

SPRINGER BRIEFS IN ELECTRICAL AND
COMPUTER ENGINEERING • SPEECH TECHNOLOGY

Pranab Kumar Dhar
Tetsuya Shimamura

Advances in Audio Watermarking Based on Singular Value Decomposition

SpringerBriefs in Electrical and Computer Engineering

SpringerBriefs in Speech Technology

Series Editor

Amy Neustein

More information about this series at <http://www.springer.com/series/10059>

Editor's Note

The authors of this series have been hand selected. They comprise some of the most outstanding scientists—drawn from academia and private industry—whose research is marked by its novelty, applicability, and practicality in providing broad-based speech solutions. The Springer Briefs in Speech Technology series provides the latest findings in speech technology gleaned from comprehensive literature reviews and *empirical investigations* that are performed in both laboratory and *real life* settings. Some of the topics covered in this series include the presentation of real life commercial deployment of spoken dialog systems, contemporary methods of speech parameterization, developments in information security for automated speech, forensic speaker recognition, use of sophisticated speech analytics in call centers, and an exploration of new methods of soft computing for improving human-computer interaction. Those in academia, the private sector, the self service industry, law enforcement, and government intelligence are among the principal audience for this series, which is designed to serve as an important and essential reference guide for speech developers, system designers, speech engineers, linguists, and others. In particular, a major audience of readers will consist of researchers and technical experts in the automated call center industry where speech processing is a key component to the functioning of customer care contact centers.

Amy Neustein, Ph.D., serves as editor in chief of the *International Journal of Speech Technology* (Springer). She edited the recently published book *Advances in Speech Recognition: Mobile Environments, Call Centers and Clinics* (Springer 2010), and serves as quest columnist on speech processing for Womensenews. Dr. Neustein is the founder and CEO of Linguistic Technology Systems, a NJ-based think tank for intelligent design of advanced natural language-based emotion detection software to improve human response in monitoring recorded conversations of terror suspects and helpline calls.

Dr. Neustein's work appears in the peer review literature and in industry and mass media publications. Her academic books, which cover a range of political, social, and legal topics, have been cited in the *Chronicles of Higher Education* and have won her a pro Humanitate Literary Award. She serves on the visiting faculty of the National Judicial College and as a plenary speaker at conferences in artificial intelligence and computing. Dr. Neustein is a member of MIR (machine intelligence research) Labs, which does advanced work in computer technology to assist underdeveloped countries in improving their ability to cope with famine, disease/ illness, and political and social affliction. She is a founding member of the New York City Speech Processing Consortium, a newly formed group of NY-based companies, publishing houses, and researchers dedicated to advancing speech technology research and development.

Pranab Kumar Dhar • Tetsuya Shimamura

Advances in Audio Watermarking Based on Singular Value Decomposition

Pranab Kumar Dhar
Graduate School of Science
and Engineering
Saitama University
Saitama, Japan

Tetsuya Shimamura
Graduate School of Science
and Engineering
Saitama University
Saitama, Japan

ISSN 2191-8112 ISSN 2191-8120 (electronic)
SpringerBriefs in Electrical and Computer Engineering
ISBN 978-3-319-14799-4 ISBN 978-3-319-14800-7 (eBook)
DOI 10.1007/978-3-319-14800-7

Library of Congress Control Number: 2015933145

Springer Cham Heidelberg New York Dordrecht London
© The Author(s) 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media
(www.springer.com)

*This book is dedicated to my Paramaradhya
Gurudev Shrimat Maharshi Sutejananda
Yoti Maharaj, and Shrimat Maharshi
Sushidhananda Yoti Maharaj for their
endless mercy.*

Abstract

In recent years, due to the rapid development of the Internet and multimedia technology, the transmission and distribution of multimedia data have become an extremely simple task. This has become a serious threat for multimedia content owners. Thus, there is a significant interest for copyright protection of multimedia data. Digital watermarking has drawn extensive attention for copyright protection of multimedia data. It is a process of embedding watermarks into the multimedia data to show authenticity and ownership. This technique has several applications such as copyright protection, data authentication, fingerprinting, data indexing, and broadcast monitoring.

This book introduces two watermarking methods for copyright protection. An audio watermarking method in discrete wavelet transform (DWT) and discrete cosine transform (DCT) domains using singular value decomposition (SVD) and quantization is presented. In this method, initially the original audio is segmented into nonoverlapping frames. DWT is applied to each frame and detail coefficients are arranged into a one-dimensional matrix formation. DCT is performed on this matrix and the obtained DCT coefficients are reshaped. SVD is applied to the DCT coefficients and watermark information is then embedded into the highest singular value by quantization. Watermark information is extracted by comparing the largest singular value of the DCT coefficients obtained from DWT subbands of each original and attacked watermarked audio frame. This method is robust against various attacks and provides good imperceptible watermarked sounds.

An audio watermarking method in fast Fourier transform (FFT) domain using SVD and Cartesian-polar transform (CPT) is presented. In this method, initially the original audio is segmented into nonoverlapping frames. FFT is applied to each frame and low-frequency FFT coefficients are selected. SVD is applied to the selected FFT coefficients of each frame represented in a matrix form. The highest two singular values of each frame are selected. The selected singular values are assumed as the components of polar coordinate system and are transformed into the components of Cartesian coordinate system. Watermark information is embedded into each of these Cartesian components using an embedding function. Watermark information is extracted by comparing the Cartesian components of the largest two

singular values obtained from the low-frequency FFT coefficients of each original and attacked watermarked audio frame. This method has high data payload and it provides good robustness against various attacks.

The performance of the proposed watermarking methods is evaluated and, finally, compared with the state-of-the-art methods. Simulation results indicate that the proposed audio watermarking methods outperform the state-of-the-art methods in terms of robustness, imperceptibility, and data payload. These results verify the effectiveness of the proposed methods as a suitable candidate for copyright protection of audio signal.

Keywords Audio watermarking, Copyright protection, Cartesian-polar transform, Discrete cosine transform, Discrete wavelet transform, Fast Fourier transform, Singular value decomposition

Acknowledgments

First of all, I would like to express my sincere gratitude to my Paramaradhyā Gurudev Shrimat Maharshi Sutejananda Yoti Maharaj and Shrimat Maharshi Sushidhananda Yoti Maharaj for their endless mercy.

I would also like to express my deep appreciation, obligation, and indebtedness to Prof. Tetsuya Shimamura, for his cooperation, encouragement, attention to details, and guidance. I am very grateful to for his valuable comments and enlightening discussion we had. It has been an honor and a pleasure to work with him during my stay at Saitama University, Japan.

I also extend many thanks to all members of Shimamura Laboratory, both past and present, for their cooperation and friendship.

I am also grateful to all my family members, especially my parents Prof. Anil Kanti Dhar and Mrs. Juthika Dhar, my father-in-law Mr. Dipal Kanti Dey, my mother-in-law Mrs. Sabita Rani Dey, my uncle Prof. Dr. Sunil Dhar, and my sister Mrs. Radha Rani Dhar for their inspiration, encouragement, and support. I am also thankful to all my friends for their moral support.

Last but not least, I am forever indebted to my wife Uma Dhar, for her love, support, patience, and divine inspiration. I am also thankful to my daughter Parama Dhar who brightens my life. This book would not have been possible without their unending love and support.

Contents

- 1 Introduction** 1
 - 1.1 Overview 1
 - 1.2 Application Areas of Digital Watermarking 2
 - 1.2.1 Copyright Protection..... 2
 - 1.2.2 Fingerprinting 3
 - 1.2.3 Content Authentication 3
 - 1.2.4 Copy Protection..... 3
 - 1.2.5 Broadcast Monitoring 3
 - 1.2.6 Information Carrier 3
 - 1.2.7 Medical Applications 4
 - 1.3 Properties of Digital Watermarking 4
 - 1.3.1 Perceptual Transparency 4
 - 1.3.2 Data Payload 4
 - 1.3.3 Robustness 5
 - 1.3.4 Blind or Informed Detection 5
 - 1.3.5 Security..... 5
 - 1.3.6 Computational Complexity..... 6
 - 1.4 Trade-off of Digital Watermarking 6
 - 1.5 Motivation 7
 - 1.6 Book Organization 9
- 2 Background Information** 11
 - 2.1 Review of Audio Watermarking Methods 11
 - 2.1.1 Time Domain Methods 11
 - 2.1.2 Transform Domain Methods 12
 - 2.1.3 Other Audio Watermarking Methods 13
 - 2.2 Signal Transformation 14
 - 2.2.1 Discrete Fourier Transform 14
 - 2.2.2 Discrete Cosine Transform 15
 - 2.2.3 Discrete Wavelet Transform..... 15
 - 2.2.4 Cartesian-Polar Transform 16

| | | |
|----------|---|-----------|
| 2.3 | Singular Value Decomposition | 16 |
| 2.4 | Summary | 16 |
| 3 | DWT-DCT-Based Audio Watermarking Using SVD | 17 |
| 3.1 | Introduction | 17 |
| 3.2 | A Brief Overview of SVD-Based Methods | 18 |
| 3.3 | Proposed Watermarking Method | 19 |
| 3.3.1 | Watermark Embedding Process | 19 |
| 3.3.2 | Watermark Detection Process | 21 |
| 3.4 | Experimental Results and Discussion | 22 |
| 3.4.1 | Imperceptibility Test | 23 |
| 3.4.2 | Robustness Test | 29 |
| 3.4.3 | Error Analysis | 30 |
| 3.4.4 | Data Payload | 33 |
| 3.4.5 | Algorithm Comparison and Discussion | 33 |
| 3.5 | Summary | 35 |
| 4 | FFT-Based Audio Watermarking Using SVD and CPT | 37 |
| 4.1 | Introduction | 37 |
| 4.2 | Proposed Watermarking Method | 38 |
| 4.2.1 | Watermark Preprocessing | 38 |
| 4.2.2 | Watermark Embedding Process | 39 |
| 4.2.3 | Watermark Detection Process | 42 |
| 4.3 | Experimental Results and Discussion | 43 |
| 4.3.1 | Imperceptibility Test | 44 |
| 4.3.2 | Robustness Test | 47 |
| 4.3.3 | Security | 48 |
| 4.3.4 | Error Analysis | 49 |
| 4.3.5 | Algorithm Comparison and Discussion | 49 |
| 4.4 | Summary | 52 |
| 5 | Conclusions | 53 |
| 5.1 | Summary of the Work | 53 |
| 5.2 | Future Research | 54 |
| | Bibliography | 55 |

List of Figures

| | | |
|-----------|--|----|
| Fig. 1.1 | An overview of a general watermarking scheme | 2 |
| Fig. 1.2 | Trade-off among imperceptibility, robustness, and data payload | 6 |
| Fig. 2.1 | Single-level DWT decomposition..... | 15 |
| Fig. 3.1 | Watermark embedding process..... | 20 |
| Fig. 3.2 | Matrix formation of the detailed coefficients D_1 and D_2 for each frame (H is used as H_i for each frame i) | 20 |
| Fig. 3.3 | Watermark detection process..... | 22 |
| Fig. 3.4 | Binary watermark..... | 23 |
| Fig. 3.5 | (a) Time domain representation of a selected frame for the signal 'Classical', (b) DWT domain representation using the coefficients A_2 , D_2 , and D_1 of the same frame for the signal 'Classical' | 24 |
| Fig. 3.6 | Representation of the matrix H_i of a selected frame for the signal 'Classical' | 24 |
| Fig. 3.7 | Representation of the matrix Y_i of a selected frame for the signal 'Classical' | 25 |
| Fig. 3.8 | Representation of the matrix Y'_i of a selected frame for the signal 'Classical' | 25 |
| Fig. 3.9 | Imperceptibility of the watermarked audio using the proposed method: (a) Original signal 'Classical', (b) Watermarked signal 'Classical', (c) Difference between original and watermarked signals | 28 |
| Fig. 3.10 | Spectrogram representation of the original audio signal and watermarked audio signal 'Classical' using the proposed method..... | 28 |
| Fig. 3.11 | Extracted watermark image with NC and BER for the audio signal 'Classical' | 30 |
| Fig. 3.12 | Probability of FPE for various values of k | 32 |

| | | |
|-----------|--|----|
| Fig. 3.13 | Probability of FNE for various values of k | 33 |
| Fig. 4.1 | Watermark embedding process..... | 39 |
| Fig. 4.2 | Watermark detection process..... | 42 |
| Fig. 4.3 | (a) Binary watermark image. (b) Encrypted image | 43 |
| Fig. 4.4 | Magnitude spectrum of a selected frame for the original audio signal 'Folk' | 44 |
| Fig. 4.5 | Magnitude spectrum of a selected frame for the watermarked audio signal 'Folk' | 45 |
| Fig. 4.6 | Imperceptibility of the watermarked audio using the proposed method: (a) Original signal 'Folk', (b) Watermarked signal 'Folk', (c) Difference between original and watermarked signals | 46 |
| Fig. 4.7 | Spectrogram representation of the original audio signal and watermarked audio signal 'Folk' using the proposed method .. | 47 |
| Fig. 4.8 | Extracted watermark image with NC and BER for the audio signal 'Folk' | 48 |
| Fig. 4.9 | Probability of FPE for various values of k | 50 |
| Fig. 4.10 | Probability of FNE for various values of k | 50 |

List of Tables

| | | |
|-----------|--|----|
| Table 3.1 | Subjective and objective difference grade | 26 |
| Table 3.2 | Subjective and objective evaluation of different watermarked sounds | 27 |
| Table 3.3 | Comparison of SNR and MOS between the proposed method and several recent methods..... | 29 |
| Table 3.4 | NC and BER of the extracted watermarks for different audio signals | 31 |
| Table 3.5 | A general comparison between the proposed method and recent audio watermarking methods sorted by data payload ... | 34 |
| Table 4.1 | Subjective and objective evaluation of different watermarked sounds | 45 |
| Table 4.2 | Comparison of SNR and MOS between the proposed method and several recent methods..... | 47 |
| Table 4.3 | NC and BER of the extracted watermarks for different audio signals | 49 |
| Table 4.4 | A general comparison between the proposed method and recent audio watermarking methods sorted by data payload ... | 51 |

List of Symbols, Notations and Abbreviations

| | |
|------------|---|
| X | Original audio signal |
| X' | Watermarked audio signal |
| X^* | Attacked watermarked audio signal |
| W_{BI} | Binary watermark image |
| f | X -coordinate of the watermark image W_{BI} |
| g | Y -coordinate of the watermark image W_{BI} |
| W_{BI}^* | Extracted watermark image |
| W_{SE} | Binary watermark sequence |
| m | Index of watermark sequence |
| W_{SE}^* | Extracted watermark sequence |
| I | Watermark length |
| M | Length of the row and column of watermark image |
| Q | Quantization coefficient |
| α | Scaling factor |
| P | Bit error rate probability of extracted watermark |
| P_{FPE} | Probability of false positive error |
| P_{FNE} | Probability of false negative error |
| T_s | Duration of audio signal (in sec.) |
| T_h | Threshold |
| K | Secret key |
| K_1 | Secret key |
| K_2 | Secret key |
| S | Singular value matrix |
| S_{ix} | X component of largest singular value |
| S_{iy} | Y component of largest singular value |
| C_1 | Constant |
| C_2 | Constant |
| C_3 | Constant |
| i | Frame number |
| F_i | i -th frame |

| | |
|--------|---|
| H | Matrix formation of detail subband D_1 and D_2 |
| L | Length of the audio signal |
| T | Transpose of the matrix |
| e | Euclidean norm |
| u | mean |
| v | variance |
| a | constant |
| DFT | Discrete Fourier transform |
| FFT | Fast Fourier transform |
| DCT | Discrete cosine transform |
| DWT | Discrete wavelet transform |
| CPT | Cartesian-polar transform |
| SVD | Singular value decomposition |
| $IFPI$ | International Federation of the Phonographic Industry |
| SDG | Subjective difference grade |
| ODG | Objective difference grade |
| MOS | Mean opinion score |
| SNR | Signal-to-noise ratio |
| BER | Bit error rate |

Chapter 1

Introduction

1.1 Overview

The recent development in computational world and the wide availability of internet have facilitated the transmission and distribution of multimedia content. As a result, the protection of intellectual property rights of digital content has been the key problem. The term ‘content’ refers to any digital information, such as digital audio, video, graphics, animation, images, text, or any combinations of these types. This digital content can be easily accessed, perfectly copied, rapidly disseminated and massively shared without losing its quality. However, the possibility of unlimited copying without the loss of fidelity has led to a considerable financial loss for copyright holders. Digital watermarking has drawn extensive attention for protecting digital contents from unauthorized copying [18, 19, 38]. It is a process of embedding watermark into the original content to show authenticity and ownership. It has been widely used for several purposes including copyright protection, information carrier, broadcast monitoring, fingerprinting, data authentication, medical safety, and so on.

Figure 1.1 shows an overview of a general watermarking scheme which consists of an watermark embedder and an watermark detector [19]. A watermark, which usually consists of a binary data sequence, is inserted into the original signal in the watermark embedder. Thus, a watermark embedder has two inputs; one is the watermark message (usually accompanied by a secret key) and the other is the original signal (audio, image, or video). The output of the watermark embedder is the watermarked signal, which cannot be perceptually discriminated from the original signal. This operation should be done in such a way that the embedded watermark should not be removed by various attacks. The amount of modification done to embed watermark is based on watermark robustness against attacks and perceptual quality of the watermarked signal. The watermarked signal is then usually broadcasted and later presented to the watermark detector. The detector extracts the embedded watermark from the watermarked signal. Watermarking methods can be

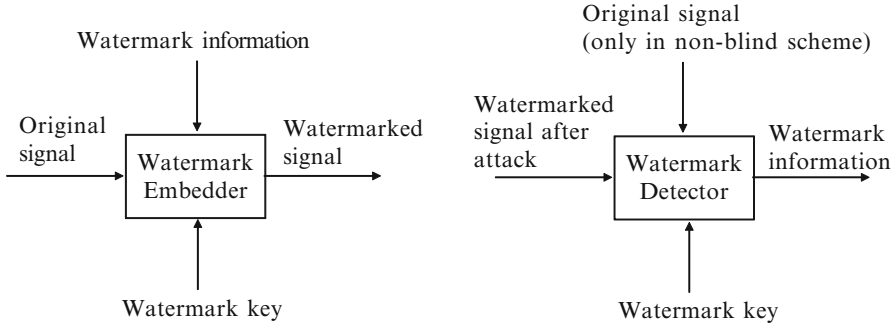


Fig. 1.1 An overview of a general watermarking scheme

classified into three categories based on watermark detection point of view. They are blind, semi-blind, and non-blind watermarking methods. Blind watermarking method does not need the original signal to extract watermark information. Semi-blind watermarking method needs some information of original signal to extract the watermark information. On the other hand, non-blind watermarking method needs the original signal to extract the embedded watermark.

1.2 Application Areas of Digital Watermarking

The principal design challenge of digital watermarking is to embed watermark information so that it is reliably detected by the watermark detector. The relative importance of the properties of digital watermarking significantly depends on the application for which the algorithm is designed [18]. For copy protection applications, the watermark must be recoverable even when the watermarked signal undergoes a considerable level of distortion, while for tamper assessment applications, the watermark must effectively characterize the modification that took place. This section briefly discusses the various application areas for digital watermarking.

1.2.1 Copyright Protection

Copyright protection is the most important application of digital watermarking. The objective is to embed information that identifies the copyright owner of the digital media, in order to prevent other parties from claiming the copyright [45].

1.2.2 Fingerprinting

The objective of fingerprinting is to convey information about the legal recipient rather than the source of digital media, in order to identify single distributed copies of digital data [18].

1.2.3 Content Authentication

The objective of content authentication is to detect the modification of digital data. This can be achieved with fragile watermarking technique that have a low robustness to certain modification [42].

1.2.4 Copy Protection

Copy protection tries to find a mechanism to disallow unauthorized copying of digital media. Copy software or device must be able to detect the watermark and allow or disallow the requested operation according to the copy status of the digital media being copied [19].

1.2.5 Broadcast Monitoring

Watermarking techniques can be useful for broadcast monitoring. Using watermarking an identification code can be embedded into the content being broadcasted. A computer-based monitoring system can detect the embedded watermark to ensure that they receive all of the airtime they purchase from the broadcasters [18].

1.2.6 Information Carrier

The embedded watermark in this application is expected to have a high capacity and to be detected using a blind detection algorithm. While the robustness against intentional attack is not required, a certain degree of robustness against common attacks may be desired [31].

1.2.7 Medical Applications

Watermarking can be used to write the unique name of the patient on the X-ray reports or magnetic resonance imaging (MRI) scan reports. This application is important because it is highly advisable to have the patients name entered on reports, and reduces the misplacements of reports which are very important during treatment [61].

1.3 Properties of Digital Watermarking

A watermarking method can be characterized by a number of properties. However, the relative importance of each property depends on the demand of the application [20, 21]. These properties are discussed below:

1.3.1 Perceptual Transparency

In almost every application, the watermark embedding process has to insert watermark information without changing the perceptual quality of the original signal. The fidelity of a watermarking algorithm is usually defined as a perceptual similarity between the original and watermarked audio signal. However, the quality of the watermarked signal is usually degraded, either intentionally by an adversary or unintentionally in the transmission process, before a person perceives it. Therefore, it is more adequate to define the fidelity of a watermarking algorithm as a perceptual similarity between the original and watermarked signal at the point at which they are presented to a consumer [20].

1.3.2 Data Payload

The data payload of a watermarking method is the number of watermark bits that are embedded within a unit of time and is usually measured in bits per second (bps). Some watermarking applications, such as copy control, require the insertion of a serial number or an author ID, with the average bit rate of 0.5 bps. In some envisioned application, like hiding speech in audio, algorithms have to be able to embed watermark information within the bit rate that is a significant fraction of the original audio bit rate, i.e., up to 150 kbps [20].

1.3.3 Robustness

The robustness of a watermarking method is defined as the ability to detect the watermark after common signal processing operations. The requirement of robustness of a watermarking method is completely application dependent. For example, in radio broadcast monitoring system, embedded watermark needs only to survive distortions caused by the transmission process, including dynamic compression and low-pass filtering, because the watermark detection is done directly from the broadcast signal. On the other hand, in some methods robustness is completely undesirable and those methods are known as fragile watermarking methods [20].

1.3.4 Blind or Informed Detection

In some applications, the detection algorithm can use the original audio signal to extract watermark from the watermarked signal (informed detection). It often significantly improves the detector performance, because the watermark information is extracted by subtracting the original signal from the watermarked signal. However, if the detection algorithm does not have access to the original signal (blind detection) and this inability substantially decreases the amount of data that can be hidden in the original signal. The complete process of embedding and extracting of the watermark is modeled as a communication channel where watermark is distorted due to the presence of strong interference and channel effects [20].

1.3.5 Security

Watermarking method must be secure in the sense that an adversary must not be able to detect the presence of embedded data and to remove the embedded data. The security of digital watermarking is interpreted in the same way as the security of encryption techniques and it cannot be broken unless the authorized user has access to a secret key that controls watermark embedding. An unauthorized user should be unable to extract the data in a reasonable amount of time even if he knows that the original signal contains a watermark and is familiar with the watermark embedding algorithm. Security requirements vary with application and the most stringent are in cover communications applications, and, in some cases, data is encrypted prior to embedding into original signal [20].

1.3.6 Computational Complexity

The implementation of a watermarking scheme is a tedious task, and it depends on the business application involved. The principal issues from the technical point of view are the computational complexity of embedding and detection algorithms and the number of embedders and detectors used in the system. For example, in broadcast monitoring, embedding and detection must be done in real time, while in copyright protection applications, time is not a crucial factor for a practical implementation. One of the economic issues in the design of embedders and detectors, which can be implemented either as hardware or software plug-ins, is the difference in processing power of different devices (laptop, mobile phone, etc.) [20].

1.4 Trade-off of Digital Watermarking

The imperceptibility, robustness and data payload are the three most important characteristics of a watermarking method. However, a trade-off is needed among these three conflicting characteristics [25]. This trade-off is shown in Fig. 1.2. A specific application of watermarking may determine what the capacity is needed. After it is determined, there needs a trade-off between imperceptibility and robustness. If more robustness against attacks is needed, then a larger modification of the signal's properties to embed the watermark will be necessary. However, this will significantly affect the imperceptibility of the watermarked signal. Another scenario is that with a predefined requirement for the imperceptibility, there will exist a trade-off between the capacity and robustness.

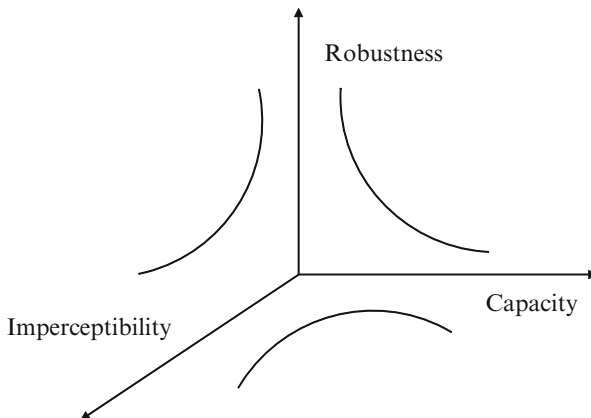


Fig. 1.2 Trade-off among imperceptibility, robustness, and data payload

1.5 Motivation

Watermarking is not the only method to protect digital content. Cryptographic encryption and digital signatures have also been studied extensively for their use in secure communication and protection of important information. However, watermarking has several important advantages compared to encryption and digital signatures. Firstly, watermarking incorporates an embedding process, preventing easy separation of the watermark from the content. Digital signatures, on the other hand, are attached to the files as headers, which can be deleted by re-recording files or be changed by format conversion. Encrypted signals, once decrypted, would have no protection against further manipulations. Secondly, watermarks usually undergo the same changes as the original content. Thus, it is possible, by examining the changes in watermark information, to detect the locations and the type of manipulations in the original signal. This could be achieved, to a certain degree, using digital signatures. However, tracing and tracking is very difficult in encrypted signals. Finally, a watermark is usually imperceptible. Therefore, it is convenient for the people who are not aware or do not use the watermark.

A significant number of watermarking methods have been proposed in the past [6, 19]. Most of the watermarking methods proposed over the last few years focus on image and video watermarking [1, 10, 15, 40, 53, 55, 69], however, this thesis focuses on audio watermarking. Audio watermarking is more challenging than image and video watermarking due to two reasons. Firstly, audio signal contains one-dimensional data, thus it is difficult to hide additional information without compromising the quality of the audio signal. Secondly, human auditory system (HAS) is significantly more sensitive than human visual system (HVS), thus a small degradation in a signal will affect the quality of the watermarked signal.

Several audio watermarking methods have been proposed in the literature in order to create robust and imperceptible audio watermarks [19, 21]. Most audio watermarking methods utilize either a time domain [5, 48, 57] or a transform domain such as discrete wavelet transform (DWT) [12–14], discrete cosine transform (DCT) [26, 60], lifting wavelet transform (LWT) [28], cepstrum domain [7, 46], and fast Fourier transform (FFT) [30, 32, 50]. The time domain methods directly insert a watermark into an audio signal in the time domain, whereas the transform domain methods embed a watermark by modifying the frequency coefficients. The time domain methods are easier to implement than the transform domain methods and require a lower computational cost; however, they are less robust against some signal processing attacks. Swanson et al. [57] proposed a watermarking scheme that embeds watermark bits by modifying the audio samples directly. Lie and Chang [48] introduced a method in which group amplitudes are modified to achieve high robustness. However, both methods have low data payload. In [14], authors presented an adaptive method using wavelet based entropy, but robustness to resampling and low-pass filtering attacks are quite low. Chen et al. [12] proposed an algorithm that embeds watermark information by energy-proportion scheme. However, the SNR results of this algorithm are not satisfactory. In [13], authors introduced an

optimization-based watermarking scheme which embeds watermark in the low frequency DWT coefficients. However, the subjective evaluation of watermarked audio signals has not been conducted in this scheme. Ercelebi and Bataki [28] proposed a watermarking method based on LWT in which a binary image is embedded as watermark. However, from the reported result, robustness to attacks of this method is quite low. Megias et al. [50] suggested a watermarking method that embeds watermark in FFT domain, but it has low data payload. Akhaee et al. [20] introduced a blind audio watermarking method in the point-to-point graph (PPG) transform domain. Data embedding is performed by shaping the configuration of the PPG points. However, an objective quality evaluation of the audio signal has not been conducted for this method. Some novel and popular audio watermarking methods use the patchwork algorithm [52, 71] and spread spectrum techniques [17, 41]. Recently, the singular value decomposition (SVD) has been used as an effective technique in digital watermarking [2, 3, 8, 27, 47, 54, 63]. Ozer et al. [54] proposed a method based on SVD and short-time Fourier transformation (STFT). In this method, a simple noise attack may lead to great difficulty in the watermark detection process. El-Samie [27] and Al-Nuaimy et al. [3] proposed an efficient SVD-based audio watermarking scheme in the transform domain which utilizes a chaotic sequence to shuffle the binary watermark to increase the confidentiality. Moreover, Al-Nuaimy et al. [3] extended the proposed watermarking method and applied it in Bluetooth-based systems and automatic speaker identification systems. However, from the reported results, the robustness needs further improvement. Bhat et al. [8] proposed a method based on DWT and quantization index modulation (QIM). This technique is robust against different attacks; however, the signal-to-noise ratio (SNR) results are slightly above 20 dB. The authors of [63] devised a method based on reduced SVD. Distortion control is utilized to control the audible distortions and to determine the strength of the peaks to represent watermark information in the embedding stage. Moreover, some other techniques such as discrete fractional sine transform (DFST) [33], time spread (TS) echo hiding [67, 68], empirical mode decomposition (EMD) [39], linear predictive coding (LPC) [62], and audio histogram techniques [65, 66] are becoming increasingly popular. The major limitation of the existing audio watermarking techniques is the difficulty in obtaining a favorable trade-off between the data payload and robustness against various attacks while maintaining the perceptual quality of the watermarked audio signal at an acceptable level. To overcome this limitation, in this book, the following audio watermarking methods are proposed.

- A DWT-DCT-based audio watermarking method using SVD and quantization is introduced [22, 23]. Initially, the original audio is segmented into non-overlapping frames. Watermark information is embedded into the largest singular value of the DCT coefficients obtained from the DWT coefficients of each frame by quantization. This method provide high imperceptible watermarked sounds as well as good robustness against various attacks.
- An audio watermarking method in frequency domain based on SVD and Cartesian-polar transform (CPT) is presented [22, 24]. In this method, watermark

information is embedded in each of the Cartesian components of the highest two singular values of the low frequency FFT coefficients of each frame. The data payload of the proposed method is relatively much higher than that of the state-of-the-art methods. In addition, it shows good robustness against various attacks.

1.6 Book Organization

This book is divided into five chapters. This chapter briefly discussed a general watermarking scheme and its properties and application areas. In addition, the motivation of this book is presented. The rest of this book is organized as follows. Chapter 2 presents a background information. Chapter 3 introduces a DWT-DCT-based audio watermarking method using SVD and quantization. Chapter 4 proposes an audio watermarking method in FFT domain based on SVD and CPT. Chapter 5 concludes this book with brief summary of the key points. A future research direction is also provided in this chapter.

Chapter 2

Background Information

The previous chapter briefly discussed the application and properties of digital watermarking, outlined the motivation, and the organization of the book. This chapter presents some existing popular audio watermarking methods and some transformation and decomposition techniques used in the proposed watermarking methods.

2.1 Review of Audio Watermarking Methods

Watermarking methods are broadly categorized into three groups: time domain methods and transform domain methods. The watermarking methods belonging to each of these two categories will be discussed briefly in the following sections.

2.1.1 Time Domain Methods

Time domain based methods embed watermark information in the time domain. These methods are simple and easy to implement. Many time domain based methods have been introduced [5, 48, 57]. However, these methods are less robust against attacks and statistical techniques are often utilized to improve the robustness. Two main methods belonging to this category are least significant bit (LSB) based method and echo hiding based method.

2.1.1.1 Audio Watermarking Using LSB Coding

LSB coding is one of the earliest techniques [70] which has been widely used in audio watermarking. The standard approach is to embed the watermark bits by altering the values of particular samples of the audio signal. The watermark bits are detected by comparing the altered values of samples with the original values of samples. The main advantage of this algorithm is that it can achieve an extremely high capacity. The main disadvantage is its extremely low robustness because random changes of the signal can destroy the watermark. In addition, the alteration of the sample values introduces a low power additive white Gaussian noise (AWGN), which makes this algorithm less perceptually transparent because listeners are very sensitive to this noise.

2.1.1.2 Audio Watermarking Using Echo Hiding

Echo hiding based watermarking method embeds a watermark bit by introducing an 'echo'. An echo is a reflection of sound, arriving at the listener some time after the original sound [35]. Four parameters of the echo are initial amplitude, decay rate of the echo amplitude, 'one' offset and 'zero' offset. When the offset between the original and the echo decreases, the two signals blend. At a certain point, the human ear does not hear the original signal and the echo, but hears a single blended signal. The point at which this happens is difficult to determine exactly. It depends on the quality of the original recording, the type of sound being echoed, and the listener. The method uses two different kernels, a 'one' kernel that is used to generate a 'one' offset echo to represent a binary '1', and a 'zero' kernel that is used to generate a 'zero' offset echo to represent a binary '0'.

2.1.2 Transform Domain Methods

2.1.2.1 Audio Watermarking Using DFT

The discrete Fourier transform (DFT) is a well known and powerful computational tool for performing frequency analysis of discrete time signals. FFT is an efficient algorithm for calculating DFT. It takes a discrete signal in the time domain and transforms this signal into the discrete frequency domain. A significant number of audio watermarking methods have been reported which utilize FFT [30, 32, 50]. Most methods embed watermark information into the magnitude of the FFT components.

2.1.2.2 Audio Watermarking Using DWT

Some DWT based watermarking methods have been proposed in [12–14]. The main idea behind the proposed methods is to segment the original audio signal into many frames first and then embed watermark bits into the low or high frequency DWT coefficients.

2.1.2.3 Audio Watermarking Using Spread Spectrum

Spread Spectrum (SS) based audio watermarking is very popular in the early days of audio watermarking. It is similar to SS communications, where a narrowband signal is modified so that its energy is spread over a much larger bandwidth. As a result, the signal energy present in any single frequency is almost undetectable. Likewise, in SS watermarking, the watermark energy is spread over many frequency bins so that the energy in any one bin is very small and is difficult to detect. Some SS watermarking algorithms have been proposed in [17, 41].

2.1.3 Other Audio Watermarking Methods

2.1.3.1 Patchwork Based Audio Watermarking

The patchwork approach was first proposed by Bender et al. [6]. The main idea of patchwork based method is to select two patches (data sets) randomly from the original signal and the mean values of these two patches are modified by a constant value, which defines the watermark strength. Consider, two randomly selected patches A and B with size N . Then each element of A and B is modified according to the following Equation:

$$a'_i = a_i + d, \quad b'_i = b_i - d \quad (2.1)$$

where a'_i, b'_i are the modified sample values and a_i, b_i are the original sample values of the patches A and B , respectively, $i = 1, 2, \dots, N$, and d is a small constant.

Then, the means of the sample values $\bar{a}, \bar{b}, \bar{a'}, \bar{b'}$ are calculated as follows:

$$\bar{a} = \frac{1}{N} \sum_{i=1}^N a_i, \quad \bar{b} = \frac{1}{N} \sum_{i=1}^N b_i, \quad \bar{a'} = \frac{1}{N} \sum_{i=1}^N a'_i, \quad \bar{b'} = \frac{1}{N} \sum_{i=1}^N b'_i \quad (2.2)$$

The expected value of sample mean difference of modified sample values $E[\bar{a'} - \bar{b'}]$ is calculated to decide whether the patch contains watermark information or not. Since two patches are used rather than one, it can detect the embedded watermark bit without the original signal.

2.1.3.2 Interpolation Based Audio Watermarking

Interpolation is a technique for constructing new data points within the range of a set of discrete data [30, 34]. Polynomial interpolation is one of the well known interpolation technique. Its advantages consist of its simplicity of implementation and the good quality of the interpolant obtained from it.

2.1.3.3 Quantization Based Audio Watermarking

The quantization based scheme is one of the simplest scheme for watermarking. In [11], a QIM function is defined as follows:

$$s(x, i) = q_{\Delta}(x + d(i)) + d(i) \quad (2.3)$$

where $s(x, i)$ is a embedding function, q_{Δ} is a uniform scalar quantizer with step size Δ , x is the original signal to be quantized, i is the watermark bit number, and $d(i)$ is the dither value. Dither value $d(i)$ is defined as follows:

$$d(i, 1) = \begin{cases} d(i, 0) + \frac{\Delta}{2}, & d(i, 0) < 0 \\ d(i, 0) - \frac{\Delta}{2}, & d(i, 0) \geq 0 \end{cases} \quad (2.4)$$

where N is the total number of watermark bits, $i = 1, 2, \dots, N$, and $d(i, 1)$, $d(i, 0)$ indicate the watermark bit '1' and '0', respectively.

2.2 Signal Transformation

In this section, some signal transformation techniques used in the proposed methods are discussed briefly.

2.2.1 Discrete Fourier Transform

The DFT is a powerful general-purpose tool for audio signal processing. Mathematically, DFT can be written as:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-2j\pi nk}, \quad k = 0, 1, \dots, N-1 \quad (2.5)$$

where $x(n)$ is the input signal in time domain and $X(k)$ is the transformed signal in frequency domain. The Fourier transform gives complex valued samples and thus both the magnitude and phase coefficients can be used for embedding watermark.

2.2.2 Discrete Cosine Transform

The DCT has been widely used in signal and image processing, especially for lossy data compression. It considers samples of the audio signal as a sum of cosine functions oscillating at different frequencies. Mathematically, the DCT can be written as

$$X(k) = c(k) \sum_{n=0}^{N-1} x(n) \cos \left(\frac{\pi(2n-1)(k-1)}{2N} \right) \quad (2.6)$$

where $k = 0, 1, \dots, N-1$, $x(n)$ is the audio signal with samples of length N , and

$$c(k) = \begin{cases} \frac{1}{\sqrt{N}}, & k = 0 \\ \sqrt{\frac{2}{N}}, & k = 1, 2, \dots, N-1 \end{cases} \quad (2.7)$$

The DCT has a sophisticated characteristic of energy compaction by collecting most of the signal energy in a few samples, leaving the other samples very small in magnitude. This characteristic can be exploited in audio watermarking to reduce the deterioration in the audio signal due to watermarking.

2.2.3 Discrete Wavelet Transform

The DWT provides a time frequency representation of a signal. A signal S can be decomposed into two sets of coefficients by the DWT as shown in Fig. 2.1. The approximate (low frequency) coefficients A_i are produced by passing the signal S through a low-pass filter G_0 . The detailed (high frequency) coefficients D_i are produced by passing the signal S through a high-pass filter H_0 . Depending on the application and length of the signal, the approximate coefficients might be further decomposed into two sets of high and low frequency coefficients.

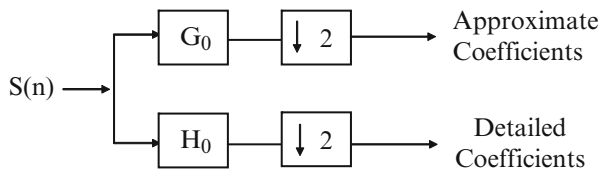


Fig. 2.1 Single-level DWT decomposition

2.2.4 Cartesian-Polar Transform

The Cartesian-polar transform (CPT) is a conformal transformation from the Cartesian coordinate system to the polar coordinate system. Consider the polar coordinate system (r, θ) , where r denotes the distance between a given point and the origin, and θ denotes the angle between a line of reference, and the line through the origin and the point. The transformation from the Cartesian coordinates to the polar coordinates is given by the following equation:

$$r = \sqrt{x^2 + y^2}, \theta = \tan^{-1} \left(\frac{y}{x} \right) \quad (2.8)$$

where (x, y) is a point in Cartesian coordinate system. The transformation from the polar coordinates to the Cartesian coordinates is given by the following equation:

$$x = r \cos \theta, y = r \sin \theta \quad (2.9)$$

2.3 Singular Value Decomposition

Let $A = \{A_{ij}\}_{N \times N}$ be an $N \times N$ matrix. The SVD of matrix A is represented in the form $A = USV^T$, where U and V are orthogonal matrices, S is a diagonal matrix with nonnegative elements, and superscript T denotes matrix transposition. The diagonal elements of S , denoted by σ_i , are called the singular values (SVs) of A and are assumed to be arranged in decreasing order $\sigma_i > \sigma_{i+1}$. The columns of U , denoted by U_i , are called the left singular vectors, while the columns of V , denoted by V_i , are called the right singular vectors of A . The SVD has some interesting properties: (i) the sizes of the matrices for SVD transformation are not fixed, and the matrices need not be square, (ii) changing SVs slightly does not affect the quality of the signal much, (iii) the SVs are invariant under common signal processing operations, and (iv) the SVs satisfy intrinsic algebraic properties.

2.4 Summary

In this chapter, some popular categories of audio watermarking methods were discussed briefly. Then some transformation and decomposition techniques utilized in the proposed methods were presented. In the rest of this book, audio watermarking methods will be introduced based on SVD, DCT, FFT, DWT, and CPT which provide better performance than the state-of-the-art methods.

Chapter 3

DWT-DCT-Based Audio Watermarking Using SVD

3.1 Introduction

This chapter presents a DWT-DCT-based audio watermarking method using SVD and quantization [22, 23]. In the proposed method, initially the original audio signal is segmented into non-overlapping frames. DWT is applied to each frame and detail coefficients are represented in matrix form. DCT is performed on the detail coefficients and the obtained DCT coefficients are reshaped. SVD is applied to the reshaped DCT coefficients of each frame. Watermark information is then embedded into the highest singular value of each audio frame by quantization. This is because (i) the DWT has spatio-frequency localization properties, (ii) the DCT has energy compression properties that improve the transparency of the watermark, (iii) changing singular values slightly does not significantly affect the quality of the audio signal, and (iv) singular values do not change significantly after various types of common signal processing attacks. Watermark information is extracted by comparing the largest singular value of original and attacked watermarked DCT coefficients obtained from the DWT sub bands of each audio frame. The major limitation of the existing SVD-based audio watermarking techniques is the difficulty in obtaining a favorable trade-off between the data payload and robustness against various attacks while maintaining the perceptual quality of the watermarked audio signal at an acceptable level. To overcome this limitation, we propose an audio watermarking method in the DWT and DCT domains using SVD and quantization, which provides better performance than the existing methods in terms of imperceptibility, robustness, and data payload. The main features of the proposed method include the following: (i) it utilizes the properties of the DWT, DCT, and SVD jointly, (ii) high-frequency DWT coefficients are arranged into a special matrix, (iii) the watermark is embedded using a new quantization function, and (iv) it achieves a good trade-off among imperceptibility, robustness, and data payload. Experimental results indicate that our proposed watermarking

method is highly robust against various attacks such as noise addition, cropping, resampling, requantization, and MP3 compression. Moreover, it outperforms state-of-the-art watermarking methods [8, 17, 27–29, 33, 54, 57, 62, 65, 66] in terms of imperceptibility, robustness, and data payload. The SNR values of the proposed method range from 38 to 41 dB, in contrast to the above state-of-the-art methods whose SNR values range from only 12 to 28 dB. In addition, the maximum bit error rate (BER) of the proposed method is 3.3203, whereas the maximum BER of the state-of-the-art methods is 51.73. Moreover, the data payload of the proposed method is 172.39 bps, which is relatively higher than that of the state-of-the-art methods.

The rest of this chapter is organized as follows. A brief overview of SVD-based methods is presented in Sect. 3.2. Then, the proposed watermarking method in DWT and DCT domains using SVD and quantization is introduced in Sect. 3.3. A comparative analysis between the proposed and state-of-the-art watermarking methods is presented in Sect. 3.4. Finally, the summary of this method is given in Sect. 3.5.

3.2 A Brief Overview of SVD-Based Methods

SVD has been used as an effective technique in digital watermarking. Most existing SVD based watermarking techniques are applied in images [4, 36, 43, 44, 49, 51, 56, 72]. Some SVD based audio watermarking techniques also exist which can be found in [2, 3, 8, 9, 27, 47, 54, 63]. All these SVD based audio watermarking algorithms can be categorized into two groups. The first group of SVD based watermarking algorithms are ‘informed’, requiring access to the original audio in order to successfully decode the embedded watermark [2, 3, 27, 54]. The scheme proposed by El-Samie [27] can be considered as a typical example to illustrate the idea used for this group. In the embedding process, the original audio is first transformed into a two dimensional matrix and then decomposed into three matrices U , S , and V by applying SVD. The watermark information is added linearly to S , resulting in a new matrix D and decomposed into three matrices U_w , S_w , and V_w using SVD. Inverse SVD is then applied to U , S_w , and V to obtain the watermarked audio. The watermark sequence can be extracted by performing the inverse operation of watermark embedding process. The second group of SVD based schemes is ‘blind’ [9, 47, 63]. These methods are based on some observations of U , S , or V matrix and do not require the original signal to extract the embedded watermark. The main limitation of the existing SVD-based audio watermarking method is the difficulty to obtain a favorable trade-off among imperceptibility, robustness, and data payload. In the next section, we propose an SVD-based audio watermarking method which provides better result than recent methods in terms of imperceptibility, robustness and data payload.

3.3 Proposed Watermarking Method

In this section, we introduce our basic watermarking method, which consists of watermark embedding and detection processes. Let $X=\{x(n), 1 \leq n \leq L\}$ be an original audio signal with L samples, $W_{BI}=\{w_{BI}(f, g), 1 \leq f \leq M, 1 \leq g \leq M\}$ be a binary image to be embedded into the original audio signal, and $w_{BI}(f, g) \in \{0, 1\}$ be the pixel value at point (f, g) .

3.3.1 Watermark Embedding Process

The proposed watermark embedding process is shown in Fig. 3.1. The embedding process is described as follows:

- Step 1:** Initially, the binary watermark image W_{BI} is converted into a one dimensional watermark sequence W_{SE} of length I , where $W_{SE}=\{w_{SE}(m), 1 \leq m \leq I\}$ and $M \times M = I$.
- Step 2:** The original audio signal X is then segmented into nonoverlapping frames $F = \{F_1, F_2, F_3, \dots, F_I\}$.
- Step 3:** A two-level DWT is performed on each frame F_i using a Haar wavelet filter. This operation produces three sets of coefficients D_1, D_2 , and A_2 , where D_1, D_2 and A_2 represent the detailed and approximate coefficients, respectively.
- Step 4:** The detailed coefficients D_1 and D_2 of each frame are arranged into a one-dimensional matrix H_i which is shown in Fig. 3.2, where i indicates the frame number. The size of the matrix is the same as that of each frame.
- Step 5:** The DCT is applied to each matrix H_i to obtain the matrix Y_i which contains the DCT coefficients. Each matrix Y_i is rearranged into an $N \times N$ square matrix R_i . This is done by dividing the matrix Y_i into N segments with N coefficients.
- Step 6:** SVD is performed to decompose each matrix R_i into three matrices: U_i, S_i , and V_i . The SVD operation is represented as follows:

$$R_i = U_i \times S_i \times V_i^T \quad (3.1)$$

- Step 7:** In order to guarantee the robustness and transparency in the proposed method, a watermark bit is embedded into the highest singular value $S_i(1, 1)$ of each matrix S_i by using a quantization function. Let $G = S_i(1, 1) \bmod Q$, where Q is a predefined quantization coefficient. A small value of Q will lead to good imperceptibility of the watermarking method but will provide low robustness to attacks, whereas a large value of Q will lead to good robustness to attacks but will provide low imperceptibility of the watermarking method. Thus, an optimal value of Q should be selected. If the bit to be embedded is a '1', we increase

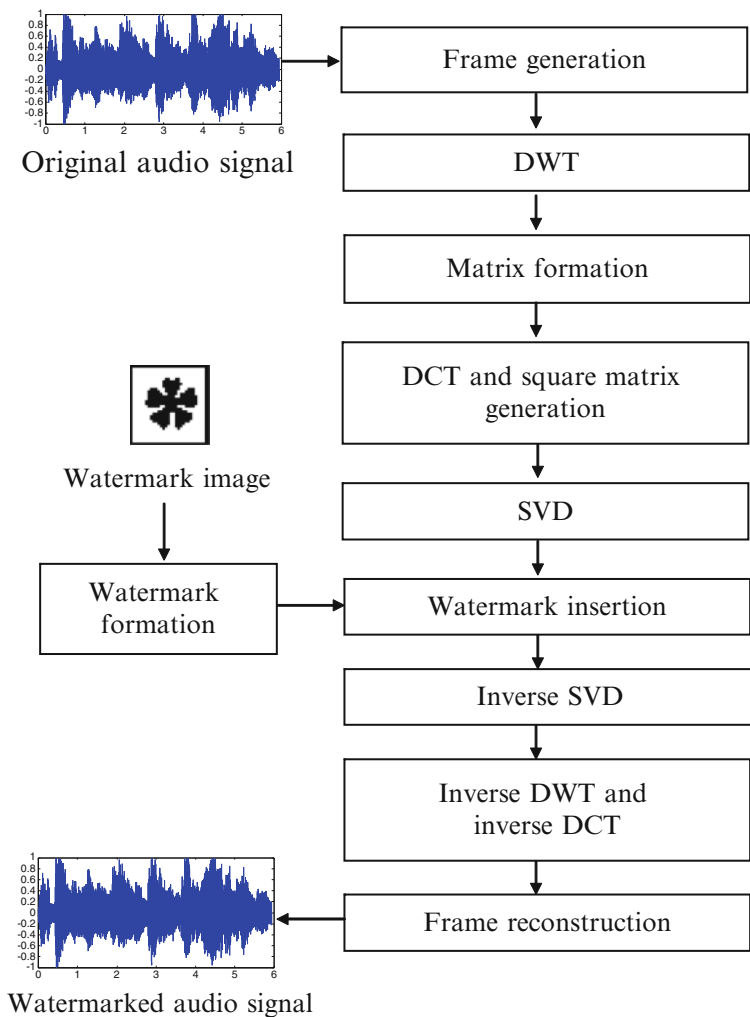


Fig. 3.1 Watermark embedding process

$$H = \begin{bmatrix} D_1 & D_2 & D_2 \end{bmatrix}$$

Fig. 3.2 Matrix formation of the detailed coefficients D_1 and D_2 for each frame (H is used as H_i for each frame i)

the value of $S_i(1, 1)$. On the other hand, if the bit to be embedded is a '0', we decrease the value of $S_i(1, 1)$. This embedding strategy can be formulated by the following quantization function:

$$S'_i(1, 1) = \begin{cases} S_i(1, 1) + \frac{Q}{C_1} + \frac{G}{C_2}, & \text{if } w_{SE}(m) = 1 \\ S_i(1, 1) - \frac{Q}{C_1} - \frac{G}{C_2}, & \text{if } w_{SE}(m) = 0 \end{cases} \quad (3.2)$$

where C_1 and C_2 are user-defined constants.

Step 8: Each modified singular value is reinserted into matrix S_i and inverse SVD is applied to obtain the modified matrix R'_i which is given by

$$R'_i = U_i \times S'_i \times V_i^T \quad (3.3)$$

Each matrix R'_i is then reshaped to create the modified matrix Y'_i by performing the inverse operation of step 5.

Step 9: The inverse DCT is performed on the modified matrix Y'_i to obtain the matrix H'_i which contains the modified detailed coefficients D'_1 and D'_2 .

Step 10: After substituting the coefficients D'_1 and D'_2 for D_1 and D_2 , respectively, a two-level inverse DWT is performed to obtain the watermarked audio frame.

Step 11: Finally, all watermarked frames are concatenated to calculate the watermarked audio signal.

3.3.2 Watermark Detection Process

The proposed watermark detection process is shown in Fig. 3.3. The detection process is described as follows:

Step 1: A two-level DWT is performed on each frame F_i^* of the attacked watermarked audio signal X^* using a Haar wavelet filter.

Step 2: The detailed coefficients D_1^* and D_2^* of each frame are arranged into a matrix H_i^* and the DCT is applied to each matrix X_i^* to obtain Y_i^* .

Step 3: Each matrix Y_i^* is rearranged to obtain R_i^* and SVD is performed on it.

Step 4: The largest singular value of each diagonal matrix S_i^* located at the same position in the pre-embedding process is calculated.

Step 5: The watermark sequence is extracted as follows:

$$w_{SE}^*(m) = \begin{cases} 1, & \text{if } S_i^*(1, 1) > S_i(1, 1) \\ 0, & \text{otherwise} \end{cases} \quad (3.4)$$

where $S_i(1, 1)$ and $S_i^*(1, 1)$ are the largest singular values of the original and attacked watermarked audio frames, respectively.

Step 6: Finally, the binary watermark image W_{BI}^* is obtained by rearranging the watermark sequence into a square matrix of size $M \times M$.

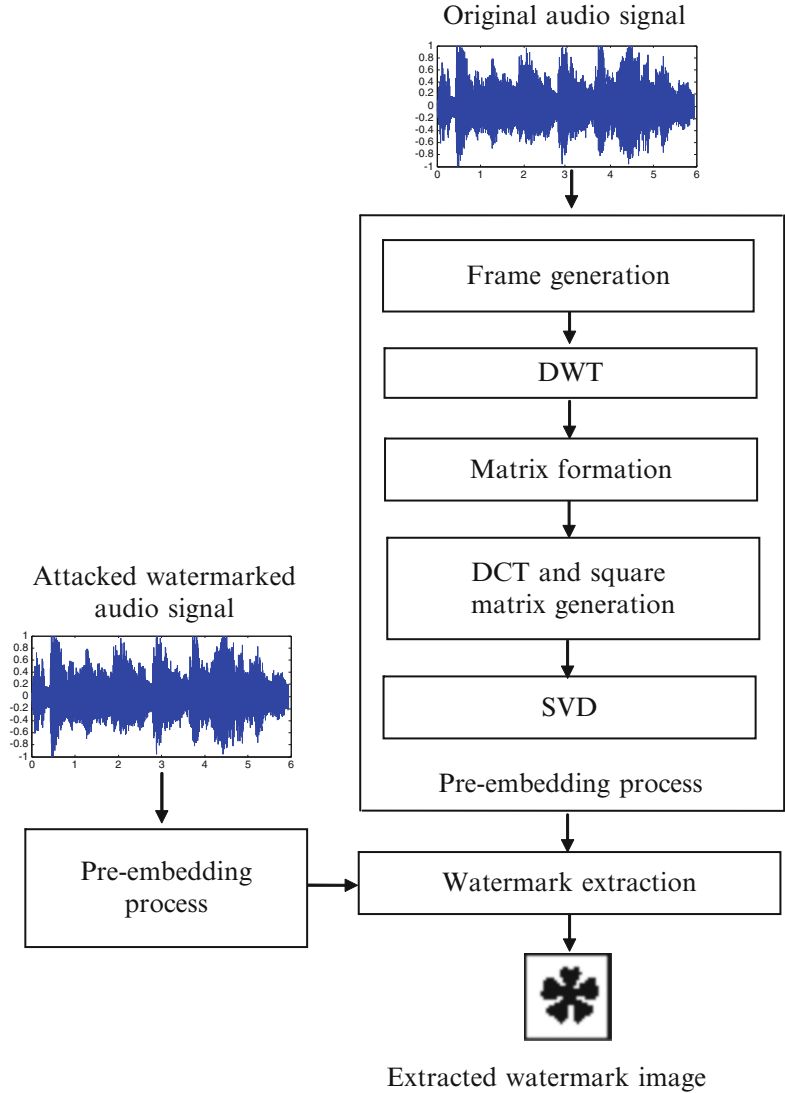


Fig. 3.3 Watermark detection process

3.4 Experimental Results and Discussion

In this section, we report several experiments carried out to demonstrate the performance of the proposed watermarking method. The performance of our method is assessed in terms of imperceptibility, robustness, and data payload. In this study, four different types of 16-bit mono audio signals (Pop, Jazz, Folk, and Classical)

Fig. 3.4 Binary watermark

in wave format sampled at 44.1 kHz were used. Each audio file contains 262,144 samples (duration 5.94 s). By using a frame size of 256 samples, we have 1,024 nonoverlapping frames for each audio signal. In each frame of the audio signal, we embed 1-bit binary watermark information. The embedded watermark is a binary logo image of size $M \times M = 32 \times 32 = 1,024$ bits, as shown in Fig. 3.4. The selected value for the quantization coefficient Q is 0.5. For convenience, the selected values for constants C_1 and C_2 are both 8. These parameters were selected in order to achieve a good trade-off among the conflicting requirements of imperceptibility, robustness, and data payload.

Usually, a lower level DWT adversely influences the robustness of a watermark; on the other hand, a higher-level DWT takes a long time to execute. Thus, a two-level DWT is applied to each audio frame. The number of coefficients D_1 and D_2 are 128 and 64, respectively. After arranging the detailed coefficients D_1 and D_2 using step 4 of the embedding process, the size of each matrix H_i is 1×256 . The DCT is applied to H_i to obtain the matrix Y_i which contains the DCT coefficients, and each Y_i is reshaped to obtain the matrix R_i of size 16×16 . Each R_i is decomposed into three matrices by applying SVD. Figure 3.5 shows the time domain representation and the corresponding DWT domain representation using the coefficients A_2 , D_2 , and D_1 of a selected frame for the original audio signal ‘Classical’. Figure 3.6 shows the matrix H_i representing the detailed coefficients D_1 and D_2 of a selected frame for the signal ‘Classical’. Figure 3.7 shows the coefficients in the DCT domain represented by matrix Y_i after applying the DCT operation to matrix H_i . Figure 3.8 shows the modified DCT coefficients represented by matrix Y'_i after embedding the watermark. The watermark information is embedded into the largest singular value of the DCT coefficients obtained from the DWT subbands of each audio frame. We observed that the variation of the largest singular value modifies all DCT coefficients slightly, which can provide high robustness against different attacks as well as good-quality watermarked sounds.

3.4.1 Imperceptibility Test

The imperceptibility of the watermarked audio signal is evaluated using a subjective listening test and an objective test.

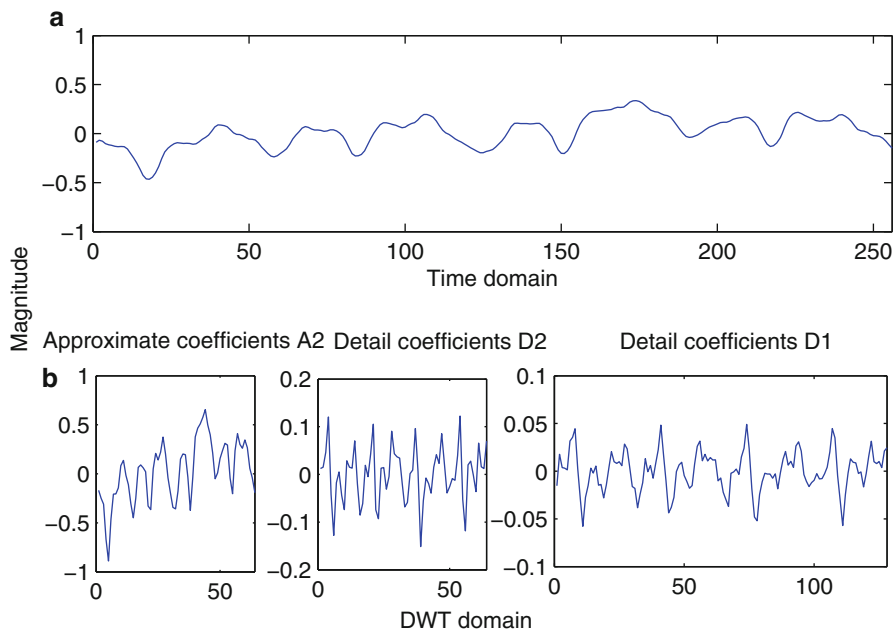


Fig. 3.5 (a) Time domain representation of a selected frame for the signal 'Classical', (b) DWT domain representation using the coefficients A_2 , D_2 , and D_1 of the same frame for the signal 'Classical'

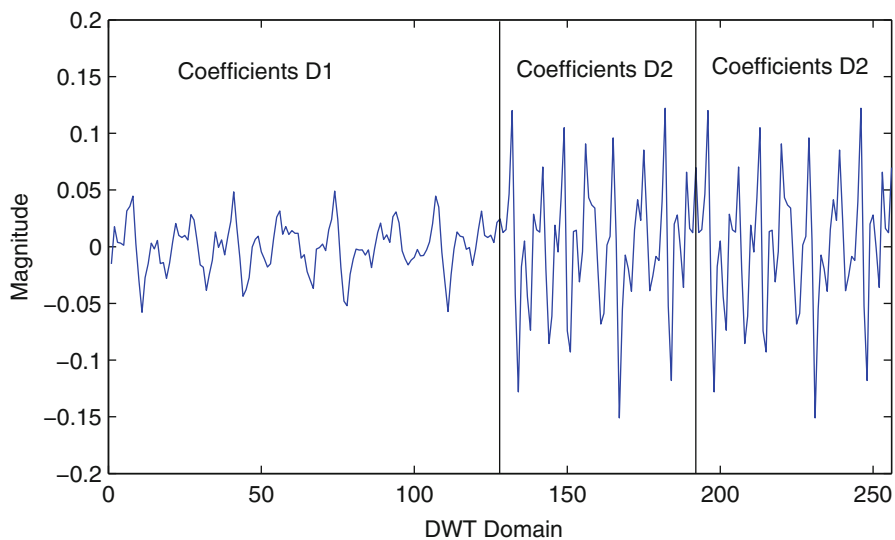


Fig. 3.6 Representation of the matrix H_i of a selected frame for the signal 'Classical'

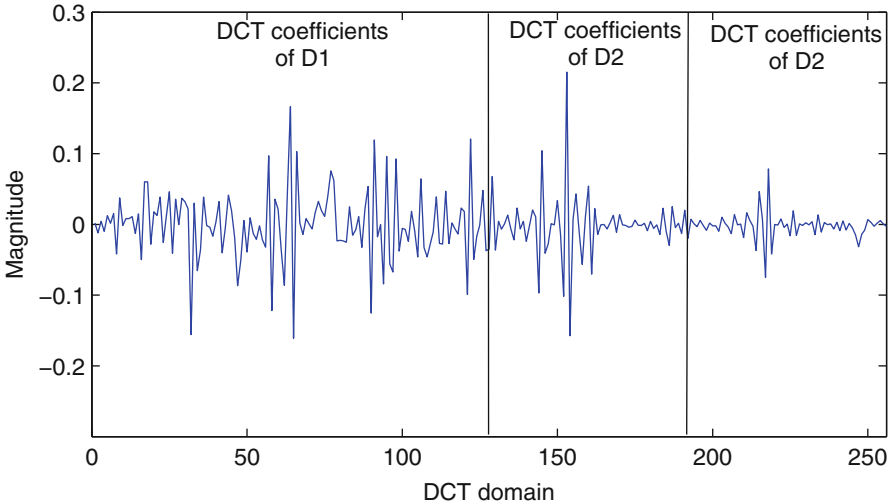


Fig. 3.7 Representation of the matrix Y_i of a selected frame for the signal ‘Classical’

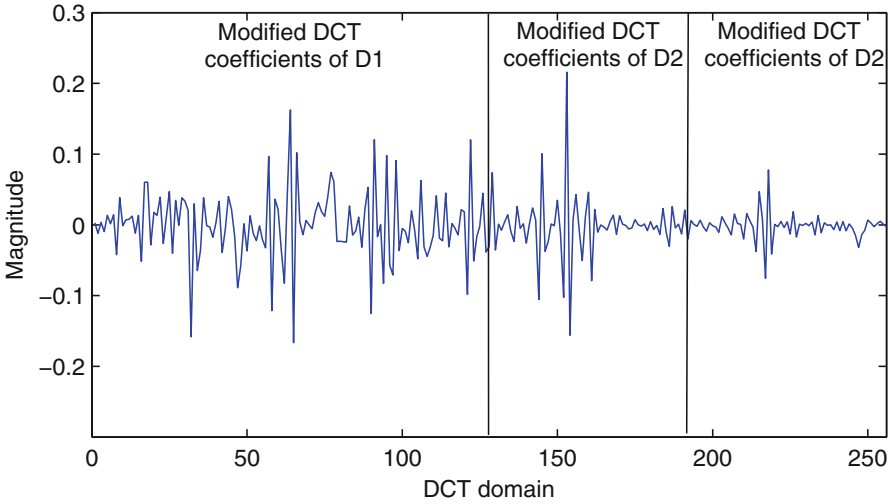


Fig. 3.8 Representation of the matrix Y'_i of a selected frame for the signal ‘Classical’

Table 3.1 Subjective and objective difference grade

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

3.4.1.1 Subjective Listening Test

The subjective difference grade (SDG) is one of the most widely used subjective methods for evaluating the quality of a watermarked audio signal. The SDG ranges from 5.0 to 1.0 (imperceptible to very annoying) as shown in Table 3.1. This test is essential for perceptual quality assessment, since the ultimate judgment is made by human acoustic perception. In the subjective listening test, the original and watermarked audio signals were given to ten participants, who were then asked to classify the difference using the SDG as shown in Table 3.1. Each audio signal was presented through headphones. Before conducting the experiment, the procedures were clearly described to the participants, who were trained in evaluating the sound quality effectively. Each participant adjusted the sound volume to their usual listening level. The participants were male and female and of different ages (22–35 years) with normal hearing ability and were suitable candidates for evaluating the sound quality effectively. The average SDG (i.e., mean opinion) scores for different watermarked sounds using the proposed method are shown in Table 3.2. From the test results, we observed that the mean opinion score (MOS) ranges from 4.9 to 5.0 for all watermarked sounds using the proposed method, indicating that original and watermarked audio signals are perceptually indistinguishable.

In addition, the ABX method, which is another subjective quality assessment technique, was employed to investigate the sound quality of the watermarked audio signal. The participants were ten male and female persons with normal hearing ability. In this test, the original audio signal ‘A’ and the watermarked audio signal ‘B’ were presented to each participant. A third audio signal ‘X’ is presented in random order, which can be either ‘A’ or ‘B’. The participants were asked to identify which of ‘A’ or ‘B’ was the same as ‘X’. One time of identification was assumed as one trial and five trials were carried out by each participant. The number of correct detections is used to determine whether the watermarked audio signal is perceptible or not. A high percentage of correct detection clearly indicates the perceptibility of the watermark in the audio signal. On the other hand, a detection percentage of 50 % suggests that the difference between the original and watermarked sounds was imperceptible. The evaluation results of all participants were summarized in terms of the percentage of correct detection and are shown in Table 3.2. We observed that the correct detection scores range from 44 to 52 %, indicating that the proposed watermarking method provides good imperceptible watermarked sound.

Table 3.2 Subjective and objective evaluation of different water-marked sounds

| Audio signal | Subjective evaluation | | Objective evaluation | |
|--------------|-----------------------|-----------------------|----------------------|-------|
| | MOS | Correct detection (%) | ODG | SNR |
| Pop | 4.9 | 52 | −0.773 | 38.02 |
| Jazz | 5.0 | 44 | −1.088 | 40.80 |
| Folk | 5.0 | 46 | −0.280 | 41.02 |
| Classical | 5.0 | 48 | −0.511 | 39.81 |
| Average | 4.97 | 47.50 | −0.653 | 39.91 |

3.4.1.2 Objective Test

The objective quality of a watermarked audio signal is measured by the SNR which is defined as

$$SNR = 10 \log_{10} \frac{\sum_{n=1}^L x^2(n)}{\sum_{n=1}^L [x(n) - x^*(n)]^2} \quad (3.5)$$

where $x(n)$ and $x^*(n)$ are the original and watermarked audio signals, respectively. After embedding a watermark, the SNR values of all selected audio signals using the proposed method are above 20 dB, conforming to the International Federation of the Phonographic Industry (IFPI) standard [8], as shown in Table 3.2.

Objective evaluation was also carried out using the objective difference grade (ODG), another objective quality assessment technique which ranges from 0.0 to −4.0 (imperceptible to very annoying) as shown in Table 3.1. The ODG is one of the output values obtained from the perceptual evaluation of audio quality (PEAQ) measurement technique specified in ITU-R BS.1387 (International Telecommunication Union-Radio-communication Sector) standard [59]. It corresponds to the subjective grade used in human based audio tests. The objective quality of the watermarked audio signals is calculated in terms of ODG and shown in Table 3.2. We observed that all ODG values range from −1.088 to −0.280, indicating that original and watermarked audio signals are perceptually indistinguishable.

Figures 3.9 and 3.10, respectively show a qualitative evaluation of the original audio with a watermarked audio, in which the watermark is imperceptible in the time domain, and a spectrogram representation using the proposed method for the audio signal ‘Classical’.

Table 3.3 shows a comparison of SNR and MOS results among the proposed method and several recent methods, which are based on the reported results in [8, 17, 27, 29, 54, 57]. From the comparison of results, we observed that our proposed method outperforms these audio watermarking methods in terms of SNR and MOS results. In other words, subjective and objective evaluations prove the high transparency of the proposed method with higher MOS and SNR than the other methods. This is because watermark information is embedded into the largest

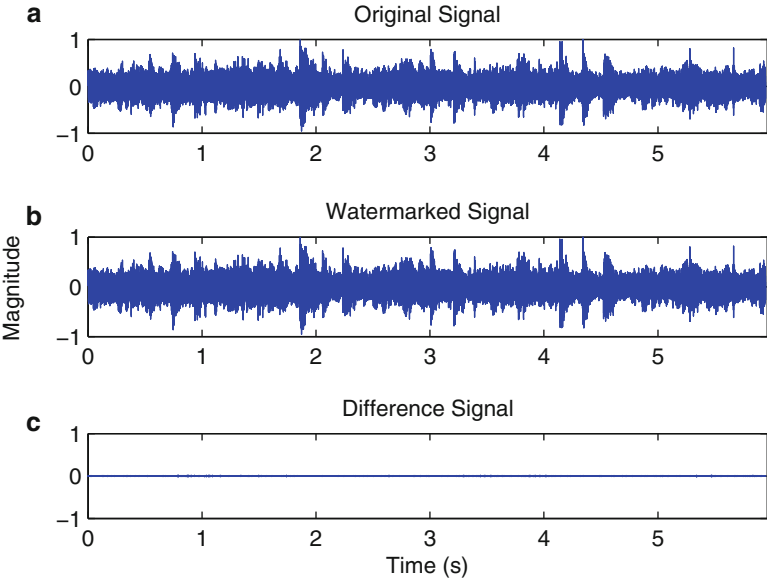


Fig. 3.9 Imperceptibility of the watermarked audio using the proposed method: (a) Original signal ‘Classical’, (b) Watermarked signal ‘Classical’, (c) Difference between original and watermarked signals

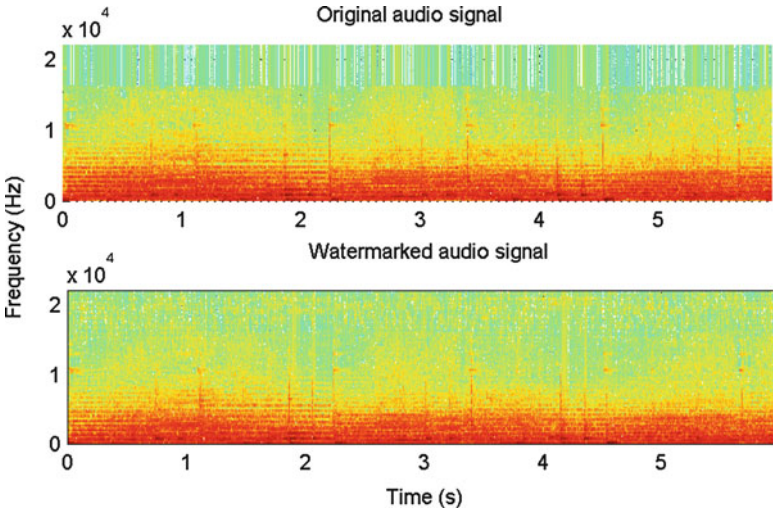


Fig. 3.10 Spectrogram representation of the original audio signal and watermarked audio signal ‘Classical’ using the proposed method

Table 3.3 Comparison of SNR and MOS between the proposed method and several recent methods

| Reference | Algorithm | SNR | MOS |
|-----------|-------------------|-------|------|
| [17] | Spread spectrum | 28.59 | 4.46 |
| [54] | STFT-SVD | 28.36 | 4.70 |
| [27] | SVD | 27.13 | 4.60 |
| [8] | DWT-SVD | 26.84 | 4.60 |
| [57] | Frequency masking | 12.87 | 2.93 |
| [29] | TS echo hiding | 22.70 | 4.70 |
| Proposed | DWT-DCT-SVD | 39.91 | 4.97 |

singular value of the DCT coefficients obtained from the detailed DWT coefficients of each audio frame and slight variations of the largest singular values do not significantly affect the quality of the sound.

3.4.2 Robustness Test

In order to test the robustness of the proposed method, the normalized correlation (NC) and BER are calculated. The NC is used to evaluate the correlation between the original watermark W_{SE} and the extracted watermark W_{SE}^* and is given by the following equation:

$$NC(W_{SE}, W_{SE}^*) = \frac{\sum_{m=1}^I w_{SE}(m) \cdot w_{SE}(m)^*}{\sqrt{\sum_{m=1}^I w_{SE}(m) \cdot w_{SE}(m)} \sqrt{\sum_{m=1}^I w_{SE}(m)^* \cdot w_{SE}(m)^*}} \quad (3.6)$$

If $NC(W_{SE}, W_{SE}^*)$ is close to 1, the similarity between W_{SE} and W_{SE}^* is very high. If $NC(W_{SE}, W_{SE}^*)$ is close to 0, the similarity between W_{SE} and W_{SE}^* is very low.





The BER is used to evaluate the accuracy of watermark detection after applying different attacks. The BER of the retrieved watermarked signal is defined as follows:

$$BER(W_{SE}, W_{SE}^*) = \frac{\sum_{m=1}^I w_{SE}(m) \oplus w_{SE}(m)^*}{I} \quad (3.7)$$

where \oplus is the exclusive-or (XOR) operation.

In order to test the robustness of the proposed method, the following signal processing attacks were performed on the watermarked audio signal.

1. Additive white Gaussian noise (AWGN): AWGN is added to the watermarked audio signal.
2. Cropping: 1,000 samples are removed from the front, middle, and end parts of the watermarked audio signal and then these samples are replaced by the watermarked samples attacked with AWGN.

| | | | | |
|---------------------|---|---|---|---|
| Attack | No attack | Noise addition | Cropping (front) | Cropping (middle) |
| NC | 1 | 0.9977 | 0.9991 | 0.9995 |
| BER | 0 | 0.2441 | 0.0977 | 0.0488 |
| Extracted watermark |  |  |  |  |





| | | | | |
|---------------------|---|---|---|---|
| Attack | Cropping (end) | Resampling | Requantization | MP3 compression |
| NC | 0.9986 | 1 | 1 | 0.9979 |
| BER | 0.1465 | 0 | 0 | 3.2227 |
| Extracted watermark |  |  |  |  |

Fig. 3.11 Extracted watermark image with NC and BER for the audio signal ‘Classical’

- 3. Resampling: The watermarked signal originally sampled at 44.1 kHz is resampled at 22.050 kHz and then restored by sampling again at 44.1 kHz.
- 4. Requantization: The 16-bit watermarked audio signal is quantized down to 8 bits/sample and again requantized back to 16 bits/sample.
- 5. MP3 compression: MPEG-1 layer 3 compression is applied. The watermarked audio signal is compressed at a bit rate of 128 kbps and then decompressed back to the wave format.

Sound Forge 8.0 was used in our experiment for resampling, requantization, and MP3 compression attacks. The AWGN and cropping operations were implemented using MATLAB 7.8.

Figure 3.11 shows the extracted watermarks along with the NC and BER values after several different attacks for the audio signal ‘Classical’. The minimum NC value and the maximum BER value are 0.9729 and 3.2227, respectively. The extracted watermark images are visually similar to the original watermark image. This clearly shows the good performance of the proposed method against different attacks. Table 3.4 summarizes the robustness results of the proposed method for the audio signals ‘Pop’, ‘Jazz’, and ‘Folk’. The NC values are all above 0.96 and the BER values are all below 4 %, indicating the high robustness of our proposed method against different attacks.

3.4.3 Error Analysis

We analyzed the performance of the proposed method in terms of the false positive error (FPE) and false negative error (FNE). It is difficult to give an exact probabilistic model for an FPE and an FNE. Here, we utilized a simplified model

Table 3.4 NC and BER of the extracted watermarks for different audio signals

| Audio signal | Attack | NC | BER (%) |
|--------------|-------------------|--------|---------|
| Pop | No attack | 1 | 0 |
| | Noise addition | 1 | 0 |
| | Cropping (front) | 0.9992 | 0.0977 |
| | Cropping (middle) | 0.9976 | 0.2930 |
| | Cropping (end) | 1 | 0 |
| | Resampling | 1 | 0 |
| | Requantization | 1 | 0 |
| | MP3 compression | 0.9870 | 1.5625 |
| Jazz | No attack | 1 | 0 |
| | Noise addition | 0.9935 | 0.7813 |
| | Cropping (front) | 0.9992 | 0.0977 |
| | Cropping (middle) | 0.9992 | 0.0977 |
| | Cropping (end) | 1 | 0 |
| | Resampling | 1 | 0 |
| | Requantization | 1 | 0 |
| | MP3 compression | 0.9810 | 2.3438 |
| Folk | No attack | 1 | 0 |
| | Noise addition | 0.9943 | 0.6836 |
| | Cropping (front) | 1 | 0 |
| | Cropping (middle) | 0.9984 | 0.1953 |
| | Cropping (end) | 0.9992 | 0.0977 |
| | Resampling | 1 | 0 |
| | Requantization | 1 | 0 |
| | MP3 compression | 0.9721 | 3.3203 |

based on a binomial probability distribution similar to that in [8] to estimate the probability of an FPE and an FNE for our proposed method.

3.4.3.1 False Positive Error

An FPE occurs when an unwatermarked audio signal is declared as a watermarked audio signal by the detector. Let k be the total number of watermark bits and h be the total number of matching bits. The FPE probability, denoted by P_{FPE} , can be calculated as follows:

$$P_{FPE} = 2^{-k} \sum_{h=\lceil 0.8k \rceil}^k \binom{k}{h} \quad (3.8)$$

where $\binom{k}{h}$ is the binomial coefficient. We observed that Eq. 3.8 is independent of the BER. This indicates that an FPE is independent of the type of attacks. As we

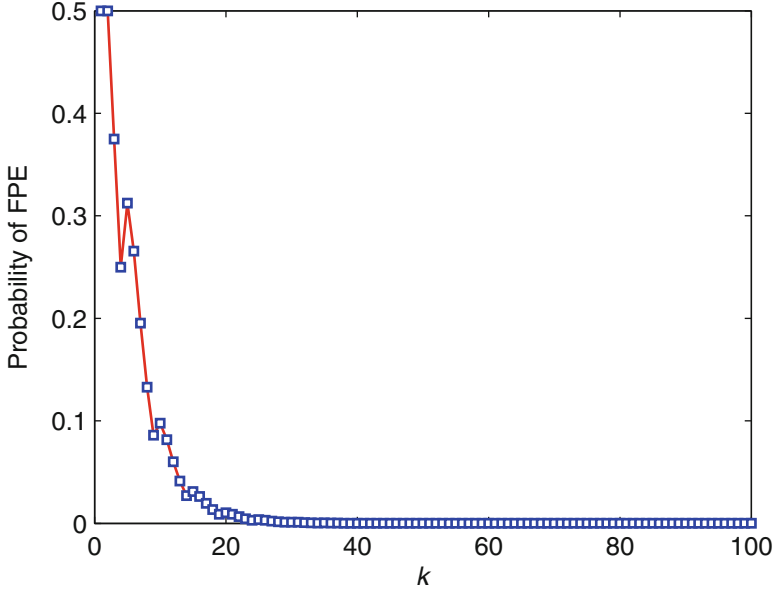


Fig. 3.12 Probability of FPE for various values of k

select $k = 1,024$, by substituting the value of h , Eq. 3.8 gives $P_{FPE} = 2.6209 \times 10^{-88}$. Figure 3.12 shows the FPE probability for $k \in (0, 100]$, which indicates that the FPE probability approaches 0 when k is larger than 20.

3.4.3.2 False Negative Error

An FNE occurs when a watermarked audio signal is declared as an unwatermarked audio signal by the detector. The FNE probability P_{FNE} can be calculated as follows:

$$P_{FNE} = \sum_{h=0}^{\lceil 0.8k \rceil - 1} \left[\binom{k}{h} (P)^h (1 - P)^{k-h} \right] \quad (3.9)$$

where P is the BER probability of the extracted watermark. The approximate value of P can be obtained from the BER under different attacks. From the simulation results, we observed that all the BER values are less than 0.04. Thus P is taken as 0.96. By substituting the values of k and P , Eq. 3.9 gives $P_{FNE} = 1.3566 \times 10^{-80}$. Figure 3.13 shows the FNE probability for $k \in (0, 100]$, which indicates that the FNE probability approaches 0 when k is larger than 20.

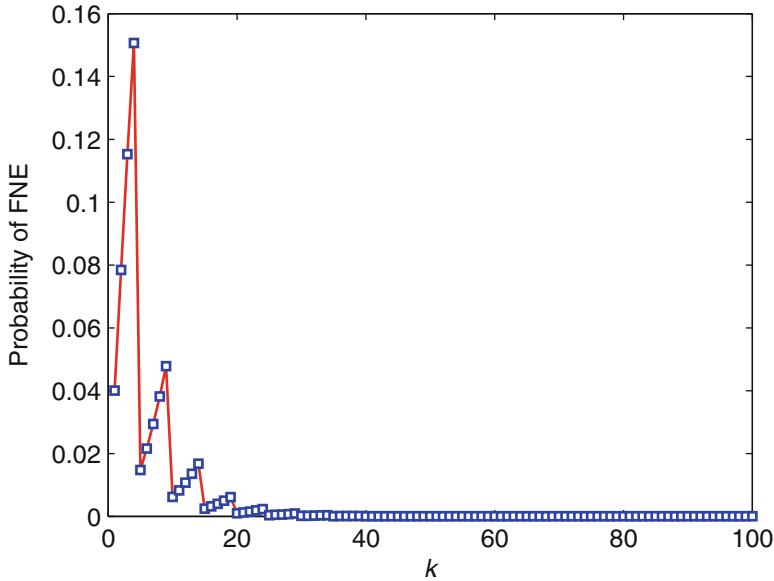


Fig. 3.13 Probability of FNE for various values of k

3.4.4 Data Payload

The data payload (also known as the capacity) of a watermarking method is defined as the number of bits that can be embedded into the original audio signal within a unit of time and is measured in bits per second (bps). If the duration of the original audio signal is T_s in seconds and the number of watermark bits to be embedded into the original audio signal is I , the data payload DP can be represented by

$$DP = \frac{I}{T_s} \text{ (bps)} \quad (3.10)$$

Usually, the data payload for any watermarking method should be more than 20 bps [8]. The data payload of the proposed method is 172.39 bps.

3.4.5 Algorithm Comparison and Discussion

Due to the diversity of watermark embedding approaches, a general comparison between the proposed method and several recent methods, sorted by data payload, is given in Table 3.5, which is based on the reported results in [8, 28, 29, 33, 62, 65, 66]. In addition, resampling, requantization, and MP3 compression are compared

Table 3.5 A general comparison between the proposed method and recent audio watermarking methods sorted by data payload

| Reference | Algorithm | Payload (bps) | Resampling BER(%) | Requantization BER(%) | MP3 Compression BER(%) |
|-----------|---------------------------|------------------|----------------------|--------------------------|---------------------------|
| Proposed | DWT-DCT-SVD | 172.3906 | 0 (22.05 kHz) | 0 (8 bits/sample) | 3.3203 (128 kbps) |
| [33] | Chaos-based DFST | 86 | 0 (22.05 kHz) | – | 3.47 (48 kbps) |
| [8] | DWT-SVD | 45.90 | 2 (22.05 kHz) | 0 (8 bits/sample) | 1 (32 kbps) |
| [62] | DWT-LPC | 10.72 | 13.64 (22.05 kHz) | 5.24 (8 bits/sample) | 5.71 (128 kbps) |
| [65] | Histogram | 3 | 0 (–) | 0 (8 bits/sample) | 15 (128 kbps) |
| [66] | DWT-based histogram | 2 | 0 (16 kHz) | 0 (8 bits/sample) | 17.5 (64 kbps) |
| [29] | TS echo hiding | – | 15 (16 kHz) | 5.5 (8 bits/sample) | 47 (–) |
| [28] | Lifting wavelet transform | – | 16.50 (36.750 kHz) | 22.09 (8 bits/sample) | 51.73 (128 kbps) |

in Table 3.5. From the comparison of results, it is seen that the proposed method achieves a higher data payload and a lower BER against several attacks than the state-of-the-art methods.

Overall, our proposed method provides superior performance to the state-of-the-art audio watermarking methods while maintaining a very good trade-off among the conflicting requirements of imperceptibility, robustness, and data payload.

3.5 Summary

In this chapter, at first a brief discussion on SVD-based watermarking methods was presented. Then, we proposed an audio watermarking method in the DWT and DCT domains based on SVD and quantization. Experimental results indicate that the proposed method provides good imperceptible watermarked sound as well as high robustness against different attacks such as noise addition, cropping, resampling, requantization, and MP3 compression. This is because the watermark is embedded into the largest singular value of the DCT coefficients obtained from the DWT subbands of each audio frame, and slight variations of the largest singular values do not significantly affect the quality of the sound. Moreover, these values change very little in the case of different attacks. Moreover, it also has very low error probability rates. In addition, the proposed method outperforms state-of-the-art audio watermarking methods in terms of imperceptibility, robustness, and data payload.

Chapter 4

FFT-Based Audio Watermarking Using SVD and CPT

4.1 Introduction

In this chapter, we present a FFT-based audio watermarking method using SVD and CPT [22, 24]. In our proposed method, watermark information is pre-processed first using a Gaussian map in order to improve the robustness and enhance the confidentiality. Then, the original audio is segmented into nonoverlapping frames. FFT is applied to each frame and low frequency FFT coefficients are selected. SVD is applied to the selected FFT coefficients of each frame represented in a matrix form. The highest two singular values of each frame are selected. The selected singular values are assumed as the components of polar coordinate system and are transformed into the components of Cartesian coordinate system. Watermark information is embedded into each of these Cartesian components using an embedding function. This is because the low frequency FFT coefficients correspond to the energy of the most perceptually significant regions in an audio signal and slight variations of the Cartesian components of the largest singular values do not significantly affect the quality of the signal. Experimental results indicate that the proposed watermarking method is highly robust against various signal processing attacks. In addition, the proposed method has a high data payload. Moreover, it outperforms state-of-the-art audio watermarking methods in terms of imperceptibility, robustness, and data payload. The main features of the proposed method include (i) it utilizes the FFT, SVD, and CPT jointly, (ii) it uses Gaussian map, containing the chaotic characteristic to enhance the confidentiality of the proposed method, (iii) the watermark bits are embedded using a new embedding function, (iv) subjective and objective evaluations reveal that the proposed method maintains a high audio quality, and (v) it achieves a good trade-off among imperceptibility, robustness, and data payload. Experimental results demonstrate that our proposed watermarking method resists various attacks such as noise addition, cropping, resampling, requantization, and MP3 compression. Moreover,

it outperforms the state-of-the-art watermarking algorithms [2, 8, 9, 12, 13, 17, 27–29, 32, 33, 39, 47, 54, 57, 62, 65, 66] in terms of imperceptibility, robustness, and data payload. The SNR and MOS values of the proposed method range from 36.19 to 37.32 dB and 4.95 to 5.0, respectively. This is in contrast to the above state-of-the-art methods whose SNR and MOS values range from only 12.87 to 29.50 dB and 2.93 to 4.7, respectively. Moreover, the BER of the proposed method ranges from 0 to 2.9297, whereas the BER of the state-of-the-art methods ranges from 0 to 51.73. Furthermore, the data payload of the proposed method is 689.56 bps which is relatively higher than that of the state-of-the-art methods.

The rest of this chapter is organized as follows. Section 4.2 introduces our proposed watermarking method including watermark preprocessing, watermark embedding process, and watermark detection process. Section 4.3 evaluates the performance of the proposed method in terms of imperceptibility, robustness, and data payload. Finally, Sect. 4.4 concludes with the brief summary of the key points.

4.2 Proposed Watermarking Method

In this section, an overview of our basic watermarking method is presented, which consists of watermark preprocessing, watermark embedding process, and watermark detection process. Let $X=\{x(n), 1 \leq n \leq L\}$ be an original audio signal with L samples, $W_{BI}=\{w_{BI}(f, g), 1 \leq f \leq M, 1 \leq g \leq M\}$ be a binary image to be embedded into the original audio signal, and $w_{BI}(f, g) \in \{0, 1\}$ be the pixel value at point (f, g) .

4.2.1 Watermark Preprocessing

Watermark should be preprocessed first in order to improve the robustness and enhance the confidentiality. The binary image will be chaotically encrypted before embedding to increase the confidentiality of the watermarking method. This paper uses a Gaussian map to enhance the confidentiality of the watermarking method which can be defined as follows:

$$y(m+1) = \exp(-a(y(m)^2)) + b \quad (4.1)$$

where $y(1) \in (0, 1)$, a and b are real parameters (map's initial condition). Then a binary sequence $z(m)$ is calculated using the following rule:

$$z(m) = \begin{cases} 1, & \text{if } y(m) > T_h \\ 0, & \text{otherwise} \end{cases} \quad (4.2)$$

where T_h is a predefined threshold. The binary watermark image W_{BI} is converted into a one-dimensional watermark sequence W_{SE} of length I , where $W_{SE} = \{w_{SE}(m), 1 \leq m \leq I\}$ and $M \times M = I$. Finally, $w_{SE}(m)$ is encrypted by $z(m)$ using the following rule:

$$u(m) = z(m) \oplus w_{SE}(m), \quad 1 \leq m \leq I \quad (4.3)$$

where \oplus is an exclusive-or (XOR) operation. Here, the value of $y(1)$, a , and b are used as secret key K_1 .

4.2.2 Watermark Embedding Process

The proposed watermark embedding process is shown in Fig. 4.1. The embedding process is described as follows:

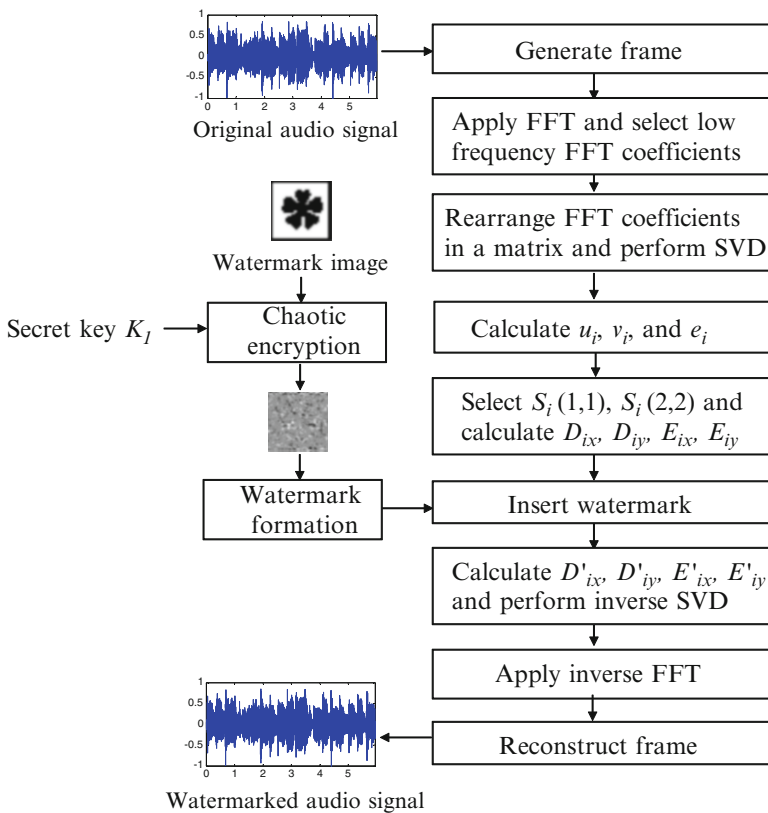


Fig. 4.1 Watermark embedding process

Step 1: The original audio signal X is first segmented into nonoverlapping frames $F = \{F_1, F_2, F_3, \dots, F_{I/4}\}$.

Step 2: Each audio frame F_i is transformed into FFT domain to calculate the FFT coefficients, where i indicates the frame number.

Step 3: The first l low frequency FFT coefficients of each frame F_i are selected and these coefficients are rearranged into an $N \times N$ square matrix R_i . This is done by dividing the coefficient set into N segments with N coefficients.

Step 4: SVD is performed to decompose each matrix R_i into three matrices: U_i , S_i , and V_i . The SVD operation is represented as follows:

$$R_i = U_i \times S_i \times V_i^T \quad (4.4)$$

Step 5: The mean u_i , variance v_i , and Euclidean norms e_i of the diagonal elements $(\sigma_1, \sigma_2, \dots, \sigma_d)$ of each matrix S_i are calculated using the following equations:

$$u_i = \frac{1}{d} \sum_{c=1}^d \sigma_c \quad (4.5)$$

$$v_i = \frac{1}{d} \sum_{c=1}^d (\sigma_c - u_i)^2 \quad (4.6)$$

$$e_i = \sqrt{\frac{1}{d} \sum_{c=1}^d \sigma_c^2} \quad (4.7)$$

where c is the index of diagonal element of each matrix S_i .

Step 6: The highest two singular values $S_i(1, 1)$ and $S_i(2, 2)$ of each matrix S_i , which are assumed as the component of polar coordinate system, are decomposed into two components D_{ix} , D_{iy} and E_{ix} , E_{iy} , respectively using the following polar-to-Cartesian transformation:

$$D_{ix} = S_i \cos \theta_1, D_{iy} = S_i \sin \theta_1 \quad (4.8)$$

$$E_{ix} = S_i \cos \theta_2, E_{iy} = S_i \sin \theta_2 \quad (4.9)$$

where θ_1 and θ_2 are the predefined angle of decomposition. For consistency between D_{ix} and D_{iy} , and E_{ix} and E_{iy} the selected value for θ_1 and θ_2 is 45° . The components D_{ix} , D_{iy} and E_{ix} , E_{iy} of the highest two singular values $S_i(1, 1)$ and $S_i(2, 2)$ of each matrix S_i are preserved to utilize in detection process and are used as secret key K_2 .

Step 7: The binary watermark image is encrypted using a Gaussian map (described in Sect. 4.2.1) which contains the chaotic characteristics.

Step 8: In order to guarantee the robustness and transparency, the proposed method embeds watermark bit into each of D_{ix} , D_{iy} , E_{ix} , and E_{iy} of the matrix S_i using u_i , v_i , and e_i . This ensures that the watermark is located at the most significant

perceptual components of the audio signal. If the bit to be embedded is '1', the following embedding equations are used:

$$D'_{ix} = D_{ix} + \frac{u_i}{C_1} + \frac{v_i}{C_2} + \frac{e_i}{C_3} \quad (4.10)$$

$$E'_{ix} = E_{ix} + \frac{u_i}{C_1} + \frac{v_i}{C_2} + \frac{e_i}{C_3} \quad (4.11)$$

where D'_{ix} and E'_{ix} are the modified x component of $S_i(1, 1)$ and $S_i(2, 2)$, respectively and C_1 , C_2 , and C_3 are the user-defined constants.

Step 9: If the bit to be embedded is '0', the following embedding equations are used:

$$D'_{ix} = D_{ix} - \frac{m_i}{C_1} - \frac{v_i}{C_2} - \frac{e_i}{C_3} \quad (4.12)$$

$$E'_{ix} = E_{ix} - \frac{m_i}{C_1} - \frac{v_i}{C_2} - \frac{e_i}{C_3} \quad (4.13)$$

where D'_{iy} and E'_{iy} are the modified y component of $S_i(1, 1)$ and $S_i(2, 2)$, respectively.

Step 10: The modified highest singular values $S'_i(1, 1)$ and $S'_i(2, 2)$ of each matrix S_i are obtained using the following Cartesian-to-polar transformation:

$$S'_i(1, 1) = \sqrt{D'^2_{ix} + D'^2_{iy}}, \theta'_1 = \tan^{-1} \left(\frac{D'_{iy}}{D'_{ix}} \right) \quad (4.14)$$

$$S'_i(2, 2) = \sqrt{E'^2_{ix} + E'^2_{iy}}, \theta'_2 = \tan^{-1} \left(\frac{E'_{iy}}{E'_{ix}} \right) \quad (4.15)$$

The modified angles θ'_1 and θ'_2 of the modified highest singular values $S'_i(1, 1)$ and $S'_i(2, 2)$ of each matrix S_i are also preserved to utilize in detection process.

Step 11: The modified singular value $S'_i(1, 1)$ and $S'_i(2, 2)$ are reinserted into matrix S_i and inverse SVD is applied to obtain the modified matrix R'_i which is given by

$$R'_i = U_i \times S'_i \times V_i^T \quad (4.16)$$

Each matrix R'_i is then reshaped to create the l modified FFT coefficients of each frame F_i by performing the inverse operation of step 3.

Step 12: After substituting the l modified FFT coefficients for first l low frequency FFT coefficients of each frame F_i , an inverse FFT is applied to the FFT coefficients of each frame F_i to obtain the watermarked audio frame F'_i .

Step 13: Finally, all watermarked frames are concatenated to calculate the watermarked audio signal X' .

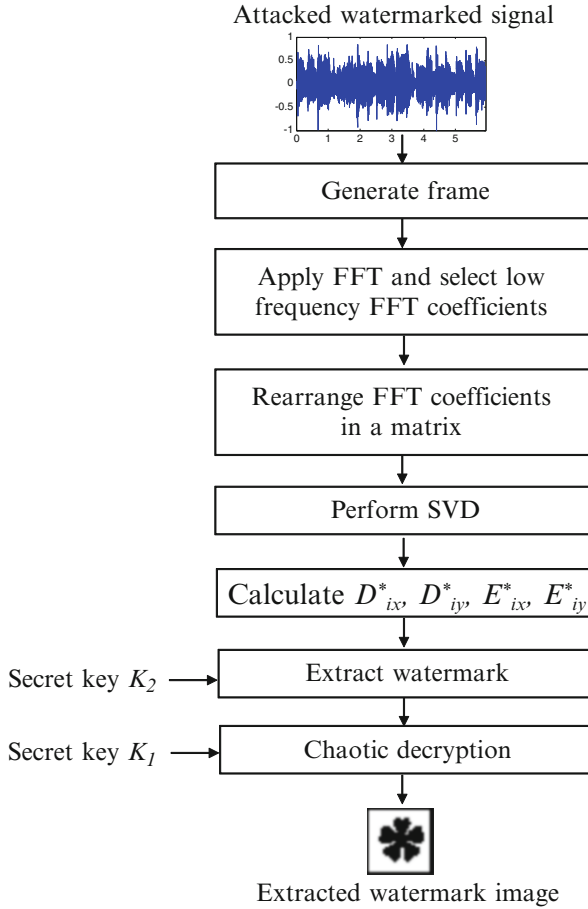


Fig. 4.2 Watermark detection process

4.2.3 Watermark Detection Process

The proposed watermark detection process is shown in Fig.4.2. The detection process is described as follows:

- Step 1:** The FFT is performed on each frame F_i^* of the attacked watermarked audio signal X^* .
- Step 2:** The first l number of low frequency FFT coefficients of each attacked watermarked frame F_i^* are selected.
- Step 3:** The selected low frequency FFT coefficients of each frame F_i^* are rearranged to obtain R_i^* and SVD is performed on it.

Step 4: Polar-to-Cartesian transformation is applied to $S_i^*(1, 1)$ and $S_i^*(2, 2)$ of each matrix S_i^* of the attacked watermarked audio frame to calculate D'_{ix} , D'_{iy} and E'_{ix} , E'_{iy} , respectively.

Step 5: Watermark sequence is extracted as follows using the secret key K_2 :

$$u(m)^* = \begin{cases} 1, & \text{if } D_{ix}^* > D_{ix} \text{ or } D_{iy}^* > D_{iy} \text{ or } E_{ix}^* > E_{ix} \text{ or } E_{iy}^* > E_{iy} \\ 0, & \text{otherwise} \end{cases} \quad (4.17)$$

where the components D_{ix} , D_{iy} and E_{ix} , E_{iy} of the highest two singular values $S_i(1, 1)$ and $S_i(2, 2)$ of each matrix S_i are used as secret key K_2 .

Step 6: The chaotic decryption is performed using the secret key K_1 to obtain the hidden binary sequence with the following rule:

$$w_{SE}^*(m) = z(m) \oplus u^*(m) \quad (4.18)$$

Step 7: Finally, watermark image is obtained by rearranging the binary sequence $w_{SE}^*(m)$ into a square matrix W_{BI}^* of size $M \times M$.

4.3 Experimental Results and Discussion

In this section, several experiments were conducted to demonstrate the performance of the proposed watermarking method. The performance of the proposed method is evaluated in terms of imperceptibility, robustness, and data payload. Here, we selected four different types of 16 bit mono audio signals (Pop, Jazz, Folk, and Classical) sampled at 44.1 kHz. Each audio file contains 262,144 samples (duration 5.94 s). By using a frame size of 256 samples, we have 1,024 nonoverlapping frames for each audio sample. We embed four binary watermark bits in each frame of audio signal. Thus, the length of the watermark sequence is 4,096. A binary logo image and the corresponding encrypted image using chaotic encryption of size $M \times M = 64 \times 64 = 4,096$ are shown in Fig. 4.3. From each frame of audio signal, we select first 36 low frequency FFT coefficients ($l = 36$). After rearranging the selected FFT

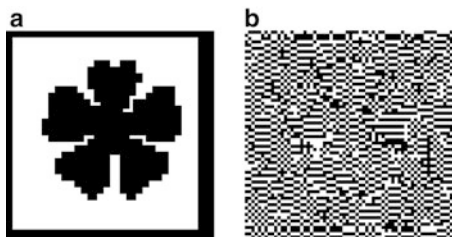


Fig. 4.3 (a) Binary watermark image.
(b) Encrypted image

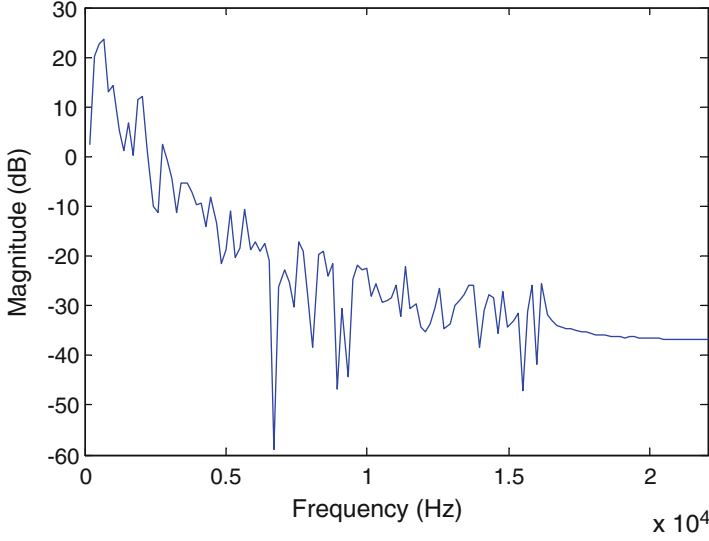


Fig. 4.4 Magnitude spectrum of a selected frame for the original audio signal ‘Folk’

coefficients of each frame in a 6×6 matrix ($N = 6$), SVD is applied to each of these matrices. In this study, the selected value for $y(1)$, a , b , and T_h are 0.4, 5.90, -0.39 , and 0.25, respectively. We experimentally found that the optimal value for θ_1 and θ_2 is 45° , which can provide a consistency between the components D_{ix} and D_{iy} , and E_{ix} and E_{iy} of the highest two singular values $S'_i(1, 1)$ and $S'_i(2, 2)$ of each matrix S_i . These parameters have been selected in order to achieve a good trade-off among the conflicting requirements of imperceptibility, robustness, and data payload.

Figures 4.4 and 4.5, respectively show the magnitude spectrum of a selected frame for the original and watermarked audio signal ‘Folk’ using the proposed method.

4.3.1 Imperceptibility Test

Generally, there are two approaches to perform perceptual quality assessment: (i) subjective evaluation test and (ii) objective evaluation test.

4.3.1.1 Subjective Evaluation Test

SDG has been used extensively for evaluating the quality of watermarked audio signal. SDG ranges from 5.0 to 1.0 (imperceptible to very annoying) shown in Table 3.1. Subjective quality evaluation of the watermarking method was carried

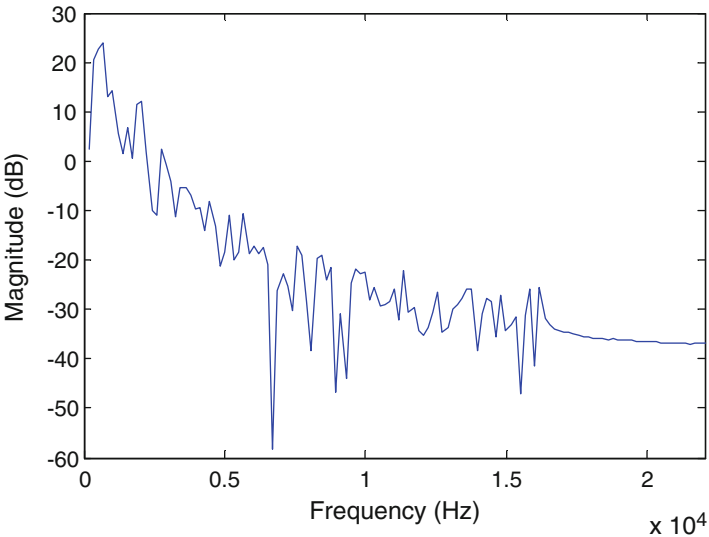


Fig. 4.5 Magnitude spectrum of a selected frame for the watermarked audio signal ‘Folk’

Table 4.1 Subjective and objective evaluation of different watermarked sounds

| Audio signal | Subjective evaluation | | Objective evaluation | |
|--------------|-----------------------|-----------------------|----------------------|-------|
| | MOS | Correct detection (%) | ODG | SNR |
| Pop | 4.9 | 54 | −0.059 | 36.75 |
| Jazz | 5.0 | 48 | −0.329 | 37.32 |
| Folk | 4.9 | 52 | −0.097 | 36.19 |
| Classical | 5.0 | 44 | −0.288 | 37.18 |
| Average | 4.95 | 49.50 | −0.193 | 36.86 |

out by blind listening tests involving ten participants of different ages (22–35 years) with normal hearing ability. Participants listened to the original and watermarked audio signals and were asked to report the dissimilarities between them using SDG. The average SDG (i.e. MOS) values for different watermarked sounds using the proposed method are shown in Table 4.1. From the test results, we observed that the average MOS values are within 4.95–5.0 for all watermarked sound using the proposed method, indicating that original and watermarked audio signals are perceptually indistinguishable.

Subjective evaluation was also conducted using ABX double blind test with ten participants. Table 4.1 shows the evaluation results of all participants in terms of percentage of correct detection. We observed that correct detection scores range from 44 to 54 %, indicating that proposed watermarking method provides good imperceptible watermarked sound.

4.3.1.2 Objective Evaluation Test

Objective evaluation was done by calculating the SNR of the watermarked audio signal. According to the IFPI recommendation, audio watermarking should be imperceptible when SNR is over 20 dB. After embedding a watermark, the SNR values of all selected audio signals using the proposed method are above 20 dB, conforming to the IFPI standard, as shown in Table 4.1.

Objective evaluation was also carried out using the ODG, another objective quality assessment technique which ranges from 0.0 to -4.0 (imperceptible to very annoying) as shown in Table 3.1. The objective quality of the watermarked audio signals is calculated in terms of ODG and shown in Table 4.1. We observed that all ODG values range from -0.059 to -0.329 , indicating that original and watermarked audio signals are perceptually indistinguishable.

Figures 4.6 and 4.7 show the time domain and a spectrogram representation of the original audio signal with a watermarked audio signal in which the watermark is imperceptible using the proposed method.

We compared the proposed method with the several recent methods in terms of the SNR and MOS. This comparison is based on the reported results in [2, 8, 12, 13, 17, 27, 29, 32, 39, 54, 57] as shown in Table 4.2. From the comparison of results, it

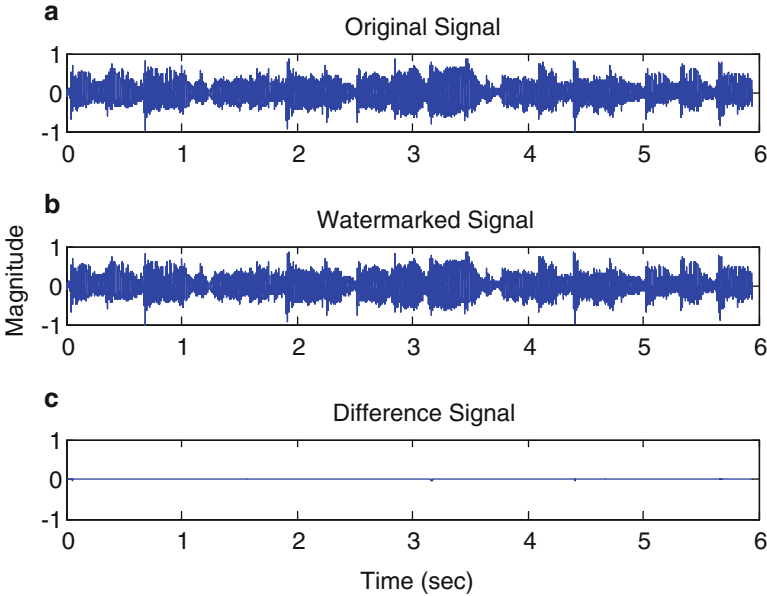


Fig. 4.6 Imperceptibility of the watermarked audio using the proposed method: (a) Original signal 'Folk', (b) Watermarked signal 'Folk', (c) Difference between original and watermarked signals

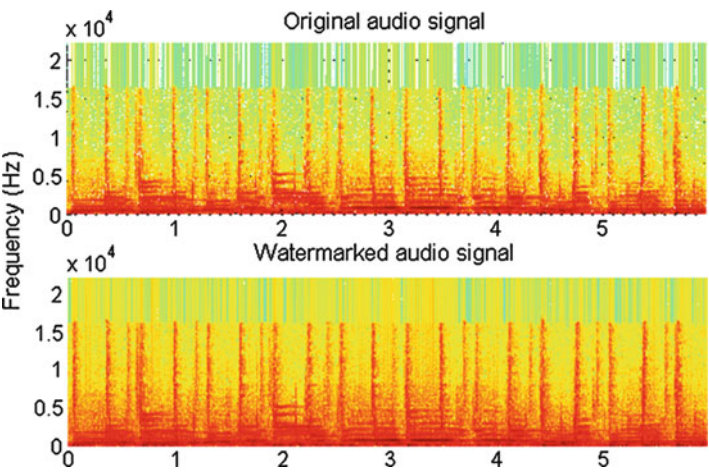


Fig. 4.7 Spectrogram representation of the original audio signal and watermarked audio signal ‘Folk’ using the proposed method

Table 4.2 Comparison of SNR and MOS between the proposed method and several recent methods

| Reference | Algorithm | SNR | MOS |
|-----------|---------------------------------|-------|------|
| [17] | Spread spectrum | 28.59 | 4.46 |
| [12] | Optimization-based quantization | 29.50 | – |
| [13] | DWT-based energy proportion | 17.95 | 4.15 |
| [32] | FFT amplitude modification | 25.70 | – |
| [54] | STFT-SVD | 28.36 | 4.70 |
| [27] | SVD | 27.13 | – |
| [2] | DWT-SVD | 28.55 | 4.33 |
| [8] | DWT-SVD | 24.37 | 4.46 |
| [57] | Frequency masking | 12.87 | 2.93 |
| [39] | EMD | 24.12 | – |
| [29] | TS echo hiding | 22.70 | 4.70 |
| Proposed | FFT-SVD-CPT | 36.86 | 4.95 |

is seen that our proposed method outperforms the recent watermarking methods in terms of SNR and MOS, indicating a high transparency of the watermarked audio signals.

4.3.2 Robustness Test

BER and NC are used measure the robustness of the proposed method. Various signal processing attacks were applied to watermarked sounds to assess the robustness









| Attack | No attack | Noise addition | Cropping (front) | Cropping (middle) |
|---------------------|---|---|---|---|
| NC | 1 | 0.9977 | 0.9991 | 0.9995 |
| BER | 0 | 0.2441 | 0.0977 | 0.0488 |
| Extracted watermark |  |  |  |  |
| Attack | Cropping (end) | Resampling | Requantization | MP3 compression |
| NC | 0.9986 | 1 | 1 | 0.9741 |
| BER | 0.1465 | 0 | 0 | 2.7588 |
| Extracted watermark |  |  |  |  |

Fig. 4.8 Extracted watermark image with NC and BER for the audio signal ‘Folk’

of the proposed method. In our experiment, we used Sound Forge 8.0 for resampling, requantization, and MP3 compression attack. AWGN and cropping were implemented using MATLAB 7.8.

Figure 4.8 shows the robustness results of the proposed method in terms of NC and BER against several attacks for the audio signal ‘Folk’. It is obvious that NC values after attacks are very high while the BER values are very low. The minimum NC value and the maximum BER value are 0.9741 and 2.7588, respectively. The extracted watermark images are visually similar to the original watermark, which further verify the good performance of the proposed method against different attacks.

The robustness results for the audio signal ‘Pop’, ‘Jazz’, and ‘Classical’, respectively are summarized in Table 4.3. We observed that the NC and BER values range from 0.9725 to 1 and 0 to 2.9272, respectively, demonstrating the high robustness of our proposed method against different attacks. This is because watermark bits are embedded into the CPT components of the highest singular values of the low frequency FFT coefficients of each audio frame.

4.3.3 Security

Robustness against attack is very important for a secured watermarking method. To enhance the security, the proposed method utilizes chaotic encryption. Since the proposed watermark embedding and detection processes depend on the secret keys K_1 and K_2 , it is impossible to maliciously detect the watermark without these keys.

Table 4.3 NC and BER of the extracted watermarks for different audio signals

| Audio signal | Attack | NC | BER (%) |
|--------------|-------------------|--------|---------|
| Pop | No attack | 1 | 0 |
| | Noise addition | 0.9943 | 0.6104 |
| | Cropping (front) | 0.9993 | 0.0732 |
| | Cropping (middle) | 1 | 0 |
| | Cropping (end) | 0.9989 | 0.1221 |
| | Resampling | 1 | 0 |
| | Requantization | 1 | 0 |
| | MP3 compression | 0.9766 | 2.4902 |
| Jazz | No attack | 1 | 0 |
| | Noise addition | 0.9883 | 1.2451 |
| | Cropping (front) | 0.9995 | 0.0488 |
| | Cropping (middle) | 1 | 0 |
| | Cropping (end) | 0.9991 | 0.0977 |
| | Resampling | 1 | 0 |
| | Requantization | 1 | 0 |
| | MP3 compression | 0.9808 | 2.0508 |
| Folk | No attack | 1 | 0 |
| | Noise addition | 0.9973 | 0.2930 |
| | Cropping (front) | 0.9991 | 0.0977 |
| | Cropping (middle) | 1 | 0 |
| | Cropping (end) | 1 | 0 |
| | Resampling | 1 | 0 |
| | Requantization | 1 | 0 |
| | MP3 compression | 0.9725 | 2.9297 |

4.3.4 Error Analysis

Two types of error may occur while searching the watermark sequence: (1) false positive error (FPE) (2) false negative error (FNE). These errors are very harmful because they impair the credibility of watermarking method. The probability of FPE, P_{FPE} , and probability of FNE, P_{FNE} , can be calculated using Eqs. (3.8) and (3.9). Figure 4.9 shows the P_{FPE} for $k \in (0, 100]$. It is noted that the P_{FPE} approaches 0 when k is larger than 30. Figure 4.10 shows the P_{FNE} for $k \in (0, 100]$. It is noted that the P_{FNE} approaches 0 when k is larger than 30.

4.3.5 Algorithm Comparison and Discussion

Table 4.4 shows a general comparison between the proposed method and the several recent methods in terms of data payload and BER against various attacks such as noise addition, resampling and MP3 compression. This comparison is based on

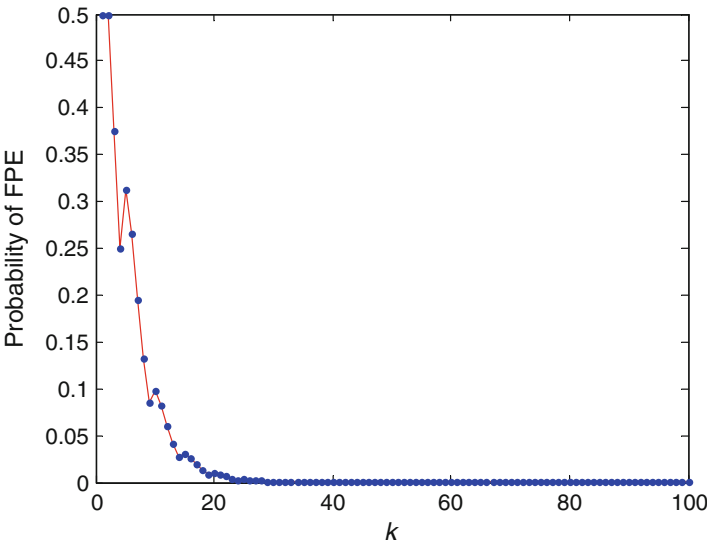


Fig. 4.9 Probability of FPE for various values of k

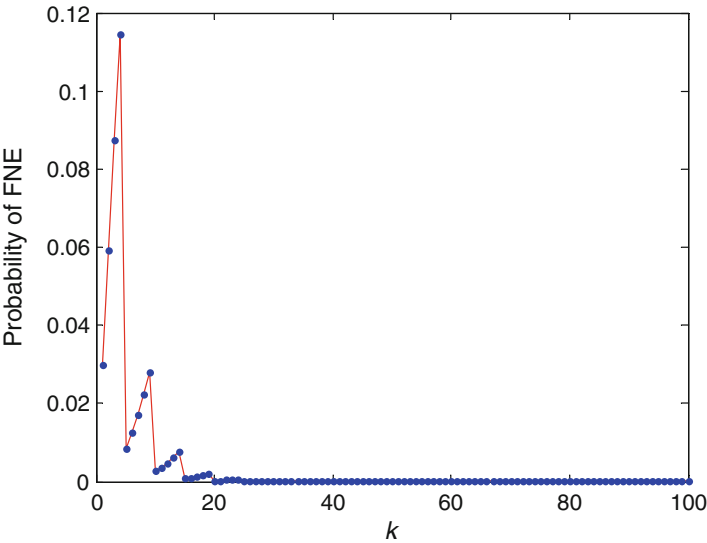


Fig. 4.10 Probability of FNE for various values of k

Table 4.4 A general comparison between the proposed method and recent audio watermarking methods sorted by data payload

| Reference | Algorithm | Payload (bps) | Noise addition BER (%) | Resampling BER (%) | MP3 Compression BER (%) |
|-----------|---------------------------------|------------------|---------------------------|-----------------------|----------------------------|
| Proposed | FFT-SVD-CPT | 689.56 | 1.25 (32 dB) | 0 (22.05 kHz) | 2.93 (128 kbps) |
| [9] | SVD-quantization | 196 | 0 (22.05 dB) | 1 (22.05 kHz) | 2 (32 kbps) |
| [12] | Optimisation-based quantization | 172.41 | – | 0.1 (22.05 kHz) | 3.93 (80 kbps) |
| [64] | Self-synchronization | 172 | 4.98 (16.12 dB) | 0 (22.05 kHz) | 24.18 (32 kbps) |
| [13] | DWT-based energy proportion | 114.82 | – | 6.92 (22.05 kHz) | 5.71 (80 kbps) |
| [33] | Chaos-based DFST | 86 | 4.22 (65 dB) | 0 (22.05 kHz) | 3.47 (48 kbps) |
| [39] | EMD | 46.9–50.3 | 0 (20 dB) | 3 (22.05 kHz) | 1 (32 kbps) |
| [8] | DWT-SVD | 45.9 | 0 (20 dB) | 3 (22.05 kHz) | 1 (32 kbps) |
| [62] | DWT-LPC | 10.72 | 5.13 (36 dB) | 13.64 (22.05 kHz) | 5.71 (128 kbps) |
| [65] | Histogram | 3 | 0 (35 dB) | 0 (–) | 15 (128 kbps) |
| [66] | DWT-based histogram | 15 (35 dB) | 0 (16 dB) | 0 (16 kHz) | 17.5 (64 kbps) |
| [28] | Lifting wavelet transform | – | – | 16.50 (36.75 kHz) | 51.73 (128 kbps) |

the reported result in [8, 9, 12, 13, 28, 33, 39, 62, 64–66]. We observed that the proposed method outperforms state-of-the-art methods in terms of data payload and BER result against various attacks. The data payload of the proposed method is 689.56 bps, whereas the data payload of these methods ranges from only 2 to 196 bps. Moreover, the BER of different attacks for the proposed method ranges from 0 to 2.9297, whereas the BER of these methods ranges from 0 to 51.73.

4.4 Summary

In this chapter, we introduced a FFT-based audio watermarking method using SVD and CPT. Experimental results indicate that the proposed watermarking method not only provides a good imperceptibility, but also shows a high robustness against various attacks such as noise addition, cropping, resampling, requantization, and MP3 compression. This is because watermark bits are embedded into each of the Cartesian components of the highest singular values of the low frequency FFT coefficients of each frame and slight variations of the Cartesian components can not significantly affect the quality of the sound and also these values can change very little against different attacks. In addition, it achieves a good tradeoff among the robustness, imperceptibility, and payload. Moreover, it also has very low error probability rates. Overall, our proposed method has higher data payload, provides higher SNR and MOS values as well as lower BER values than the state-of-the-art audio watermarking methods.

Chapter 5

Conclusions

This chapter concludes this book with a brief summary of our research work. The future research work is also discussed in this chapter.

5.1 Summary of the Work

Digital watermarking is a process of embedding a secret signal called the watermark within the original signal to show authenticity and ownership. It has been utilized effectively to provide solutions for ownership protection, copyright protection, content authentication, speech quality evaluation, secret communication. Depending on the watermark application and purpose, two important issues need to be addressed for digital audio watermarking. One is to provide trustworthy evidence to protect rightful ownership and the other is to achieve an appropriate trade-off among imperceptibility, robustness, and data payload. The audio watermarking methods proposed in this dissertation is summarized as follows:

- A DWT-DCT-based audio watermarking method using SVD and quantization has been introduced. In this method, watermark information is embedded into the largest singular value of the DCT coefficients obtained from the DWT coefficients of each frame by quantization. This method provide high imperceptible watermarked sounds as well as good robustness against various attacks.
- An audio watermarking method in FFT domain based on SVD and CPT has been presented. In this method, watermark information is embedded in each of the Cartesian components of the highest two singular values of the low frequency FFT coefficients of each frame. The data payload of the proposed method is relatively much higher than that of the state-of-the-art methods. In addition, it shows good robustness against various attacks.

- The proposed audio watermarking methods can be effectively utilized for various applications of watermarking, specially, for audio copyright protection.

5.2 Future Research

There are several directions for future research on the proposed methods introduced in this dissertation. In the future work, synchronization code and error correcting codes will be incorporated to improve the robustness of the proposed methods. In addition, the adaptive selection of quantization parameter Q might further improve the performance of the proposed methods. Some modern attacks such as channel fading, jitter, and packet drop will be considered, because these attacks are particularly relevant in various networks such as GSM (Global System for mobile Communications) and CDMA (Code Division Multiple Access). Moreover, psychoacoustic model can be adopted to improve the imperceptibility of the proposed methods. Furthermore, computational complexity of the proposed methods will be carried out.

Bibliography

1. Alghoniemy M, Tewfik AH (2004) Geometric invariance in image watermarking. *IEEE Trans. Image Process* 13(2):145–153
2. Ali AH, Ahmad M (2010) Digital audio watermarking based on the discrete wavelet transform and singular value decomposition. *Eur J Sci Res* 39(1):6–21
3. Al-Nuaimy W, El-Bendary MAM, Shafik A, Shawki F, Abou-El-Azm AE, El-Fishawy NA, Elhalafawy SM, Diab SM, Sallam BM, El-Samie FEA, Kazemian HB (2011) An SVD audio watermarking approach using chaotic encrypted images. *Digit Signal Process* 21(6):764–779
4. Aslantas V (2008) A singular-value decomposition-based image watermarking using genetic algorithm. *AEU Int J Electron Commun* 62(5):386–394
5. Bassia P, Pitas I, Nikolaidis N (2001) Robust audio watermarking in the time domain. *IEEE Trans Multimed* 3(2):232–241
6. Bender W, Gruhl D, Morimoto N, Lu A (1996) Techniques of data hiding. *IBM Syst J* 35 (3–4):313–336
7. Bhat VK, Sengupta I, Das A (2008) Audio watermarking based on mean quantization in cepstrum domain. In: *Proceedings of the 16th international conference on advanced computing communications*, Tamilnadu, India, pp 73–77
8. Bhat VK, Sengupta I, Das A (2010) An adaptive audio watermarking based on the singular value decomposition in the wavelet domain. *Digit Signal Process* 20(6):1547–1558
9. Bhat VK, Sengupta I, Das A (2011) An audio watermarking scheme using singular value decomposition and dither modulation. *Multimed Tools Appl* 52(2–3):369–383
10. Chan PW, Lyu MR, Chin RT (2005) A novel scheme for hybrid digital video watermarking: approach, evaluation and experimentation. *IEEE Trans Circuits Syst Video Technol* 15(12):1638–1649
11. Chen B, Wornell GW (2001) Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. *IEEE Trans Inf Theory* 47(4):1423–1443
12. Chen ST, Huang HN, Chen CJ, Wu GD (2010) Energy-proportion based scheme for audio watermarking. *IET Signal Process* 4(5):576–587
13. Chen ST, Wu GD, Huang HN (2010) Wavelet-domain audio watermarking scheme using optimisation-based quantization. *IET Signal Process* 4(6):720–727
14. Chen ST, Huang HN, Chen CJ, Tseng KK, Tu SY (2013) Adaptive audio watermarking via the optimization point of view on the wavelet-based entropy. *Digit Signal Process* 23(3):971–980
15. Chu WC (2003) DCT-based image watermarking using subsampling. *IEEE Trans Multimed* 5(1):34–38

16. Cox IJ, Miller ML (2002) The first 50 years of electronic watermarking. *J Appl Signal Process* 56(2):225–230
17. Cox I, Killian J, Leighton F, Shamoon T (1997) Secure spread spectrum watermarking for multimedia. *IEEE Trans Image Process* 6(12):1673–1687
18. Cox IJ, Miller ML, Bloom JA (2001) *Digital watermarking*. The Morgan Kaufmann Publishers, San Francisco, USA
19. Cox IJ, Miller ML, Bloom J, Fridrich J, and Kalker T (2007) *Digital watermarking and steganography*. The Morgan Kaufmann series in multimedia information systems. Elsevier, USA
20. Cvejic N (2004) *Algorithms for audio watermarking and steganography*, Ph.D thesis, University of Oulu
21. Cvejic N, Seppanen T (2007) *Digital audio watermarking techniques and technologies: application and benchmarks*. IGI Global, USA
22. Dhar PK (2014) *Studies on digital audio watermarking using matrix decomposition*, Ph.D thesis, Saitama University
23. Dhar PK, Shimamura T (2013) A DWT-DCT-based audio watermarking method using singular value decomposition and quantization. *J Signal Process* 17(3):69–79
24. Dhar PK, Shimamura T (2013) Audio watermarking in transform domain based on singular value decomposition and Cartesian-polar transformation. *Int J Speech Technol* 17(2):133–144
25. Dittmann J, Megías D, Lang A, Joancomarti JH (2006) Theoretical framework for a practical evaluation and comparison of audio watermarking schemes in the triangle of robustness, transparency and capacity. In: Shi YQ (ed) *Transactions on data hiding and multimedia security*, vol 4300. Springer, Berlin/Heidelberg, pp 1–40
26. Dutta MK, Gupta P, Pathak VK (2014) A perceptible watermarking algorithm for audio signals. *Multimed Tools Appl* 73(2):691–713
27. El-Samie FEA (2009) An efficient singular value decomposition algorithm for digital audio watermarking. *Int J Speech Technol* 12(1):27–45
28. Ercelebi E, Batakci L (2009) Audio watermarking scheme based on embedding strategy in low frequency components with a binary image. *Digit Signal Process* 19(2):265–277
29. Erfani Y, Siahpoush S (2009) Robust audio watermarking using improved TS echo hiding. *Digit Signal Process* 19(5):809–814
30. Fallahpour M, Megias D (2009) High capacity audio watermarking using FFT amplitude interpolation. *IEICE Electron Express* 6(14):1057–1063
31. Fallahpour M, Megias D (2010) DWT-based high capacity audio watermarking. *IEICE Trans Fundam* E93-A(1):331–335
32. Fallahpour M, Megias D (2010) Robust high-capacity audio watermarking based on FFT amplitude modification. *IEICE Trans Inf Syst* E93-D(1):87–93
33. Fan M, Wang H (2009) Chaos-based discrete fractional sine transform domain audio watermarking scheme. *Comput Electr Eng* 35(3):506–516
34. Fujimoto R, Lwaki M, Kiryu T (2006) A method of high bit-rate data hiding in music using spline interpolation. In: *Proceedings of the international conference on intelligent information hiding and multimedia signal processing*, California, USA, pp 11–14
35. Gruhl D, Lu A, Bender W (1996) Echo hiding. In: *Proceedings of the 1st information hiding workshop*, Cambridge, UK, vol 1174, pp 295–315
36. Guo JM, Prasetyo H (2013) Security attack on the wavelet transform and singular value decomposition image watermarking. In: *Proceedings of the IEEE international symposium on consumer electronics*, Hsinchu, pp 217–218
37. Kang H, Yamaguchi K, Kurkoski B, Yamaguchi K, Kobayashi K (2008) Full-index-embedding patchwork algorithm for audio watermarking. *IEICE Trans Inf Syst* E91-D(11):2731–2734
38. Katzenbeisser S, Petitcolas FAP (2000) *Information hiding techniques for steganography and digital watermarking*. Artech House, Norwood
39. Khaldi K, Boudraa AO (2013) Audio watermarking via EMD. *IEEE Trans Audio Speech Lang Process* 21(3):675–680

40. Kim HS, Lee HS (2003) Invariant image watermark using Zernike moments. *IEEE Trans Circuits Syst Video Technol* 13(8):766–775
41. Kirovski D, Malvar HS (2003) Spread spectrum watermarking for audio signal. *IEEE Trans Signal Process* 51(4):1020–1033
42. Kundur D, Hatzinakos D (1999) Digital watermarking for telltale tamperproofing and authentication. *Proc IEEE* 87(7):1167–1180
43. Lai CC (2011) A digital watermarking scheme based on singular value decomposition and tiny genetic algorithm. *Digit Signal Process* 21(4):522–527
44. Lai CC, Tsai CC (2010) Digital image watermarking using discrete wavelet transform and singular value decomposition. *IEEE Trans Instrum Meas* 59(11):3060–3063
45. Langelaar G, Setyawan I, Lagendijk R (2000) Watermarking digital image and video data: a state of the art overview. *IEEE Signal Process Mag* 17:20–46
46. Lee SK, Ho YS (2000) Audio watermarking in cepstrum domain. *IEEE Trans Consum Electron* 46(3):744–750
47. Lei BY, Soon IY, Li Z (2011) Blind and robust audio watermarking scheme based on SVD-DCT. *Signal Process* 91:1973–1984
48. Lie WN, Chang LC (2006) Robust high quality time domain audio watermarking based on low frequency amplitude modification. *IEEE Trans Multimed* 8(1):46–59
49. Liu R, Tan T (2002) An SVD-based watermarking scheme for protecting rightful ownership. *IEEE Trans Multimed* 4(1):121–128
50. Megias D, Serra-Ruiz J, Fallahpour M (2010) Efficient self-synchronized blind audio watermarking system based on time domain and FFT amplitude modification. *Signal Process* 90(12):3078–3092
51. Mohammad AA, Alhaj A, Shaltaf S (2008) An improved SVD-based watermarking scheme for protecting rightful ownership. *Signal Process* 88:2158–2180
52. Natgunanathan I, Xiang Y, Rong Y, Zhou W, Guo S (2012) Robust patchwork-based embedding and decoding scheme for digital audio watermarking. *IEEE Trans Audio Speech Lang Process* 20(8):2232–2239
53. Noorkami M, Mersereau RM (2008) Digital video watermarking in p-frames with controlled video bit rate increase. *IEEE Trans Inf Forensics Secur* 3(3):441–455
54. Ozer H, Sankur B, Memon N (2005) An SVD-based audio watermarking technique. In: *Proceedings of the 7th ACM workshop multimedia security*, New York, pp 51–56
55. Ridzon R, Levicky D (2007) Robust digital watermarking based on the log-polar mapping. *Radioengineering* 16(4):76–81
56. Rykaczewski R (2007) Comments on SVD-based watermarking scheme for protecting rightful ownership. *IEEE Trans Multimed* 9(2):412–423
57. Swanson MD, Zhu B, Tewfiq AH, Boney L (1998) Robust audio watermarking using perceptual masking. *Signal Process* 66(3):337–355
58. Sweldens W (1996) The lifting scheme: a custom-design construction of biorthogonal wavelets. *Appl Comput Harmon Anal* 3(2):186–200
59. Thiede T, Treurniet WC, Bitto R, Schmidmer C, Sporer T, Beerens JG, Colomes C, Keyhl M, Stoll G, Brandenburg K, Feiten B (2000) PEAQ – the ITU standard for objective measurement of perceived audio quality. *J Audio Eng Soc* 48(1–2):3–29
60. Tsai HH, Cheng JS, Yu PT (2003) Audio watermarking based on HAS and neural networks in DCT domain. *EURASIP J Adv Signal Process* 3:252–263
61. Ulutas M, Ulutas G, Nabyev VV (2011) Medical image security and EPR hiding using Shamir's secret sharing scheme. *J Syst Softw* 84(3):341–353
62. Wang R, Xu D, Chen J, Du C (2004) Digital audio watermarking algorithm based on linear predictive coding in wavelet domain. In: *Proceedings of the 7th IEEE international conference on signal processing*, Beijing, China, vol 1, pp 2393–2396
63. Wang J, Healy R, Timoney J (2011) A robust audio watermarking scheme based on reduced singular value decomposition and distortion removal. *Signal Process* 91(8):1693–1708
64. Wu S, Huang J, Huang D, Shi YQ (2005) Efficiently self-synchronized audio watermarking for assured audio data transmission. *IEEE Trans Broadcast* 51(1):69–76

65. Xiang S, Huang J (2007) Histogram based audio watermarking against time scale modification and cropping attacks. *IEEE Trans Multimed* 9(7):1357–1372
66. Xiang S, Kim HJ, Huang J (2008) Audio watermarking robust against time scale modification and MP3 compression. *Signal Process* 88(10):2372–2387
67. Xiang Y, Peng D, Natgunanathan I, Zhou W (2011) Effective pseudonoise sequence and decoding function for imperceptibility and robustness enhancement in time-spread echo-based audio watermarking. *IEEE Trans Multimed* 13(1):2–13
68. Xiang Y, Natgunanathan I, Peng D, Zhou W, Yu S (2012) A dual-channel time-spread echo method for audio watermarking. *IEEE Trans Inf Forensics Secur* 7(2):383–392
69. Xing Y, Tan J (2010) A color image watermarking scheme resistant against geometrical attacks. *Radioengineering* 19(1):62–67
70. Yeh C, Kuo C (1999) Digital watermarking through quasi m-arrays. In: *Proceedings of the IEEE workshop on signal processing system, Taipei, Oct 1999*, pp 456–461
71. Yeo IK, Kim HJ (2003) Modified patchwork algorithm: a novel audio watermarking scheme. *IEEE Trans Speech Audio Process* 11(4):381–386
72. Zhang XP, Li K (2005) Comments on an SVD-based watermarking scheme for protecting rightful ownership. *IEEE Trans Multimed* 7(3):593–594