermarking schemes. Due to this, most of the fragile watermarking schemes are in the time domain. Many methods have been proposed to improve the

drawback of weak robustness.

2.5.2 Transform Domain Watermarking

The transform based watermarking scheme first transforms the cover object into the frequency domain [44]. It then selects a number of frequency bands according to the user-specified key and modifies these bands to carry the watermark bits. Compared with time domain based watermarking techniques, transform domain based watermarking techniques have become main stream and more popular in the area of digital watermarking, since such schemes not only provide good watermarked quality but also stronger robustness under general attacks or noise addition. Generally, the transform domain approach is more robust against most of the attacks but one major drawback is the high computational requirement.

Therefore, many researchers have already proposed a variety of schemes and applications. The most popular and common transform domain techniques are discrete Fourier transform (DFT) calculated by the fast Fourier transform (FFT) algorithm [48-50], discrete cosine transform (DCT) [12, 51], discrete wavelet transform (DWT) [52-53], discrete fractional sine transform (DFRST) [54], LWT [55] and singular value decomposition (SVD) [56-63]. Meanwhile, hybrid transforms are also becoming more popular in the audio watermarking field. Typical hybrid transforms include DWT-DCT [41], support vector

regression (SVR)-DWT-DCT [64], DWT-SVD [58] and SVD-DCT [61].

2.5.3 Compressed Domain Watermarking

Sequences of raw host signal need a lot of space. Hence, there are several compression methods that compress the raw host signal to reduce storage requirements. Often digital audio is already in a compressed format when watermarking is applied. It is desirable to embed the watermark directly into the compressed bitstream without going through a full decoding, watermarking and re-encoding steps, which will add considerable complexity and additional time delay.

2.6 Watermarking Attacks and Evaluations

Actually, the watermark signal is usually attacked before extraction/detection by the watermark detector. When the watermark is attacked, it may remove the inserted watermark, lower the detection ratio, or lower the watermark security. The watermarked signal may be attacked multiple times too. Removal attacks can destroy or damage an embedded watermark. An ideal watermarking scheme should be able to withstand any attacks. Depending on the application and watermarking requirements, the list of distortions and attacks to be considered includes: signal processing attacks, additive and multiplicative noise (Gaussian, uniform), linear filtering (low-pass, high-pass filtering), lossy

compression (e.g. ISO/MPEG-1 Audio Layer III (MP3), advanced audio coding (AAC)), synchronization attacks (time scale modification (TSM), pitch scale modification (PSM), cropping, shifting, jittering).

In the development of digital watermarking methods, a significant and often neglected issue is the design of the proper evaluation and benchmarking method [65]. In the early days, the lack of a formal evaluation mechanism for designing watermarking methods caused difficulty or confusion among researchers and hindered the use of digital watermarking in various applications. By using a well-defined watermarking benchmark, a table of test results that provide a trustable and reliable summary of the performances of the proposed watermarking scheme can be made available for end users by watermarking researchers or related software manufacturers. To address this problem, some watermarking benchmarks for still digital image watermarking have been first designed, e.g. StirMark, Certimark, Checkmark and Optimark. The benchmark tool for audio signal is StirMark for audio (SMBA) [66]. These benchmarks can be used to evaluate robust watermarking algorithms that are designed to survive normal signal processing attacks. Thus these benchmarks include many attacks that are encountered in various applications. The availability of such benchmarks contributed significantly to the advancement of

watermarking algorithms.

2.7 Chaotic Sequences in the Watermarking Field

In the mathematical field, chaos is often defined as the deterministic behavior which is very sensitive to its initial conditions, that is, very small perturbations of initial conditions of a chaotic dynamic system can result in large variations in its behavior. A chaotic map that exhibits some sort of chaotic behavior can be fully described by

$$\{x(n) : x(n) = F(x(n-1), r)\}, \quad (2.5)$$

where r is a function seed and $F(\bullet)$ is a nonlinear transformation function. A discrete-time chaotic signal $x(n)$ can be derived from a chaotic system with a single state variable by applying the recursion

$$x(n) = F(x(n-1), r) = F^n(x(0), r), \quad (2.6)$$

where $x(0)$ is the system initial condition. A chaotic sequence may be easily reproduced given the same initial conditions and initial value $x(0)$. A slight change in the initial conditions of a chaotic function will lead to significant changes in the resultant mapping but their auto-correlation values may change slightly which can be verified in Figure 2.2. Chaotic maps often occur in the study of dynamic systems and fractal generation. A fractal may be constructed by an iterative procedure that generates them. This is often because there are

several different iterative procedures to generate the same fractal.

Figure 2.2: The chaotic sequences with slight change and auto-correlation.

The simplest chaotic map is the logistic map which is defined as:

$$x(n+1) = r \times x(n) \times (1 - x(n)), \quad (2.7)$$

where $x(n)$ is the current value of the mapping at time n with an initial value, $x(0)$, and r is the function seed. The periodic and bifurcation regimes (the chaotic regime) are two distinct regimes for the logistic map. A bifurcation diagram is shown in Figure 2.3, where the horizontal axis denotes the values of the parameter, r , while the vertical axis denotes the possible long-term values of x . When $r = 3$, $x(n)$ oscillates between two values and never converges. As r increases, $x(n)$ goes through bifurcations and eventually becomes chaotic.

When $r \approx 3.5699$, $x(n)$ becomes random. When $r > 3.5699$, the behavior is in chaotic state.

Figure 2.3: Bifurcation diagram of logistic map.

To date, a number of chaotic functions have been proposed in the literature for the purpose of watermark generation, the most prominent and popular ones are the tent map, Bernoulli map, logistic map, Markov map, Chebyshev map and Baker map. The pseudo random sequences generated by the chaotic dynamic system can be used to randomize coefficients in the watermarking field. A chaotic function can produce almost uncountable random sequences that have good auto-correlation and cross-correlation characteristics and are extremely sensitive to the initial secret keys. From Figure 2.4, it can be seen that different time series of different chaotic sequences have different auto-correlation and cross-correlation properties.

However their auto correlations are very similar.

Figure 2.4: Time series, auto- and cross- correlation of different chaotic maps.

Recently, a large number of watermark generation algorithms have been developed to date, ranging from the use of personal logos to the use of pseudo noise (PN) sequence of numbers [67]. Some watermark generation schemes proposed recently [2, 7, 11-12] use a PN generator to produce a watermark sequence which is inserted in the cover object. These sequences can be precisely modeled as independent, identically distributed random variables which follow the uniform distribution. Obviously, as an alternative to PN

sequence, there are three obvious advantages of the chaotic watermark over the commonly used PN sequence [68]. Firstly, it is easy to generate the watermark with different chaotic maps such as logistic, tent, Bernoulli and Chebyshev maps. Secondly, different chaotic system models, parameters and initial values can generate watermark sequence with zero auto-correlation. The third is the security: the watermark cannot be broken without the chaotic model and its corresponding key. Moreover, it is noted that there are differences from their auto- and cross-correlations as shown in Figure 2.5. The auto-correlation of chaotic and PN sequence is very similar. However, the cross-correlation property of chaotic sequence is slightly different from the PN sequences.

Figure 2.5: Comparison of auto- and cross-correlations of chaotic and PN sequence.

Meanwhile, the chaotic map has been shown to produce lowpass watermarks and also white watermarks. These white watermarks are similar to those generated by a PN generator. Since there are irrational numbers close to every rational number, the map exhibits sensitive dependence on initial conditions. As the number of iteration times increases, chaotic sequences are characterized by white spectrum. The control of spectral properties of the generated sequences of these chaotic functions offers a distinct advantage over the sequences generated by the pseudo random generators. For example, if the watermarked audio will be subjected to lowpass manipulations in general, a lowpass watermark is generated to be more robust to these attacks. By simply altering the function seed in these chaotic functions, one can generate watermarks with different spectral properties. Instead of seeding a pseudorandom number generator and then utilizing a filter to the resultant signal to produce colored noise, a single seed can decide the properties of the generated sequence from the chaotic function. Thus chaotic sequences are adopted to enhance the security, anti-counterfeit and noninvertibility of the proposed watermarking schemes.

Above all, the main idea for adopting a chaotic function to obtain a watermark is that a single variable which seeds the chaotic function, will

always lead to the same output (mapping) when certain constraints or initial conditions are applied on the mapping. The adoption of chaotic functions for the watermark generation has also been proposed in [13-19, 68]. For example, in [16] and [17], the Bernoulli and Markov maps are utilized as the chaotic functions. Two important values, the initial value and the function seed, are required to recreate the same watermark in these types of watermark generation schemes. The high robustness to lowpass attacks is one advantage of these watermarks schemes. With an additive or multiplicative technique, these watermarks are typically distributed over the full host signal. In [68], due to the unique features of chaos, chaotic sequences have shown superior robustness when compared to the widely used PN sequences in watermarking applications.

The process of generating a watermark derived from a chaotic map involves several steps. A value for the function seed and an initial starting value must first be selected. The holder of the copyright can utilize random real numbers between 0 and 1 as the initial secret keys in order to obtain the satisfactory imperceptible quality in the embedding process. The knowledge of the algorithm itself is insufficient to extract watermarks without the correct keys.

From the chaos characteristics, obviously, chaotic watermark can be

applied in the watermarking field to replace the most popularly used PN sequence. The initial values and chaotic function parameters can be used as the keys to control watermarking security.

2.8 SVD Principles and Properties

The traditional transform techniques such as DFT, DCT and DWT just decompose a signal in terms of a standard basis set, which is not an optimal representation in some sense. Owing to the unique features and attractive properties such as stability with little disturbance, SVD [69] has been used in many signal processing applications. As a kind of orthogonal transform and a numerical technique for diagonalizing matrix, SVD is a technique for linear algebraic in the transformed domain comprising basis states that are optimal in some sense. Recently, SVD is widely used in the watermarking field as a well-known numerical analysis tool in the sense that a slight modification of the large SVs will not affect the transparency of the cover object. The 1-D audio signal can be rearranged as a matrix. Thus audio signal also can take advantage of SVD property to tradeoff the robustness and transparency requirement. The SVD of $\mathbf{A}$ is usually defined as

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T = \begin{bmatrix} u_{1,1} & \cdots & u_{1,r} \\ \vdots & \ddots & \vdots \\ u_{m,1} & \cdots & u_{m,r} \end{bmatrix} \begin{bmatrix} s_{1,1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & s_{r,r} \end{bmatrix} \begin{bmatrix} v_{1,1} & \cdots & v_{1,r} \\ \vdots & \ddots & \vdots \\ v_{n,1} & \cdots & v_{n,r} \end{bmatrix}^T = \sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^r u_{i,k} \times s_{k,k} \times v_{k,j}, \quad (2.8)$$

where $\mathbf{U}$ is a $m \times r$ matrix, $\mathbf{V}$ is a $r \times n$ matrix and $\mathbf{S}$ is a $r \times r$ diagonal matrix with positive elements; the superscript T denotes matrix transposition and r is the rank of matrix $\mathbf{A}$. The diagonal elements of $\mathbf{S}$ are called SVs of $\mathbf{A}$, which are nonnegative and assumed to be arranged in a decreasing order, that is, $s_{1,1} > s_{2,2} > \dots > s_{r,r}$.

In SVD based watermarking, a signal can be treated as a matrix and decomposed into three matrices. Calculating the SVD consists of finding the eigenvalues and eigenvectors of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$. The eigenvectors of $\mathbf{A}^T\mathbf{A}$ consist of the columns of matrix $\mathbf{V}$ and the eigenvectors of $\mathbf{A}\mathbf{A}^T$ consist of the columns of matrix $\mathbf{U}$. The SVs in matrix $\mathbf{S}$ are square roots of eigenvalues from $\mathbf{A}\mathbf{A}^T$ or $\mathbf{A}^T\mathbf{A}$. The SVs indicate the energy intensity in its corresponding eigenlayer. The attractive properties of the SVs [69-70] are as follows:

Stability: Let $\mathbf{A}, \mathbf{B} \in R^{m \times n}$ and their corresponding SVs are $s_1, s_2, \dots, s_n$ and

$r_1, r_2, \dots, r_n$, a relation between them is established as:

$|s_i - r_i| < \|\mathbf{A} - \mathbf{B}\|_2, i = 1, 2, \dots, n$. This indicates that the SVs have very good

stability, i.e., when there is a little disturbance with the host audio, the variation

of its SV is not greater than 2-norm of disturbance matrix.

Proportionality: The SVs of $k\mathbf{A}$ are $|k|$ times of the SVs of $\mathbf{A}$.

Transpose: $\mathbf{A}$ and its transposed versions $\mathbf{A}^T$ have the same non-zero SVs.

Flipping: $\mathbf{A}$ and its flipped versions about vertical or horizontal axis have the same non-zero SVs.

Rotation: $\mathbf{A}$ and its rotated versions obtained through rotating $\mathbf{A}$ by an arbitrary angle have the same non-zero SVs.

Scaling: If $\mathbf{A} \in R^{m \times n}$, then its scaled version $\mathbf{A}_s$ has the SVs equal to $\sqrt{L_r L_c}$ times of the SVs of $\mathbf{A}$, where L_r and L_c are the scaling factors of rows and columns, respectively.

Actually, the abovementioned SVD properties are very suitable for developing robust watermarking approaches because the watermarked signal won't be easily corrupted by attacks such as rotation, noise addition and scaling. Furthermore, the embedded watermark is expected to be extracted effectively owing to these unique properties. SVD is a powerful and desirable transformation technique for robust watermarking and it has been widely applied in robust image watermarking. Recently, SVD has been applied to robust audio watermarking [56-63] by transforming the host audio signal into a matrix with non-negative scalar entries from the perspective of linear algebra.

2.9 Literature Review

A large number of audio watermarking techniques can be found in the literature. The most widely used watermarking techniques are echo hiding, SS, patchwork and content based methods. Echo hiding method [47, 71] embeds a watermark by adding an echo signal to increase robustness. Echo, defined as delay or offset, is imperceptible if the delay time is short. However, it is signal dependant or offset dependant and is not suitable for speech signal with frequent silent intervals. SS method is the mainstream audio watermarking technique which is weighted in the time (or frequency) domain by adopting PN or chaotic sequence to generate and spread the watermark [72-73]. In this scheme, computational complexity and synchronization are the main challenges. Patchwork method is implemented by adding and subtracting a constant value from two corresponding sample sets [74]. It is a typical watermarking method which takes advantage of the host signal to tradeoff between robustness and imperceptibility. Increasing the watermark strength within the constraints of masking threshold is a good way to find the balance point in this method. Content-based method explores dynamic and unique audio features for watermarking approaches [3, 75]. This scheme is signal dependant or adaptive by modifying audio frames or scale factor (SF). Besides,

HAS model is usually adopted in this technique.

Meanwhile, for the typical audio watermarking scheme, it can be divided into watermarking system with or without synchronization in a wide sense. The synchronization and self-synchronization technique have the ability to resist cropping, shifting, TSM, PSM and jittering attacks. For example, Lie *et al.* [32] proposed a time domain audio watermarking scheme based on HAS by exploiting the relation of the average of the absolute amplitude differences. However, this group amplitude quantization method obtains a relatively low detection ratio under different attacks. Meanwhile, the histogram based audio watermarking algorithms against TSM and cropping attacks are introduced in [76] and [42]. In [76], Xiang and Huang recommended a histogram based audio watermarking scheme in the time domain to withstand the desynchronization attacks as histogram and mean are invariant to TSM, PSM and jittering attacks. The multibit watermark is inserted with the control of histogram intensity. However, the data payload is very low. Besides, there is no security measure in this method as the watermark position is not shuffled in this method. In [42], Xiang *et al.* extended the histogram method to the DWT domain and scramble the watermark by a PN sequence to improve the security. However, the capacity of this approach is still very low as it is only 2 bps. As we know,

transform domain algorithm can improve robustness, thus most self-synchronization based audio watermarking method is performed in the transform domain. In [52], Wu *et al.* adopted a quantization index modulation (QIM) method to embed watermark and achieve self-synchronization with the use of the DWT localization property. Although the DWT technique reduces the computation time significantly, this technique suffers from TSM and amplitude scaling attacks due to the single coefficient quantization. Furthermore, in [50], Megias *et al.* proposed the recent self-synchronization audio watermarking algorithm by modifying the FFT amplitude. The time domain insertion of synchronization code is combined with the FFT domain informative watermark embedding. This scheme is very fast and can be applied in realtime.

In [54], another new transform, DFRST, is integrated with chaos technique in the audio watermarking algorithm. In this method, security is enhanced by using the chaotic sequence. DFRST properties are also exploited for audio watermarking and security issue. However, no synchronization code technique is employed. The imperceptibility of this scheme is also not very good and needs further improvement.

In the recent state-of-the-art publications on audio watermarking

technique, SVD related audio watermarking is a very hot topic and has been widely developed and extensively studied due to the superior advantage of SVD over other transforms. For instance, Abd El-Samie *et al.* [77] and Al-Nuaimy *et al.* [63] suggested an efficient SVD-based audio watermarking in the transform domain and used the chaotic sequence to shuffle the binary watermark to increase confidentiality. Furthermore, Al-Nuaimy *et al.* [63] extended the proposed SVD audio watermarking to be applied in bluetooth based systems and automatic speaker identification systems. However, from the reported results, the robustness needs further improvement. Besides, it will not be robust to TSM and amplitude scaling modification attacks as there is no built-in synchronization.

In [61], Lei *et al.* proposed a very robust SVD-DCT audio watermarking method. The audio watermarking is better than the selected SVD based method in terms of robustness and imperceptibility as claimed by the authors.

At the same time, Wang *et al.* [62] proposed a reduced SVD (RSVD) and distortion removal audio watermarking scheme. Rather than using the popular SVD watermarking method by modifying the SVs, this method adopts a different approach by taking advantage of the SV distortion, that is, RSVD. The $\mathbf{U}$ matrix is modified to embed the watermark bits. The audio fidelity is

preserved by a threshold-based distortion control. However, in this method, synchronization is not provided and there is no security measure.

Furthermore, there are some watermarking schemes based on chaos and HAS in the compressed domain in the literature too. For instance, a novel watermarking method introduced in [13] used the chaotic sequence and MPEG-1 psychoacoustic model to improve robustness in the frequency domain. Besides, this scheme has relatively high robustness, flexibility and supports multiple watermarking. However, listening test was not conducted to test the imperceptibility. There is also no performance comparison with other methods. Bassia *et al.* [30] developed an algorithm that modified each audio sample by adding a signal-dependent, low-pass shaped watermark signal. This amplitude modification watermarking scheme generates watermark by thresholding a chaotic map to form a bipolar sequence. A non-blind scheme can resist most of the attacks. However, it is not robust to TSM and synchronization attacks.

Finally, a description of perceptual audio watermarking schemes can be found in many other publications too [3, 7, 31]. In 1996, Boney [7] *et al.* put forward an algorithm to utilize the MPEG psychoacoustic model in order to obtain the necessary frequency-masking values to achieve audio transparency. The watermark is generated by permuting a PN sequence with a filter which

implements the HAS frequency masking characteristics. Thus these watermarks are signal dependent and different for other audio signals. This technique was further explored by Swanson *et al.* [3] which took advantage of perceptual coding techniques in order to embed the watermark efficiently and appropriately. Bassia and Pitas proposed an audio watermarking with the SS method in the time domain [30].

Chapter 3

A Perception-based Audio Watermarking Scheme in the Compressed Bitstream

3.1 Introduction

In the last few years, with the advancement in audio standards, audio techniques and features can be exploited to protect copyright with watermarking method. Therefore, multitudes of watermarking schemes are available for the protection of the audio formats like MP3, MPEG-2 and MPEG-4 AAC to prevent copyright pirating and malicious modifications. The widespread use of the compressed MPEG-4 SLS audio format on internet, the commercial importance of copyright protection and the potential loss of profits caused by illegal copying necessitate a proposal for the MPEG-4 SLS audio copyright protection. Therefore, for the newest standard released in 2006 by ISO/IEC, MPEG-4 scalable lossless audio (SLS) [78] coding, introduction of watermarking schemes for MPEG-4 scalable to lossless audio is very significant. Actually, embedding watermarks for MPEG-4 SLS audio in the compressed domain based on integer modified discrete cosine transform

(IntMDCT) is very feasible.

The goal of this chapter is to design and analyze a robust, feasible and applicable watermarking scheme based on chaos and HAS model to protect MPEG-4 SLS audio clip in the IntMDCT domain. Finding the right balance between robustness and transparency will be an important problem that has to be addressed in this chapter. The watermark should be embedded into the MPEG-4 SLS audio compressed bitstream and extracted directly. The chaotic watermark data or copyright information is embedded into the IntMDCT coefficients and detected from the coded bitstream directly. The modifications should be as small as possible to ensure inaudibility of the watermark. Moreover, the watermark should also be robust to various attacks like signal processing, and decoding/recoding attacks. Lastly, the proposed method should introduce some possibility of commercial realization and be tailored for a wide range of applications with the exploitation of MEG-4 SLS audio features, encoding and decoding process. All in all, the main contributions of the proposed watermarking framework are threefold. Firstly, it is compatible to the SLS bitstream without much audio quality distortion and is robust to MPEG-4 SLS audio compression and coding process. Secondly, it is based on psychoacoustic model and significant state to ensure imperceptibility. There is

no additional psychoacoustic model needed as the SLS perceptual model is used directly. Chaotic sequence rather than PN sequence is adopted to improve the security. The proposed method is simple and convenient to implement. Thirdly, it is applied in the IntMDCT domain and adaptive SS method is used. The watermark can be adaptively modified in a way to achieve a good tradeoff between robustness and transparency.

The rest of this chapter is organized as follows. A brief review of the recent watermarking schemes is introduced in Section 3.2. Section 3.3 describes MPEG-4 SLS structure and related characteristics. Psychoacoustics and the significant state applied in the proposed watermarking scheme are presented in Section 3.4. Section 3.5 introduces chaotic sequences in the watermarking field. The watermark embedding and extraction are given in Sections 3.6 and 3.7, respectively. Preliminary performance analysis is conducted in Section 3.8. Experimental results are discussed in Section 3.9 followed by the chapter summary in Section 3.10.

3.2 Related Work

Since most audio products are distributed in the compressed format, there are a lot of watermarking methods implemented in the compressed domain

especially for MPEG audio format in state-of-the-art publications. As the most successful and popular format, MP3 has attracted the most interest [5, 42, 79]. Some efforts are centered on embedding extra information into MP3 frames by modifying the SFs. This method is simple but affects the sound quality significantly. MP3Stego software, which is developed by Petitcolas *et al.* for MP3 audio data protection, inserts watermark information during MP3 compression process [5]. This method is not robust to the coding process attack and the hidden bits will be removed after MP3 compression attack. Moreover, the assigned bits will be changed after recompression. Another audio watermarking scheme proposed by Yang *et al.* [79] made use of the audio “linbits” features to be robust to MP3 coding attacks. Though this technique is simple and convenient, it is very sensitive to attacks and has a low degree of robustness. The recent developed watermarking scheme [42] suggested by Xiang *et al.* focused on TSM and compression attack by shaping the histogram to insert watermark. This data protection scheme solved the desynchronization problem with histogram-based algorithm. Moreover, this scheme can survive some attacks though it is not very robust to common signal processing attacks.

At the same time, there are some recent watermarking techniques based on MPEG-2 AAC audio [67, 80-82]. For example, watermarking prototype for

MPEG-2 AAC advocated in [82] was a solution to the pulse coding modulation (PCM) watermarking problem. The weak point of the method is that the overall performance of the system is not optimal and the bit error ratio (BER) is still very high. Besides, the watermark is not a blind detection system. Tachibana *et al.* [80] proposed a bitstream watermarking scheme for MPEG-2 AAC audio based on a two-dimensional pseudo random array and detected the watermark with the correlation method. It has lower robustness as there is no HAS model. At the same time, the algorithm still has much room for improvement because no listening test was performed to verify the imperceptibility. Another novel watermarking scheme for MPEG-compressed audio designed by Quan *et al.* [81] employed wet paper codes and inserted data directly by modifying the MPEG audio quantization process. The novel enhanced SS based watermarking scheme for AAC proposed by Cheng *et al.* was very fast for realtime application as it used the quantization indices directly [67]. The general scheme applied in the DCT domain is also suitable for other domain and other audio formats. This robust watermarking algorithm with low complexity adopts spectral filtering to reduce noise and improve the detection bitrate.

Apart from MP3 and MPEG-2 AAC data protection framework, the latest high data rate audio watermarking techniques are designed in the latest

publications [83-84] as well. In [83], Li *et al.* introduced three techniques to embed watermark adaptively based on Advanced Audio Zip (AAZ). The masking threshold is used in order to hide a watermark signal. Under the constraints and control of the masking threshold, the watermark is adaptively and transparently embedded. The system is highly dependent on the threshold which makes it too computationally intensive for realtime application. A similar data hiding scheme for audio property protection based on IntMDCT was suggested in [84] for high data rate audio such as MPEG-4 AAC. The data hiding technique using the IntMDCT time-frequency transforms is also applicable to other compressed audio data and spatial sound information. The disadvantage of this hiding technique is that the perceptual model is too simple to achieve higher data rate with less audio distortion. In other words, higher data rate has higher audible degradation which is not suitable for real application.

3.3 MPEG-4 SLS Structure

MPEG-4 SLS codec adopts the IntMDCT based lossless coding approach [85-86]. The input integer PCM format is converted losslessly into the frequency domain with IntMDCT. IntMDCT is a lossless integer to integer transform

which is similar to the normal MDCT transform. The IntMDCT spectral data are coded with a core MPEG-4 AAC layer and another complementary Lossless Enhanced (LLE) layer. Bit-plane coding method is utilized to produce the fine grain scalability needed for the lossless portion of the lossless bitstream. The error mapping process manages to preserve the probability distribution skew of the original IntMDCT coefficients. The resulting coefficients are then coded by the entropy coder used in the LLE layer to generate the core layer AAC bitstream.

Figure 3.1 illustrates the basic coding principle of the SLS encoder. Either bitplane Golomb coding (BPGC) or context-based arithmetic coding (CBAC) and Low energy mode coding (LEMC) can be used to code the residual spectrum. BPGC is adopted in SLS as the major arithmetic coding scheme. It uses a probability assignment rule that is derived from statistical properties (Laplacian distributed) of the residual spectrum in SLS. CBAC complements BPGC to further improve the coding efficiency. LEMC is adopted for coding signals from low energy regions to improve the coding efficiency of BPGC by further incorporating more sophisticated probability assignment rules. The low energy bit-planes will be coded at last using LEMC until it reaches the plane of the least significant bit (LSB) for all scale factor bands (sfb). Finally, the LLE

bitstream is multiplexed with the core AAC bitstream to produce the final SLS bitstream.

Figure 3.1: Block diagram of SLS encoder.

To this end, AAC core encoder can be taken as a quantization and coding process where the IntMDCT spectral data are first grouped into different sfbs, which are then quantized with different quantization step sizes and coded to produce the AAC bitstream. Assuming that an IntMDCT coefficient, $c(k)$, from a particular sfb, s , is quantized at the AAC core encoder, for $k = \{0, 1, \dots, N-1\}$ where N is the dimension of IntMDCT and generates an output $i(k)$, the following property holds

$$|thr[i(k)]| \leq |c(k)| < |thr[i(k)]| + \Delta[i(k)], k \in s, \quad (3.1)$$

where $thr[i(k)]$ is the next quantized value closer to zero with respect to $i(k)$ and is calculated via table lookup and linear interpolation to ensure the

deterministic behavior necessary for lossless coding. $i(k)$ is the quantized IntMDCT spectral data vector produced by the AAC quantizer, that is, $Q[c(k)] = i(k)$, where $Q(\bullet)$ is the non-uniform quantizer in the SLS core layer. $\Delta[i(k)]$ is the quantization step size for $i(k)$. If the SF of the band s to which $[0,1]$ belongs is denoted as $scale_factor(s)$, then $thr[i(k)]$ is computed

$$thr[i(k)] = \begin{cases} \text{sgn}[i(k)] \sqrt[4]{2^{scale_factor(s)}} (|i(k)| - C)^{4/3}, & i(k) \neq 0, \\ 0, & i(k) = 0, \end{cases} \quad (3.2)$$

$$\Delta[i(k)] = thr[|i(k)| + 1] - thr[|i(k)|]. \quad (3.3)$$

Here, $scale_factor(s)$ is the SF that determines the quantization step size for sfb s , $\text{sgn}[\bullet]$ is the sign operator and the rounding offset, C , is equal to 0.4054 in the standard.

Intuitively, given the SLS quantized value, $i(k)$, it is sufficient in most cases to use the minimum mean square error (MMSE) residual obtained by subtracting the IntMDCT coefficient $c(k)$ from its MMSE reconstruction, i.e., $\hat{c}(k) = E\{c(k) | i(k)\}$, where $E\{\bullet\}$ is the expectation operation. However, such an error mapping process will generally produce uniformly distributed error, $e(k)$, which is virtually incompressible. In order to fully utilize the statistical properties of $c(k)$ implied in Eq. (3.1), MPEG-4 SLS adopts a somewhat anti-intuitive and different approach in order to produce a residual signal $e(k)$ with

the smallest entropy, where the residual signal $e(k)$ is computed by

$$e(k) = \begin{cases} c(k) - \lfloor thr[i(k)] \rfloor, & i(k) \neq 0 \\ c(k), & i(k) = 0 \end{cases} \quad k = 1, \dots, 1024. \quad (3.4)$$

Here $\lfloor \bullet \rfloor: \mathbb{R} \rightarrow \mathbb{Z}$ is the flooring operation that rounds a floating point value to its nearest integer with smaller amplitude. The residual spectrum is then coded to generate the scalable LLE layer bitstream. In the final step of the encoding process, the output of LLE bitstream is multiplexed with the core AAC bitstream to produce the final SLS bitstream.

3.4 Psychoacoustic Model and Significant State

3.4.1 Psychoacoustics

Most recent audio coding standards are based on psychoacoustics which incorporate HAS properties and exploits the “undetectable” signal information. The undetectable information is detected using signal analysis by integrating into the coder several psychoacoustic properties such as absolute hearing thresholds (ATH), simultaneous masking, the spread of masking along the basilar membrane and temporal masking. Masking is a fundamental property of HAS and a basic element of MPEG-4 SLS audio. The masking phenomenon occurs when a sound is rendered imperceptible by any signal like a noise, a "masker", or unwanted sound of the same duration as the original sound. HAS

model can generate masking threshold which quantifies the largest amount of bearable distortion at each point so that no audible artifacts are introduced by the quantization of the time-frequency parameters. Integrating psychoacoustic notions with basic properties of signal quantization and coding can lead to the theory of transparent audio watermarking.

Actually, the frequency-masking model can be used to obtain the spectral characteristics of a watermark based on HAS inaudible information. With psychoacoustic signal analysis which can estimate the signal masking power based on psychoacoustic principles, perceptual distortion control is obtained. Signal, mask and noise energy plot for one frame in an SLS test sequence excerpt [87] is illustrated in Figure 3.2, where the excerpt is coded by MPEG-4 SLS reference codec at 128kbps. It can be observed that the watermark material can be transparently embedded when the quantization and distortion is controlled to be lower than the masking threshold.

Figure 3.2: Signal, mask energy and allowed distortion in one frame.

3.4.2 Significant State

The probability distribution of bit-plane symbols is usually correlated with their frequency location and the significant state of the adjacent spectral lines [85].

The amplitudes of adjacent spectral lines are correlated due to the leakage of the IntMDCT filterbank. In addition, the amplitude of the IntMDCT spectrum is also highly correlated with the quantization interval of SLS core quantizer if it is present. The significant states of the adjacent spectral lines are designed to capture the correlations among the amplitude of the current IntMDCT spectral data, those of the adjacent IntMDCT spectral lines and the quantization interval of the AAC core quantizer. Besides, a vector, $sig_cx(k, j)$, is defined to represent the significant states for adjacent IntMDCT spectral line of bit-plane

symbol, where $k = 1, \dots, N$ and j is the bit-plane, that is,

$$sig_cx(k, j) = \{sig_cx(k-2, j), sig_cx(k-1, j), sig_cx(k+1, j), sig_cx(k+2, j)\}, \quad (3.5)$$

where the significant state, $sig(k, j)$, is denoted as

$$sig(k, j) = \begin{cases} 0, & \hat{c}_j(k) = 0 \\ 1, & \hat{c}_j(k) \neq 0 \end{cases}. \quad (3.6)$$

and $sig(k, j)$ is taken as 0 if k is outside the IntMDCT spectrum. Here $\hat{c}_j(k)$ is

the partial reconstruction for $c(k)$ up to bitplane j ,

$$\hat{c}_j(k) = \sum_{f=j+1}^{M-1} b(k, f)2^f + thr[i(k)], \quad (3.7)$$

where $b(k, f) \in \{0, 1\}$ is a bit-plane symbol which is either 0 or 1. The

$sig_cx(k, j)$ context is only used in coding bit-plane symbols from

insignificant IntMDCT spectral lines. IntMDCT coefficient, $c(k)$, is

insignificant when $i(k)$ is equal to zero in the core layer. For those from

significant ones (i.e. $i(k) \neq 0$ or is near the DC components), the sig_core

context is further introduced. The value of sig_core context is given by

$$sig_core(k) = \begin{cases} 0, & c(k) \text{ is from an insignificant sfb,} \\ 1, & c(k) \text{ is from a significant sfb,} \end{cases} \quad (3.8)$$

Furthermore, for $sig_core(k) = 1$, the following rule holds

$$0 \leq e(i) \leq \Delta(i(k)), \quad (3.9)$$

Actually, the proposed watermarking algorithm explores both the

perceptual model and significant state theories. In this method, the selected coefficients under the constraints of perceptual model and significant state can be modified slightly and adaptively to embed watermark in a way that does not produce any perceived effect.

3.5 Chaotic Watermarking

As chaotic functions can generate similar pseudo random sequences with almost identical performance, the simplest chaotic map, logistic map, is chosen to produce the chaotic sequence that encrypts the binary watermark because it is the simplest to implement. As the sequence obtained by the map is composed of real numbers, the output chaotic sequence is quantized into binary stream by the following rule:

$$\beta(n) = \begin{cases} 1, & x(n) \geq 0.5 \\ 0, & x(n) < 0.5 \end{cases} \quad (3.10)$$

Finally, the watermark to be inserted into the original audio signal in this scheme is generated by

$$w(n) = \sum_{n=1}^{N_w} b(n) \oplus \beta(n), \quad (3.11)$$

where N_w is the length of the watermark, $b(n)$ is binary data which can be any digital signature for copyright or other purpose. All types of digital data (text, images etc) that are represented using binary numbers in the computer can be

used as $b(n)$. $\oplus$ denotes the Exclusive-OR operation.

3.6 Watermark Embedding

Generally, in the proposed watermark embedder, a SS based approach is implemented in the compressed domain. The binary digital signature is spread and encrypted by a chaotic sequence with a secret key. After the non-uniform quantization in the core layer, the watermark is inserted into the quantized IntMDCT coefficients, which is usually with different quantization steps in different sfbs to shape the quantization noise so that it can be best masked. In the proposed scheme, adaptive watermark embedding is employed to shape the local watermark strength under the control of the psychoacoustic model. The allowed distortion control is also adopted to improve the system performance. Finally, the multiplexer produces the bitstream including data and side information to complete the watermarking process. The proposed watermark embedding system is summarized in Figure 3.3.

In order to embed the watermark, the original input audio is segmented into different frames and each frame is decomposed into different subbands. The IntMDCT transform in the SLS encoder is performed to obtain the transform coefficients. Besides, in this watermarking scheme, the transformed

coefficients are grouped into sfbs which are generally fixed at 49 for a sampling rate of 48 kHz.

Figure 3.3: Proposed watermark embedding diagram.

In order to enhance the audio fidelity, watermark embedded into the original audio is conditional on the just noticeable distortion constraints and perceptual shaping. The watermark is chosen to be added to the perceptually significant scalefactor bands of the spectrum in such a way that the watermark is not easily and illegally removed from the original audio without causing too much audio distortion. The psychoacoustic model is applied on those subbands to determine the masking thresholds for each subband. The masking threshold is the allowed distortion for each IntMDCT coefficient. Here it is assumed that similar to perceptual audio coding, if the distortion created in the scalable

coding is below the mask, then transparent quality can be achieved. Actually, the allowed distortion (masking threshold), $M(s)$, can be obtained by MPEG-4 SLS psychoacoustic model [88]. Therefore, in the proposed watermarking scheme, $M(s)$ for each sfb s in terms of dB is computed as:

$$M(s) = \begin{cases} \frac{E(s)}{SMR(s)}, & E(s) > 70\text{dB} \text{ and } SMR(s) > 1 \\ E(s), & E(s) > 70\text{dB} \text{ and } SMR(s) \leq 1. \\ E(s) \times 1.1, & E(s) \leq 70\text{dB} \end{cases} \quad (3.12)$$

Note that Eq.(3.12) is the default mask implementation in the MPEG-4 SLS, where $s(0 \leq s \leq S)$ and S is the total number of sfb, SMR is the signal to mask ratio. $SMR(s)$ for each sfb s is calculated using psychoacoustic model in the MPEG-4 SLS reference model encoder. $E(s)$ is the signal energy (in terms of dB) for sfb s and is computed by

$$E(s) = \sum_{k=O[s]}^{O[s+1]-1} c^2(k), \quad (3.13)$$

where $O(s)$ is the starting index of spectrum coefficient for sfb s and $c(k)$ is the IntMDCT coefficient. k is the index of the coefficients. The watermark energy is compared with $M(s)$, if the watermark energy is far below $M(s)$, the introduced watermark will be imperceptible. As a result, there will be watermark embedded. Otherwise, if the watermark energy happens to be higher than $M(s)$ in some spectral range, or they are at a similar level, there will be no

watermark embedded. This adherence to the psychoacoustic model makes it very difficult to “hear” the hidden data and maintains the transparency of the scheme at the same time. Therefore, the disadvantage of not embedding the watermark in every frame is that special design is needed to adhere to psychoacoustic model to make the watermark insertion imperceptible.

In an SLS codec, the bit-plane coding is performed in a sequential order. The plane of the most significant bit (MSB) for spectral data from the lowest sfb to the highest sfb is coded first. To be more specific, the illustration of the watermarking scheme for MPEG-4 SLS audio exploring the perceptual model and significant state context information is shown in Figure 3.4, where n/w means watermark will be embedded adaptively. For each sfb s , if it is declared as significant and under the constraints of allowed distortion, the watermark can be embedded. Otherwise, there will be no watermark data embedded in this sfb. The significant state and allowed distortion determine the region to be embedded and how much can be embedded. As a result, this selection makes the watermark embedding adaptive in accordance to the contents of the audio and MPEG-4 SLS psychoacoustic model. To further enhance the adaptiveness of the proposed algorithm, the embedding mechanism also employs an adaptive method to determine watermark embedding intensity.

Suppose that the watermark data to be embedded is represented by $w(k)$ and the host signal to be embedded is represented by $S(k)$. The method of embedding watermark into the host signal is expressed as

$$S'(k) = S(k) + a(k) \times w(k), \quad (3.14)$$

where $S'(k)$ is the watermarked signal. Generally, if $S'(k)$ is the watermarked signal and $S(k)$ is the original signal, then the SNR (R_{SNR}) is usually defined as:

$$R_{SNR} = 10 \log_{10} \frac{\sum_k S^2(k)}{\sum_k (S'(k) - S(k))^2}, \quad (3.15)$$

Figure 3.4: Watermark embedding with perceptual model.

Substitute Eq. (3.16) to Eq. (3.15), then the scaling factor, $a(k)$, can be obtained

by

$$a(k) = \sqrt{\left(\sum_k S^2(k) \right) 10^{R_{SNR}/10}}. \quad (3.16)$$

Thus the adopted SS method can be expressed as

$$S'(k) = S(k) + \sqrt{\left(\sum_k S^2(k)\right) 10^{R_{SNR}/10}} w(k). \quad (3.17)$$

For a predefined threshold R_{SNR} , the scaling parameter, $a(k)$, can be adaptively adjusted by audio signals to control watermarking strength. Consequently, the watermarking system can be adjusted by changing R_{SNR} , which is an advantage of the proposed method. Different R_{SNR} is chosen for optimizing the tradeoff between robustness and imperceptibility of the proposed scheme.

3.7 Watermark Extraction

Figure 3.5 describes the watermark extraction process. After SLS watermarked bitstream is parsed, it is entropy decoded to recover $S'(k)$. As the SS method is adopted in the proposed scheme, it is a non-blind method, thus the extraction and detection procedure needs the original signal. The watermark extraction is the reverse process of the watermark embedding process. Specifically, the watermark extraction rule is:

$$w'(k) = \frac{S'(k) - S(k)}{a(k)}. \quad (3.18)$$

The obtained $w'(k)$ are de-spread with the same quantized chaotic sequence in Eq. (3.11) in order to detect and recover the hidden binary data.

The inserted binary data can be obtained by:

$$b'(k) = \sum_{k=1}^{N_w} w'(k) \oplus \beta(k). \quad (3.19)$$

Figure 3.5: Diagram of watermark message extraction.

In order to detect the existence of the inserted binary data, a similarity measure is used. The similarity is defined as follows:

$$Sim(b, b') = \frac{\sum_{k=1}^{N_w} b(k) \times b'(k)}{\sqrt{\sum_{k=1}^{N_w} b(k)^2} \sqrt{\sum_{k=1}^{N_w} b'(k)^2}}. \quad (3.20)$$

If $Sim(b, b') > \delta$, where δ is the decision threshold, then the watermark exists and the watermark message is output, otherwise the watermark is considered to be absent.

3.8 Performance Analysis

3.8.1 Error Analysis

Two types of errors may occur while searching for the watermark sequence:

false positive error (i.e., false watermark detection) and false negative error (i.e., failure to detect an existing watermark). It is rather difficult to give an

exact probabilistic model of false positive and negative errors. Here, a simplified model based on binomial probability distribution is utilized to provide an analysis to estimate the probabilities of false positive and negative error for the proposed technique. To determine the degree of similarities between the original and extracted watermark signals, the normalized cross-correlation coefficient (NC) is used, which is computed as

$$NC(w, w') = \frac{\sum_i w(i)w'(i)}{\sqrt{\sum_i w^2(i)} \sqrt{\sum_i w'^2(i)}}. \quad (3.21)$$

1) False Positive Error

The probability of false watermark detection is defined as:

$$P_{fp} = P\{NC(w, w') \geq T_p \mid \text{no watermark}\}. \quad (3.22)$$

where $P\{A|B\}$ is the probability of event A given event B , T_p is a threshold.

Since $w(i)$ and $w'(i)$ are either 0 or 1, correspondingly, $w^2(i)$ and $w'^2(i)$ are either 0 or 1. NC can be rewritten as

$$NC(w, w') = \frac{\sum_i w(i)w'(i)}{\sqrt{\sum_i w^2(i)} \sqrt{\sum_i w'^2(i)}} \geq \frac{\sum_i w(i)w'(i)}{M}. \quad (3.23)$$

where M is the length of the binary watermark in 1-D format. To decide whether the watermark is present or not, NC is compared with a threshold. If NC is larger than a minimum threshold, then the watermark is present. Suppose

n bits of errors occur when extracting watermark data, then $m = M - n$ bits are identical to the original watermark data. It is easy to derive:

$$\sum_i w(i)w'(i) = m - n = M - 2n. \quad (3.24)$$

If Eq. (3.23) and Eq. (3.24) are substituted into Eq. (3.22), the probability of false positive error can be further derived as

$$\begin{aligned} P_{fp} &= P\{NC(w, w') \geq T_\rho \mid \text{no watermark}\} \\ &= \left\{ \frac{\sum_i w(i)w'(i)}{M} \geq T_\rho \mid \text{no watermark} \right\} \\ &= \left\{ \frac{M - 2n}{M} \geq T_\rho \mid \text{no watermark} \right\}. \end{aligned} \quad (3.25)$$

When $n \leq M(1 - T_\rho)/2$, the false positive error will occur, which is calculated as:

$$\begin{aligned} P_{fp} &= \sum_{n=0}^{\lceil M(1-T_\rho)/2 \rceil} P\left\{ \sum_i w(i)w'(i) = M - 2n \mid \text{no watermark} \right\} \\ &= \sum_{n=0}^{\lceil M(1-T_\rho)/2 \rceil} \binom{M}{n} p^n (1-p)^{M-n}, \end{aligned} \quad (3.26)$$

where $\binom{M}{n} = \frac{M!}{n!(M-n)!}$, p is the probability of error in BER when some

watermark extraction is performed, which is defined as

$$BER = \frac{1}{M} \sum_{i=1}^M w(i) \oplus w'(i). \quad (3.27)$$

where $\oplus$ is the exclusive or (*XOR*) operator.

Since watermark values are either 0 or 1, thus $p = 0.5$, then

$$P_{fp} = \sum_{n=0}^{\lceil M(1-T_p)/2 \rceil} \binom{M}{n} 0.5^M. \quad (3.28)$$

2) False Negative Error

False negative error is the probability of declaring a watermarked audio as an unwatermarked one. Less false negative errors imply a better watermarking scheme. Similarly, the probability of false negative error can be determined as:

$$\begin{aligned} P_{fn} &= P\{NC(w, w') < T_p \mid \text{no watermark}\} \\ &= \sum_{n=\lceil M(1-T_p)/2 \rceil+1}^M \binom{M}{n} p^n (1-p)^{M-n}, \end{aligned} \quad (3.29)$$

where p is the BER of the extracted watermark.

In the proposed scheme, the false positive and negative rate is nonlinear with respect to the watermark length. From the above analysis, when the watermark length is 1024 in the proposed scheme, the probability of the false positive and negative error is approximately zero, which indicates the good performance of this scheme.

3.8.2 Security Analysis

The security of the embedded watermark should be considered in a watermarking system for copyright protection due to the fact that, if the positions of the embedded watermark are guessed by an attacker successfully, the embedded watermark can be easily detected and altered. This section describes the security analysis as security is an important component of a

secure watermarking scheme. To improve the confidentiality, the key space should be large enough to make the attacks by brute force infeasible. In the proposed scheme, key or initial condition is represented using a floating-point number between 0 and 1. In a single-precision floating-point number, 24 bits are used to represent the significant number while 8 bits are used to represent the exponent. Hence the exponent is fixed while the significant number can be varied. Hence the total number of possible keys or initial conditions is 2 to the power of 24 which is more than 16 million. This number can also be greatly increased if so desired through the use of double precision floating point numbers. In fact, the security of a watermarking scheme usually depends on keys rather than the privacy of the scheme. In this proposed audio watermarking scheme, key, K_1 , is used to generate the logistic sequence for enhancing the security of the proposed scheme. Therefore, the size of the key value space affects the confidentiality of the proposed scheme. The key value space K_1 is calculated as follows: Suppose $K_1 = \{0 < K_1(i) < 1 \mid i = 1, 2, \dots, N_1\}$, N_1 is an integer which is used to generate the logistic sequence, hence it should be large enough to produce chaotic sequence, $Y = \{y(i, j) \mid i = 1, 2, \dots, N_1, j = 1, 2, \dots, N_2\}$, where N_1 denotes the number of chaotic sequences and N_2 represents the length of each chaotic sequence.

When $K_1' = \{0 < K_1(i) + d < 1 \mid i = 1, 2, \dots, N_1\}$, it will generate another group of chaotic sequences $Y' = \{y'(i, j) \mid i = 1, 2, \dots, N_1, j = 1, 2, \dots, N_2\}$. The function $f_k = SS(d)$ is used to test key space of K_1 as follows:

$$f_k = SS(d) = \frac{\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} |y(i, j) - y'(i, j)|}{N_1 \times N_2}. \quad (3.30)$$

Figure 3.6 plots the function of $f_k = SS(d)$. It can be seen that f is equal to 0 when $d_0 = 10^{-17}$ in this method. Thus the key space of K_1 is $1/d_0 = 10^{17}$, which means that there is enough key space to guarantee high confidentiality of the proposed watermarking system. Based on this security analysis, it can be concluded that the embedded watermarks are secure to attackers who try to exhaustively or statistically detect and read them. All in all, the proposed scheme with such a long key is enough for reliability and practical usage.

Figure 3.6: Key space under different initial value difference.

3.9 Experimental Tests and Results

The main audio attacking and editing tools for watermarking employed in the experiments are GoldWave (v5.25), CoolEditPro (v2.1) and EAQUAL v0.1.3 alpha (for measuring the objective difference grade (ODG) value). Some signal processing functions are implemented in MATLAB. A total of 15 MPEG-4 standard test sequences (16 bits/sample, 48 kHz and WAV format) listed in Table 3.1 are utilized to analyze the watermarking algorithm's performance.

Table 3.1 MPEG-4 audio test sequences (16bits stereo).

Index	Item(.wav)	Index	Item(.wav)	Index	Item(.wav)
1	avemaria	6	cymbal	11	haffner
2	blackandtan	7	dcymbals	12	mfv
3	broadway	8	etude	13	unfo
4	cherokee	9	flute	14	violin
5	clarinet	10	fouronsix	15	waltz

3.9.1 Time and Spectrum Waves

Tests were run to evaluate whether the watermarked data in MPEG-4 SLS audio can be detected through more qualitative methods and the characteristics of the changes. This research analyzes the sound waveforms in the time and frequency domains. The original SLS audio waveforms are compared with the watermarked audio containing the hidden data. The resulting waveforms in the time domain and their spectrum plots in the frequency domain are shown in Figures 3.7 and 3.8, respectively. Both analyses do not show any distinguishable differences between the original audio and the watermarked

audio. However, there seems to be a slight amount of signal loss and a slight increase in distortion due to the watermark insertion but the overall distortion is acceptable.

Figure 3.7: Original, watermarked and residue audio waveform in time domain.

Figure 3.8: Spectrogram of original, watermarked and residue audio.

3.9.2 Watermark Robustness

In the experiment, the watermark robustness is evaluated in terms of BER after the watermark extraction. The robustness performance after the following common signal processing manipulations are evaluated and tested: (1) MP3, AAC and SLS compression of watermarked PCM audio attacks; (2) low pass filtering (LPF) (with cut-off frequency of 44.1kHz); (3) requantization test (16bits-8bits-16bits); (4) delay (500ms, 10%); (5) echo addition (500ms, 10%); (6) resampling (44.1kHz-22.05 kHz-44.1 kHz); (7) equalization; (8) noise addition (add white noise to 20dB SNR). The BER results after these common signal processing attacks are shown in Table 3.2. It can be seen from the table that the BER results of the proposed scheme are quite satisfactory.

Table 3.2 BER of watermark extraction under attacks.

Item(.wav)	1	2	3	4	5	6
no manipulation	0	0	0	0	0	0
MP3@128kbps	0	0	0	0	0	0
MP3@64kbps	0.002	0.002	0.001	0.04	0.05	0
AAC@128kbps	0	0	0	0	0	0
AAC@64kbps	0	0	0	0	0.003	0
SLS@128kbps	0	0	0	0	0	0
SLS@96kbps	0	0	0	0	0	0
SLS@64kbps	0	0	0	0	0	0
LPF	0.001	0.001	0	0	0.001	0
requantization	0	0	0	0	0	0
delay	0.005	0.005	0	0.007	0.008	0
echo addition	0	0	0	0	0	0
resampling	0.021	0.022	0.02	0.036	0.048	0
equalization	0	0	0	0	0	0
noise addition	0	0	0	0	0	0

3.9.3 Objective Evaluation Test

ITU-R Recommendation standard BS.1387 [89] is used to measure the audible distortion of the embedding process and provides the definition for perceptual evaluation of audio quality (PEAQ). PEAQ is used to rate the quality of the audio signal after modifications performed by a watermark embedder. EAQUAL tool [90] implements PEAQ and computes the ODG and noise mask ratio (NMR) by means of a neural network. The computed ODG values are in the range of $[-4, 0]$, where -4 is the worst (very annoying) and 0 is the best (imperceptible). Specifically, the embedding distortion caused by the proposed watermarking algorithm expressed with the ODG value is very close to the threshold of becoming audible as shown in Table 3.3. The relatively high NMR indicates the robustness of the watermarking scheme. The main reason of having satisfactory ODG is due to the fact that only a small part of samples are modified for watermarking and only small modifications of the magnitudes are being made.

Table 3.3 Objective Evaluation Test Results.

Item(.wav)	ODG	NMR	Item(.wav)	ODG	NMR
1	-0.10	-30.49	8	-0.12	-29.92
2	-0.11	-30.18	9	-0.11	-30.15
3	-0.09	-31.67	10	-0.10	-30.87
4	-0.14	-28.22	11	-0.14	-28.94
5	-0.15	-26.78	12	-0.11	-32.11
6	-0.08	-34.23	13	-0.09	-33.02
7	-0.11	-30.28	14	-0.13	-29.67

3.9.4 Subjective Listening Test

In the subjective listening tests, the subjects are asked to discern the watermarked and host audio clips. The evaluation of the subjective quality of the watermarked material was done by a MUSHRA listening test, which stands for “Multi Stimulus test with Hidden Reference and Anchors” and is described in ITU-BS.1534 [91]. The watermarked signal is graded with respect to the host signal according to a five-grade impairment scale defined in ITU-R BS.562, called subjective difference grade (SDG), which is the differences between the subtraction between the subjective ratings given separately to the host and watermarked signals. The test sequences have been used extensively for assessment of subjective sound quality in the MPEG-4 audio development process. To give an anchor for comparison, the quality of the watermarked items is compared with the quality of the original items. Besides, a total of ten subjects who are both experienced and familiar with the set of test items are invited to perform the test. The results of the listening test are shown in Figure 3.9. As can be seen from the above results, the quality degradation of the bitstream watermarking system is very small for the vast majority of the test items. For all items, the small variations of the signals indicate that there is no significant distortion introduced by this scheme.

Figure 3.9: Subjective listening test results.

3.9.5 Watermark Payload

Table 3.4 shows the data payload results of all the test sequences. As different test sequences have different features which determine how much watermark data can be embedded, the data payload differs accordingly.

Table 3.4 Watermark Payload for standard test sequences.

Item (.wav)	Payload (bps)	Item (.wav)	Payload (bps)	Item (.wav)	Payload (bps)
1	4	6	9	11	4
2	5	7	10	12	2
3	8	8	3	13	5
4	5	9	3	14	4
5	3	10	5	15	6

3.10 Chapter Summary

In this chapter, the perception and chaos based audio watermarking scheme in the compressed bitstream is proposed to provide a feasible and effective audio copyright protection scheme for MPEG-4 SLS audio signal. Performance

analysis shows good property of the watermarking system. Several objective and subjective tests are carried out to analyze the performance of the proposed scheme for various signal manipulations. The experimental results show that the proposed scheme is inaudible and is robust against common signal processing attacks.

Chapter 4

Blind and Robust Audio Watermarking Scheme Based on SVD-DCT

4.1 Introduction

It is a well-known fact that watermarks are generally not placed in perceptually insignificant regions of the audio signal because many common signal and geometric processing affect these components [92]. Hence, the watermark information is embedded only in the SVs which correspond to the energy of the most perceptually significant regions in the original audio. Robust watermarking methods for audio often combine short-time Fourier transform (STFT), DCT, discrete sine transform (DST) or DWT with SVD [26, 56, 58, 77, 93]. However, in these methods, SVD is only an auxiliary tool to modify the transform coefficients of STFT, DCT, DST or DWT. Effectively, such methods are essentially STFT, DCT, DST or DWT based in that these transforms are performed on the cover object directly. Therefore, the superior characteristics of SVD for robust watermarking are not fully exploited. The difficulty of applying SVD to the host audio directly is due to the fact that SVD

transformed values are very sensitive and difficult to be manipulated by other transforms.

In this chapter, a blind and robust audio watermarking scheme is proposed to capitalize on the combined features of the SVD, DCT and synchronization code technique. As a slight distortion in the largest coefficients will not cause much effect and can hold against general signal processing attacks, it would be sufficient to embed the watermark with only a slight modification. Watermark detection is efficient and blind in the sense that only the secret key for encrypting the embedded locations but not the original audio is required. Unlike the traditional SVD based watermarking schemes where the watermark bits are embedded directly in the SVs, the proposed scheme is based on bit embedding in the high-frequency DCT coefficients of the block of SVs obtained by SVD transformed audio sub-blocks. To improve the fidelity and the perceptual quality of the watermarked audio, the adaptive frequency mask which takes into account the HAS [3, 94] to control the strength of the embedded local watermark is also implemented.

The remainder of this chapter is organized as follows. In Section 4.2, related work of typical SVD watermarking algorithms are introduced. Some basic SVD principles are given in Section 4.3. In Section 4.4, the proposed

robust watermarking method is described in detail followed by a brief description of the watermark extraction in Section 4.5. Section 4.6 discusses the error probability issues. The experimental results and performance analysis are provided in Section 4.7. Finally, Section 4.8 concludes the chapter.

4.2 Related Work

For robust SVD based audio watermarking scheme, most algorithms are non-blind or semi-blind [56, 58, 77]. The typical schemes amongst them are the semi-blind watermarking method based on STFT [56] and non-blind watermarking method adaptable to different domains in [77]. For instance, in [56], Ozer *et al.* first proposed a non-oblivious robust watermarking based on SVD of the spectrogram of the signal. Though the audio watermarking technique is robust to some common signal processing attacks, it was a semi-blind technique and did not adopt any synchronization technique. Therefore, its robustness performance to synchronization attacks such as Flipsample, Copysample and Cutsample attacks is relatively poor and unsatisfactory.

An efficient SVD based algorithm for digital audio watermarking is presented in [77], which is domain adaptive as it can be applied in DCT, DWT and DST domains according to different requirements. Moreover, the proposed segment-by-segment implementation method has enhanced detectability

compared to the simple implementation on the whole host audio signal directly. Security is increased with the use of chaotic encrypted watermarks. However, this permutation and SS based audio watermarking scheme is not blind as the original signal and other information are needed in the watermark detection process.

The recent SVD based blind audio watermarking method proposed by Bhat K. *et al.* [58] is based on QIM in the wavelet domain using the synchronization code technique. This quantization technique is found to be high in capacity and robust to a lot of attacks. However, the imperceptibility is not very good as shown by its SNR and ODG results. The SNR results are just slightly above 20 dB. Furthermore, results on the detailed synchronization attacks such as jittering, amplitude variation and TSM are not reported in this scheme.

There are other SVD based audio watermarking schemes applied to the other domains in the literature for different purposes [60]. For example, Zezula and Misurec proposed an audio watermarking algorithm in the modulated complex lapped transform (MCLT) domain based on SVD [60].

Actually, compared to image and video signals, audio signals are represented by fewer samples per time interval. Moreover, an additional

problem in audio watermarking is that inaudibility is much more difficult to achieve than invisibility in image watermarking [3]. Therefore, the image watermarking schemes based on SVD have gained more attention and interest and there are a lot of SVD and SVD-DCT based image watermarking algorithms [95-99]. For example, in [93], the semi-blind pure SVD based watermarking scheme is proposed by Liu and Tan in 2002. In this scheme, the SVD matrices of the reference watermark have been found to be easily discredited so this scheme cannot be used for copyright protection [100-101]. There are also some blind watermarking methods which embed watermarks in SVs of the transformed coefficients [95]. In both [96] and [97], the intrinsic property of orthogonality of the SVD matrices is altered. As a result, the extracted watermark cannot be guaranteed to be correct even if no attacks are performed.

4.3 Watermark Embedding

4.3.1 Synchronization Code

The synchronization code [41-42, 48, 51, 94] is used to locate the position of hidden informative bits, thus resisting the cropping and shifting attacks. Suppose $\{Syn(k)\}$ is an original synchronization code and $\{Seq(k)\}$ is an unknown sequence with the same length. If the number of different bits

between $\{Syn(k)\}$ and $\{Seq(k)\}$, when compared bit by bit, is less than or equal to a predefined threshold, T , the sequence $\{Seq(k)\}$ will be determined as the synchronization code. The advantage of embedding the synchronization code in the time domain is the low cost incurred in searching. Therefore, synchronization code is embedded in the time domain to reduce the computation cost. Before embedding, the synchronization code $\{Syn(k)\}$ should be arranged into a binary data sequence.

In this scheme, the host signal is partitioned into two portions for synchronization code insertion and watermark embedding. Synchronization insertion involves splitting the audio host signal into proper segments according to the length of $\{Syn(k)\}$. This algorithm utilizes a chaotic sequence as the synchronization code in front of the watermark to locate the position where the watermark is to be embedded. The logistic chaotic sequence with initial value in the interval $[0,1]$ is selected to generate the synchronization code and denoted as:

$$y(k+1) = g(y(k)) = \lambda \times y(k) \times (1 - y(k)). \quad (4.1)$$

where $3.57 < \lambda \leq 4$, $y(k)$ is mapped into the synchronization sequence $\{Syn(k) | k = 1, \dots, L_{syn}\}$ with the following rule:

$$Syn(k) = \begin{cases} 1, & \text{if } y(k) > 0.5 \\ 0, & \text{otherwise} \end{cases}. \quad (4.2)$$

The synchronization code insertion part is cut into L_{syn} audio segments and each audio segment has P samples denoted as:

$$SA(k) = A(k \times P + u), 1 \leq k \leq L_{syn}, 1 \leq u \leq P. \quad (4.3)$$

Then each bit of the synchronization code is embedded into each $SA(k)$ as follows:

$$SA'(k) = \begin{cases} \text{round}(\frac{SA(k)}{Q}) \times Q, & \text{if } Syn(k) = 1 \\ \text{floor}(\frac{SA(k)}{Q}) \times Q + \frac{Q}{2}, & \text{if } Syn(k) = 0 \end{cases}, \quad (4.4)$$

where Q denotes the embedding strength, $\text{round}(\bullet)$ means rounding to the nearest integer, $\text{floor}(\bullet)$ is rounding to minus infinity.

After embedding, the embedded and attacked signal $SA''(k)$ is also split into L_{syn} segments, and then the synchronization code is extracted by the following rule:

$$Syn'(k) = \begin{cases} 0, & \text{if } \frac{Q}{4} \leq \text{mod}(SA''(k), Q) < \frac{3Q}{4} \\ 1, & \text{otherwise} \end{cases}, \quad (4.5)$$

where $\text{mod}(\bullet)$ denotes modulus after division.

4.3.2 Embedding Method

The combined features of SVD and DCT are exploited in the proposed watermarking algorithm. The detailed steps of the watermark embedding procedure are shown in Figure 4.1. The watermark embedding algorithm proposed here hides several bits of the watermark in every SVD-DCT block in

the randomly permuted positions selected on a pseudo-random basis. At first, the SVD-DCT block is acquired by the following steps:

Figure 4.1: Diagram of watermark embedding process.

Step 1: Rearrange the 1-D audio signal into a 2-D signal, $H(m_1, n_1)$.

Step2: Partition the 2-D host signal into non-overlapping subblocks, $H_j (j=1, 2, \dots, M_1 \times N_1)$ in the row-column order. $M_1 \times N_1$ is the total number of blocks in the 2-D host audio signal.

Step 3: For each block, $H_j (j=1, 2, \dots, M_1 \times N_1)$, apply 8×8 block based SVD to get U_j , S_j and V_j , respectively,

$$H_j(r, c) = U_j(r, c)S_j(r, c)V_j(r, c) \quad (1 \leq r \leq M_1, 1 \leq c \leq N_1). \quad (4.6)$$

As we know, the SVs represent the energy distribution of the watermark in different SVD layers and most energy is concentrated in the bigger SVs. The first SVD coefficient, $S_j(1, 1)$ in every block is grouped together to form a new matrix, DC , which is of size $M_1 \times N_1$,

$$DC(r, c) = S_j(1, 1) \quad (1 \leq r \leq M_1, 1 \leq c \leq N_1). \quad (4.7)$$

Step 4: The DC matrix is in turn sub-divided into 4×4 subblocks and the DCT is then applied to each subblock. Hence after the SVD-DCT transform, a 4×4 SVD-DCT coefficient block is obtained from the cover signal. A SVD-DCT block in the zigzag order is shown in Figure 4.2, where the shaded positions are potential locations for watermark embedding.

1	2	6	7
3	5	8	13
4	9	12	14
10	11	15	16

Figure 4.2: Sketch of SVD-DCT macroblock.

Then the watermark is embedded into the host audio through the following steps:

Step 1: Select the coefficients pairs to be modified in the potential positions of the SVD-DCT block.

Step 2: Compute the frequency mask. To determine the level of tolerance against distortions caused by the watermark embedding, a linear model for frequency mask is generated as follows

$$mask = \alpha \times edge + \beta, \quad (4.8)$$

where the parameter, α , is for controlling the local embedding strength of the watermark and β is the base strength of the watermark which is needed to resist the effect of the rounding and truncation operations in the spatial domain. In the

experiment, α and β are set to 0.125 and 4, respectively. Since blocks with edge characteristics often have a lot of frequency components, the parameter *edge* is introduced to reduce artifacts: *edge* is the sum of the absolute values of the DCT-coefficients of indexes 9-16 that represent the higher DCT frequencies as marked in Figure 4.2. High values in these components in the blocks of the first SVs indicate that the macroblocks have edge characteristics. In fact, the parameter *edge* is formulated as follows

$$edge = \sum_{i=9}^{16} |F(i)|, \quad (4.9)$$

where F is the SVD-DCT block matrix in the zigzag order.

Step 3: Use the frequency mask to weight the watermark amplitude.

Step 4: Modify the relationship between the selected coefficients pairs according to the watermark embedding rule.

In fact, at the start of each SVD-DCT block embedding, a pseudo-random position sequence is generated based on a secret key for the purpose of choosing the coefficients pairs to be modified. In this way, the probability for an attacker to detect or alter the watermark through identification of the embedded positions is significantly reduced. In the embedding procedure, the difference between the magnitudes of two coefficients in a selected pair is computed as follows

$$D(x_1, y_1, x_2, y_2) = |F(x_1, y_1)| - |F(x_2, y_2)|, \quad (4.10)$$

where F is the SVD-DCT block matrix and $(x_1, y_1), (x_2, y_2)$ are the coordinates of the selected pairs. Assuming that the high frequencies in the SVD-DCT block are close enough so that $D(x_1, y_1, x_2, y_2)$ is positive if the watermark bit to be embedded is “1” and negative otherwise.

To embed a watermark bit “1”, the procedure is as follows:

If $D(x_1, y_1, x_2, y_2) \geq \text{mask}$, no operation is needed.

If $D(x_1, y_1, x_2, y_2) < \text{mask}$, the following operations are needed:

If $F(x_1, y_1) \neq 0$ or $F(x_2, y_2) \neq 0$, then

$$\begin{aligned} F'(x_1, y_1) &= \text{sign}(F(x_1, y_1)) \times [(|F(x_1, y_1)| + |F(x_2, y_2)|) / 2 + 0.5 \times \text{mask}], \\ F'(x_2, y_2) &= \text{sign}(F(x_2, y_2)) \times [(|F(x_1, y_1)| + |F(x_2, y_2)|) / 2 - 0.5 \times \text{mask}], \end{aligned} \quad (4.11)$$

If $F(x_1, y_1) = F(x_2, y_2) = 0$, then

$$\begin{aligned} F'(x_1, y_1) &= 0.5 \times \text{mask}. \\ F'(x_2, y_2) &= -0.5 \times \text{mask}. \end{aligned} \quad (4.12)$$

The parameter mask in Eq. (4.11) and (4.12) is the frequency mask for changing the watermarking strength according to the sharpness of the macroblocks.

To embed a watermark bit “0”, the newly computed difference

$D(x_1, y_1, x_2, y_2)$ should be negative, thus the procedure is as follows:

If $D(x_1, y_1, x_2, y_2) \geq \text{mask}$, no operation is needed;

If $D(x_1, y_1, x_2, y_2) < mask$, the following operations are needed

If $F(x_1, y_1) \neq 0$ or $F(x_2, y_2) \neq 0$, then

$$\begin{aligned} F'(x_1, y_1) &= sign(F(x_1, y_1)) \times [(|F(x_1, y_1)| + |F(x_2, y_2)|) / 2 - 0.5 \times mask], \\ F'(x_2, y_2) &= sign(F(x_2, y_2)) \times [(|F(x_1, y_1)| + |F(x_2, y_2)|) / 2 + 0.5 \times mask], \end{aligned} \quad (4.13)$$

If $F(x_1, y_1) = F(x_2, y_2) = 0$, then

$$\begin{aligned} F'(x_1, y_1) &= -0.5 \times mask. \\ F'(x_2, y_2) &= -0.5 \times mask. \end{aligned} \quad (4.14)$$

To achieve an appropriate tradeoff between the robustness and transparency after watermark embedding, the potential locations of macroblocks are in the high frequency band. This differs from the traditional method of working in the mid-frequency band of DCT blocks because in this work, the DCT block is derived from the perceptual important components of the signal concerned.

The watermarking is achieved by changing the difference of the magnitudes of a selected pair to a defined extent. In this scheme, only 2 pairs of coefficients in the 8 potential positions in Figure 4.2 are selected for embedding watermark bits. There are two reasons for this selection. One is to maintain the imperceptibility and the other is to achieve high security. Since the attacker does not know which 2 pairs are embedded with watermark bits, attacking all the 8 coefficients will cause intolerable distortions to the signal concerned.

After the watermark embedding, inverse DCT of the macroblock is performed on the modified SVD-DCT blocks and then the blocks consisting of 4×4 SVs are reconstructed. Only the first SVs of the original 8×8 blocks contain watermark information and other SVs are kept unchanged.

4.4 Watermark Extraction

The watermark extraction procedure is rather simple and described as follows:

Step 1: Retrieve the positions of the coefficients pairs in each SVD-DCT block according to the secret key, the host signal characteristics and the positions of different SVD-DCT blocks which are similar in the watermark embedding procedure.

Step 2: The difference between coefficients of each selected pair is computed by

$$D'(x_1, y_1, x_2, y_2) = |F'(x_1, y_1)| - |F'(x_2, y_2)|, \quad (4.15)$$

where $|F'(x_1, y_1)|$ and $|F'(x_2, y_2)|$ are the selected coefficients of the watermarked signal.

Step 3: Sum the value of the difference $D'(x_1, y_1, x_2, y_2)$ corresponding to all pairs of coefficients where the same bit is repeatedly embedded as follows

$$H'(i) = \sum_{\Phi(i)} D'(x_1, y_1, x_2, y_2), \quad (4.16)$$

where $\Phi(i)$ is the set of the selected pairs for the i -th bit.

Step 4: The watermark is extracted with the following rules

$$w'(i) = \begin{cases} 1, & \text{if } H'(i) \geq 0 \\ 0, & \text{if } H'(i) < 0 \end{cases} \quad (4.17)$$

4.5 Error Analysis

It is often known that the false positive error analysis is defined as

$$P_{fp} = \sum_{n=0}^{\lceil M(1-T_p)/2 \rceil} \binom{M}{n} 0.5^M. \quad (4.18)$$

A plot of the probability of false positive error for $M \in (0,100]$ is shown in Figure 4.3. It can be seen that probability of the false positive error approaches 0 when M is larger than 30. In fact, in the scheme, $M = 256, T_p = 0.5$, then $P_{fp} = 2.4486 \times 10^{-16}$.

Figure 4.3: Probability of false positive error under various M .

Similarly, the probability of false negative error can be determined as

$$P_{fn} = \sum_{n=\lceil M(1-T_\rho)/2 \rceil + 1}^M \binom{M}{n} p^n (1-p)^{M-n}. \quad (4.19)$$

Figure 4.4 plots the probability of false negative error for $M \in (0, 100]$. It is noted that the probability of false negative error is close to zero when M is larger than 30. In this scheme $M = 256$, thus the false negative error in this scheme is near to 0. In fact, if $M = 256$, $T_\rho = 0.5$, $p = 0.9$, then $P_{fn} = 2.9347 \times 10^{-321}$.

Generally, if a desired probability of false positive error is given, the detection threshold T_ρ can be determined by P_{fp} . The determination of threshold should consider the trade-off between false positive and negative errors. An increase in T_ρ will reduce P_{fp} but will increase P_{fn} . In the simulation, T_ρ is set to 0.5.

Figure 4.4: Probability of false negative error under various M .

4.6 Experimental Results and Discussions

In this section, some experimental results demonstrating the performance of the proposed SVD-DCT algorithm will be presented. In this experiment, the well-known SQAM files [102] are selected as test samples to analyze the performance of the proposed watermarking algorithm. Each audio signal is a 16-bit mono file in the WAV format with 44.1 kHz sampling rate. A total of 12 host audio signals including music, percussive instruments, tonal instruments and speech signals are used. The detailed description of each signal is listed in Table 4.1.

Table 4.1 Test sequences of SQAM database (16 bits).

Index	Signal description	Item (.wav)	Index	Signal description	Item(.wav)
1	Bass	bass47_1	7	Quartet	quar48_1
2	Electronic tune	frer07_1	8	Sopranor	sopr44_1
3	Glockenspiel1	gspl35_1	9	Trumprt	trpt21_2
4	Glockenspiel2	gspl35_2	10	Violoncello	vioo10_2
5	Harpsichord	harp40_1	11	Female speech	spfe49_1
6	Horn	horn23_2	12	Male speech	spme50_1

A 16×16 binary image is used as watermark signal for all audio signals. The original watermark signal and the extracted one without any attack are shown in Figure 4.5. Actually, Figure 4.5 verifies that the watermark can be extracted clearly without any attack. Performance of audio watermarking algorithms is usually evaluated with respect to imperceptibility (inaudibility) and robustness. All the parameters have been experimentally selected to achieve a good

tradeoff among the conflicting requirements of imperceptibility, robustness and capacity.

Figure 4.5: Watermark signal; (a) original watermark; (b) extracted watermark without attack.

4.6.1 Perceptual Audio Quality Assessment

Audio watermarking intends to embed an unperceivable and secure watermark into host signals, thus it is essential that the watermarking scheme should be perceptually transparent (imperceptibility). Generally, there are two approaches to perform the perceptual audio quality assessment [103]: (1) Objective evaluation tests by perceptual evaluation of audio signals. (2) Subjective listening tests by human's acoustic perception.

4.6.1.1 Objective Evaluation Test

Based on the results in the existing literature, there are several audio objective quality measures such as SNR, segmental signal to noise ratio (SSNR), cepstral distance (CD) [77] and log-likelihood ratio (LLR) measures [104]. These objective metrics are selected to evaluate the perceptual performance between the host and watermarked signals.

Signal-to-Noise-Ratio (SNR) and Segmental SNR (SSNR)

SNR is a statistical difference metric which is used to measure the perceptual similarity between the undistorted original and the distorted watermarked audio signal. SNR is computed by

$$SNR = 10 \log_{10} \left(\frac{\sum_{i=1}^L S^2(i)}{\sum_{i=1}^L (S'(i) - S(i))^2} \right), \quad (4.20)$$

where $S(i)$ and $S'(i)$ correspond to the original and watermarked signal, respectively.

SSNR is defined as the average of the SNR values of short segments of the watermarked signal and used to give an objective indication of the watermarking effect on the original audio. It is a good estimator for audio signal quality which is calculated as

$$SSNR = \frac{10}{K} \sum_{m=0}^{K-1} \log_{10} \frac{\sum_{i=1}^r S^2(i)}{\sum_{i=1}^r (S'(i) - S(i))^2}, \quad (4.21)$$

where K is the number of frames in the watermarked audio signal and r is number of samples per frame.

Cepstral Distance (CD)

CD is easy to implement with few computational requirements and has the capability for accurately predicting subjective listening scores. The CD measures cepstral coefficients computed from the original and watermarked signals on a frame-by-frame basis. For the k -th frame, the CD is defined as

$$CD = \frac{1}{N} \sum_{k=1}^N \sqrt{\sum_{i=1}^{16} [real(IDFT\{\log |DFT(S(k))|\}) - real(IDFT\{\log |DFT(S'(k))|\})]^2}$$
(4.22)

where $real(IDFT\{\log |DFT(S(k))|\})$ and $real(IDFT\{\log |DFT(S'(k))|\})$ are the i -th real cepstral coefficients of the original and watermarked signals, respectively. DFT and IDFT are the discrete Fourier transform and its inverse. The *zero*-th coefficient is not generally included in Eq. (4.22) since it is primarily affected by system gain rather than system distortion. Furthermore, by excluding high order coefficients, CD represents a cepstrally smoothed spectral-difference measure.

Log Likelihood Ratio (LLR)

The LLR metric for an audio segment is based on the assumption that the segment can be represented by a p -th order all-pole linear predictive coding model of the form:

$$S(n) = \sum_{m=1}^P a(m)S(n-m) + G_x u(n), \quad (4.23)$$

where $S(n)$ is the n -th audio sample, $a(m)$ (for $m = 1, 2, \dots, P$) are the coefficients of an all-pole filter. G_x is the gain of the filter and $u(n)$ is an appropriate excitation source for the filter. The LLR metric is then defined as

$$LLR = \left| \log \left(\frac{\vec{a}_s \vec{R}_s \vec{a}_s^T}{\vec{a}_{s'} \vec{R}_{s'} \vec{a}_{s'}^T} \right) \right|, \quad (4.24)$$

where $\vec{a}_s$ and $\vec{a}_{s'}$ are the LPC coefficient vectors for the original and

watermarked audio signals, respectively, $\vec{R}_s$ is the auto-correlation matrix of the watermarked audio signal. The closer to zero the LLR is, the better is the quality of the watermarked audio signal.

Classical objective measures such as SNR, SSNR, CD and LLR do not use the characteristics of the HAS. On the other hand, the perceptual evaluation of audio quality (PEAQ) which is specified in ITU-Recommendation standard BS.1387 incorporates a psychoacoustic model. The PEAQ was originally proposed in 1998 and last updated in 2001. The output of the PEAQ algorithm is the ODG which corresponds to the subjective grade used to measure the differences between the original and watermarked audio signals. To implement the PEAQ and compute the ODG, EAQUAL software is utilized [90]. The computed ODG values are in the range of $[-4, 0]$ and the details are shown in Table 4.2.

Table 4.2 MOS, SDG and ODG descriptions and results from the proposed scheme.

MOS	SDG	ODG	Description	Quality
5	0	0	Imperceptible	Excellent
4	-1	-1	Perceptible, but not annoying	Good
3	-2	-2	Slightly annoying	Fair
2	-3	-3	Annoying	Poor
1	-4	-4	Very annoying	Bad

4.6.1.2 Subjective Listening Test

Besides the objective test, the essential subjective quality assessment is also

utilized to evaluate the audibility of the watermarked audio as the ultimate judgment is made by human acoustic perception. In the subjective listening tests, MUSHRA described in ITU-BS.1534, stands for “Multi Stimulus test with Hidden Reference and Anchors” are selected to evaluate the scheme. The watermarked signal is graded with respect to the host signal according to a five-grade impairment scale defined in ITU-R BS.562, called subjective difference grade (SDG). SDG is the difference between the absolute quality grade assigned to the distorted and the original signals, both measured on a five-grade impairment scale. If all listeners identify the hidden reference correctly, the SDG is identical to the grade on the five-grade impairment-scale assigned to the processed signal. The value of an SDG should therefore never be outside the range of minus four to zero. SDG details are also specified in Table 4.2. A listening test was actually performed with ten listeners who were both experienced and familiar with the used set of test items to estimate the SDG grades of the watermarked signals. The ten people listened to each test sample for 10 times and gave a grade. Then the average grade for each test sample from all listeners is the final SDG grade.

Mean opinion score (MOS), specified by ITU-T recommendation P.800, is also a good indication of the perceived audio quality after watermarking

distortion. Thus the MOS scores, which are described in Table 4.2, will be used to measure the performance as well.

4.6.1.3 Test Results

A short portion of an original violoncello audio signal and the corresponding watermarked audio signal are shown in Figure 4.6, which demonstrates that the imperceptibility of the proposed scheme is good. The corresponding SNR, SSNR, CD, LLR values, along with MOS, SDG and ODG grades obtained by conducting the subjective and objective tests are tabulated in Table 4.3. It can be noted that the embedding distortion caused by the proposed watermarking algorithm expressed with the ODG value is very close to the threshold of becoming audible. The average ODG score is -0.57 from all test items which confirms that the watermarked audio signals are perceptually similar to the original audio signals. Besides, it is evident that all the MOS scores are within [4.51, 4.90], and high MOS scores indicate that the proposed scheme provides good imperceptibility of the watermarked audio signals. The LLR is almost close to zero and the average is 0.019, which is obviously an indication of the good performance of the proposed scheme. The CD results in the experiment are very good for all test samples which can be seen from Table 4.3. It is clear that the average result of CD is satisfactory. Furthermore, as can be seen from

the above results, the audio quality distortion of the watermarking system is very small. The average results demonstrate that there is no significant distortion introduced by this scheme. The main reason of having satisfactory imperceptibility results is due to the fact that only a small number of samples are modified in the watermarking algorithm and all parameters are carefully chosen to give a better performance.

(a)

(b)

Figure 4.6: Waveform and spectrogram of the original audio signal and the watermarked host audio signal; (a) time domain waveforms; (b) frequency domain spectrograms (SNR=36.98, SSNR=41.05, MOS=4.86, ODG=-0.53, SDG=-0.51, LLR=0.012, CD=0.44).

Table 4.3 Objective and subjective assessment results of the proposed scheme.

Index	SNR	SSNR	MOS	ODG	SDG	LLR	CD
1	25.90	28.62	4.51	-0.79	-0.73	0.039	0.63
2	26.18	27.44	4.54	-0.77	-0.74	0.033	0.58
3	41.96	44.36	4.90	-0.28	-0.28	0.006	0.36
4	37.93	40.04	4.86	-0.36	-0.35	0.018	0.42
5	35.39	37.88	4.83	-0.31	-0.31	0.016	0.49
6	34.21	35.59	4.81	-0.52	-0.50	0.015	0.47
7	28.71	30.78	4.59	-0.69	-0.65	0.018	0.53
8	27.49	30.84	4.56	-0.72	-0.69	0.021	0.51
9	39.37	39.93	4.89	-0.42	-0.41	0.009	0.38
10	36.98	41.05	4.86	-0.53	-0.51	0.012	0.44
11	28.81	31.63	4.61	-0.68	-0.64	0.021	0.52
12	27.39	29.59	4.58	-0.71	-0.67	0.023	0.50
Average	32.53	34.81	4.71	-0.57	-0.54	0.019	0.48

4.6.2 Robustness

Watermarked audio signals may undergo common signal processing operations which may affect the perceived quality of the host signal and corrupt the watermark image embedded within the signal. In order to illustrate the robustness of the watermarking scheme, a great number of common attacks on audio signals in the experiments are implemented to assess the robustness of the scheme. Some attacks are performed using MATLAB 2008, Adobe Audition 1.0, and GoldWave 5.18, which are all popular tool-sets for professional audio processing and editing. The common signal processing attacks are described as follows:

Additive noise: White Gaussian noise is added to the watermarked signal until the resulting signal has a SNR of 20 dB.

Amplitude variation: The watermarked signal is attenuated up to 120%.

Cropping: Segments of 100 samples are removed from the watermarked audio signal at three randomly selected positions.

Denoising: The watermarked audio signal is denoised by using the “Hiss removal” function of GoldWave.

Echo addition: An echo signal with a delay of 50 ms and a decay of 10% is added to the original audio signal.

Expanding: Expand the watermarked signal with increment of 3 dB and -3dB, respectively.

Jittering: The watermarked audio signal undergoes small rapid variations.

LPF: The low-pass filter with the 44.1 kHz cut-off frequency is applied to the watermarked audio signal.

MP3 Compression: The coding/decoding is performed using a software implementation of the MP3 coder with different bit rates (32kbps, 48 kbps, 64 kbps, 96 kbps, and 128 kbps) and then decoded back to the WAV format.

Pitch shifting: Tempo-preserved pitch shifting is a tough attack for audio watermarking algorithms as it causes frequency fluctuation. In the experiment,

the pitch is shifted 1% higher and then 1% lower.

Re-quantization: 16-bit watermarked signal is quantized to 8-bits and then back to 16-bits.

Re-sampling: As the original audio signal is sampled with a sampling rate of 44.1 kHz, thus the watermarked audio signal is down-sampled to 11.025 kHz, 22.05 kHz, and then up-sampled back to 44.1 kHz. Watermarked audio signal is also up-sampled to 88.2 kHz and then down-sampled back to 44.1 kHz.

Reverse amplitude: Reverse the signs of the sample amplitudes.

Time-scale modification: The watermarked audio signal is lengthened by 4% while preserving the pitch.

Table 4.4 summarizes the watermark detection results against various common signal processing attacks. It is obvious that NC values after attacks are very high while the values of BER (%) are very low. As BER (%) is used in [58], BER (%) is adopted in Table 4.4 for easy comparison with the algorithm in [58]. The average values of NC and BER after all attacks are 0.9992 and 0.2000, respectively, which are better than the values of NC and BER (%) in [58]. The better results in the scheme indicate the robustness of the proposed audio watermarking scheme. Besides, the extracted watermark images are very visually similar to the original watermark, which further verify the good

performance of the scheme. It goes without saying that the proposed algorithm is slightly better than the algorithm in [58] as can be seen from the robustness test results.

Table 4.4 Results of robustness against common signal processing attacks achieved by the proposed scheme vs. that in [58].

Attacks	Algorithm in [58]		The proposed		Extracted watermarks
	BER (%)	NC	BER (%)	NC	
No attack	0	1	0	1	EE
Additive noise	0	1	0	1	EE
Amplitude variation	-	-	0	1	EE
Cropping:	0	1	0	1	EE
Denoising	5	0.9634	0	1	EE
Echo addition	2	0.9847	0	1	EE
Expanding	-	-	0	1	EE
Jittering	-	-	0	1	EE
LPF	0	0.9993	0	1	EE
MP3 (32kbps)	1	0.9907	3	0.9878	EE [•]
MP3 (48kbps)	-	-	1	0.9959	EE
MP3 (64kbps)	0	0.9980	0	1	EE

MP3 (96kbps)	-	-	0	1	EE
MP3 (128kbps)	-	-	0	1	EE
Pitch shifting	-	-	0	1	EE
Re-quantization:	0	1	0	1	EE
Re-sampling(44.1-22.05-44.1)	2	0.9854	0	1	EE
Re-sampling(44.1-11.025-44.1)	2	0.9854	0	1	EE
Re-sampling(44.1-88.2-44.1)	2	0.9854	0	1	EE
Time-scale modification	-	-	0	1	EE
Average	1.0667	0.9918	0.2000	0.9992	

4.6.3 Robustness to Stirmark Attacks

Performance of the proposed scheme is evaluated under the standardized Stirmark attacks for audio. SMBA is a standard and common robustness evaluation tool for evaluating the robustness of audio watermarking method [66]. In this experiment, different attacks are applied to the watermarked and un-watermarked host signal by using the Stirmark software. The parameters of the Stirmark attacks are the default parameters included in the version of the tool available on the Web.

The Stirmark attack tests are performed on the test set and the results are presented in Table 4.5. From the results of the SMBA test, it can be observed

that the scheme can resist those common attack operations. The BER values are very low after attacks, so this scheme is robust and secure. Furthermore, the watermarks extracted after various attacks on the watermarked signal are given in Table 4.5. It should be noted that since synchronization techniques are adopted in the proposed algorithm, it is robust to synchronization attacks such as Flipsample, Copysample and Cutsample. Besides, the proposed method demonstrates its superiority over the algorithm in [12] and [56]. The BER values after attacks are compared with the two published algorithms: the SVD-STFT based algorithm in [56] and the DCT algorithm in [12]. It is obvious that the performance of the proposed algorithm is much better than the method in [12] and slightly better than the method in [56].

Table 4.5 Results of robustness against Stirmark attacks for audio achieved by the proposed scheme vs. SVD based [56] and DCT based [12] schemes.

Attacks	DCT based method in [12]	SVD based method in [56]	The proposed	Extracted Watermarks
Addbrumm	1.25	0	0	EE
AddDynNoise	1.56	0	0	EE
AddFFTNNoise	51.25	0	0	EE
Addnoise	0.78	0	0	EE
Addsinus	0.77	0	0	EE
Amplify	52.32	0.75	0	EE

Bassboost	0	0	0	EE
Compressor	0	0	0	EE
Copysample	100	0.5	0.195	ĖĖ
Cutsamples	100	0	0	EE
Echo	23.43	0	0	EE
Exchange	0	0	0	EE
Extrastereo	0	0	0	EE
Fft_hlpass	0.31	0	0	EE
Fft_invert	52.6	0	0	EE
Fft_real_reverse	0.78	0	0	EE
Fft_stat1	19.84	0.5	0.0234	ĖĖ
Fft_test	19.80	0.4	0.0078	EE
Flipsample	21.66	0.75	0.0156	ĖĖ
Invert	52.42	0	0	EE
Lsbzero	0	0	0	EE
Normalize	0	0	0	EE
Nothing	0	0	0	EE
Rc_highpass	2.03	0	0	EE

Rc_lowpass	0	0	0	EE
Smooth	0	0	0	EE
Stat1	0	0	0	EE
Stat2	0	0	0	EE
Voiceremove	52.1	0	0	EE
Zerocross	0	0	0	EE
Zerolength	60.5	0	0	EE
Zeroremove	100	0	0	EE
Average of all attacks	22.2937	0.0906	0.0076	

4.6.4 Embedding Capacity

From the watermark capacity analysis and formula, it could be argued that the true capacity of the proposed scheme is 43 bps, given the sampling rate 44.1 kHz and the duration time of the test audio signals is 6s. For the embedding capacity, the proposed scheme is also compared with other recent audio watermarking scheme with and without synchronization methods [41-42, 48, 51, 54, 105]. Table 4.6 shows the capacity performance comparison results of the proposed hybrid SVD-DCT based audio watermarking scheme with several state-of-the-art audio watermarking strategies with and without synchronization techniques. Many other methods are not listed as they do not report the

embedding capacity. It can be seen that the capacity of the method is satisfactory compared to other methods. The capacity of the algorithm is the same as the algorithm in [105]. However, the imperceptibility of the proposed method is better than that in [105].

Table 4.6 Comparison results of capacity performance achieved by our proposed scheme vs. state-of-the-art audio watermarking algorithms.

Algorithms	SNR	Synchronization	Capacity
in [42]	42.8	Yes	2
in [48]	29.5	Yes	4.27
in [51]	43.0	Yes	36
in [41]	43.1	Yes	-
in [105]	25	No	43
in [54]	30-45	No	86
The proposed	32.53	Yes	43

4.6.5 Discussions

A great variety of traditional and state-of-the-art watermarking approaches such as echo hiding, SS, LSB and frequency masking methods have been presented in the literature. Recently, some SVD based audio watermarking methods have been developed too. Usually, SNR is only a simple measure of the signal power relative to the noise power and it is not closely correlated to human perception as SNR does not take into account perceptual models. For example, the SNR of a signal which is amplified by 10% (i.e., $1.1x$) is 20 dB. However, the difference in MOS between signals x and $1.1x$ would be 0 (no error). The reason is that the amplitude change only implies a volume amplification of the signal, but there is no significant degradation of audio quality which means

there is no difference in MOS. Therefore, in order to make an appropriate comparison with the traditional watermarking methods and the SVD based methods, SNR results and MOS grades are both selected. The comparison results among the proposed method, the traditional and SVD based methods are given in Table 4.7 which are based on the reported results in the literature [2-3, 12, 56, 58, 77]. From the comparison results, it can be seen that the proposed method outperforms some traditional methods and SVD based methods in terms of SNR and MOS. Only LSB method is better than the proposed method but the proposed method is obviously better than the SVD based methods.

Table 4.7 Comparison results of SNR and MOS values achieved by the proposed scheme vs. traditional and SVD based methods.

Reference	Algorithm	SNR	MOS
in [2]	Echo	21.47	3.60
in [2]	Phase	12.20	2.44
in [2]	LSB	67.91	4.90
in [3]	Frequency Masking	12.87	2.93
in [12]	SS	28.59	4.46
in [56]	SVD-STFT	28.36	4.70
in [58]	SVD-DWT	26.84	4.60
in [77]	SVD	27.13	4.60
The proposed	SVD-DCT	32.53	4.71

Finally, it should be pointed out that the proposed algorithm satisfies the desired features of optimal audio watermarking, which have been set by the IFPI. IFPI states that the watermarking scheme should not degrade the perception of audio. Referring to the figures and tables above, it is easy to conclude that the performance of the proposed algorithm fulfills the IFPI

performance requirement as the SNR of the proposed method is far beyond 20dB.

4.7 Chapter Summary

In this chapter, a blind and robust audio watermarking technique which incorporates SVD, DCT and synchronization code technique is proposed. Extensive experimental works have shown that the proposed watermarking scheme possesses strong robustness to common signal processing operations. Moreover, the proposed scheme achieves very low error probability rates. The comparison of the proposed algorithm with traditional and SVD based algorithms shows better performance from the proposed algorithm. The obtained simulation results verify the effectiveness of the audio watermarking as a reliable solution to the copyright protection problem.

Chapter 5

A Robust Audio Watermarking Scheme Based on LWT and SVD

5.1 Introduction

It is found that the conventional wavelet transform actually has very good performance because of its multi-resolution property and perfect reconstruction [52, 58]. However, the classic wavelet transform is mainly computed by convolution which results in high computation load. Besides, the generated floating point numbers increase storage requirements. As a result, a new wavelet is designed and developed to increase the efficiency. First proposed by Sweldens [106], LWT is the second generation wavelet which is based on the traditional wavelet. In fact, the lifting scheme has several unique properties in comparison with the traditional wavelet: 1) LWT allows for an in-place implementation of the fast wavelet transform and the construction of wavelets without using Fourier transform. Hence LWT can be calculated more efficiently and needs less memory space. 2) It is particularly easy to build non-linear wavelet transforms and LWT has the time-frequency localization capability. 3) LWT coefficients are integers and therefore are not subjected to

quantization errors unlike the traditional wavelet transform. Consequently, LWT is widely used in the audio watermarking field [55, 107-108].

Apart from the traditional transform technique, LWT and lifting based method have also been proven to have good performance in audio watermarking in recent years as lifting can further improve the computation time of the wavelet transform. For example, in [108], a LWT based audio watermarking is suggested for fast implementation of watermarking due to time saving of the lifting scheme. The binary watermark scrambled by PN sequence is inserted into the LWT approximate coefficient in the low frequency domain. This algorithm achieves a relatively good imperceptibility and robustness to filter attack. However, the robustness results after resampling, requantization and MP3 compression attack are not very good and the result after MP3 compression attack is hardly acceptable too. In [55], Tao *et al.* proposed a robust audio watermarking scheme in the LWT frequency domain based on the statistical characteristics of sub-band coefficients. In this scheme, the watermarking implementation efficiency is improved by the adoption of LWT.

Currently, in the perspective of linear algebra, SVD is extensively applied in robust watermarking to withstand attacks due to its unique and special characteristics. As a factorization of a real matrix, SVD transform has been

applied widely in image watermarking [93] first for ownership protection and quickly extended to audio watermarking [56-63]. Moreover, QIM [109] is also a very popular method for watermark embedding and data hiding. If LWT is combined with QIM method and SVD, it can reduce operation times and very robust results can be achieved.

In this chapter, an efficient and robust watermarking algorithm for copyright protection based on LWT, SVD and QIM with synchronization code technique to withstand desynchronization attacks is proposed. The reason to combine LWT, SVD and QIM is that superior performance can be achieved. The host audio signal is decomposed by LWT and a meaningful binary image as watermark data is scrambled by chaotic signal. The scrambled watermark is embedded in the low frequency subband of the original host audio through the modification of the singular values (SVs).

The organization of this chapter is as follows. Section 5.2 presents the principle of LWT. Section 5.3 discusses the embedding method. Watermark extraction is described in Section 5.4. The experimental results and algorithm comparison are discussed in Section 5.5. Finally, Section 5.6 summarizes this chapter.

5.2 Wavelet Lifting

The lifting scheme is proposed to reduce computation time and memory requirement as lifting scheme adopts an in-place implementation of wavelet transform. Lifting wavelet simplifies the problem by directly analyzing the problem in integer domain. In addition, the lifting wavelet saves time and has the frequency localization feature which overcomes the weakness of the traditional wavelet.

Actually, the main principle of the lifting wavelet is to construct a new wavelet with better characteristics based on a simple wavelet, which is also the basic idea of lifting. As the basis of integer wavelet transform, lifting wavelet algorithm is generally composed of three steps: split/merge, prediction and update. The detail reasoning and proof of the lifting scheme is given in [106].

1) Split step: The split step is also called the lazy wavelet transform. The operation just splits the input signal $x(n)$ into even and odd samples: $Xe(n)$ and $Xo(n)$.

$$\begin{aligned}Xe(n) &= x(2n). \\Xo(n) &= x(2n+1).\end{aligned}\tag{5.1}$$

2) Prediction step: Keep even samples unchanged and use $Xe(n)$ to predict $Xo(n)$. The two signal subsets from split process should be closely correlated.

The difference between the prediction value of $P[Xe(n)]$ and the real value of

$Xo(n)$ is defined as detail signal $d(n)$

$$d(n) = Xo(n) - P[Xe(n)], \quad (5.2)$$

where $P[\bullet]$ is the prediction operator. The detail signal, $d(n)$, denotes the high-frequency component of the original signal $x(n)$.

3) Update step: Introduce the update operator $U[\bullet]$ and use detail signal, $d(n)$, to update even samples $Xe(n)$. Then the approximate signal, $c(n)$, denotes the low-frequency component of the original signal

$$c(n) = Xe(n) + U[d(n)]. \quad (5.3)$$

In fact, the reconstruction of LWT is an inverse process of decomposition.

The lifting scheme of decomposition and reconstruction is illustrated in Figure

5.1.

Figure 5.1: Decomposition and reconstruction of lifting wavelet.

5.3 Embedding Method

5.3.1 Watermark Preprocessing

Watermark should be first preprocessed in order to improve the robustness and enhance the confidentiality. The binary image will be chaotically scrambled before embedding to increase the security of watermarking technique. Watermark binary image is scrambled by a chaotic map which is reproduced in a permuted matrix. This chapter uses a Skew tent map to enhance the confidentiality of the watermarking method. The skew tent map is defined as follows:

$$x(n+1) = \begin{cases} \frac{1}{\alpha} \times x(n), & 0 \leq x(n) < \alpha \\ \frac{1}{\alpha-1} \times x(n) + \frac{1}{1-\alpha}, & \alpha \leq x(n) \leq 1 \end{cases}, \quad (5.4)$$

where $\alpha \in (0,1)$ is the system parameter. The initial value is adopted as key K_1 .

Then the binary image logo or signature, $b(n)$, is scrambled by $x(n)$ with the following rule:

$$w(n) = b(n) \oplus x(n), \quad 1 \leq n \leq N_w, \quad (5.5)$$

where N_w is length of the watermark, $\oplus$ is the *XOR* operator. After this random chaotic sequence encryption, the watermark is permuted and cannot be guessed by random search.

5.3.2 Synchronization Code

In the audio watermarking field, quite a number of watermarking schemes exploit the synchronization code technique [51-52]. The watermark will have dislocation of the watermark regions due to the desynchronization attacks.

Synchronization code is an effective way to locate the position of hidden informative bits after desynchronization attacks. Besides, such localized synchronization codes eliminate false alarm error due to data modification on watermark embedding. The synchronization code is often put in front of the watermark and used to identify the position of the inserted watermark. The data structure of this method is shown in Figure 5.2. It is shown that the cover signal is divided into proper segments according to the length of $Syn(k)$ and $W(k)$.

Figure 5.2: Data structure of hidden bitstream.

In the proposed method, a pseudo random sequence generated by chaotic signal is exploited as the synchronization code to increase the security of the synchronization code. By using a strongly chaotic nature, the system is cryptographically secure. The synchronization code is generated by thresholding the Bernoulli shift map. The Bernoulli shift map is one of the simplest deterministic chaotic maps which contain many chaotic characteristics. A binary shift Bernoulli Map can be defined as:

$$x(k+1) = \begin{cases} 2x(k), & \text{if } 0 \leq x(k) < \frac{1}{2} \\ 2x(k)-1, & \text{if } \frac{1}{2} \leq x(k) \leq 1 \end{cases}, \quad (5.6)$$

where $x(0) \in (0,1)$ (map's initial condition) is adopted as a secret key, K_2 , which must be specified. $x(k)$ is mapped into the synchronization sequence $C = \{c(k), 1 \leq k \leq L_{\text{syn}}\}$ with the following rule

$$c(k) = \begin{cases} 1, & \text{if } x(k) > \tau \\ 0, & \text{otherwise} \end{cases}. \quad (5.7)$$

τ is a predefined threshold for synchronization code. Time domain embedding has the strength that it is less computation intensive in finding the synchronization code. Thus the synchronization code is hidden in the time domain to reduce the calculation times. Before embedding, the synchronization code should be arranged into a binary data sequence.

As the first portion of the audio signal S is used to embed the synchronization code, it is split into proper segments according to the length of synchronization code. In the proposed method, the portion A of the signal is divided into LA audio segments with N samples, which are denoted as

$$SA(k)(j) = S(k \times N + j), \quad 1 \leq k \leq LA, 1 \leq j \leq N. \quad (5.8)$$

It is obvious that a slight change of the audio signal will not modify the statistical average value greatly. In order to achieve good tradeoff between inaudibility and robustness of the watermarking algorithm, the proposed

method embeds synchronization code into the statistical average value of audio samples. Thus the mean value of each segment $SA(k)$ is first calculated by

$$\overline{SA(k)} = \frac{1}{N} \sum_{j=1}^N SA(k)(j), \quad (1 \leq k \leq LA). \quad (5.9)$$

Then synchronization code is inserted into $SA(k)$ bit by bit as follows:

$$SA'(k)(j) = \begin{cases} SA(k)(j) + 2 \times \lfloor \overline{SA(k)} \rfloor, & \text{if } Syn(k) = 1 \\ SA(k)(j) - 2 \times \lfloor \overline{SA(k)} \rfloor, & \text{if } Syn(k) = 0 \end{cases}. \quad (5.10)$$

The reason for adding and subtracting the mean value is that the mean value is made positive when the bit “1” is embedded and it is made negative when the bit “0” is embedded. Suppose $SA''(k)$ is the embedded and attacked signals and the mean value of each segment $\overline{SA''(k)}$ is calculated by

$$\overline{SA''(k)} = \frac{1}{N} \sum_{j=1}^N SA''(k)(j), \quad (1 \leq k \leq LA). \quad (5.11)$$

In the synchronization code extraction procedure, the audio samples also follow the rules in the embedding process and the mean value of each segment is calculated. Actually, the extraction is based on the popular odd-even parity rule. That is: if the mean value is greater than or equal to zero, a bit “1” is detected. If the mean is less than zero, a bit “0” is detected. Therefore, the rule to extract the synchronization code is as follows:

$$Syn'(k) = \begin{cases} 1, & \text{if } \overline{SA''(k)} \geq 0 \\ 0, & \text{if } \overline{SA''(k)} < 0 \end{cases}. \quad (5.12)$$

Finally, the normalized correlation coefficient (NC) is adopted to calculate

the similarity between the original and extracted synchronization codes as follows:

$$NC(Syn(k), Syn'(k)) = \frac{\sum_{k=1}^n Syn'(k) \times Syn(k)}{\sqrt{\sum_{k=1}^n Syn'(k)^2} \sqrt{\sum_{k=1}^n Syn(k)}}, \quad (5.13)$$

where $Syn(k)$ and $Syn'(k)$ are the original and extracted synchronization codes, respectively. If the NC between $Syn(k)$ and $Syn'(k)$ is greater than or equal to a specified threshold T , then $Syn'(k)$ is considered to be a synchronization code. Otherwise, there is no synchronization code detected and extracted.

5.3.3 Watermark Embedding

As LWT has good space localization characteristics and the low frequency component concentrates more than 90% of the signal energy, the anti-interference and anti-compression ability of audio watermarking scheme will be very good by using LWT technique. Therefore, the signal is directly embedded into the low frequency component of the LWT subband. Figure 5.3 presents the block diagram of the watermark embedding algorithm.

Figure 5.3: Diagram of watermark embedding process.

In this watermarking technique, the popular QIM method is exploited in the embedding process because of its good robustness and blind nature. As a result, the proposed method is blind and does not need the original audio for the data extraction. The second part of the host audio, SB , is used to embed the watermark.

Specifically, the embedding process is described by the following steps:

Step 1: Perform LWT on the audio segment, SB , of the host audio signal.

$$I = LWT(SB). \quad (5.14)$$

Step 2: The approximate coefficients after the LWT decomposition are divided into non-overlapping blocks. The length of audio blocks depends on the amount of data that need to be embedded and the number of LWT decomposition

levels. The watermark sequence is embedded successively into the blocks in the low-frequency subband.

Step 3: Scrambling the watermark image with the method mentioned in Section 5.1.

Step 4: For each block, perform SVD transform to obtain the SVs and first SV, $S(1,1)$,

$$I = USV^T. \quad (5.15)$$

Step 5: Embedding the watermark into SVs with the QIM method. The encrypted watermark $w(i)$ is added to the first SVs, $S(1,1)$, of each block. The watermark embedding method is based on the popular odd and even parity rule: Let $Q = \text{round}(S(1,1)/\beta)$, $D = \text{mod}(Q, 2)$, where β is the quantization step. A small value of β will lead to good imperceptibility of the watermarking scheme but low robustness to the attacks. Thus an optimal β is chosen to tradeoff between inaudibility and robustness of the watermark. The embedding rule is that:

If D is 0 and $w(i)$ is 1, then $Q = Q + 1$; if D is 1 and $w(i)$ is 0, then $Q = Q + 1$. In other cases, Q remains unchanged.

Step 6: The first SVs is further modified by the updated Q as follows

$$S_w(1,1) = \beta \times \text{round}(Q). \quad (5.16)$$

Step 7: $S_w(1,1)$ is used to build the watermarked block, I_w , by applying inverse SVD as follows:

$$I_w = US_w V^T. \quad (5.17)$$

Step 8: Inverse LWT is applied to reconstruct the watermarked signal

$$SB_w = LWT^{-1}(I_w). \quad (5.18)$$

5.4 Watermark Extraction

The main step of watermark extraction is as follows:

Step 1: Perform LWT on the watermarked signal

$$I_e = LWT(SB_w). \quad (5.19)$$

Step 2: For the obtained wavelet approximation coefficient, block based method is also used, that is, the LWT approximate coefficients are divided into different blocks too.

Step 3: SVD is performed in each block.

$$I_e = U_e S_e V_e^T. \quad (5.20)$$

Step 4: Let $Q_e = \text{round}(S_e(1,1) / \beta)$ and $D_e = \text{mod}(Q_e, 2)$, then the extraction rule is

$$w'(n) = \begin{cases} 1, & D_e = 1 \\ 0, & D_e = 0 \end{cases} \quad (5.21)$$

Step 5: Perform the decryption with the same chaotic sequence to get the

hidden binary image or signature

$$b'(n) = w'(n) \oplus x(n). \quad (5.22)$$

5.5 Experimental Results

In this section, several experiments are conducted to demonstrate the performance of the proposed LWT-SVD based audio watermarking approach.

The performance of the proposed scheme is assessed in terms of robustness and imperceptibility. The proposed scheme has been tested under different scenarios, such as the sound quality assessment material (SQAM) [102] clips, full songs (classical, pop and rock music) and human voice signal. The test audio signal in the proposed scheme is 44.1 kHz sampled, with 16 bits/sample. 32×32 binary image logo is used in the proposed scheme to conduct performance evaluation. LWT decomposition level is 3. In the experiment, SNR and SSNR are used for the evaluation of the quality of the watermarked audio signals. BER is used for evaluating the reliability of the extracted watermarks. As for the proposed scheme, $N_w = 1024$ bits is embedded in the 6s host audio, thus the payload of the proposed method is 170.67 bps. This is a comparatively high payload as the normal payload is 20-50 bps.

5.5.1 Imperceptibility Test

For the imperceptibility test, the time domain waveforms and the spectra in the

frequency domain are given for performance evaluation. The time domain waveforms show the difference between the original and watermarked waveforms in time domain while the spectrum can illustrate the differences in the frequency domain. Figure 5.4 presents the waveforms of the original, watermarked, and residual signals, respectively. Figure 5.5 shows the spectrum of the original and watermarked signals, respectively. The average SNR and SSNR results of all the test signals versus different quantization steps are shown in Figure 5.6. From the waveforms and spectra, it can be observed that there is not much distinguishable difference between the original and watermarked audio, which is also verified by the SNR and SSNR results in Figure 5.6, as SNR and SSNR results are all above 20dB even when the quantization step is 1. The SNR and SSNR results can totally satisfy the IFPI requirements.

Figure 5.4: Waveform of original, watermarked and residual signals

Figure 5.5: Spectrogram of original and watermarked signal.

Figure 5.6: SNR and SSNR results versus quantization step.

5.5.2 Subjective Listening Test

Generally, SNR is just a simple way to present the sense of imperceptibility by measuring the signal distortion caused by watermarking. However, the human perception does not correlate well with the SNR measure. Consequently,

subjective quality evaluation of the watermarking methods must be conducted to provide a better test of inaudibility from the human perspective. In this experiment, an informal subjective listening test is performed to evaluate the perceptual quality of the watermarked audio. Ten listeners involved in the listening tests are asked to classify the difference between the original and watermarked audio in terms of 5-point MOS impairment scale (5: imperceptible, 4: perceptible but not annoying, 3: slightly annoying, 2: annoying 1: very annoying). In this way, the inaudibility of the proposed watermarking schemes has been tested through listening tests. Figure 5.7 plots the MOS scores of the 15 test sequences. The average MOS for the tested audio excerpts is 4.82 for our algorithm, which means that the watermarked audio and the original audio are perceptually indistinguishable.

Figure 5.7: MOS scores of the 15 test sequences.

5.5.3 Robustness to Common Signal Processing Attacks

The watermarking scheme evaluation should comprise blind attacks (such as MP3 compression, noise addition, etc.) and intentional attacks. The non-intentional attacks do not know whether the watermark exists and where the watermark is, while the intentional attacks know where and what are embedded and try to remove it. The blind attacks (common signal processing) and desynchronization attacks are used to test the robustness of the proposed scheme. In this experiment, the detailed parameters for these general signal processing attacks can be found in Section 4.6.2.

The robustness test is evaluated in terms of the above mentioned attacks. BER results and extracted watermarks after attacks are used to show the robustness results as the robustness is of great significance in watermarking. Table 5.1 demonstrates the robustness results of the proposed scheme and comparison results with DCT domain method in [51], DWT domain method in [52] and DWT-DCT domain method in [41]. It is obvious that the robustness results of the proposed method are much better than the algorithms in [52] and [51], and slightly better than the algorithm in [41]. The reason why our method outperforms all these selected methods [41], [52] and [51] in robustness is mainly because of the advantageous properties of LWT-SVD over single DWT or DCT transform, and the hybrid DWT-DCT transform. Both LWT and SVD

have very attractive features for robust watermarking and the combination of them enhances the robustness too.

Table 5.1 Robustness to common signal processing attacks and comparison.

Attack	No attack	Requantization	Resampling 22.05 kHz	Resampling 11.025 kHz	Resampling 8kHz	Additive noise	Low-pass filtering 4kHz
Huang [51]
BER	0	0	0.501	0.489	0.482	0.026	0.491
Wu [52]
BER	0	0	0.5029	0.4902	0.5020	0.0195	0.5019
Wang [41]
BER	0	0	0	0	0.0586	0	0
The proposed
BER	0	0	0	0	0	0	0
Attack	Echo addition	Equalization	MP3-256kbps	MP3-128kbps	MP3-112kbps	MP3-96kbps	MP3-64kbps
Huang [51]
BER	0.190	0.159	0	0	0	0.502	0.498
Wu [52]
BER	0.1914	0.1504	0	0	0	0	0.5020
Wang [41]
BER	0.0078	0.0039	0	0	0	0.0107	0.0225
The proposed
BER	0.002	0.0027	0	0	0	0	0
Attack	Cropping(10% (front))	Cropping(10% (middle))	Adding(10% (front))	Amplitude scaling up to 150%	Amplitude scaling down to 50%	Pitch shift 1 degree higher	Pitch shift 1 degree lower
Huang [51]
BER	0	0	0.500	0.498	0.508	0.501	0.504

Wu [52]
BER	0	0	0.5078	0.5049	0.4815	0.5156	0.4580
Wang [41]
BER	0	0	0	0.2363	0.2549	0.5166	0.501
The pro- posed
BER	0	0	0	0.158	0.162	0.064	0.059
Attack	TSM(+1%)	TSM(-1%)	TSM(-2%)	TSM(-3%)	TSM(-4%)	Jittering (1/10000)	Jittering (1/5000)
Huang [51]
BER	0.503	0.482	0.484	0.431	0.454	0.045	0.140
Wu [52]
BER	0.4785	0.5195	0.4854	0.4932	0.4766	0.0488	0.1406
Wang [41]
BER	0.001	0	0.1543	0.4102	0.4717	0	0.0215
The pro- posed
BER	0	0	0.1455	0.3877	0.4580	0	0.0186

5.5.4 Robustness to Stirmark Attacks

The robustness of the proposed scheme is also benchmarked by SMBA software which is a very popular benchmark tool for audio watermarking. The parameters of these standardized attacks are the default in the software configuration. The detailed benchmark attack results are summarized in Table 5.2. The proposed method is also compared with the selected state-of-the-art watermarking methods in literature [12] and [56]. From this comparison results,

it is noted that the proposed method is slightly better than the watermarking scheme in [56] but much better than the watermarking scheme in [12] regarding for the Stirmark attacks.

Table 5.2 Robustness to the Stirmark attacks.

Attacks	Cox [12]	Ozer [56]	The proposed	Extracted watermarks of the proposed method	Attacks	Cox [12]	Ozer [56]	The proposed	Extracted watermarks of the proposed method
Addbrumm	1.25	0	0		Fft_stat1	19.84	0.5	0.13
AddDynNoise	1.56	0	0		Fft_test	19.80	0.4	0.026
AddFFTNoise	51.25	0	0		Flipsample	21.66	0.75	0.029
Addnoise	0.78	0	0		Invert	52.42	0	0
Addsinus	0.77	0	0		Lsbzero	0	0	0
Amplify	52.32	0.75	0		Normalize	0	0	0
Bassboost	0	0	0		Nothing	0	0	0
Compressor	0	0	0		Rc_highpass	2.03	0	0
Copysample	100	0.5	0.2		Rc_lowpass	0	0	0
Cutsamples	100	0	0		Smooth	0	0	0

Echo	23.43	0	0	EEE EEE	Stat1	0	0	0	EEE EEE
Exchange	0	0	0	EEE EEE	Stat2	0	0	0	EEE EEE
Extrastereo	0	0	0	EEE EEE	Voiceremo ve	52.1	0	0	EEE EEE
Fft_hlpass	0.31	0	0	EEE EEE	Zerocross	0	0	0	EEE EEE
Fft_invert	52.6	0	0	EEE EEE	Zerolength	60.5	0	0	EEE EEE
Fft_real_rever se	0.78	0	0	EEE EEE	Zeroremove	100	0	0	EEE EEE
Average of all attacks	22.293 7	0.090 6	0.012						

5.5.5 Security Analysis

For a secure audio watermarking scheme, robustness against attacks is an important issue. To enhance the security, the key space should be large enough to make brute force attack infeasible. Moreover, secret keys are adopted for security purpose. The digital computer often stores floating-point number using 32 bits, which consist of an exponent being represented using eight bits and a significant number being represented using 24 bits. In our scheme, keys or initial conditions are floating-point numbers. Hence the exponent is fixed while the significant number may be varied. Hence the total number of possible initial conditions is 2 to the power of 24 which is more than 16 million. This number can also be greatly increased if so desired through the use of double precision

floating point numbers. Actually, the security of information system depends on keys rather than the privacy of the scheme.

In the proposed audio watermarking scheme, keys K_1 and K_2 are employed to generate chaotic sequences for enhancing the security of the proposed scheme. Hence, the size of key value space influences the security of the proposed scheme. As keys K_1 and K_2 are both used in this scheme, K_1 is taken for example to compute its key value space. Suppose $K_1 = \{0 < K_1(i) < 1 \mid i = 1, 2, \dots, N_1\}$, N_1 is an integer which is used to generate the sequence and should be large enough to produce sequences $Y = \{y(i, j) \mid i = 1, 2, \dots, N_1, j = 1, 2, \dots, N_2\}$, where N_1 denotes the number of chaotic sequences and N_2 represents the length of each chaotic sequence. When $K_1' = \{0 < K_1(i) + d < 1 \mid i = 1, 2, \dots, N_1\}$, generate another group of chaotic sequences $Y' = \{y'(i, j) \mid i = 1, 2, \dots, N_1, j = 1, 2, \dots, N_2\}$. The function $f_k = SS(d)$ is used to test the key space of K_1 as follows:

$$f_k = SS(d) = \frac{\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} |y(i, j) - y'(i, j)|}{N_1 \times N_2}. \quad (5.23)$$

Figure 5.8 plots the function of $f_k = SS(d)$. It can be seen that f is equal to 0 when $d_0 = 10^{-17}$ in the proposed method. Thus the key space of K_1 is $1/d_0 = 10^{-17}$, which means that there is enough key space to guarantee high confidentiality of the proposed watermarking system. Similarly, the key spaces

of K_2 can be computed in the same way as K_1 as shown in Figure 5.8. Therefore, the total key space of the watermarking scheme is 10^{34} , which means that there is enough key space to guarantee high confidentiality of the proposed watermarking system. Furthermore, the employed synchronization codes can be used to prevent the signal from being detected by the intelligent attackers and the proposed audio watermarking has different combinations of the secret key. Based on this security analysis, it can be seen that the embedded watermarks are secure to attackers who try to exhaustively or statistically detect and read them. All in all, the proposed scheme with such a long key is enough for reliable and practical use.

Figure 5.8: Key space under different initial value difference.

5.5.6 Algorithm Comparisons and Discussions

Table 5.3 shows the algorithm comparison with DWT-SVD method in [58] regarding the robustness test results as they both adopt the wavelet transform

and the SVD techniques. From this table, it can be argued that the proposed LWT-SVD method outperforms the DWT-SVD method in [58] since the average BER (%) and NC results are better than the method in [58].

Table 5.3 Algorithm comparison with DWT-SVD method regarding the robustness.

Attacks	DWT-SVD [58]		The proposed	
	BER (%)	NC	BER (%)	NC
No attack	0	1	0	1
Additive noise	0	1	0	1
Amplitude variation	-	-	0	1
Cropping:	0	1	0	1
Denoising	5	0.9634	0	1
Echo addition	2	0.9847	0	1
Expanding	-	-	0	1
Jittering	-	-	0	1
LPF	0	0.9993	0	1
MP3 (32kbps)	1	0.9907	3.25	0.9802
MP3 (48kbps)	-	-	0.97	0.9958
MP3 (64kbps)	0	0.9980	0	1
MP3 (96kbps)	-	-	0	1
MP3 (128kbps)	-	-	0	1
Pitch shifting	-	-	0	1
Re-quantization:	0	1	0	1
Re-sampling (44.1-22.05-44.1)	2	0.9854	0	1
Re-sampling (44.1-11.025-44.1)	2	0.9854	0	1
Re-sampling (44.1-88.2-44.1)	2	0.9854	0	1
TSM	-	-	0	1
Average	1.0667	0.9918	0.2110	0.9988

Furthermore, the proposed scheme is compared with other related audio watermarking schemes in the literature. The selected watermarking schemes are typical audio watermarking schemes with self-synchronization or without synchronization, with and without SVD transform, and other conventional audio watermarking schemes. The detail comparisons among these audio watermarking schemes in terms of robustness, imperceptibility, payload and

other features are listed in Table 5.4. Note that if the detail subjective results (MOS score) are not provided in the references, N/A is reported in Table 5.4 regarding the subjective test. There are a few findings from this table are summarized as follows:

- 1) Most of the recent publications focus on one or two features, such as synchronization, efficiency, robustness and imperceptibility.
- 2) Few of them have provided the security analysis; however, the proposed algorithm provides detailed error analysis and key space analysis.
- 3) A higher embedding rate often leads to unsatisfactory robustness for the same amount of attacking disturbance. A suitable embedding rate should be between 20bps and 50bps.

From the comparison results in Table 5.4, it can be seen that the proposed algorithm can obtain a relatively high payload and good transparency results. It is capable of achieving a moderately high SNR results since SNR results in the proposed scheme is not as high as the schemes in [41-42, 52, 76], but higher than most of the selected schemes. Besides, the payload in the proposed method is 170.67 bps, which is lower than those in [52] and [41]. However, the payload of the proposed scheme is relatively high compared to the rest of the selected schemes. It is also above 20 bps which satisfy the IFPI requirement. Apart from

the security consideration, execution time of the algorithm is also an important aspect for a good watermarking scheme. The proposed approach is not computation intensive because of the adoption of LWT. The error and security analysis are also provided as well as the detailed Stirmark attack and it has blind nature. All in all, it can be concluded that the proposed algorithm not only achieves a satisfactory compromise between robustness, payload, imperceptibility and time complexity but it also meet all the IFPI requirements.

Table 5.4 Algorithm comparisons.

Ref.	Method	Synchr- nization	SNR (dB)	Payload (bps)	MOS	Blind	Audio content	Secret keys used	Embedding domain	Stirmark attack test	Error analysis reported	Security analysis provided
Bassia [30]	Spread spectrum	No	22	44.1	N/A	Yes	4 audio excerpts	Yes	Time	No	No	No
Huang [51]	DCT quantization	Yes	43.0	36	No	Yes	2 audio signals	BCH coding	DCT	No	No	No
Cvejec [49]	Spread spectrum	Yes	N/A	27.1	4.71	No	21 audio pieces	PN sequence	FFT	No	No	No
Wu [52]	DWT QIM	Yes	30.64	172.41	No	Yes	2 audio signals	No	DWT	No	Yes	No
Lie [32]	Amplitude modification	Yes	24.5	43.1	N/A	Yes	3 music signals	section length	Time	No	Yes	Yes
Wang [41]	DWT-DCT quantization	Yes	43.1	N/A	No	Yes	Music& speech	Yes	DWT- DCT	Yes	No	No
Li [48]	Spread spectrum	Yes	29.5	4.27	4.9	Yes	5 music signals	Frame length	FFT subband	Yes	Yes	No
Xiang [76]	Histogram	Yes	>40	3	4.07	Yes	6 music signals	No	Time	Yes	No	No
Xiang [42]	DWT-based Histogram	Yes	42.8	2	3.34	Yes	6 SQAM clips	PN sequence	DWT	Yes	No	No
Fan [54]	Chaos-based DFRST	No	30	86	No	Yes	6 music signals	Chaotic sequence	DRFST	No	Yes	Yes
Ercelebi [108]	LWT	No	25.93	N/A	N/A	No	9 audio signals	pseudo random sequence	Transform	No	No	No
Megias [50]	Amplitude modification	Yes	25.7	33.09	4.69	Yes	6SQAM clips & songs	Yes	Time	No	No	No
Ozer [56]	STFT-SVD	No	28.36	32	4.7	Semi	3 audio clips	No	Transform	Yes	No	No
Ah.Ali [57]	DWT-SVD	No	28.55	N/A	4.33	No	Music& speech	No	DWT transform	Yes	No	No
Abd El-Samie [77]	SVD	No	27.13	N/A	N/A	No	1 music clip	Chaotic sequence	Transform	No	No	No
Al- Nuaimy [63]	SVD	No	27.13	N/A	No	No	1 music clip	Chaotic sequence	Tansform	No	No	No
Bhat K [58]	DWT-SVD	Yes	24.37	45.9	N/A	Yes	5 music files	Yes	DWT	No	Yes	No
Wang [62]	FFT-RSVD	No	27.23	187	4.36	Yes	25 music files	No	FFT-SVD transform	No	No	No
Lei [61]	SVD-DCT	Yes	32.53	43	N/A	Yes	15SQAM clips	Chaotic sequence	Transform	Yes	Yes	No
The proposed	LWT-SVD	Yes	40	170.67	4.82	Yes	15 QAM clips& songs	Chaotic sequence	LWT- SVD	Yes	Yes	Yes

Note: N/A means not reported in the reference.

5.6 Chapter Summary

In this chapter, a very robust and blind audio watermarking scheme is proposed which makes good use of the features of SVD, LWT, synchronization code technique and QIM. The proposed method is similar to SVs modification in the low-frequency subband. The robustness of our scheme is validated by common signal processing and Stirmark attacks. The performance analyses and comparison results indicate that the proposed watermarking scheme maintains good audio quality and high robustness against various attacks, including MP3 compression, low-pass filtering, amplitude scaling, time scaling, cropping, jittering, sampling rate change, bit resolution transformation, additive noise, echo addition and equalization.

Chapter 6

An Optimal Audio Watermarking Scheme Based on Particle Swarm Optimization

6.1 Introduction

Transparency, robustness and capacity requirements are conflicting requirements that mutually influence one another. Therefore, the fundamental design issue is to achieve a good tradeoff or compromise between these requirements.

Recently, PSO has attracted increasing attention because of its simple concept, easy implementation, and quick convergence [110]. These advantages allow it to be applied successfully in a variety of fields including watermarking. PSO attempts to mimic the natural process of group communication and can be used to solve many different problems in a wide range of domains from neural network training to cost minimization. The main motivation of the classical PSO algorithm is the clever swap of information about the local and global optimal values. It leads to a global optimum by updating generations based on movement and intelligence in an evolutionary system. Compared with other

global optimization techniques, such as genetic algorithm (GA) [111-113] and adaptive tabu search (ATS) [59, 114], simplicity, shorter CPU run time and fewer parameters to adjust are the advantages of PSO. As a result, a lot of image and video watermarking schemes based on PSO have been developed and implemented [112, 115-118]. However, to the best of our knowledge, PSO has yet been applied in digital audio watermarking field for performance improvement and optimization.

In fact, the performance of the watermarking scheme can be improved by way of the artificial intelligence technique in the sense that watermarking scheme can be regarded as an optimization issue [115]. As a result, there are a lot of watermarking schemes that have been improved by SVM [119] and GA [111-113] in the literature. For example, in [120], a robust watermarking algorithm for audio signal using ATS technique in the wavelet domain is suggested by Sriyingyong and Ataakitmongcol. The tabu search is developed for preserving good audio quality and maintaining the watermark robustness. In [119], the SVM method is used to select the optimal embedding position rather than optimize the audio watermarking performance. In [121], Kumsawat *et al.* proposed an audio watermarking scheme based on digital multi-wavelet transform and GA to find the optimal watermarking quantization steps to

optimize the quality of the watermarked audio and the watermark robustness. With the GA optimization of the quantization steps, the watermarking performance is improved and optimized. The quantization steps are optimally chosen to obtain the most suitable performance with different audio properties. The experimental results also demonstrate that the watermarking performance with GA optimization obviously outperforms the performance without GA optimization. Actually, this GA based method for the audio watermarking is derived from the image watermarking technique using GA algorithm [122], which is actually based on computational intelligence optimization. However, the robustness results are not very satisfactory and the imperceptibility results are not very good as the SNR values are just slightly higher than 20dB without any attack. Moreover, there is no subjective evaluation of the watermarking scheme.

SVD has been attracting a lot of interest and becoming a focal point in the watermarking field as an important and popular transform [61, 112-113]. In the proposed scheme, the SVs are modified without changing much perceptual property of the signal according to the SVD properties. SVD related watermarking also combines this approach with recently developed stochastic and bio-inspired PSO algorithm to optimize the performance of watermarking

procedure. In this chapter, the watermark capacity is fixed, thus PSO is employed to find a tradeoff only between watermark imperceptibility and robustness. All in all, an optimal audio watermarking scheme is proposed based on SVD and PSO for audio intellectual right protection by using the popular SS algorithm in this chapter. This chapter is organized as follows. Section 6.2 introduces the related preliminaries and the proposed method. The experimental results are described in Section 6.3. Section 6.4 summarizes this study.

6.2 Methodology

6.2.1 PSO

PSO is a population based stochastic optimization technique proposed by Kennedy and Eberhart [110] to solve multidimensional optimization problems with a computational model. PSO is an algorithm inspired by social behavior of bird flocking or fish schooling. A complete theoretical analysis of this algorithm can be found in [123].

The purpose of PSO algorithm is to find the optimal solution by using a population of particles. This algorithm is initialized with a group of random particles, each of which represents a candidate solution to the problem. Let D denote the swarm size. Each individual particle i ($1 \leq i \leq D$) has its own position, p_i , and velocity, v_i . These particles search for optimal value of a given

objective function iteratively, then keep track of their individual best position p_i^{best} (pbest) and find the global position p_{gi}^{best} (gbest) from all the best positions through a search space. With respect to the two best values, the velocity and position of particle i ($1 \leq i \leq D$) in iteration $t + 1$ are updated by

$$v_i(t+1) = w_i \times v_i(t) + c_1 r_1 (p_i^{best}(t) - x_i(t)) + c_2 r_2 (p_{gi}^{best}(t) - x_i(t)), \quad (6.1)$$

$$p_i(t+1) = p_i(t) + v_i(t+1), \quad (6.2)$$

$$x_i(t+1) = x_i(t) + v_i(t+1), \quad (6.3)$$

where c_1 and c_2 are positive constants referred to as cognitive and social parameters, respectively, r_1 and r_2 are random numbers uniformly distributed in the range of $[0,1]$. w_i is a parameter called the inertia weight which controls the momentum of the particle and tunes changes in these values. It is a number in the range of $[0,1]$. Generally, w_i is denoted by [124]

$$w_i = w_{\max} - \frac{w_{\max} - w_{\min}}{iter_{\max}} \times iter, \quad (6.4)$$

where $iter_{\max}$ is the maximum number of iterations, $w_{\max}$ and $w_{\min}$ are the maximum and minimum weights, respectively, and $iter$ is the current number of iterations. These coefficients control how far a particle moves in a single iteration. v_i is the moved distance in one-step for a particle i and is limited within $[v_{\min}, v_{\max}]$, where $v_{\min}$ and $v_{\max}$ are the minimum and maximum velocity in one-step, respectively. The flowchart of PSO is illustrated in Figure 6.1. The main steps of PSO algorithm are given as follows:

Figure 6.1: The PSO flowchart.

- 1) Initialization;
- 2) Evaluation and selection;
- 3) Updating the particles;
- 4) Repeat steps 2–3 until the stopping criterion is reached.

The qualitative measure of the selection of PSO algorithm parameter can be found in [125]. From the PSO equation, it is obvious that the tradeoff is related to the PSO parameters such as $w_{\max}$, $w_{\min}$, $iter_{\max}$, c_1 and c_2 . Based on the analysis in [125], the parameter couples close to the center of the stability

triangle lead to quick convergence, while parameter couples close to its borders need many iterations to converge. Fitness evaluation goes through these processes iteratively until it reaches the desired termination criterion.

6.2.2 Watermark Preprocessing

In the recent decade, most watermarking schemes adopt watermarks which are generated from pseudo random number sequences. Recent works have also recommended the use of chaotic functions such as Markov map, Bernoulli map, skew tent map and logistic map [18, 126]. These maps have perfect dynamic properties and have big key space with good security function. In order to enhance the security, chaos is utilized to permute the embedding position of each watermark bit for the purpose of watermark encryption.

Watermark should be first preprocessed in order to improve the robustness and enhance the confidentiality. The binary image will be chaotically scrambled before embedding to increase the safety of watermarking technique. The watermark binary image is scrambled by a chaotic map which is reproduced in a permuted matrix. The chaotic map used to enhance the confidentiality of the proposed watermarking method is defined as follows

$$c_x = \begin{cases} 1-2x^2, & -1 \leq x < -0.5 \\ 1-0.5(-2x)^{1.2}, & -0.5 \leq x < 0 \\ 1-2x, & 0 \leq x < 0.5 \\ -(2x-1)^{0.7}, & 0.5 \leq x \leq 1 \end{cases} \quad (6.5)$$

Then c_x is mapped into c_n using a threshold of 0.5, that is, if $c_x > 0.5$, $c_n = 1$, or else $c_n = 0$. Therefore, c_n are the binary values which belong to $\{0,1\}$. Then the binary image logo or signature, b_n , is scrambled by c_n with the following rule

$$w_n = b_n \oplus c_n, \quad 1 \leq n \leq N_w. \quad (6.6)$$

After this random chaotic sequence encryption, the watermark is permuted and cannot be guessed by random search.

6.2.3 Watermark Embedding

In this section, the block-oriented audio watermarking scheme based on SVD is proposed instead of performing SVD on the original un-watermarked host signal as a whole. Block-oriented SVD technique partitions the cover object signal as a whole. Figure 6.2 shows the watermark embedding procedure. The watermark embedding procedure can be described as follows:

Figure 6.2: Diagram of the watermark embedding process.

Step 1: Reshape the original 1-D cover audio signal to 2-D host signal I for the

block-oriented method.

Step 2: Partition the 2-D host signal I into blocks I_{ij} .

Step 3: Carry out DWT on each block to transform them into different low frequency and high frequency subbands.

Step 4: Perform SVD on the low frequency subband

$$I_{ij} = U_{ij} D_{ij} V_{ij}^T. \quad (6.7)$$

Step 5: Embed one bit of the watermark in each block:

$$I_{ij}^w = U_{ij} D_{ij} (1 + k w_{ij}) V_{ij}^T. \quad (6.8)$$

Here, w_{ij} is the watermark bit embedded into block D_{ij} and $w_{ij} \in \{0, 1\}$. k is the SF or watermark strength of the embedded signal which will be optimized in the PSO training. In the proposed method, the popular SS watermarking method is adopted. Therefore, all the advantages of the SS method hold here. The selections of watermark strength can tradeoff watermarking fidelity and robustness, thus increasing the flexibility of this algorithm. In addition, one can embed watermark bits $w_{ij} = 1$ in blocks I_{ij} which have larger SVs while embedding the zero bits in blocks with smaller SVs.

6.2.4 Watermark Extraction

Figure 6.3 gives a block diagram of the watermark extraction procedure. The extraction sequence consists of the following steps:

Figure 6.3: Diagram of the watermark extraction process.

Step 1: Rearrange the watermarked and potentially attacked 1-D audio signal to 2-D signal.

Step 2: Divide the 2-D signal into blocks in the same fashion used for embedding the original audio.

Step 3: Carry out DWT on each block to transform them into different low frequency and high frequency subbands.

Step 4: Perform SVD on the low frequency subbands

$$I_{ij}^* = U_{ij}^* D_{ij}^* V_{ij}^{*T}. \quad (6.9)$$

Step 5: The extracted watermark bit is obtained by

$$w'_{ij} = \frac{1}{kD_{ij}^*} (U_{ij}^{*T} I_{ij}^w V_{ij}^* - D_{ij}^*). \quad (6.10)$$

6.2.5 Performance Improvement Using PSO

Generally, the imperceptivity and robustness are mutually contradictory, thus development of a watermarking system needs optimization of these two requirements to improve performance. PSO approach can serve this purpose in that it searches for optimal points in order to achieve the optimum performance. As to the performance improvement based on PSO, two issues need to be considered. Firstly, the fitness function (also known as evaluation function or

objective function) should be well designed. Secondly, the parameters related to the PSO techniques should be carefully selected.

As can be seen from the PSO equations (Eq. (6.1)-(6.3)), different velocities and initial values can be defined for each particle. Thus the optimization algorithm can be run with random velocities and initial values as the velocities are generated from random initial values in Eq. (6.1). As for verifying the robustness of watermarking, applying attacks to the watermarked cover object is necessary. Therefore, the attacks in the optimization problem need to be properly selected so that the attacked signal still preserves its significance and practical value after attacks. Different attacks may have different effects on the host signal. For audio watermarking, resampling, MP3 compression, cropping and additive noise are more frequently encountered in attacks than echo addition and delay. Hence these four kinds of attacks, namely, resampling, MP3 compression at 64 kbps, cropping and additive noise are selected in the fitness evaluation function. Furthermore, their coefficients are also taken into consideration when designing the final fitness function. The PSO algorithm tries to find the places which are more robust to withstand the attack. Normally, the fitness function should be the function of both robustness and imperceptibility, which is usually denoted as

$$fitness = f(imperceptibility, robustness). \quad (6.11)$$

In the proposed method, NC defined in Eq. (3.22) is used to measure the robustness. Moreover, the SNR, SSNR and LLR metric [104] are also used for measuring the audio inaudibility.

The four above mentioned attacks: resampling, MP3 compression, cropping and additive noise attacks, are also added to the training procedure as shown in Figure 6.4, each of which corresponds to a NC value. PSO overcomes the local optimal problem by using a fitness function because it chooses the highest evaluation solution at each iteration. From the above discussions, it is obvious that better imperceptibility, more robustness and reasonable capacity are all required for designing the watermarking algorithm. Given the preliminaries in Section 6.1, the watermark capacity is fixed so that both the watermark imperceptibility and watermark robustness are improved after PSO.

Figure 6.4: Watermarking method with PSO training.

As we know, the main objective of developing an audio watermarking technique is to achieve both imperceptibility and robustness requirements. To obtain this objective, the fitness function is defined as the function of imperceptibility and robustness within PSO-based methods. Because SNR and SSNR values are usually more than 20 dB, parameters, α and η , are used to weight the effects. 20 dB is adopted as a balancing value. Another weighting factors β , is also employed to be included in the fitness function to weight NC values. Moreover, the weighting factors, δ is used to weight LLR metric. λ_i is also introduced into the fitness function to balance the four mentioned attacks.

Actually, the final fitness function in the proposed method is defined as

$$fitness = \alpha(SNR - 20) + \beta \sum_{i=1}^4 \lambda_i NC_i + \eta(SSNR - 20) + \delta \times LLR, \quad (6.12)$$

where SNR , $SSNR$ and LLR represent the quality of the watermarked audios and NC represents the watermark robustness. All the weighting parameters are set in order to balance the effects and contributions from these factors. Though there are some similarities between the proposed algorithm and the existing methods [117, 121], the proposed approach still has significant innovation. The fitness function also shares some similar concepts for optimizing different goals.

During iteration, all particles will be updated according to the PSO

formulas. This PSO implementation allows us to have better parameter selection for use in realtime applications. Objectively, higher SNR, SSNR and LLR values lead to better imperceptibility and higher NC values lead to better robustness. In fact, the goal of the optimization algorithm is to find the maximum value in Eq. (6.12).

Furthermore, reasonable computation time or rational number of iterations is an important factor too. To perform this task, the parameters should be tuned properly by taking advantage of a PSO cycle. In fact, the parameter selection of PSO algorithm is mostly rooted in experience rather than pure logic despite recent research efforts. It is easy to run PSO algorithm with different coefficients several times and finally obtain the best parameters (until convergence). In fact, when different results are obtained in most runs, or when the same result is obtained consistently, but no better points are found after exhaustive iterations, the convergence rate is not suitable. In this way, a slower or faster convergence rate should be selected. On the other hand, if consistent results are achieved in most runs with continuous improvement over iterations, then the convergence parameter is desirable and that parameter set should be used in the optimization process, which is further verified in the simulation.

Furthermore, there is also a need to choose the number of particles

carefully. If too many particles are chosen, more places can be searched to find the global optimal solution easily. However, it will increase the computation time per iteration and may result in memory allocation problems. On the contrary, if too few particles are chosen, the output result might only be a locally optimal one, which is a problem generally encountered in optimization. The reasonable number of particles will be validated by the simulation results.

6.2.6 False Positive Error Analysis

The watermark detection threshold, τ_{nc} , can be commonly determined by false positive error rate. For convenience of the analysis, suppose that the distribution of detection values obtained from unwatermarked signals is approximately Gaussian distributed. According to the false positive analysis of normalized correlation proposed in Cox *et al.*'s book [1], the standard deviation, σ_{nc} , is derived as

$$\sigma_{nc} = \frac{|w'| \sigma_w}{|w| |w'|} = \frac{1}{\sqrt{N_w}}, \quad (6.13)$$

where w is the original watermark and w' is the extracted watermark. σ_w is the standard deviation of w , which is equal to 1. N_w is the length of w . By the Gaussian assumption of the normalized correlation value, the probability of the false alarm in the local host signal P_{FA} can be derived as

$$P_{FA} \approx \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{N_w}{2}} \tau_{nc} \right), \quad (6.14)$$

where $\text{erfc}(\bullet)$ is complementary error function. By viewing the watermarked system as an independent communication channel, the existence of watermark is claimed. Then the false alarm probability, P_{FA} , can be calculated as

$$P_{FA} = \sum_{i=K}^{N_w} \binom{N_w}{i} P_E^i (1 - P_E)^{N_w - i} = 1 - (1 - P_E)^{N_w}, \quad (6.15)$$

where N_w denotes the watermark length, P_E is probability of $w \neq w'$. From the above derivation, it can be seen that the false alarm probability P_{FA} for a host signal depends not only on the correlation detection threshold, τ_{nc} , but also on K and N_w . $N_w = 1024$ in the experiment. According to the watermarking standard, the desired false alarm rate must be smaller than 10^{-8} . Therefore, P_{FA} is set at 0.5×10^{-8} , where the correlation detection threshold, τ_{nc} , is calculated as 0.68 accordingly.

6.3 Experimental Results

In this section, some experiments are conducted to demonstrate the effectiveness of the proposed digital watermarking scheme. The inaudibility and robustness of the watermark are validated by watermarking the original audio with the parameters optimized from PSO training. Experiments are performed to evaluate the watermarked audio quality and the robustness of the watermark under typical attacks. The parameters of PSO training and

watermarking should be specified. As mentioned before, the parameters are important in PSO training. The optimal parameters should be found according to the results of many simulation experiments, the statistical analysis of the parameters and algorithm performance. In the proposed method, the weighting factors are set based on numerous experimental results and validated empirically, which is closely related to the tradeoff between robustness and imperceptibility. Besides, the SNR, SSNR, LLR and NC are equally important. Therefore, the weighting factors are all set to 1 (i.e., $\alpha = \beta = \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \eta = \delta = 1$). In the experiment, a 32×32 binary image is used as watermark, thus N_w is 1024. A total of 12 test sequences (16 bits/sample, 48 kHz and WAV format) including classical, instrumental and popular music are adopted to analyze the watermarking algorithm's performance, which are listed in Table 3.1. For easy and fair comparison, the duration of the test signal is 30s. The capacity of the proposed method is 34.13 bps.

6.3.1 Results of PSO Training

In the PSO training, the attacks in the training are first specified. In the simulation, the resampling in the training process is first downsampling to 8 kHz and then upsampling to 48 kHz. MP3 compression bitrate is 64 kbps. The

cutting attack is cutting 10% of the samples from 5 random positions. Additive noise attack is performed by adding white Gaussian noise at 4% of the power of the audio signal.

In order to analyze PSO training parameters effect on optimal performance of the watermarking algorithm, simulations must be conducted with different cognitive and social parameters (c_1 and c_2) and inertia weights ($w_{\max}$). In this simulation, $w_{\min} = 0.1$ is set and different values $w_{\max}$ are used to do the performance evaluation. Table 6.1 shows the fitness value (Fit) and SF under the effect of different cognitive and social parameters (c_1 and c_2) and inertia weights ($w_{\max}$) on the avemaria audio signal. Table 6.2 tabulates the values of Fit and SF under the effect of different cognitive and social parameters (c_1 and c_2) and inertia weights ($w_{\max}$) on the cymbal audio signal. In these tables, the upper limit for the number of maximum iteration and particles are chosen to be 50 and 30, respectively. From Tables 6.1 and 6.2, it is found that $w_{\max} = 0.5$, $c_1 = c_2 = 1.8$ are good parameter settings for this optimization process and achieve the optimal performance of the proposed scheme in both avemaria and cymbal audio signals, where the highest values of Fit occur. Therefore, $w_{\max} = 0.5$, $c_1 = c_2 = 1.8$ is chosen as the best parameters in the PSO training and used in all the test sequences. Table 6.3 and 6.4 shows the optimal

values of SF, Fit, SNR, SSNR, LLR and NC_m with different particles of the avemaria and cymbal audio signals. It is shown that with the increase of the particles the optimal values approach to a stationary values. Therefore, 20 particles are used in our scheme to lower the computation time and speedup the whole system without affecting the performance.

Table 6.1 The effect of the cognitive coefficients and inertia weights of PSO algorithm on the avemaria audio signal.

Inertia weight (w_{max})	Cognitive and social parameters											
	$c_1 = c_2 = 1.0$		$c_1 = c_2 = 1.2$		$c_1 = c_2 = 1.4$		$c_1 = c_2 = 1.6$		$c_1 = c_2 = 1.8$		$c_1 = c_2 = 2.0$	
	SF	Fit	SF	Fit	SF	Fit	SF	Fit	SF	Fit	SF	Fit
0.9	0.1710	14.8190	0.1710	14.8190	0.1710	14.8191	0.1710	14.8190	0.1710	14.8190	0.1710	14.8178
0.8	0.1710	14.8190	0.1710	14.8190	0.1710	14.8190	0.1710	14.8191	0.1710	14.8191	0.1710	14.8203
0.7	0.1710	14.8190	0.1710	14.8190	0.1710	14.8190	0.1710	14.8190	0.1710	14.8190	0.1710	14.8190
0.6	0.1710	14.8190	0.1710	14.8190	0.1710	14.8191	0.1710	14.8190	0.1710	14.8190	0.1710	14.8201
0.5	0.1710	14.8190	0.1710	14.8190	0.1710	14.8192	0.1710	14.8191	0.1710	14.8402	0.1709	14.8259
0.4	0.1710	14.8190	0.1710	14.8192	0.1710	14.8188	0.1711	14.8113	0.1711	14.8183	0.3695	8.9461
0.3	0.1711	14.8136	0.1711	14.8102	0.1710	14.8215	0.1712	14.7994	0.1724	14.6912	0.4860	1.2087
0.2	0.1712	14.8113	0.1788	14.0107	0.2693	11.9672	0.2857	7.7143	0.6253	-2.6878	0.4436	2.2846
0.1	0.2920	7.4409	0.3668	4.5731	0.3560	4.9392	0.2575	9.0918	0.4820	1.3026	0.4856	1.2184

Table 6. 2 The effect of the cognitive coefficients and inertia weights of PSO algorithm on the cymbal audio signal.

Inertia weight (w_{max})	Cognitive and social parameters											
	$c_1 = c_2 = 1.0$		$c_1 = c_2 = 1.2$		$c_1 = c_2 = 1.4$		$c_1 = c_2 = 1.6$		$c_1 = c_2 = 1.8$		$c_1 = c_2 = 2.0$	
	SF	Fit	SF	Fit	SF	Fit	SF	Fit	SF	Fit	SF	Fit
0.9	0.4590	14.9446	0.4590	14.9446	0.4590	14.9446	0.4590	14.9446	0.4591	14.9397	0.4676	14.6916
0.8	0.4590	14.9446	0.4590	14.9446	0.4590	14.9446	0.4590	14.9446	0.4590	14.9440	0.4601	14.9117
0.7	0.4590	14.9446	0.4590	14.9446	0.4590	14.9446	0.4590	14.9446	0.4590	14.9443	0.4590	14.9446
0.6	0.4590	14.9446	0.4590	14.9446	0.4590	14.9446	0.4590	14.9446	0.4590	14.9446	0.4588	14.9486
0.5	0.4590	14.9446	0.4590	14.9446	0.4590	14.9446	0.4589	14.9453	0.4590	14.9790	0.5042	13.6821
0.4	0.4590	14.9446	0.4589	14.9454	0.4590	14.9446	0.4586	14.9542	0.4593	14.9343	0.6427	10.5655
0.3	0.4589	14.9453	0.4591	14.9417	0.4591	14.9410	0.4588	14.9494	0.4632	14.8205	0.7684	8.3838
0.2	0.5656	12.1783	0.4591	14.9422	0.4635	14.8123	0.4804	14.3273	0.6260	10.8937	0.7961	7.9608
0.1	0.8149	7.6747	0.7037	9.4470	0.5196	13.2860	0.9369	6.0550	0.6638	10.1645	0.8482	7.2134

Table 6.3 Optimal values under different particles (avemaria.wav).

Particles	Fit	SF	SNR	SSNR	LLR	NC _m
10	14.8378	0.1710	34.0834	34.5137	0.0402	0.9982
20	14.8383	0.1710	34.0840	34.5142	0.0402	0.9982
30	14.8402	0.1710	34.0847	34.5148	0.0402	0.9982
40	14.8402	0.1710	34.0848	34.5149	0.0402	0.9982
50	14.8402	0.1710	34.0848	35.5149	0.0402	0.9982

Table 6.4 Optimal values under different particles (cymbal.wav).

Particles	Fit	SF	SNR	SSNR	LLR	NC _m
10	14.9729	0.4590	31.2451	37.6238	0.0123	0.9961
20	14.9786	0.4590	31.2495	37.6279	0.0123	0.9961
30	14.9790	0.4590	31.2512	37.6295	0.0123	0.9961
40	14.9790	0.4590	31.2512	37.6295	0.0123	0.9961
50	14.9790	0.4590	31.2564	37.6344	0.0123	0.9961

Figure 6.5 and Figure 6.6 show SF, SNR, SSNR, NC_m and LLR values of avemaria and cymbal audio signals with PSO training using 30 particles and 50 iterations, which also shows the convergence of PSO training after 50 iterations of the avemaria and cymbal audio signals. It is noted that the SF, SNR, SSNR, NC_m and LLR values gradually converge to fixed values as the number of iteration increases, which confirms the effect of PSO training. These parameters are optimally varied to achieve the most desirable effect for the original audios according to their different characteristics. Besides, it should be noted that 50 iterations are enough to achieve the optimal values from these figures. Therefore, 50 iterations are used in the simulation.

(a)

(b)

Figure 6.5: PSO training results of avemaria audio signal; (a) Optimal SF results (optimal SF =0.171); (b) Optimal SNR and SSNR results (optimal SNR=34.0847 and optimal SSNR=34.5148); (c) Optimal NC_m results (optimal NC_m =0.9982); (d) Optimal LLR results (optimal LLR=0.0402).

Figure 6.6: PSO training results of cymbal audio signal; (a) Optimal SF results (optimal SF =0.4590);(b) Optimal SNR and SSNR results (optimal SNR=31.2512) and optimal SSNR=37.6295); (c) Optimal NC_m results (optimal NC_m =0.9961); (d) Optimal LLR results (optimal LLR=0.0123).

Table 6.5 shows the simulation results of the Fit, SF, SNR, SSNR, LLR

and average NC of the four attacks (NC_m) by applying the PSO-based watermarking scheme with 30 particles and 50 iterations. Table 6.6 tabulates the NC_m , NC_1 (resampling), NC_2 (MP3 compression), NC_3 (Cropping) and NC_4 (additive noise) results after the PSO training. The good NC results (average value of NC_m of the 12 test audio signals is 0.9978) verify the robustness of the proposed method.

Table 6.5 Optimal parameters after PSO of 12 test audios.

Index	Fit	SF	SNR	SSNR	LLR	NC_m
1	14.8402	0.1710	34.0847	34.5148	0.0402	0.9982
2	14.8774	0.2724	34.2231	34.5119	0.0267	0.9980
3	14.7242	0.9145	33.3379	35.0611	0.0603	0.9980
4	14.7384	0.2553	34.2124	34.2051	0.0584	0.9983
5	14.9790	0.4590	31.2512	37.6295	0.0123	0.9961
6	14.9653	0.0540	32.2747	36.6745	0.0062	0.9931
7	14.8719	0.0778	34.3143	34.4124	0.0277	0.9962
8	14.8009	0.8233	32.8335	35.7346	0.0433	0.9984
9	14.6617	0.4447	34.1195	34.1444	0.0736	0.9997
10	14.7066	0.2623	34.2605	34.0970	0.0643	0.9989
11	14.8511	0.2729	34.2848	34.3933	0.0322	0.9999
12	14.8189	0.1766	33.4512	35.1541	0.0397	0.9982
Average	14.8196	0.3486	33.5540	35.0444	0.0404	0.9978

Table 6.6 NC results of 12 test audios.

Index	NC_m	NC_1	NC_2	NC_3	NC_4
1	0.9982	1	1	0.9978	0.9948
2	0.9980	1	1	1	0.9922
3	0.9980	1	1	1	0.9922
4	0.9983	1	1	0.9995	0.9936
5	0.9961	0.9922	1	1	0.9922
6	0.9931	0.9812	1	0.9921	0.9994
7	0.9962	1	1	0.9896	0.9952
8	0.9984	1	1	1	0.9936
9	0.9997	1	1	1	0.9988
10	0.9989	1	1	1	0.9957
11	0.9999	1	1	1	0.9994
12	0.9982	1	1	0.9940	0.9936
Average	0.9978	0.9978	1	0.9977	0.9951

6.3.2 Imperceptibility

Figure 6.7 plots the time domain waveform of the original and watermarked signal of the two representative test signals, avemaria and cymbal. The little visual difference between the original and watermarked signal indicates that the transparency of the proposed method is very good. Besides, SNR of the two signals are as high as 31.2512 dB and 34.0847 dB, which is relatively high and can demonstrate good inaudibility in the sense of objective evaluation.

As mentioned above, SNR, SSNR and LLR are objective metrics to measure the imperceptivity. However, these classical objective measures do not employ the characteristics of the HAS. The perceptual audio quality measure (PAQM) [127], developed by PTT Research in 1992, is considered as an audio quality measure and has been presented to measure the audio quality based on a psychoacoustic representation. In [127], it shows that PAQM yields improved correlation with the subjectively perceived quality as compared to other perceptual audio quality models. Thus the SDG, such as “0” (insensitive), “1” (audible), “2” (slightly annoying), “3” (annoying), “4” (very annoying), and “5” (catastrophic), is used in the proposed algorithm to measure the perceptual performance as well. The listening test involving 10 listeners is carried out to obtain the SDG. Table 6.7 shows the SDG results of the test signals. It can be seen that the perceptual quality of the proposed method is very good.

Figure 6.7: Waveforms and spectrograms of the original and watermarked audio; (a) avemaria audio signal (SNR=34.0847dB, SSNR=34.5148dB, LLR=0.0267);(b)cymbal audio signal (SNR=31.2512dB, SSNR=37.6295dB, LLR=0.0123); (c)spectrogram of avemaria audio signal; (d) spectrogram of cymbal audio signal.

Table 6.7 Perceptual quality of 12 test audio signals.

Index	SDG	Index	SDG	Index	SDG
1	0.29	5	0.45	9	0.29
2	0.28	6	0.41	10	0.28
3	0.32	7	0.27	11	0.21
4	0.28	8	0.38	12	0.26

6.3.3 Robustness

The robustness of the proposed algorithm is tested under typical signal

processing manipulations. The attacks are selected to demonstrate the capabilities and shortcomings of the scheme and include some attacks that are not adopted in the training. Moreover, the robustness of the proposed algorithm is compared with the ATS method in [120] and GA method in [121] as they both have the optimization technique to improve audio watermarking performance. For fair comparison, the parameter specifications of the common signal processing attacks are the same as that reported in [121].

It should be noted that the attack specification of the four attacks in the training is slightly different from the above specification. The cropping, additive noise and resampling attacks in the training are stronger than the above specification. The robustness test results of the above mentioned various attacks are shown in Table 6.8, which also displays BER (%) results from the watermark extraction. It can be noted that the proposed method is very robust against these attacks. Besides, the proposed method is also compared with the methods in [120] and [121]. From the algorithm comparison results, it can be noted that both the algorithm in [121] and the proposed algorithm are very robust to resampling, requantization, additive noise and MP3 compression at 96, 64 and 56kbps and the latter outperform the algorithm in [120]. However, the proposed method is more robust against the MP3 compression, additive

noise, resampling and cropping attacks than the GA algorithm in [120] and [121], as the BER (%) results for the proposed algorithm are 0, while the BER (%) results in [120] and [121] are non-zero. From the robustness test results, it is obvious that the robustness results of the proposed method are very good.

Table 6.8 Robustness test and algorithm comparison.

Attacks	No attack	Resampling	Requantization
ATS method in [120]
BER(%)	0	0	0.1953
GA method in [121]
BER(%)	0	0	0
The proposed
BER(%)	0	0	0
Attacks	LPF	Additive noise	Cropping
ATS method in [120]
BER	3.3203	2.6367	5.3711
GA method in [121]
BER (%)	1.1719	0	0.1953
The proposed
BER(%)	0	0	0
Attacks	MP3 96kbps	MP3 64kbps	MP3 56kbps
ATS method in [120]
BER	4.1013	5.3711	5.7617

GA method in [121]	EEE EEE	EEE EEE	EEE EEE
BER(%)	0	0	0
The proposed	EEE EEE	EEE EEE	EEE EEE
BER(%)	0	0	0

Table 6.9 shows the rough comparison of the proposed method with PSO (after PSO training) and without PSO (before PSO training). It should be noted that “with PSO” method means the proposed method, while “without PSO” indicates that the scheme employs fixed SF for all the test sequences. It can be observed that the average NC results after all these attacks are 0.9978 and 1, respectively, and BER (%) results after all these attacks are 0.3574 and 0, respectively. It should be noted that the NC results here are different from NC_m results in the PSO training as the parameter specifications are different. Based on these results, it can be concluded that the proposed method with PSO achieves more satisfactory results for the NC and BER (%) metrics, which validate that PSO training can improve the performance. Actually, Table 6.9 shows that the proposed method achieves more satisfactory results in terms of robustness as compared to the method without PSO.

Table 6.9 Robustness results with and without PSO.

Attack type	Without PSO		With PSO	
	NC	BER (%)	NC	BER (%)
Attack free	1	0	1	0
Resampling	1	0	1	0
Requantization	1	0	1	0
Low-pass filtering	1	0	1	0
Additive noise	0.9977	0.1564	1	0
Cropping	0.9998	0.1247	1	0
MP3-56kbps	0.9977	0.1443	1	0
MP3-64kbps	0.9825	3.1481	1	0
MP3-96kbps	1	0	1	0
MP3-128kbps	1	0	1	0
Average	0.9978	0.3574	1	0

6.3.4 Algorithm Comparison

Table 6.10 shows the rough comparison of the proposed approach with ATS approach in [120] and GA approach in [121]. It can be seen that the proposed method clearly outperforms ATS approach in [120] and GA approach in [121], as the average values of SNR and NC are higher than those in [120] and [121] while BER (%) is lower than that in [120] and [121]. The reason is that PSO technique can achieve global optimal values and has fast convergence. Moreover, the use of SVD method in the transform domain enhances the robustness of the proposed method. The modification of the SVs of the original audio signal can achieve good transparency at the same time.

6.4 Chapter Summary

In this chapter, an optimal audio watermarking scheme combining PSO and

SVD with chaotic map is proposed. The proposed method adopts SS method to embed the watermark by modifying the SVs of the host signal. The robustness and imperceptibility of the watermark is improved by PSO technique. Moreover, the experimental results demonstrate that the proposed method is very robust against a large number of attacks such as additive noise, resampling, MP3 compression, requantization and cropping while imperceptibility is maintained. Moreover, the watermarked audio retains good quality with SNR values larger than 30dB. Finally, the algorithm comparison results also demonstrate that the proposed scheme outperforms the schemes developed by GA method and ATS optimization techniques in detection and imperceptibility results with the same capacity.

Table 6.10 Comparison results with ATS method and GA method.

Attack type	ATS [120]			GA [121]			The proposed		
	SNR	NC	BER (%)	SNR	NC	BER (%)	SNR	NC	BER (%)
Attack free	24.1209	1	0	24.5518	1	0	33.55	1	0
Resampling	23.5533	1	0	23.5052	1	0	31.67	1	0
Requantization	20.0936	0.9982	0.1953	20.9320	1	0	30.25	1	0
LPF	10.3506	0.9810	3.3203	10.3884	0.9393	6.9336	20.42	1	0
Additive noise	19.5607	0.9669	2.6367	19.5684	1	0	28.16	1	0
Cropping	10.8289	0.9447	5.3711	16.3075	0.9983	0.1953	19.41	1	0
MP3-56kbps	18.5400	0.9787	3.6133	18.6666	0.9992	0.0977	27.42	1	0
MP3-64kbps	18.5348	0.9589	4.1016	18.8265	1	0	27.83	1	0
MP3-96kbps	19.5537	0.9488	5.3711	19.5920	1	0	28.48	1	0
MP3-128kbps	19.5607	0.9425	5.7617	19.7552	1	0	28.99	1	0
Ave.	18.4697	0.9720	3.0371	19.2094	0.9937	0.7227	27.62	1	0

Chapter 7

Conclusions and Recommendations

7.1 Conclusions

In this thesis, there are four main parts of original research. The conclusions of this thesis are as follows:

Firstly, an audio watermarking scheme with high robustness against malicious attacks is put forward for MPEG-4 SLS audio. Significant state is adopted to select embedding points for random embedding position, and then the selected data is spread by chaotic sequence to increase robustness against watermark attacks during the embedding process. In addition, the masking characteristics of psychoacoustic model are used to avoid introducing audible noise in the MPEG-4 SLS compression, which is validated by listening tests. Consequently, the proposed method has higher robustness against malicious attack. The performance analyses and simulation results show that the proposed scheme not only guarantees robustness against common operations attacks but also keeps the watermark imperceptible and inaudible.

Secondly, a new and effective hybrid SVD-DCT domain based watermarking scheme for audio copyright protection is presented, where the

watermark bits are embedded in the largest SVs of the original audio by modifying their high frequency DCT coefficients. The frequency mask for invisible watermarking is modeled according to the statistics of SV blocks and is thus adaptive to the individual audio. The proposed scheme exhibits high robustness and imperceptibility performance, and provides high security through use of chaotic sequence encryption. The watermark detection is also efficient and blind, i.e., only the secret key rather than the original host audio is needed. Moreover, the adoption of the synchronization code technique renders the proposed approach very robust to desynchronization attacks. The experiments show that this method is highly robust against general signal processing attacks such as MP3 compression, cropping and resampling. The robustness of the proposed watermarking scheme against various attacks is much higher than the state-of-the-art SVD-based method and demonstrates its potential application in covert communication. This novel scheme can also be applied to a variety of common network multimedia, e.g., image, video and 3D mesh.

Thirdly, a novel and oblivious audio watermarking scheme based on LWT and SVD is proposed. With QIM quantization method, SVs of the low frequency LWT coefficients are modified to insert the watermark data. The

synchronization and chaotic signals are adopted in the proposed scheme to improve the performance. Experimental results demonstrate that the proposed LWT-SVD method is not only robust to both general signal processing and desynchronization attacks but also outperforms the selected previous studies.

Lastly, unlike the above three mentioned audio watermarking schemes that have a fixed and maybe non-optimized parameters, the PSO technique is introduced to improve and optimize the tradeoff between robustness and imperceptibility. The SS method is adopted to embed the watermark using the optimized watermark strength. The experimental results validate the practicality of the proposed method and show that the proposed algorithm yields a watermark that is visually recognizable and very robust to typical signal manipulations. Furthermore, the results of comparison with other optimization techniques indicate that the proposed algorithm is more robust against most common attacks.

7.2 Recommendations for Further Research

In this section, a few unresolved issues are brought up and some possible research directions are proposed.

Apart from the time and transform domain methods that are implemented in this thesis, the use of other transforms can also be investigated. Promising

transforms for robust audio/speech watermarking are cepstrum domain, discrete fractional Fourier/sine/cosine transform and hybrid transform schemes such as SVD-DWT-DCT, LWT-DCT, SVD-cepstrum. Hybrid transforms may provide more security and better performance as the strong points of different transform techniques can be integrated.

Other intelligent systems such as genetic algorithm and differential evolution algorithm can also be used to further improve and optimize the watermarking scheme. One possible direction of research will be to combine it with improved DM and QIM methods.

Finally, the use of error correction codes for digital audio watermarking should be investigated. Error correction codes are first proposed for protecting digital data against noisy transmission lines. Extra bits are introduced so that transmission errors can be detected and even corrected in some cases. Future research should look into what kind of error correction codes are more robust against the various types of attacks that are commonly encountered in audio watermarking. The optimal way to embed the error correction codes in the digital bitstream is also an interesting topic to examine.

List of Author's Publications

Journal Papers

- [1] **B. Lei**, I. Y. Soon, and Z. Li, "Blind and robust audio watermarking scheme based on SVD-DCT," *Signal Processing*, vol. 91, pp. 1973-1984, 2011.
- [2] **B. Lei**, I. Y. Soon, F. Zhou and Z. Li, "A Robust Audio Watermarking Scheme Based on Lifting Wavelet Transform and Singular Value Decomposition," *Signal Processing*, vol. 92, no.9, pp. 1985–2001, 2012.
- [3] **B. Lei** and I. Y. Soon, "A New Multipurpose Audio Watermarking Scheme Algorithm with Synchronization and Encryption," *Journal of Zhejiang University SCIENCE C (Computers & Electronics)* , vol.13, no.1, pp.11-19, 2012.
- [4] **B. Lei** and I. Y. Soon, "Comments on "an SVD audio watermarking approach using chaotic encrypted images", *Digital Signal Processing, A Review Journal*, under revision, 2012.
- [5] **B. Lei** and I. Y. Soon, "Robust SVD-based Audio Watermarking Scheme with Differential Evolution Optimization", *IEEE Transactions on Audio, Speech and Language Processing*, under revision, 2012.

Conference Papers

- [1] **B. Lei**, I. Y. Soon, F. Zhou, and Z. Li, "An Optimal Audio Watermarking Scheme Based on Particle Swarm Optimization," in *Proc. of Eighth Int'l Conference on Information, Communications, and Signal Processing* (ICICS 2011), pp.1-4, Singapore, 13-16, Dec. 2011.
- [2] **B. Lei**, I.Y. Soon, F. Zhou, and Z. Li, "A Robust Audio Watermarking Scheme Based on Lifting Wavelet Transform and Singular Value Decomposition," in *Proc. of 10th Int'l Workshop on Digital-forensics and Watermarking* (IWDW2011), pp.86-96, Atlantic City, New Jersey, USA, 23-26 Oct. 2011.
- [3] Z. Li, K. H. Yap, and **B. Lei**, "An Oblivious Robust Image Watermarking Scheme in the SVD-DCT Domain," in *Proc. of Int'l Conference on Image Processing* (ICIP 2011), pp.2813-2816, Brussels, Belgium, 11- 14, Sep. 2011.
- [4] **B. Lei** and I. Y. Soon, "A Perception-based Audio Watermarking Scheme," in *Proc. of IEEE Region 10 Conference on TENCON* (TENCON 2009), pp.1-4, Singapore, 23-26, Nov. 2009.
- [5] **B. Lei** and I. Y. Soon, "Audio watermarking for DRM based on chaotic map", in *Proc. of Int'l Conference on Computer Engineering and Technology* (ICCET2009), pp.55-59, Singapore, 22-24, Jan. 2009.

- [6] **B. Lei** and I. Y. Soon, "Audio Content Protection Based on Integer Transform", in *Proc. of IEEE Int'l Conference on Communications, Circuits and Systems (ICCCAS 2009)*, pp. 489-492, Chengdu, China, 23-25, Jul. 2009.

References

- [1] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*: Morgan Kaufmann Publishers, 2002.
- [2] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Systems Journal*, vol. 35, pp. 313-36, 1996.
- [3] M. D. Swanson, B. Zhu, A. H. Tewfik, and L. Boney, "Robust audio watermarking using perceptual masking," *Signal Processing*, vol. 66, pp. 337-355, 1998.
- [4] G. A. R. A. Z. Tirkel, R. M. van Schyndel, W. J. Ho, N. R. A. Mee, and C. F. Osborne, , "Electronic Watermark," in *Proc.of Digital Image Computing: Techniques and Applications*, 1993.
- [5] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Information hiding-a survey," *Proceedings of the IEEE*, vol. 87, pp. 1062-1078, 1999.
- [6] M. Arnold, "Audio watermarking: features, applications and algorithms," in *Proc.of IEEE Int'l Conference on Multimedia and Expo*, 2000, pp. 1013-1016.

- [7] L. Boney, A. H. Tewfik, and K. N. Hamdy, "Digital watermarks for audio signals," in *Proc. of the Int'l Conference on Multimedia Computing and Systems*, 1996, pp. 473-480.
- [8] I. J. Cox and M. L. Miller, "The first 50 years of electronic watermarking," *Eurasip Journal on Applied Signal Processing*, pp. 126-132, 2002.
- [9] I. J. Cox, M. L. Miller, J. A. Bloom, J. Fridrich, and T. Kalker, "Digital Watermarking and Steganography," *Morgan Kaufmann*, 2008.
- [10] X. He, "Watermarking in Audio: Key Techniques and Technologies," *Cambria Press*, 2008.
- [11] I. J. Cox, J. Kilian, T. Leighton, and T. Shamon, "Secure spread spectrum watermarking for images, audio and video," in *Proc. of Int'l Conference on Image Processing*, 1996, pp. 243-246
- [12] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, vol. 6, pp. 1673-1687, 1997.
- [13] A. Giovanardi, G. Mazzini, and M. Tomassetti, "Chaos based audio watermarking with MPEG psychoacoustic model I," in *Proc. of Joint Conference of the Fourth International Conference on Information,*

- Communications and the Fourth Pacific Rim Conference on MultimediaSignal Processing* 2003, pp. 1609-1613 vol.3.
- [14] G. C. M. Silvestre, N. J. Hurley, G. S. Hanau, and W. J. Dowling, "Informed audio watermarking scheme using digital chaotic signals," in *Proc. of IEEE Int'l Conference on Acoustics, Speech, and Signal Processing*, 2001, pp. 1361-4.
- [15] H. Xiang, L. Wang, H. Lin, and J. Shi, "Digital watermarking systems with chaotic sequences," in *Proc. of the Int'l Society for Optical Engineering*, USA, 1999, pp. 449-457.
- [16] S. Tsekeridou, V. Solachidis, N. Nikolaidis, A. Nikolaidis, A. Tefas, and I. Pitas, "Bernoulli shift generated chaotic watermarks: Theoretic investigation," in *Proc. of IEEE Int'l Conference on Acoustics, Speech and Signal Processing* 2001, pp. 1989-1992.
- [17] A. Tefas, A. Nikolaidis, N. Nikolaidis, V. Solachidis, S. Tsekeridou, and I. Pitas, "Markov chaotic sequences for correlation based watermarking schemes," in *Chaos, Solitons and Fractals*, 2003, pp. 567-573.
- [18] A. Tefas, A. Nikolaidis, N. Nikolaidis, V. Solachidis, S. Tsekeridou, and I. Pitas, "Performance analysis of correlation-based watermarking

- schemes employing Markov chaotic sequences," *IEEE Transactions on Signal Processing*, vol. 51, pp. 1979-94, 2003.
- [19] S. Tsekeridou, V. Solachidis, N. Nikolaidis, A. Nikolaidis, A. Tefas, and I. Pitas, "Statistical analysis of a watermarking system based on Bernoulli chaotic sequences," *Signal Processing*, vol. 81, pp. 1273-93, 2001.
- [20] Borko Furht, Edin Muharemagic, and D. Socek, *Multimedia Encryption and Watermarking*: Springer, 2005.
- [21] S. G. Lian, "Commutative encryption and watermarking in video compression," *IEEE transactions on circuits and systems for video technology*, vol. 17, pp. 774-778, 2007.
- [22] B. Rosenblatt, B. Trippe, and S. Monney, "Digital Rights Management: Business and Technology," *M&T Books*, 2002.
- [23] R.-S. Run, S.-J. Horng, J.-L. Lai, T.-W. Kao, and R.-J. Chen, "An improved SVD-based watermarking technique for copyright protection," *Expert Systems with Applications*, vol. 39, pp. 673-689, 2012.
- [24] W.-H. Lin, S.-J. Horng, T.-W. Kao, R.-J. Chen, Y.-H. Chen, C.-L. Lee, and T. Terano, "Image copyright protection with forward error

- correction," *Expert Systems with Applications*, vol. 36, pp. 11888-11894, 2009.
- [25] I. Kamel and Q. Albluwi, "A robust software watermarking for copyright protection," *Computers & Security*, vol. 28, pp. 395-409, 2009.
- [26] A. A. Mohammad, A. Alhaj, and S. Shaltaf, "An improved SVD-based watermarking scheme for protecting rightful ownership," *Signal Processing*, vol. 88, pp. 2158-80, 2008.
- [27] S. Katzenbeisser and F. A. P. Petitcolas, *Information Hiding Techniques for Steganography and Digital Watermarking*: Artech House, Norwood, Mass, USA, 2000.
- [28] F. Y. Shih, "Digital Watermarking and Steganography: Fundamentals and Techniques," *CRC Press*, 2008.
- [29] S. Katzenbeisser, Petitcolas, F.A.P., *Information Hiding Techniques for Steganography and Digital Watermarking*. Norwood, Mass, USA: Artech House, 2000.
- [30] P. Bassia, I. Pitas, and N. Nikolaidis, "Robust audio watermarking in the time domain," *IEEE Transactions on Multimedia*, vol. 3, pp. 232-41, 2001.

- [31] H. Malik, R. Ansari, and A. Khokhar, "Robust audio watermarking using frequency-selective spread spectrum," *IET Information Security*, vol. 2, pp. 129-150, 2008.
- [32] W. N. Lie and L. C. Chang, "Robust and high-quality time-domain audio watermarking based on low-frequency amplitude modification," *IEEE Transactions on Multimedia*, vol. 8, pp. 46-59, 2006.
- [33] F. Di Martino and S. Sessa, "Fragile watermarking tamper detection with images compressed by fuzzy transform," *Information Sciences*, vol. 195, pp. 62-90, 2012.
- [34] P.-C. Su, C.-S. Wu, I.-F. Chen, C.-Y. Wu, and Y.-C. Wu, "A practical design of digital video watermarking in H.264/AVC for content authentication," *Signal Processing: Image Communication*, vol. 26, pp. 413-426, 2011.
- [35] D. Xiao and F. Y. Shih, "An improved hierarchical fragile watermarking scheme using chaotic sequence sorting and subblock post-processing," *Optics Communications*, vol. 285, pp. 2596-2606, 2012.

-
- [36] J. Li, R. Wang, D. Yan, and Y. Li, "A multipurpose audio aggregation watermarking based on multistage vector quantization," *Multimedia Tools and Applications*, pp. 1-23.
- [37] R. Ni, Q. Ruan, and H. D. Cheng, "Secure semi-blind watermarking based on iteration mapping and image features," *Pattern Recognition*, vol. 38, pp. 357-368, 2005.
- [38] C.-C. Chang, K.-N. Chen, C.-F. Lee, and L.-J. Liu, "A secure fragile watermarking scheme based on chaos-and-hamming code," *Journal of Systems and Software*, vol. 84, pp. 1462-1470, 2011.
- [39] X. He, A. I. Illiev, and M. S. Scordilis, "A high capacity watermarking technique for stereo audio," in *Proc. of IEEE Int'l Conference on Acoustics, Speech, and Signal Processing*, 2004, pp. 393-396.
- [40] A. Mansouri, A. M. Aznavah, F. Torkamani-Azar, and F. Kurugollu, "A Low Complexity Video Watermarking in H.264 Compressed Domain," *IEEE Transactions on Information Forensics and Security* vol. 5, pp. 649-657, 2010.
- [41] X.-Y. Wang and H. Zhao, "A novel synchronization invariant audio watermarking scheme based on DWT and DCT," *IEEE Transactions on Signal Processing*, vol. 54, pp. 4835-40, 2006.

- [42] S. Xiang, H. J. Kim, and J. Huang, "Audio watermarking robust against time-scale modification and MP3 compression," *Signal Processing*, vol. 88, pp. 2372-87, 2008.
- [43] C. I. Podilchuk and E. J. Delp, "Digital watermarking: Algorithm and application," *IEEE Signal Processing Magazine*, vol. 18, pp. 33-46, 2001.
- [44] C. S. Shieh, H. C. Huang, F. H. Wang, and J. S. Pan, "Genetic watermarking based on transform-domain techniques," *Pattern Recognition*, vol. 37, pp. 555-565, 2004.
- [45] W.-N. Lie and L.-C. Chang, "Robust and high-quality time-domain audio watermarking based on low-frequency amplitude modification," *IEEE Transactions on Multimedia*, vol. 8, pp. 46-59, 2006.
- [46] D. Gruhl, A. Lu , and W. Bender, "Echo hiding " in *Proc. of the First Int'l Workshop on Information Hiding*, 1996, pp. 295-315.
- [47] O. T. C. Chen and W. Wen-Chih, "Highly Robust, Secure, and Perceptual-Quality Echo Hiding Scheme," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, pp. 629-638, 2008.

- [48] W. Li, X. Xue, and P. Lu, "Localized audio watermarking technique robust against time-scale modification," *IEEE Transactions on Multimedia*, vol. 8, pp. 60-69, 2006.
- [49] N. Cvejic and T. Seppanen, "Spread spectrum audio watermarking using frequency hopping and attack characterization," *Signal Processing*, vol. 84, pp. 207-213, 2004.
- [50] D. Megias, J. Serra-Ruiz, and M. Fallahpour, "Efficient self-synchronised blind audio watermarking system based on time domain and FFT amplitude modification," *Signal Processing*, vol. 90, pp. 3078-3092, 2010.
- [51] J. Huang, Y. Wang, and Y. Q. Shi, "A blind audio watermarking algorithm with self-synchronization," in *Proc.of IEEE Int'l Symposium on Circuits and Systems*, 2002, pp. 627-630.
- [52] S. Wu, J. Huang, D. Huang, and Y. Q. Shi, "Efficiently self-synchronized audio watermarking for assured audio data transmission," *IEEE Transactions on Broadcasting*, vol. 51, pp. 69-76, 2005.
- [53] X.-Y. Wang and H. Zhao, "A Novel Synchronization Invariant Audio Watermarking Scheme Based on DWT and DCT," *IEEE Transactions on Signal Processing*, vol. 54, pp. 4835-4840, 2006.

- [54] M. Fan and H. Wang, "Chaos-based discrete fractional Sine transform domain audio watermarking scheme," *Computers & Electrical Engineering*, vol. 35, pp. 506-516, 2009.
- [55] Z. Tao, H.-M. Zhao, J. Wu, J.-H. Gu, Y.-S. Xu, and D. Wu, "A lifting wavelet domain audio watermarking algorithm based on the statistical characteristics of sub-band coefficients," *Archives of Acoustics*, vol. 35, pp. 481-491, 2010.
- [56] H. Özer., B. Sankur., and N. Memon., "An SVD-based audio watermarking technique," in *Proc. of the 7th Workshop on Multimedia and Security*, 2005, pp. 51-56.
- [57] A. Al-Haj and A. Mohammad, "Digital Audio Watermarking Based on the Discrete Wavelets Transform and Singular Value Decomposition," *European Journal of Scientific Research*, vol. 39, pp. 6-21, 2010.
- [58] V. Bhat K, I. Sengupta, and A. Das, "An adaptive audio watermarking based on the singular value decomposition in the wavelet domain," *Digital Signal Processing*, vol. 20, pp. 1547-1558, 2010.
- [59] S. Vongpraphip and M. Ketcham, "An intelligence audio watermarking based on DWT-SVD using ATS," in *Proc. of WRI Global Congress on Intelligent Systems*, 2009, pp. 150-154.

- [60] R. Zezula and J. Miurec, "Audio digital watermarking algorithm based on SVD in MCLT domain," in *Proc. of 3rd Int'l Conference on Systems*, 2008, pp. 140-143.
- [61] B. Y. Lei, I. Y. Soon, and Z. Li, "Blind and robust audio watermarking scheme based on SVD-DCT," *Signal Processing*, vol. 91, pp. 1973-1984, 2011.
- [62] J. Wang, R. Healy, and J. Timoney, "A robust audio watermarking scheme based on reduced singular value decomposition and distortion removal," *Signal Processing*, vol. 91, pp. 1693-1708, 2011.
- [63] W. Al-Nuaimy, M. A. M. El-Bendary, A. Shafik, F. Shawki, A. E. Abou-El-azm, N. A. El-Fishawy, S. M. Elhalafawy, S. M. Diab, B. M. Sallam, F. E. Abd El-Samie, and H. B. Kazemian, "An SVD audio watermarking approach using chaotic encrypted images," *Digital Signal Processing*, vol. 21, pp. 764-779, 2011.
- [64] X. Wang, W. Qi, and P. Niu, "A new adaptive digital audio watermarking based on support vector regression," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, pp. 2270-2277, 2007.

- [65] N. Cvejic and T. Seppanen, "Digital Audio Watermarking Techniques and Technologies: Applications and Benchmarks," *Information Science Reference*, 2008.
- [66] A. Lang, J. Dittmann, R. Spring, and C. Vielhauer, "Audio watermark attacks: from single to profile attacks," in *Proc. of the 7th workshop on Multimedia and Security*, New York, NY, USA, 2005.
- [67] S. Cheng, H. Yu, and Z. Xiong, "Enhanced spread spectrum watermarking of MPEG-2 AAC audio," in *Proc. of IEEE Int'l Conference on Acoustics, Speech and Signal Processing* 2002, pp. IV/3728-IV/3731.
- [68] A. Mooney, J. G. Keating, and I. Pitas, "A comparative study of chaotic and white noise signals in digital watermarking," *Chaos, Solitons and Fractals*, vol. 35, pp. 913-921, 2008.
- [69] H. C. Andrews and C. L. Patterson, "Singular Value Decomposition and Digital Image Processing " *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-24, pp. 26-53, 1976.
- [70] E. Biglieri and K. Yao, "Some properties of singular value decomposition and their applications to digital signal processing," *Signal Processing*, vol. 18, pp. 277-289, 1989.

- [71] D. Gruhl, A. Lu , and W. Bender, "Echo hiding " *Lecture Notes in Computer Science*, vol. 1174, pp. 295-315, 1996.
- [72] B. S. Ko, R. Nishimura, and Y. Suzuki, "Time-spread echo method for digital audio watermarking," *IEEE Transactions on Multimedia*, vol. 7, pp. 212-219, 2005.
- [73] I. K. Yeo and H. J. Kim, "Modified patchwork algorithm: a novel audio watermarking scheme," *IEEE Transactions on Speech and Audio Processing*, vol. 11, pp. 381-6, 2003.
- [74] D. Kirovski and H. S. Malvar, "Spread-spectrum watermarking of audio signals," *IEEE Transactions on Signal Processing*, vol. 51, pp. 1020-33, 2003.
- [75] J. Seok, J. Hong, and J. Kim, "A Novel Audio Watermarking Algorithm for Copyright Protection of Digital Audio," *ETRI Journal*, vol. 24, p. 3, 2002.
- [76] S. Xiang and J. Huang, "Histogram-based audio watermarking against Time-Scale Modification and cropping attacks," *IEEE Transactions on Multimedia*, vol. 9, pp. 1357-1372, 2007.

-
- [77] F. E. Abd El-Samie, "An efficient singular value decomposition algorithm for digital audio watermarking," *International Journal of Speech Technology*, vol. 12, pp. 27-45, 2009.
- [78] ISO./IEC.23003-1:2007, "Information technology -MPEG audio technologies-Part 1: MPEG Surround," 2007.
- [79] S.-J. Yang, Kim,Do-Hyung,and Chung ,Jae-Ho "Watermark Insertion Into MP3 Bitstream Using the linbits Characteristics " *Proc.of the 115th AES ConventionPaper 5991*, 2003.
- [80] R. Tachibana, S. Shimizu, S. Kobayashi, and T. Nakamura, "An audio watermarking method using a two-dimensional pseudo-random array," *Signal Processing*, vol. 82, pp. 1455-1469, 2002.
- [81] X. Quan and H. Zhang, "Data Hiding in MPEG Compressed Audio Using Wet Paper Codes," in *Proc. of 18th International Conference on Pattern Recognition*, 2006, pp. 727-730.
- [82] C. Neubauer and J. Herre, "Audio watermarking of MPEG-2 AAC bitstream," *Proc. of108th AES Convention*, 2000.
- [83] Z. Li, Q. Sun, and Y. Lian, "Design and Analysis of a Scalable Watermarking Scheme for the Scalable Audio Coder," *IEEE Transactions on Signal Processing* vol. 54, pp. 3064-3077, 2006.

-
- [84] R. Geiger, Y. Yokotani, and G. Schuller, "Audio Data Hiding with High Data Rates Based on Intmdct," in *Proc. of IEEE Int'l Conference on Acoustics, Speech and Signal Processing*, 2006, pp. V-V.
- [85] R. Yu, S. Rahardja, C.-C. Lin, and C. C. Ko, "A fine granular scalable to lossless audio coder," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, pp. 1352-1363, 2006.
- [86] R. Geiger, R. Yu, J. Herre, S. Rahardja, S. W. Kim, X. Lin, and M. Schmidt, "ISO/IEC MPEG-4 high-definition scalable advanced audio coding," *AES: Journal of the Audio Engineering Society*, vol. 55, pp. 27-43, 2007.
- [87] T. Li, S. Rahardja, and S. N. Koh, "Frequency Region-Based Prioritized Bit-Plane Coding for Scalable Audio," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, pp. 94-105, 2008.
- [88] "Scalable Lossless Coding (SLS)," *ISO/IEC 14496-3:2005/Amd 3* 2006.
- [89] ITU. *ITU-R BS.1387-Method for objective measurements of perceptual audio quality*. Available: http://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1387-1-200111-I!!PDF-E.pdf
- [90] ITU. (2012, May 1). *Evaluation Of Audio Quality*. Available: <http://wiki.hydrogenaudio.org/index.php?title=EAQUAL>

- [91] F. Kozamernik and G. Stoll, "EBU listening tests on Internet audio codecs," *EBU Technical Review*, pp. 20-33, 2000.
- [92] A. Robert and J. Picard, "On the use of masking models for image and audio watermarking," *IEEE Transactions on Multimedia*, vol. 7, pp. 727-39, 2005.
- [93] R. Liu and T. Tan, "An SVD-based watermarking scheme for protecting rightful ownership," *IEEE Transactions on Multimedia*, vol. 4, pp. 121-128, 2002.
- [94] S. Wu, J. Huang, D. Huang, and Y. Q. Shi, "Self-synchronized audio watermark in DWT domain," 2004, pp. 712-715.
- [95] P. Bao and X. Ma, "Image adaptive watermarking using wavelet domain singular value decomposition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, pp. 96-102, 2005.
- [96] F. Huang and Z.-H. Guan, "A hybrid SVD-DCT watermarking method based on LPSNR," *Pattern Recognition Letters*, vol. 25, pp. 1769-1775, 2004.
- [97] K.-L. Chung, W.-N. Yang, Y.-H. Huang, S.-T. Wu, and Y.-C. Hsu, "On SVD-based watermarking algorithm," *Applied Mathematics and Computation*, vol. 188, pp. 54-57, 2007.

- [98] Z. Peng and W. Liu, "Color image authentication based on spatiotemporal chaos and SVD," *Chaos, Solitons and Fractals*, vol. 36, pp. 946-952, 2008.
- [99] Y. Wu, "On the security of an SVD-based ownership watermarking," *IEEE Transactions on Multimedia*, vol. 7, pp. 624-627, 2005.
- [100] X.-P. Zhang and K. Li, "Comments on "an SVD-based watermarking scheme for protecting rightful ownership"," *IEEE Transactions on Multimedia*, vol. 7, pp. 593-594, 2005.
- [101] R. Rykaczewski, "Comments on "an SVD-based watermarking scheme for protecting rightful ownership"," *IEEE Transactions on Multimedia*, vol. 9, pp. 421-3, 2007.
- [102] EBU. (2011, May 30). *SQAM - Sound Quality Assessment Material*. Available: <http://sound.media.mit.edu/mpeg4/audio/sqam/>
- [103] Y. Lin and W. H. Abdulla, "Perceptual evaluation of audio watermarking using objective quality measures," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing* 2008, pp. 1745-1748.
- [104] R. Crochiere, J. Tribolet, and L. Rabiner, "An interpretation of the log likelihood ratio as a measure of waveform coder performance," *IEEE*

- Transactions on Acoustics, Speech and Signal Processing*, vol. 28, pp. 318-323, 1980.
- [105] H. Kang, K. Yamaguchi, B. Kurkoski, K. Yamaguchi, and K. Kobayashi, "Full-Index-Embedding Patchwork Algorithm for Audio Watermarking," *IEICE Transactions on Information and Systems*, vol. E91.D, pp. 2731-2734, 2008.
- [106] W. Sweldens, "The lifting scheme: A custom-design construction of biorthogonal wavelets," *Applied and Computational Harmonic Analysis*, vol. 3, pp. 186-200, 1996.
- [107] D. Kundur and D. Hatzinakos, "Digital watermarking using multiresolution wavelet decomposition," in *Proc.of IEEE Int'l Conference on Acoustics, Speech and Signal Processing* 1998, pp. 2969-2972.
- [108] E. Ercelebi and L. Batakci, "Audio watermarking scheme based on embedding strategy in low frequency components with a binary image," *Digital Signal Processing*, vol. 19, pp. 265-77, 2009.
- [109] G. W. W. B. Chen "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding,"

- IEEE Transactions on Information Theory*, vol. 47, pp. 1423-1443, 2001.
- [110] L. Kennedy and R. C. Eberhart, *Swarm intelligence*: CA: Morgan Kaufmann, 2001.
- [111] A. Al-Hajl and A. Abu-Errub, "Performance optimization of discrete wavelets transform based image watermarking using genetic algorithms," *Journal of Computer Science*, vol. 4, pp. 834-841, 2008.
- [112] V. Aslantas, "A singular-value decomposition-based image watermarking using genetic algorithm," *AEU - International Journal of Electronics and Communications*, vol. 62, pp. 386-394, 2008.
- [113] C. C. Lai, "A digital watermarking scheme based on singular value decomposition and tiny genetic algorithm," *Digital Signal Processing: A Review Journal*, vol. 21, pp. 522-527, 2011.
- [114] H. C. Huang, S. C. Chu, J. S. Pan, C. Y. Huang, and B. Y. Liao, "Tabu search based multi-watermarks embedding algorithm with multiple description coding," *Information Sciences*, vol. 181, pp. 3379-3396, 2011.

- [115] V. Aslantas, S. Ozer, and S. Ozturk, "Improving the performance of DCT-based fragile watermarking using intelligent optimization algorithms," *Optics Communications*, vol. 282, pp. 2806-2817, 2009.
- [116] M. Ishtiaq, B. Sikandar, M. Arfan Jaffar, and A. Khan, "Adaptive watermark strength selection using Particle Swarm Optimization," *ICIC Express Letters*, vol. 4, pp. 2403-2409, 2010.
- [117] Y.-R. Wang, W.-H. Lin, and L. Yang, "An intelligent watermarking method based on particle swarm optimization," *Expert Systems with Applications*, vol. 38, pp. 8024-8029, 2011.
- [118] C. H. Wu, Y. Zheng, W. H. Ip, C. Y. Chan, K. L. Yung, and Z. M. Lu, "A flexible H.264/AVC compressed video watermarking scheme using particle swarm optimization based dither modulation," *AEU - International Journal of Electronics and Communications*, vol. 65, pp. 27-36, 2011.
- [119] X.-Y. Wang, P.-P. Niu, and H.-Y. Yang, "A robust, digital-audio watermarking method," *IEEE Multimedia*, vol. 16, pp. 60-69, 2009.
- [120] N. Sriyingyong and K. Attakitmongcol, "Wavelet-based audio watermarking using adaptive tabu search," in *Proc. of 1st Int'l Symposium on Wireless Pervasive Computing*, 2006, pp. 1-5.

- [121] P. Kumsawat, "A Genetic Algorithm Optimization Technique for Multiwavelet-Based Digital Audio Watermarking," *EURASIP Journal on Advances in Signal Processing*, pp. 471842-471852, 2010.
- [122] P. Kumsawat, K. Attakitmongcol, and A. Srikaew, "A new approach for optimization in image watermarking by using genetic algorithms," *IEEE Transactions on Signal Processing*, vol. 53, pp. 4707-4719, 2005.
- [123] A. P. Engelbrecht, *Fundamentals of computational swarm intelligence*: UK: Wiley, 2005.
- [124] S. Kemmoe Tchomte and M. Gourgand, "Particle swarm optimization: A study of particle displacement for solving continuous and combinatorial optimization problems," *International Journal of Production Economics*, vol. 121, pp. 57-67, 2009.
- [125] I. C. Trelea, "The particle swarm optimization algorithm: convergence analysis and parameter selection," *Information Processing Letters*, vol. 85, pp. 317-325, 2003.
- [126] A. Mooney, J. G. Keating, and D. M. Heffernan, "Performance analysis of chaotic and white watermarks in the presence of common watermark attacks," *Chaos, Solitons & Fractals*, vol. 42, pp. 560-570, 2009.

- [127] J. G. Beerends and J. A. Stemerdink, "Perceptual audio quality measure based on a psychoacoustic sound representation," *AES: Journal of the Audio Engineering Society*, vol. 40, pp. 963-978, 1992.