الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
وزارة التعليم العالي والبحث العلمي
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
جامعة فرحات عباس سطيف 1
UNIVERSITE FERHAT ABBAS SETIF1
UFAS1 (ALGERIE)

THESE

Présenté à la Faculté de Technologie

Pour l'Obtention du Diplôme de

Doctorat

Domaine : Science et Technologie
Filière : Electronique
Option : Traitement du signal

Par

Mr. TERCHI Younes

Développement de nouvelles techniques fréquentielles de tatouage du signal audio

Soutenue le : 20/10/2018 devant le jury composé de :

AMARDJIA Nourredine	Professeur	Université Sétif 1	Président
BOUGUEZEL Saad	Professeur	Université Sétif 1	Rapporteur
BENZID Redha	Professeur	Université Batna 2	Examinateur
KACHA Abdellah	Professeur	Université de Jijel	Examinateur

الجمهورية الجزائرية الديمقراطية الشعبية
PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
وزارة التعليم العالي والبحث العلمي
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

جامعة فرحات عباس سطيف 1
FERHAT ABBAS UNIVERSITY - SETIF1
UFAS1 (ALGERIA)

THESIS

Presented to the Faculty of Technology

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

Domain : Science and Technology
Sector : Electronics
Option : Signal Processing

By

Mr. TERCHI Younes

Development of new frequency techniques for audio signal watermarking

Defended on : 20/10/2018 before the jury composed of :

AMARDJIA Nourredine	Professor	University Sétif 1	President
BOUGUEZEL Saad	Professor	University Sétif 1	Supervisor
BENZID Redha	Professor	University Batna 2	Examiner
KACHA Abdellah	Professor	University of Jijel	Examiner

To my loving parents

Acknowledgments

First and foremost, I like to express my deepest gratitude and heartfelt thanks to my thesis supervisor, Prof. Bouguezel Saad, for his invaluable guidance and mentorship throughout the span of this research. Without his insight, inspiration, valuable advice and encouragement, it would not have been possible to complete this thesis. I feel honored and privileged for having the opportunity to work under his supervision.

It is my great pleasure to thank Prof. Amardjia Nourredine for his acceptance to be the president of the jury.

I thank Prof. Benzid Redha and Prof. Kacha Abdellah for accepting to be members of the examining committee.

Finally, I owe special thanks to my family members for their encouragement, love, patience, and sacrifices. Their support and understanding have contributed a great deal for completing this work.

Table of contents

List of figures

List of tables

List of acronyms

AQIM	Angular Quantization Index Modulation
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
CDF	Cumulative Density Function
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DM	Dither Modulation
DSSS	Direct Sequence Spread Spectrum
DWT	Discrete Wavelet Transform
EAQUAL	Evaluation of Audio Quality
FHSS	Frequency Hopping Spread Spectrum
HAS	Human Auditory System
i.i.d.	Independent Identically Distributed
IAQIM	Improved Angular Quantization Modulation
IFPI	International Federation of Phonographic Industry
LQIM	Logarithmic Quantization Index Modulation
MPEG	Moving Picture Experts Group
PAPR	Peak to Average Power Ratio
PDF	Probability Density Function
PEAQ	Perceptual Evaluation of Audio Quality
PEMO-Q	Perceptual Model-Quality Assessment
PRN	Pseudo-Random Noise
PRNG	Pseudo-Random Noise Generator
PRP	Pseudo-Random Permutation
PSM	Pitch-Scale Modification
QIM	Quantization Index Modulation
RDM	Rational Dither Modulation
SDG	Subjective Difference Grade
SDMI	Secure Digital Music Initiative
SNR	Signal to Noise Ratio

SS	Spread Spectrum
SWR	Signal to Watermark Ratio
TSM	Time-Scale Modification

General introduction

Watermarking refers to the practice of imperceptibly modifying an object to insert in it a message (*watermark*) [1], i.e., hiding a specific information about the object without causing perceivable distortion. It has a long history going back to the late thirteenth century when watermarks first introduced by paper mills in Italy to indicate the paper producer or its brand, and thus served as the basis of confirming the paper source. By the eighteenth century, watermarks began to be used as anti-counterfeiting measures on money and other documents. Up to now, paper watermarks are still of great importance to authenticate documents and the most common form of them is that of the bill. The first technical example similar to the notion of paper watermarks was a patent filed for *watermarking* musical signals by Emil Hembrooke in 1954 [2]. He proposed a method to insert Morse code in audio signals in order to identify the ownership of music, and hence any disputation about the ownership of the music could be justly settled. The term *digital watermarking*, which has been used first by Komatsu and Tominaga in 1988 [3], is the outcome of the digital era, i.e., digital watermarking is the extension of analog watermarking to digital content. Since 1995, digital watermarking has gained a lot of attention and is rapidly evolving [4]. This is due to the rise of many problems related to multimedia security that are caused by the rapid development of communication systems and social networks. Digital watermarking has been primarily used for copyright protection and ownership identification. Nevertheless, it provides solutions to many other multimedia security problems and has been successfully used in a wide range of applications [5]. Nowadays, digital watermarking is mostly related to media signals and is defined as the art of embedding a signature (watermark) in a host digital media signal (image, video, or audio) without altering its perceptual value.

Watermarking techniques can be classified according to their robustness against content preserving attacks and manipulations into fragile, semi-fragile and robust [6–8]. Fragile techniques are sensitive to the slightest change in the host signal, whereas semi-fragile techniques can detect tamper and offer some robustness to content preserving attacks and manipulations, and finally robust techniques provide a high level of robustness to content preserving attacks and manipulations. Regarding their embedding domain, watermarking techniques can also be classified into time (spatial) and transform domain techniques [6]. Time domain techniques have low computational complexity and poor robustness, whereas, transform domain techniques achieve better robustness at the cost of higher computational complexity. The discrete wavelet transform (DWT), the discrete cosine transform (DCT), and the discrete Fourier transform (DFT) are widely used by the transform domain techniques due to their suitable properties [1,4,6,9,10], which ensure the embedding of the watermark in the most significant part of the host signal, and thus the watermark can survive content preserving attacks and manipulations. From the viewpoint of the watermark extraction process, watermarking techniques can be categorized into blind, non-blind, and semi-blind techniques. The original host signal is not required in the extraction process of the blind techniques and is mandatory for that of the non-blind techniques, but only a partial information about the original host signal (known as the side information) is required in the extraction process of the semi-blind techniques [6]. Non-blind techniques are not of great interest, semi-blind techniques are appropriate for applications such as copy protection and fingerprinting [11,12], and blind techniques are suitable for applications such as covert communication [12]. Although several blind [13–15] and semi-blind [12,16–18] watermarking techniques have been considered separately in the literature, it is highly desirable to develop a single watermarking technique that is applicable for both blind and semi-blind watermarking.

Due to the fact that the human auditory system (HAS) is more sensitive to distortions than the human visual system (HVS) [9,19], audio watermarking is a more challenging task than image and video watermarking, and according to the international federation of the phonographic industry (IFPI), the signal to watermark ratio (SWR) of any practical audio watermarking technique should be higher than 20 dB to have acceptable transparency [20,21]. Several embedding methods have been proposed in the literature for robust blind and semi-blind digital watermarking in general, and others for robust blind and semi-blind digital audio watermarking in particular such as the least significant bit (LSB) modification, spread spectrum (SS), patchwork, echo hiding, and the well-known quantization index modulation (QIM) [22]. Among these embedding methods, QIM has well attracted the attention of researchers in recent years [9,23–25]. It is a host interference free class of watermarking methods [15], and provably optimal in terms of channel capacity [22]. QIM also offers strong robustness against additive noise attacks, while having simplicity and low computational complexity [22]. Owing to its favorable attributes, QIM has been extensively used in the development of robust blind and semi-blind audio watermarking techniques dedicated to multimedia security-oriented applications. However, QIM has various realizations with different characteristics, and generally, the use of a given realization is neither justified nor optimized for a given watermarking technique. Besides, the conventional QIM severely suffers from gain attacks and the existing QIM-based solutions to gain attacks are vulnerable to noise addition.

As mentioned earlier, although digital watermarking has originally been devised for the purpose of copyrights protection, it has many other fields of application. Among these applications, audio fingerprinting for audio signals identification is gaining more and more attention in the literature due to its importance in many modern tasks such as music indexing, audio connecting, song identification, etc. [26]. Audio watermarking and robust-hashing, each with its advantages and drawbacks, have been both separately and effectively applied for the development of audio fingerprinting techniques [2,26,27]. Robust-hashing techniques have the advantage of extracting a fingerprint without modifying the host audio samples, and the drawback of high storage requirement and low robustness, whereas watermarking techniques modify the host audio samples in order to embed a fingerprint, which leads to better robustness and lower storage requirement. Therefore, regarding the common fingerprinting application of robust-hashing and watermarking, it is interesting and desirable to develop a joint hashing-watermarking technique that reduces the amount of modifications caused by applying watermarking separately and achieve higher robustness and lower storage requirements. However, this task is difficult, due to the conceptual differences between existing robust-hashing and watermarking methods. Thus, to achieve this objective, one must develop a robust-hashing method that has a similar concept of watermarking or vice versa.

This thesis is concerned with the development of new audio watermarking techniques in the frequency domain. For this purpose, by taking into account the interesting attributes of QIM and its disadvantages, we propose two QIM-based audio watermarking techniques [28,29] to solve the problems of QIM mentioned earlier and achieve better rate-robustness-transparency trade-offs than those of the existing audio watermarking techniques. Moreover, to apply the proposed watermarking techniques in audio fingerprinting, we propose two novel audio fingerprinting techniques, namely robust-hashing [30] and joint hashing-watermarking [31].

The thesis is divided into four chapters, and the rest of the thesis is organized as follows:

Chapter 1 provides an overview of digital audio watermarking where the system framework, applications, classifications, and requirements of digital audio watermarking are given. Moreover, the state of the art of audio watermarking methods and their advantages and disadvantages are discussed.

Chapter 2 proposes a blind audio watermarking technique in the DCT domain by introducing a parametric QIM, then finding closed-form expressions of metrics that measure robustness and transparency. Afterward, we formulate watermarking as a mathematical optimization problem, i.e., maximizing robustness at any fixed level of transparency using the derived expressions. The solution of the optimization problem serves to find values for which the proposed parametric QIM is optimal. Moreover, we develop an approach for selecting the coefficients that carry the watermark in a manner that it survives filtering attacks. Finally, we propose a fast implementation of the proposed technique to reduce its computational complexity.

Chapter 3 proposes a QIM-based technique that is robust against both gain and additive noise attacks. Furthermore, the proposed technique unifies blind and semi-blind audio watermarking in a single framework. For its semi-blind mode of operation, we develop for the proposed technique an efficient side information recovery procedure. Moreover, we theoretically demonstrate the robustness of the proposed technique to both gain and additive noise attacks. Furthermore, the closed form expressions for the watermarking distortion, probability of error under an additive white Gaussian noise attack, and the probability of error recovery of the developed side information recovery procedure are derived and verified.

Chapter 4 proposes a joint hashing-watermarking technique for fingerprinting applications. We first develop a quantization based robust-hashing technique for audio fingerprinting by using the DWT to summarize the audio signal samples followed by a QIM minimum distance decoder to extract robust hashes. Moreover, we propose a discrimination procedure for which we derive the theoretical expressions of the false acceptance and false rejection rates. Subsequently, we exploit the proposed hashing technique to develop a joint hashing-watermarking technique for audio fingerprinting that significantly enhances the transparency and robustness while reducing the storage requirement.

Finally, we conclude the thesis and highlight some important issues that may be taken as avenues for further research in audio watermarking.

Chapter 1

An overview of digital audio watermarking

1.1 Introduction

The recent explosion of communication systems and the Internet as collaborative mediums has opened the door for companies or people who want to share or sell their multimedia products. Nonetheless, the advantages of such open mediums can lead to very serious problems for digital media owners who do not want their products to be distributed without their consent. This is due to the ease of illegal reproduction, manipulation and distribution of digital media that is exchanged through communication networks or carried out on multimedia devices. Digital watermarking is an efficient solution for these and other information security problems [32–35]. It provides means for embedding a message or signature (watermark) proving the ownership in an image, video or audio signal without destroying its perceptual value [36–38]. Despite the fact that digital watermarking has many applications, we focus on copyrights protection in view of its dominance; digital watermarking has been devised specifically to face copyrights protection problems. In addition, watermarking techniques designed for copyrights protection can easily be applied in many other applications.

In this chapter, we provide an overview of digital watermarking, specifically its system framework, applications and classifications. Subsequently, we consider the requirements of digital audio watermarking techniques and their benchmarking for copyrights protection. Finally, we discuss the state of the art of audio watermarking methods and present the concepts as well as the advantages and drawbacks of the most prominent methods.

1.2 Overview of digital watermarking

1.2.1 Digital watermarking systems

In general, a digital watermarking system consists of an embedder and an extractor. A generic, yet complete model of digital watermarking systems is shown in Fig. 1.1. The embedder combines the host (cover) media signal X, a secret key K and the watermark W in order to produce the watermarked signal X_W. Such procedure can be described mathematically as

$$X_W = E(X, W, K) \tag{1.1}$$

where $E(\cdot,\cdot,\cdot)$ is the embedding function. Furthermore, X_W should perceptually be identical or at least similar to X. Note that for some digital watermarking applications, the inputs indicated by dashed lines in Fig. 1.1 are optional, whereas these inputs are compulsory for others. During the course of transmission, the watermarked signal is likely to be modified either by common signal processing manipulations performed by communication systems (e.g., lossy compression, filtering, noise, etc.) or altered by malicious attempts to remove the watermark. The aim of the extraction process is to recover the watermark from the attacked watermarked signal. By comparing the extracted watermark with the original watermark, it can be verified whether the received signal has been watermarked or not.

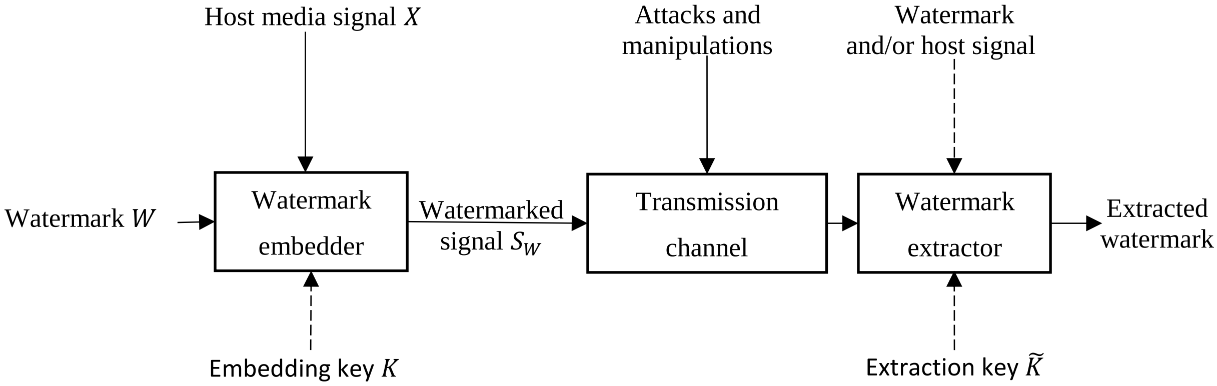


Fig. 1.1 A generic digital watermarking system.

1.2.2 Digital watermarking applications

During recent years, digital watermarking has been used in a wide range of applications. We list and briefly describe here the major digital watermarking applications.

1.2.2.1 Copyrights Protection

The investigation of digital watermarking was initially driven by the desire for copyrights protection. The idea is to embed a watermark with copyright information into the media. When proprietorial disputes happen, the watermark can be extracted as reliable proof that makes an assertion about the ownership of the media. For this purpose, the watermark must be inseparable from the host media and robust against various attacks intended to destroy it, i.e., the only way to foil the watermark extraction is by destroying the host media. Moreover, the system requires a high level of security to prevent unauthorized detection. These features enable the media owner to prove the existence of his watermark and thus claim the ownership of the disputed media. In addition, since it is not necessary for the watermark to be very long, the data payload for this application does not have to be very high [1].

1.2.2.2 Content Authentication

In authentication application, the objective is to verify whether the content has been tampered with or not. Since the watermark experience the same manipulations as the host media signal, it is possible to learn about the occurred manipulations by examining the watermark that is extracted from a manipulated watermarked signal. For this purpose, fragile and semi-fragile watermarks are commonly employed. If the content is manipulated in an illegal fashion, the watermark will be changed to reveal that the content is not authentic [39]. Fragile and semi-fragile watermarks are generally not only employed to indicate whether the data has been altered or not, but also to supply localization information as to where the data was altered.

1.2.2.3 Broadcast Monitoring

The target of broadcast monitoring is to collect information about the broadcasts of a specific content (e.g., broadcasted by radio or television stations), namely, the time, duration, and number of broadcasts [40]. This information is then used to verify whether the content was broadcasted as agreed or not, which is essential for parties such as advertisers who want to be sure that the ads they pay for are broadcasted as agreed. In this application, the watermark robustness is not the main concern due to the low risk of distortion. Instead, transparency is more required [41].

1.2.2.4 Copy Control

The applications of digital watermarking discussed to this point have an effect only after the violation has happened. In the copy control application, the aim is to prevent the violation from been happening by preventing people from making illegal copies of copyrighted media. The idea is to embed a watermark in the media signal that indicates the copy status of the content for copyright compliant devices, which have been proposed by the Secure Digital Music Initiative (SDMI). For example, a compliant device would not copy a watermarked media that carries a “copy never” watermark [1].

1.2.2.5 Fingerprinting

Fingerprinting is best known for its ability to identify signals and thus used to link signals to their corresponding meta-data (e.g. linking artist and song name to an audio signal). For this purpose, a unique watermark is embedded in a host signal as a fingerprint, which serves as an effective tool for signal identification. In general, fingerprinting requires a high level of security and transparency, as well as strong robustness.

1.2.2.6 Other applications

In addition to the applications mentioned above, there exist many other emerging applications of digital watermarking, such as error detection, quality of service assessment in multimedia communications, subjective signal quality measurement, bandwidth extension, air traffic control, etc. [42].

1.2.3 Classification of digital watermarking

Digital watermarking can be categorized according to different characteristics into several categories as summarized in Table 1.1 [6–8,42].

Table 1.1 Classifications of digital watermarking techniques.

Basis for classification	Categories
Type of media	Image/ audio/ video
Perceptibility	Imperceptible/ perceptible
Robustness	Robust/ semi-fragile/ fragile
Need of the host in the detection	Blind/ semi-blind/ non-blind
Embedding domain	Time/ transform
Reversibility	irreversible/ reversible

1.2.3.1 Image, audio, video, or text watermarking

Digital watermarking has been successfully applied to image, audio, and video signals. Image watermarking has been well-developed since the beginning of watermarking research. With relation to image watermarking, most current video watermarking techniques treat video frames as a sequence of still images and watermark each of them accordingly. Compared with image and video watermarking, audio watermarking is more challenging because of the less redundancy in audio signals and the high sensitivity of the human auditory system (HAS), which is higher than that of the human visual system (HVS). Audio watermarking has attracted more and more attention in recent years, which is a result of the rapid development of audio compression techniques and communication systems that have led to the ease of unauthorized exchanging of copyrighted audio files.

1.2.3.2 Imperceptible or perceptible

For images, perceptible watermarks are visual patterns such as logos merged into one corner of the image, visual but not obstructive. Although perceptible watermarking is practically easy to implement, it is not the focus of digital watermarking. As mentioned earlier, the aim of digital watermarking is to imperceptibly embed the watermark into digital media [42].

1.2.3.3 Robust, semi-fragile or fragile

Robustness is the capability of the watermark to survive various attacks and manipulations, that is, a robust watermark is hard to remove without destroying the host media. Therefore, robustness is a mandatory requirement for copyrights protection, ownership verification, and other security-oriented applications. Conversely, a fragile watermark is a watermark that is vulnerable to the slightest modification, and thus it is mainly used for the purpose of authentication, whereas a semi-fragile watermark is used for the same purpose as fragile watermark, yet it is only robust against some attacks and manipulations and sensitive to others [6–8].

1.2.3.4 Blind, semi-blind or non-blind

From the viewpoint of the watermark extraction process, audio watermarking techniques are categorized into blind, non-blind and semi-blind. The original host signal is not required in the extraction process of the blind techniques and is mandatory for that of the non-blind techniques, but only a partial information about the original host signal (known as the side information) is required in the extraction process of the semi-blind techniques [6]. The side information is of great importance for semi-blind watermarking and must be transmitted over a high-fidelity channel [18]. Non-blind techniques are of less interest in practical applications, whereas semi-blind techniques are appropriate for applications such as copy protection and fingerprinting [11,12], and blind techniques are suitable for applications such as covert communication [12].

1.2.3.5 Time or transform domain

Watermarking techniques are basically classified as time (spatial) domain techniques or transform domain techniques [6]. Time domain techniques have low computational complexity and robustness, whereas transform domain techniques achieve better robustness at the cost of higher computational complexity. Many discrete transforms have been successfully employed by digital watermarking techniques, such as the discrete cosine transform (DCT), the discrete Fourier transform (DFT), and the discrete wavelet transform (DWT).

1.2.3.6 Non-reversible or reversible

In reversible watermarking, the watermark can completely be removed from the watermarked signal, and thus the possibility of the host signal perfect reconstruction. However, the price of such reversibility involves some loss of robustness, security, and blindness. On the other hand, Non-reversible watermarking usually introduces a small but irreversible degradation of the original signal quality [43,44].

1.3 Digital audio watermarking

Compared to digital image and video watermarking, digital audio watermarking is a more difficult task. Generally, the human auditory system (HAS) is much more sensitive than the human visual system (HVS). Therefore, inaudibility for audio is a lot more difficult to achieve than invisibility for images. Moreover, audio signals are represented by far fewer samples per time interval, and hence the amount of information that can be embedded robustly and inaudibly is much lower than that for visual media [45]. Due to the fact that the requirements of audio watermarking techniques are the most difficult to be satisfied compared with those for the other applications of digital watermarking, and in view of the dominance of the copyrights protection application, we give in this section the requirements of audio watermarking techniques and their benchmarking regarding the copyrights protection application.

1.3.1 Requirements of audio watermarking techniques for copyrights protection

Audio watermarking systems for copyrights protection have to comply with the following main requirements: excellent imperceptibility for preserving the perceptual quality of the audio signal, strong robustness against various attacks, and high-level of security to prevent unauthorized detection/removal. Furthermore, Data payload and computational complexity are two additional criteria that are desirable in most cases [1].

- Imperceptibility is a prerequisite of practicality. The process of audio watermarking is considered to be imperceptible or transparent if no differences between the host and watermarked signals are perceivable. Otherwise, it is perceptible or nontransparent. In the process of watermark embedding, to preserve the perceptual quality of the host audio signal, a psychoacoustic model derived from the auditory masking phenomenon or at least a knowledge about psychoacoustics would be relied on to deceive the human perception [46,47]. Consequently, it appears as if there is nothing added to the host media and the watermarked audio signal is identical to the original one from the perspective of the ear.

- Robustness is a measure of reliability and refers to the capability of resisting a variety of unintentional and intentional attacks. In other words, the watermark detector should be able to extract the watermark from the attacked watermarked signal. Examples of attacks on audio watermarking include many kinds of signal processing manipulations, such as noise addition, resampling, re-quantization, MPEG (Moving Picture Experts Group) compression, random samples cropping, timescale modification (TSM), and pitch-scale modification (PSM). The last three attacks belong to desynchronization attacks, which introduce displacement and heavily threaten the survival of the watermark.
- Since the watermarking embedding and extraction algorithms are likely to be open to the public, security is essential for copyrights protection. Therefore, we should guarantee that the watermarks cannot be ascertained even by reversing the embedding process or performing statistical detection [48]. Usually, pseudo-random encryption and/or scrambling operations can be adopted to add randomness into the embedding and detection processes so that the digital watermarking system is secured.
- Data payload (watermarking capacity) refers to the number of bits carried within a unit of time [49]. In digital audio watermarking, it is defined as the number of bits embedded in a one-second audio fraction and expressed in bit per second (bit/s or bps). The data payload of the audio watermarking system varies greatly depending on the embedding parameters and the embedding algorithm. Besides, different applications require different embedding payloads. For instance, copyrights protection application does not require a very high data payload [1].
- Computational complexity measures the required number of operations taken by a processor to embed and extract the watermark, and a low computational complexity is always desirable [1].

In practice, no one system can fully satisfy all the requirements and some trade-offs always exist among criteria. Typically, an audio watermarking system operates with a compromise between excellent imperceptibility and strong robustness. In order to ensure the robustness, one would embed the watermark into perceptually important regions or increase the strength of watermarking. However, such strategies are liable to cause perceivable distortion to the host signal, which is against the property of imperceptibility. Moreover, both robustness and imperceptibility are in close connection with the embedding payload. If one embeds more bits into an audio signal, the imperceptibility would become worse and even robustness might be affected [20].

1.3.2 Benchmarking on audio watermarking techniques

Along with the advancement of audio watermarking techniques, the necessity for benchmarking various algorithms effectively and comprehensively becomes imperative. Since appropriate assessment criteria always depend on the application, it is impractical and inaccurate to develop a universal benchmark for all kinds of digital watermarking systems [50]. As discussed above, imperceptibility, robustness and security are key principles in designing any audio watermarking scheme for security-oriented applications. Accordingly, performance evaluation is focused in our research on those three aspects.

1.3.2.1 Perceptual Quality Assessment

Similar to evaluating the quality of perceptual codecs in the audio, image, and video fields, perceptual quality assessment on the watermarked audio files is usually classified into two categories: subjective listening tests by human acoustic perception and objective evaluation tests by perception modeling or quality measures. Both of them are indispensable to the perceptual quality evaluation of audio watermarking. As perceptual quality is essentially decided by human opinion, subjective listening tests on audiences from different backgrounds are required in many applications. In subjective listening tests, the subjects are asked to discern the watermarked and host audio clips. Two popular models are the ABX test [51] and the MUSHRA test (i.e., MUlti Stimuli with Hidden Reference and Anchors) [52], derived from ITU-R Recommendation BS.1116 [53] and BS.1534 [54], respectively. Moreover, the watermarked signal is graded relative to the host signal according to a five-grade impairment scale as shown in Table 1.2. It is known as the subjective difference grade (SDG), which equals to the subtraction between subjective ratings given separately to the watermarked and host signals. Therefore, SDG near 0 means that the watermarked signal is perceptually undistinguished from the host signal, whereas SDG near 4 represents a seriously distorted version of the watermarked signal. However, such audibility tests are not only costly and time-consuming, but also heavily depend on the subjects and surrounding conditions. Therefore, the industry desires the use of objective evaluation tests to achieve automatic perceptual measurement. Currently, the most commonly used objective evaluation is perception modelling, i.e., assessing the perceptual quality of audio data via a stimulant ear, such as evaluation of audio quality (EAQUAL) [21,55], perceptual evaluation of audio quality (PEAQ) [55], and perceptual model-quality assessment (PEMO-QA) [55]. Moreover, objective quality measures are exploited as an alternative approach to quantify the dissimilarities caused by audio watermarking. For instance, a widely used quality measure is the signal to watermark ratio (SWR), calculated as follows [55]:

$$\mathrm{SWR} = \frac{\text{power of the host signal}}{\text{watermarking distortion}} \tag{1.2}$$

Table 1.2 Subjective difference grade (SDG).

Difference grade	Description of impairments
0	Imperceptible
-1	Perceptible but not annoying
-2	Slightly annoying
-3	Annoying
-4	Very annoying

1.3.2.2 Robustness Test

The goal of the robustness test is to examine the ability of a watermarking system to resist signal modifications in real-life applications. In the robustness test, various attacks are applied on the watermarked signal to produce several attacked signals. Subsequently, watermark detection is performed on each attacked signal to check whether the embedded watermark survives or not. The detection rate is measured by bit error rate (BER) defined as

$$\mathrm{BER} = \frac{\text{number of successfully extracted watermark bits}}{\text{total number of watermark bits}} \tag{1.3}$$

A competent robustness test should comprise an extensive range of possible attacks. Tens of attacks are employed in some popular audio watermarking evaluation platforms, i.e., secure digital music initiative (SDMI) standard, STEP2001, and StirMark for Audio [56]. In summary, typical signal manipulations on audio watermarking schemes are classified into three categories: common signal manipulations (such as noise addition, resampling, re-quantization, amplitude scaling, low-pass filtering, echo addition, reverberation, MP3 compression, DA/AD conversion, and combinations of two or more), desynchronization attacks (such as random samples cropping, jittering, zeros inserting, time-scale modification and pitch-scale modification), and advanced attacks (such as collisions and multiple watermarking). In most cases, a robustness test on an audio watermarking system includes the first two kinds of attacks, while the last kind is only taken into consideration for some specific applications. Moreover, desynchronization attacks are more challenging for most audio watermarking systems. Loss of synchronization would cause a mismatch in positions between watermark embedding and detection, which is disastrous to watermark retrieval [20].

It is worth to notice that there is a limit for taking a robustness test, that is, the degree of deterioration by attacks should be kept within an acceptable limit because it is needless for detection to proceed on a watermarked signal that is already severely destroyed. Therefore, attack parameters should control the amplitude of noise added and the extent of stretching or shifting within certain limits.

1.3.2.3 Security Analysis

Security analysis is performed to evaluate the characteristics of security for audio watermarking systems. Since security is attributed to the randomness merged by sequences of pseudorandom numbers and/or scrambling operations, an intuitive method of security analysis is to calculate the largeness of the key space, i.e., number of possible embedding ways. If there are more possible ways of embedding, it would be difficult for unauthorized detection to ascertain the embedded watermark. This indicates that the system has a high level of security.

Note that in the performance evaluation, a variety of audio signals have to be involved to truly verify the properties of the audio watermarking system. The test set should be representative of a typical range of audio content, such as classical, rock and folk music, vocal and instrumental music, and so on.

1.4 Literature review of watermarking methods

In recent years, there has been a considerable interest in the development of audio watermarking methods. To clarify the essential principles underlying a diversity of sophisticated watermarking methods, this section gives an overview of basic methods for audio watermarking, such as least significant bit (LSB) modification, patchwork, spread spectrum, echo hiding, and quantization index modulation.

1.4.1 *Least significant bit (LSB) modification*

One of the earliest attempts of information hiding and watermarking for digital media is the least significant bit (LSB) coding/modification [50]. In the simplest implementation of LSB modification, the least significant bit of the host signal sample is replaced by the to-be-hidden watermark bit. In a more secure scenario, the watermark encoder uses a secret key to choose a pseudo-random subset of the host signal samples. Then, the replacement of watermark is performed on those chosen samples. In the decoder side, the same secret key is used to select the same subset to extract from it the watermark bits.

The obvious advantage of LSB is its the high watermarking capacity. For example, when using only the least significant bit of the CD quality (44.1 kHz sampling rate, 16 bits per sample) host signal, the encoder can achieve 44100 bits per second (bps) watermark capacity. Some audio watermarking system uses the least 3 or even 4 significant bits of the host audio signal for watermarking embedding, achieving super high 132.3 kbps to 176.4 kbps watermark capacity. Another advantage of LSB coding is its simplicity because it requires very little computation cost for both the watermark encoder and decoder, and making real-time watermark embedding and extraction possible even for computationally limited devices. However, despite its advantages, LSB has several drawbacks:

- LSB coding has very weak robustness; the simplest attacks like random cropping or small noise addition would destroy the embedded watermark.
- The depth of LSB is limited. In order to minimize the possible audible distortion, only the least 4 significant bits of the 16 bits per sample host audio signal can be used for watermark coding purpose.

1.4.2 *Patchwork*

Patchwork was originally developed for image watermarking [57] and later being used for audio watermarking as well [58,59]. The patchwork method uses statistical hypothesis on two sets of large samples for information hiding, which makes it a good method for audio watermarking due to the huge number of digital samples in audio signals. In the simple patchwork encoding scenario, a secret key is used to pseudo-randomly select two sets of samples, i.e. patches denoted by A and B. The amplitudes of each sets are slightly changed in the opposite way, i.e., the amplitudes of one set of samples are increased by a small amount d and the amplitudes of the other set of samples are decreased by the same amount d, which is carefully chosen such that it is not too small so that it is robust to possible added noise during transmission, nor it is too large to introduce perceivable audible distortion. This can be illustrated as

$$\begin{cases} a_k^w = a_k + d \\ b_k^w = b_k - d \end{cases} \tag{1.4}$$

where a_k and b_k are the k^{th} samples of the selected sets A and B, respectively, and a_k^w and b_k^w are the k^{th} samples of the watermarked sets $A^{(W)}$ and $B^{(W)}$.

At the decoder side, the same secret key is employed to choose two sets $A^{(R)}$ and $B^{(R)}$ from the received signal. Then the expectation difference of these sets is computed. If it is equal to $2d$, the received signal is indicated as watermarked, i.e., the received signal is indicated as watermarked if $E\{A^{(R)} - B^{(R)}\} \approx 2d$, with $E\{\cdot\}$ denoting the statistical average. To understand this decision rule, let us consider the case where the received signal is the watermarked signal, i.e., $A^{(R)} = A^{(W)}$ and $B^{(R)} = B^{(W)}$.. In this case

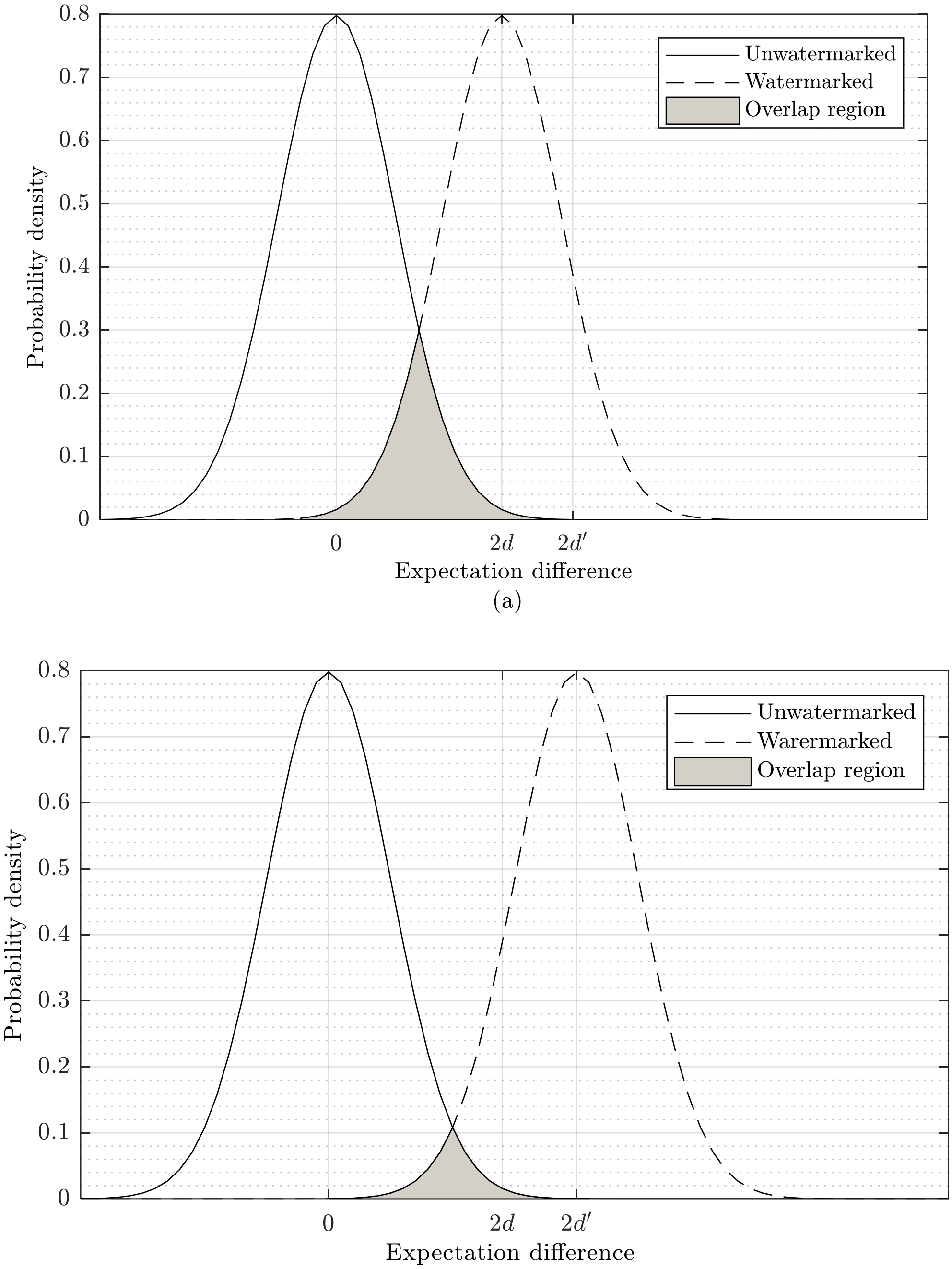
$$E\{A^{(R)} - B^{(R)}\} = \frac{1}{N}\sum_{k=1}^{N}(a_k^w - b_k^w) \tag{1.5}$$

by replacing (1.4) in (1.5) we find

$$E\{A^{(R)} - B^{(R)}\} = 2d + \frac{1}{N}\sum_{k=1}^{N}(a_k - b_k) \tag{1.6}$$

Due to the pseudo-random selection of $A^{(R)}$ and $B^{(R)}$, the last portion of the right-hand side in (1.6) is expected to be zero, and thus $E\{A^{(R)} - B^{(R)}\}$ is expected to be near $2d$ for the case where the received signal is watermarked. Similarly, $E\{A^{(R)} - B^{(R)}\}$ is expected to be near zero for the case where the received signal is not watermarked

The problem for patchwork is that in real application systems, the mean difference between two randomly selected data sets is not always zero; the last portion of the right-hand side in (1.6) is expected to be zero but its value is not zero in most cases. Although the distribution of the mean difference of those two watermarked patches is shifted to the right of the unwatermarked version by $2d$, there is still some overlap between the two distributions as illustrated in Fig. 1.2. Therefore, there lies probability of wrong detection. It is possible to make the detection more accurate by increasing the amount of modification d. However, this would lead to a higher risk of audible distortions.



(a)

(b)

Fig. 1.2 Patchwork expectation difference distribution for watermarked and unwatermarked signals, (a) amount of modification $\boldsymbol{d}$, and (b) amount of modification $\boldsymbol{d'}$, such that $\boldsymbol{d'} > \boldsymbol{d}$.

1.4.3 Spread spectrum

Spread spectrum (SS) watermarking is considered as one of the most popular methods for digital watermarking [60–62]. It spreads the watermark throughout the spectrum of the host signal, and thus the signal energy present in every frequency bin is very small and hardly noticeable. In this way, the embedded watermark can possess a large measure of security as well as robustness [60,61]. However, the process of watermarking may easily introduce perceivable distortion to audio signals. Therefore, amplitude shaping by the masking threshold from psychoacoustic models is often employed to keep the watermark inaudible [62].

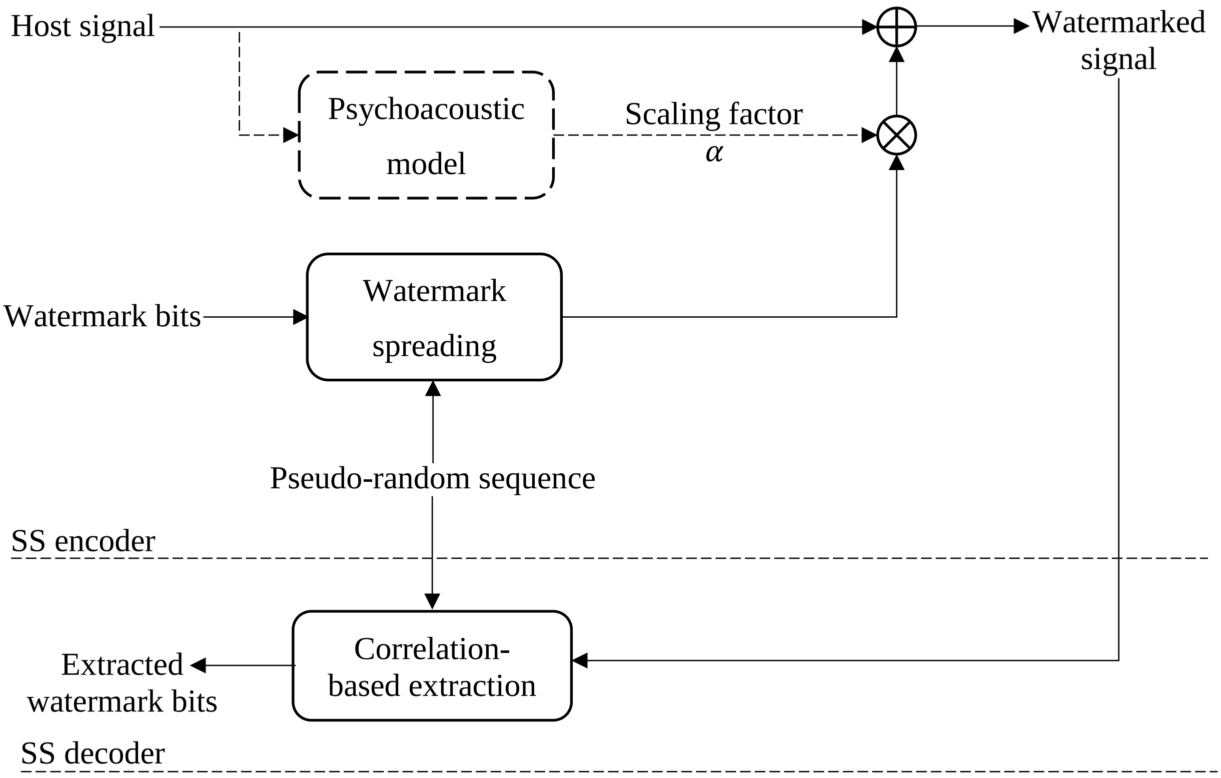


Fig. 1.3 Block diagram of SS encoder and decoder.

There are two main forms of SS watermarking, namely, direct sequence spread spectrum (DSSS) and frequency hopping spread spectrum (FHSS). The DSSS-based audio watermarking method is more commonly used and its basic scheme is shown in Fig. 1.3. In the watermark embedding process, the watermark w is modulated by the pseudo-random sequence r_s to produce the modulated watermark w_m. To keep w_m inaudible, scaling factor α is used to control the amplitude of w_m. Then the watermarked signal s_w is produced by adding $\alpha\ w_m$ to the host signal s_0. In the watermark detection process, the watermark w_e is extracted by correlating the watermarked signal s_w with the same pseudo-random sequence r_s used in the embedding.

Note that the watermark can be spread not only in the time domain but also in various transformed domains. Discrete Fourier transform (DFT), discrete cosine transform (DCT), and discrete wavelet transform (DWT) are some examples of transforms that are frequently used in SS-based techniques.

SS can achieve good robustness-transparency trade-off by using psychoacoustic models. However, the usage of such models enormously increases the computational complexity. Besides, the major disadvantage of SS is its low embedding rates and host interference because the pseudo-random sequence employed by SS has a small but non-zero correlation with the host signal, which leads to a probability of error when extracting the watermark even if the signal is not attacked. The probability of error can be reduced by using larger values of α, however, a larger value of α leads to poorer perceptual quality.

1.4.4 Echo hiding

Echo hiding embeds the watermark into host signals by introducing different echoes. With well-designed amplitudes and delays (offset), the echoes are perceived as resonance to host audio signals and would not produce uncomfortable noises [36].

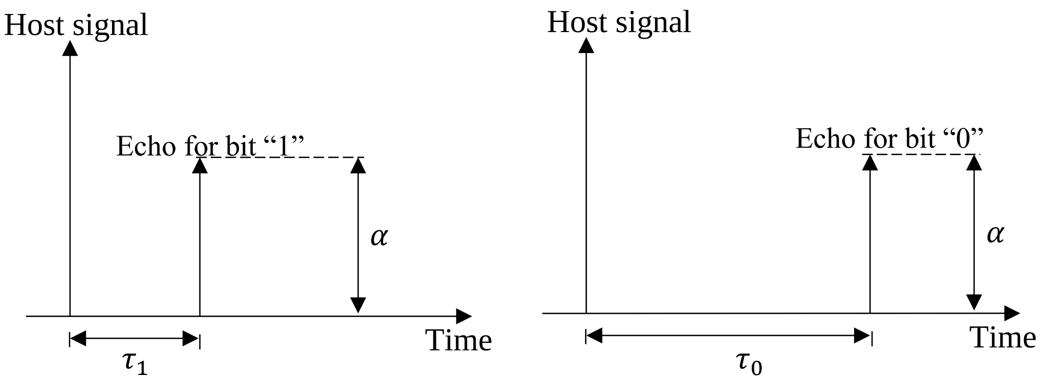


Fig. 1.4 Impulse response of echo kernels. (a) "One" kernel, (b) "Zero" kernel.

In the embedding process, the watermarked signal $x^{(w)}(t)$ is generated in the time domain by the convolution between the host signal $x(t)$ and the echo kernel $h(t)$, where t represents time. The basic echo hiding scheme employs a single echo kernel whose impulse response is expressed as

$$h(t) = \delta(t) + \alpha \cdot \delta(t - \tau) \tag{1.7}$$

where α is echo amplitude and τ is the delay. To represent the bits "1" and "0," echo kernels are created with different delays, i.e., $\tau = \tau_1$ if the watermark bit value is "1", and $\tau = \tau_0$ if the watermark bit value is "0", with $\tau_1 \neq \tau_0$, as shown in Fig. 1.4. Usually, the maximum allowable delay offset for 44.1 kHz sampled audio signals is about 100∼150 samples (about 2.3 ∼ 3.4 ms) [36]. Consequently, the watermarked signal is described as follows [5]

$$\begin{aligned} x^{(w)}(t) &= x(t) * h(t) \\ &= x(t) + \alpha \cdot x(t - \tau) \end{aligned} \tag{1.8}$$

In order to detect the watermark, cepstrum analysis is utilized to discern the value of delay. The complex cepstrum of the watermarked signals $X^{(w)}(t)$ is defined as

$$X^{(w)}(t) = \mathcal{F}^{-1}\{\ln(\mathcal{F}\{x^{(w)}(t)\})\} \tag{1.9}$$

where $\mathcal{F}^{-1}\{\cdot\}$ and $\mathcal{F}\{\cdot\}$ denote the Fourier transform and its inverse, respectively, and is used to calculate the auto-cepstrum function [5,32] given by

$$c_a(t) = \mathcal{F}^{-1}\left\{\ln\left(\mathcal{F}\left\{\left(X^{(w)}(t)\right)^2\right\}\right)\right\} \tag{1.10}$$

that is employed to extract the watermark bit as [5]

$$w = \begin{cases} 1, & c_a(\tau_1) \geq c_a(\tau_0) \\ 0, & \text{otherwise} \end{cases} \tag{1.11}$$

The performance of echo hiding depends on echo kernels, and hence different echo kernels have been introduced to improve the imperceptibility and robustness of the embedded echoes [32,36]. By selecting the proper amplitude and delay of echo kernels, the echoes embedded as the watermark can be imperceptible and robust against most attacks. However, echo hiding suffers from two deficiencies. One deficiency is its weak security because obvious cepstrum peaks might be tampered with intentionally, e.g., an unauthorized person can remove the watermark easily by modifying the cepstrum peaks. The other deficiency is about inherent echoes contained in natural sound, which might result in false-positive errors.

1.4.5 Quantization index modulation

Quantization index modulation (QIM) has been proposed by Chen and Wornell as a class of provably good methods for digital watermarking [22]. It has a very good rate-robustness-transparency trade-offs and is a widely popular class of watermarking methods that hide information by quantizing samples. In QIM, each sample is quantized with a predetermined quantization step. Then, a slight modification is made to each quantized sample according to the values of the watermark bits. In [63], the author introduced a simple implementation of QIM as follows: suppose the input host sample is X, the quantization step is Δ, and the watermark bit to be embedded is $w \in \{0,1\}$, then the watermarked sample $\hat{X}$ is obtained as

$$\hat{X} = Q(X,\Delta) + (2w-1)\frac{\Delta}{4} \tag{1.12}$$

where $Q(X,\Delta)$ is the standard quantization function given by

$$Q(X,\Delta) = \left[\frac{X}{\Delta}\right]\Delta \tag{1.13}$$

with $[\cdot]$ denoting the rounding of a value to the nearest integer.

Fig. 1.5 illustrates the watermark embedding of this implementation. The sample X is first quantized to the $Q(X,\Delta)$ (black circle). If the to be embedded watermark bit is 1, then $\Delta/4$ is added to the quantized sample value, which moves the sample up to the while circle. Otherwise, $\Delta/4$ is subtracted from the quantized sample value, which moves the sample down to the cross.

At the decoder side, the absolute value of the difference between the received sample and its quantized value is computed. If it is between 0 and $\Delta/4$, then the extracted watermark bit is “1”. If the difference lies between $-\Delta/4$ and 0, then the extracted watermark bit is “0”. Otherwise, the received signal is not watermarked. This can be illustrated as

$$w = \begin{cases} 1, & 0 \le \left|\hat{X} - Q(\hat{X},\Delta)\right| < \dfrac{\Delta}{4} \\ 0, & \dfrac{\Delta}{4} \le \left|\hat{X} - Q(\hat{X},\Delta)\right| \le \Delta \end{cases} \tag{1.14}$$

or equivalently, by the minimum distance decoder as

$$w = \underset{\omega \in \{0,1\}}{\operatorname{argmin}} \left| Q(\hat{X},\Delta) + (2\omega - 1)\frac{\Delta}{4} - \hat{X} \right| \tag{1.15}$$

QIM is very simple to implement, optimal in terms of channel capacity, and robust to noise addition attacks. As long as the introduced noise at transmission channel is less than $\Delta/4$, the detector can always correctly extract the watermark. However, if the noise exceeds $\Delta/4$, the watermark might not be precisely detected. This is a tradeoff between watermark robustness and transparency. A larger quantization step Δ leads to more robustness against noise addition with the risk of creating audible distortion to the host audio signal. Nevertheless, despite its simplicity and attractive attributes, QIM’s major drawback is its diversity and high fragility against gain attacks [64].

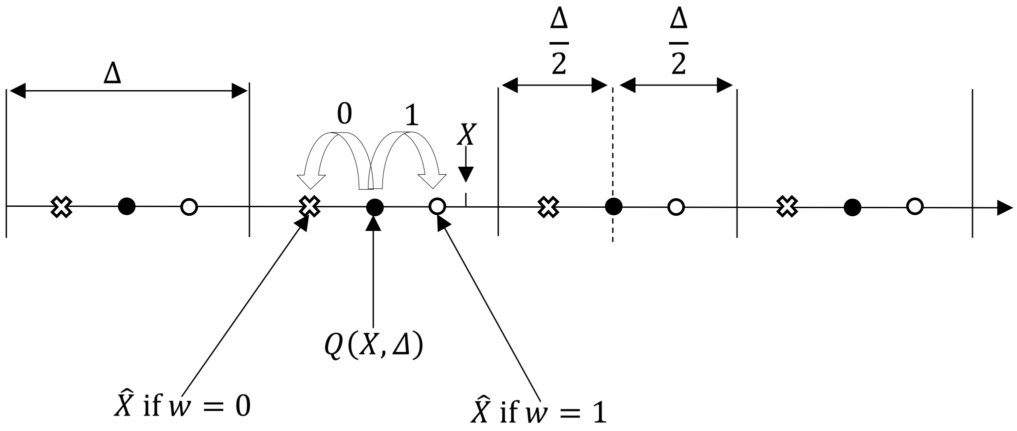


Fig. 1.5 A generic QIM embedding procedure.

1.5 Conclusion

In this chapter, an overview of digital watermarking along with its system framework, applications and classifications have been provided. Moreover, the requirements of digital audio watermarking techniques and their benchmarking for copyrights protection have been highlighted. Furthermore, state of the art of audio watermarking methods has also been studied and the concepts, advantages and drawbacks of different existing methods have been discussed. From these study and discussion, it can be concluded that QIM-based audio watermarking methods are more attractive than other methods for copyrights protection due to the optimality of QIM and its good rate-robustness-transparency trade-offs. Therefore, we thoroughly and objectively exploit QIM in the subsequent chapters.

Chapter 2

Proposed blind audio watermarking technique based on a parametric QIM

2.1 Introduction

Due to its advantages, QIM has received a wide attention from researchers in the field of watermarking. Wu et al. proposed in [20] a QIM-based watermarking technique in the discrete wavelet transform (DWT) domain and a baseline mechanism for its self-synchronization. In [65], the authors presented a QIM-based audio watermarking technique for quantizing a weighted group amplitude of the lowest DWT approximation band, where the coefficients of the approximation band are then altered in a way that transparency is optimized. Chen et al. proposed in [13] a QIM-based audio watermarking technique that quantizes the absolute amplitude of the lowest 8-level DWT frequency band, where the signal coefficients are then modified in a manner that maximizes transparency. Hu et al. proposed in [14] a QIM-based audio watermarking technique by quantizing the standard deviations of the discrete cosine transform (DCT) bands. The main idea of this technique is to exploit the HAS in order to devise an adaptive quantization step, which results in a non-uniform quantization. In [66], the authors proposed a QIM-based audio watermarking technique, which embeds the watermark bits by quantizing the low-frequency band of the host signal. The concept of this technique is to first identify the sample positions, which offer a good transparency when quantized. The sample positions vary from a frame to another and thus are stored as a side information to be transmitted separately over a high-fidelity channel to the decoder. Secondly, embed the watermark in a host signal many times then perform numerous attacks on the watermarked signal followed by several watermark extractions to serve the calculations performed by the employed stochastic optimization. This is to find a quantization step providing a good robustness to the considered attacks.

The above QIM-based audio watermarking techniques have many drawbacks, namely, the vulnerability to high-pass filtering-like attacks and lossy compression of the quantization of the considered global frame characteristic (e.g. the absolute amplitude in [13,65] and the standard deviation in [14]). Moreover, the optimization in [65] and [13] is performed to maximize the transparency without considering the robustness, and the 8-level DWT in [13] restricts the maximum achievable capacity, which renders the technique in [13] not suitable for applications requiring high embedding capacity. The adaptive quantization steps in [14] serve for higher capacity at the expense of robustness due to the extra error caused by recovering the quantization steps in the extraction procedure, and the sensitivity of the extracted quantization steps to additive noise. Although [66] considers both robustness and transparency, the technique is not blind as the side information is essential to perform the watermark extraction, and time-consuming because of the use of stochastic optimization, which is known for its downsides compared to analytical optimization. Finally, the most important deficiency in the techniques reported in [13,14,65,66] is that the used QIM realization is neither justified nor optimized for a given technique. An exhaustive literature review shows that other watermarking techniques such as those reported in [10,15,67–71] also have such a deficiency. Therefore, it is highly desirable to find a QIM realization that is more appropriate or optimal for a given watermarking technique.

In this chapter, we propose a new blind audio watermarking technique based on the transform domain [28]. It consists of (1) segmenting the audio signal into frames, (2) transforming each frame using an orthogonal transform, (3) embedding one watermark bit per transformed frame, and (4) applying the inverse transformation on each of the resulting watermarked frames. In order to find a QIM realization that is optimal for the proposed watermarking technique, we (1) propose a new parametric QIM, which reduces for some values of the parameters to the QIM realizations used in [10,13–15,20,65–71], (2) derive the theoretical expressions for the signal to watermark ratio (SWR) and the probability of error, and (3) propose an optimization technique based on the Lagrange multipliers method to find the optimal values for the parameters of the proposed parametric QIM, which is used to quantize a single coefficient per frame. The principle of the proposed optimization technique is to set the SWR to a fixed value SWR_0, which is according to the international federation of photographic industry (IFPI) higher than 20 dB [72], in order to guarantee the imperceptibility while minimizing the probability of error under an additive white Gaussian noise (AWGN) attack to maximize the robustness. Although any orthogonal transform can be used in the proposed technique, the DCT is chosen for its high energy compaction capability. As mentioned earlier, the transform-domain watermarking techniques suffer from higher computational complexities. Fortunately, this is not the case with the proposed technique for which we propose a very fast scheme to pass from the time domain to transform domain and vice versa. The idea behind this scheme is to appropriately exploit the fact that the proposed watermarking technique embeds only one watermark bit per frame by adopting the single coefficient quantization approach. We also propose a procedure leading to the best interval from which an embedding position can be selected to provide a good trade-off between the effects of high and low pass filtering attacks. This would be feasible due to the adopted single coefficient quantization approach, which offers a good interpretation for the high and low pass filtering attacks. The embedding positions and parameters of the obtained optimal parametric QIM can be used as a secret key in the proposed watermarking technique. Finally, we experimentally show the validity of the theoretical analysis developed in this chapter and investigate the efficiency of the proposed watermarking technique in terms of robustness and imperceptibility, and compare it with those of the techniques reported in [13,14,20,65].

The rest of this chapter is organized as follows. Section 2.2 presents the proposed watermarking technique along with the proposed parametric QIM. Theoretical expressions for the SWR and the probability of error are derived in Section 2.3. The optimal parametric QIM is obtained in Section 2.4 using an optimization method based on the Lagrange multipliers. This section proposes also a closed-form expression approximating the derived analytical expression of the probability of error using the particle swarm optimization (PSO) [73–75]. Section 2.5 suggests a procedure to find the best interval for selecting watermark embedding positions. An algorithm for a fast implementation of the proposed watermarking technique is proposed in Section 2.6. Experimental results along with comparisons and discussions are presented in Section 2.7. Section 2.8 gives some conclusions.

2.2 Proposed watermarking technique

In this section, we propose an efficient blind audio watermarking technique based on the transform domain. We first briefly review the standard QIM and some of its realizations and then devise a parametric QIM to be used in the proposed watermark embedding and extraction algorithms.

2.2.1 *Existing QIM realizations*

In this subsection, we present some existing QIM realizations. It has been suggested in [76] to obtain the watermarked sample $\hat{X}$ by using a quantizer that depends on w as

$$\hat{X} = Q(X, \Delta, w) \tag{2.1}$$

Since then, various realizations of QIM have been proposed. For instance, the QIM used in [13,20,65] has the form

$$\hat{X} = \left[\frac{X}{\Delta}\right]\Delta + w\frac{\Delta}{2} + \frac{\Delta}{4} \tag{2.2}$$

whereas that considered in [10,14,67] has the form

$$\hat{X} = \left[\frac{X - w\frac{\Delta}{2}}{\Delta}\right]\Delta + w\frac{\Delta}{2} \tag{2.3}$$

The realization of the QIM is adopted in [66] and [68] as

$$\hat{X} = \left[\frac{X - \frac{\Delta}{2} - w\frac{\Delta}{2}}{\Delta}\right]\Delta + w\frac{\Delta}{2} + \frac{\Delta}{2} \tag{2.4}$$

In [69], the realization is

$$\hat{X} = \left[\frac{X - \frac{\Delta}{2}}{\Delta}\right]\Delta + w\frac{\Delta}{2} \tag{2.5}$$

The form used in [70] is

$$\hat{X} = \left[\frac{X - \frac{\Delta}{2}}{\Delta}\right]\Delta + w\frac{\Delta}{2} + \frac{\Delta}{2} \tag{2.6}$$

Another interesting realization of the QIM is the well-known dither-modulation introduced by Chen and Wornell in [76] and considered in [71], which can be formulated in the case of uniform scalar quantization as

$$\hat{X} = Q(X, \Delta, w) = \left[\frac{X + d(w)}{\Delta}\right]\Delta - d(w) \tag{2.7}$$

with

$$d(1) = \begin{cases} d(0) + \frac{\Delta}{2}, & d(0) < 0 \\ d(0) - \frac{\Delta}{2}, & d(0) \geq 0 \end{cases} \tag{2.8}$$

where $d(0)$ can be chosen pseudo-randomly with a uniform distribution over the interval $[-\Delta/2, \Delta/2]$.

2.2.2 Proposed parametric QIM

The diversity of the QIM realizations found in the literature suggests a sort of unification, which can be achieved by introducing a parametrization approach. Moreover, this approach can be exploited to find the optimal QIM realization for a given watermarking technique. Therefore, we define and propose a new parametric realization of the QIM as

$$\begin{aligned}\hat{X} &= Q_{\alpha,\beta,\gamma}(X,\Delta,w)\\ &= \left[\frac{X+\beta+\gamma w}{\Delta}\right]\Delta + w\frac{\Delta}{2} + \alpha\end{aligned} \tag{2.9}$$

where α, β and γ are parameters to be defined and w is the watermark bit to be embedded. The proposed watermark extraction process is given by

$$\begin{aligned}\hat{w} &= Q_{\alpha}^{-1}(\hat{X},\Delta)\\ &= \begin{cases}1 & \text{if } \dfrac{\Delta}{4} \leq \operatorname{mod}(\hat{X}-\alpha,\Delta) < \dfrac{3\Delta}{4}\\ 0 & \qquad\qquad \text{otherwise}\end{cases}\end{aligned} \tag{2.10}$$

where $\hat{w}$ is the extracted watermark bit.

It is clear that for a suitable choice of the parameters α, β and γ, the proposed parametric QIM defined in (2.9) reduces to any of the QIMs given by (2.2)-(2.6). For instance, for $\beta = \gamma = 0$ and $\alpha = \Delta/4$, the parametric QIM reduces to the QIM given by (2.2), which is used in [13,20,65], and for $\alpha = \beta = 0$, $\gamma = -\Delta/2$, reduces to the QIM given by (2.3), which is used in [10,14,15,67].

2.2.3 Proposed watermark embedding

The objective of the proposed watermark embedding is to embed a binary image in a digital audio signal as shown in Fig. 2.1. Hence, the image is firstly mapped into a vector W whose elements are the watermark bits w_m. The original digital audio signal is segmented into non-overlapping frames $x^{(m)}$ each of length L, and an orthogonal linear transformation T is applied on each frame to obtain $X^{(m)}$ as

$$X_r^{(m)} = \sum_{l=0}^{L-1} x_l^{(m)} \phi_l^*(r), r = 0, 1, \dots, L-1 \tag{2.11}$$

where $\phi_l(r)$, $0 \leq l, r \leq L-1$, is a set of linearly independent orthogonal basis constituting the kernel of the transform T and $(\cdot)^*$ denotes the complex conjugate operation.

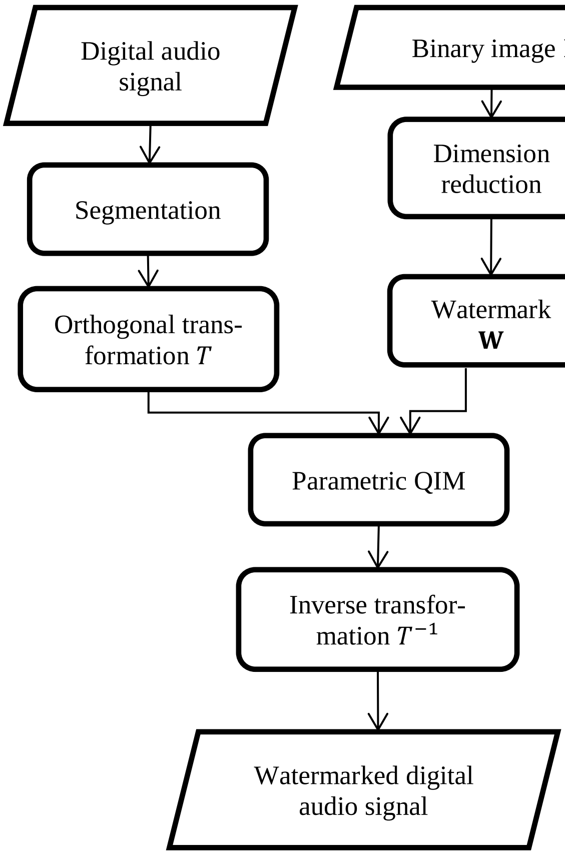


Fig. 2.1 Proposed watermark embedding algorithm.

We embed one bit per frame in the transform domain. Hence, the watermark bit w_m indexed by m is embedded according to the proposed parametric QIM given by (2.9) in a chosen coefficient $X_k^{(m)}$ of the transformed frame $X^{(m)}$. Thus, the watermarked coefficient $\hat{X}_k^{(m)}$ is obtained as

$$\begin{aligned} \hat{X}_k^{(m)} &= Q_{\alpha,\beta,\gamma}\left(X_k^{(m)}, \Delta, w_m\right) \\ &= \left[\frac{X_k^{(m)} + \beta + \gamma w_m}{\Delta}\right]\Delta + w_m\frac{\Delta}{2} + \alpha \end{aligned} \tag{2.12}$$

where α, β, γ and Δ to be defined in a way to obtain the optimal robustness while ensuring imperceptibility (or inaudibility). Then, the inverse transformation T^{-1} is applied on each watermarked frame $\hat{X}^{(m)}$ in the transform domain to obtain the watermarked frame $\hat{x}^{(m)}$ in the time domain as

$$\hat{x}_l^{(m)} = \sum_{r=0}^{L-1} \hat{X}_r^{(m)} \phi_l(r), l = 0, 1, \dots, L-1 \tag{2.13}$$

The resulting watermarked frames are all joined to construct the watermarked digital audio signal.

2.2.4 Proposed watermark extraction

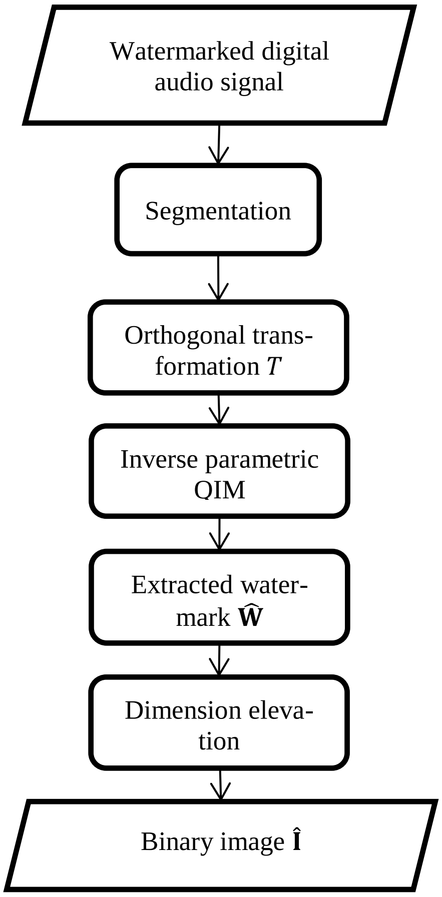

Watermarked digital audio signal
Segmentation
Orthogonal transformation T
Inverse parametric QIM
Extracted watermark $\hat{\mathbf{W}}$
Dimension elevation
Binary image $\hat{\mathbf{I}}$

Fig. 2.2 Proposed watermark extraction algorithm.

The proposed watermark extraction process is illustrated in Fig. 2.2. The watermarked digital audio signal is firstly segmented into non-overlapping frames $\hat{x}^{(m)}$ each of length L, and the same orthogonal linear transformation T given by (2.11) used in the embedding process is applied on each frame to obtain $\hat{X}^{(m)}$. The watermark bit $\widehat{w}_m$ indexed by m is extracted from the coefficient $\hat{X}_k^{(m)}$ of the transformed frame $\hat{X}^{(m)}$. For this purpose, we apply the inverse parametric QIM given by (2.11) as

$$\begin{aligned}\widehat{w}_m &= Q_\alpha^{-1}\left(\hat{X}_k^{(m)}, \Delta\right) \\ &= \begin{cases} 1 & \text{if } \dfrac{\Delta}{4} \leq \mathrm{mod}\left(\hat{X}_k^{(m)} - \alpha, \Delta\right) < \dfrac{3\Delta}{4} \\ 0 & \text{otherwise} \end{cases}\end{aligned} \tag{2.14}$$

Finally, extract all the bits of the watermark and arrange them in a vector $\widehat{W}$, which is then converted to a two-dimensional image, i.e., the extracted binary image.

2.3 Performance analysis

In this section, we derive the theoretical expressions for the SWR and the probability of error in order to use them for finding the values of the parameters that optimize the proposed watermarking technique.

2.3.1 *Signal to watermark ratio*

The signal to watermark ratio is defined by

$$\mathrm{SWR} = \frac{P_S}{P_W} \tag{2.15}$$

where P_S is the signal power, which is the sum of the squared samples of the host audio signal divided by the signal length, and is given by

$$P_S = \frac{1}{L_w L} \sum_{m=1}^{L_w} \sum_{k=0}^{L-1} \left(x_k^{(m)} \right)^2$$

with L_w being the total number of the frames. The watermark power P_W is the power of the noise added to the signal due to the embedding distortion and is given by

$$P_W = \left\langle \frac{1}{L} \sum_{l=0}^{L-1} \left(x_l^{(m)} - \hat{x}_l^{(m)} \right)^2 \right\rangle \tag{2.16}$$

where $x_l^{(m)}$ and $\hat{x}_l^{(m)}$ are the samples of the host and watermarked frames $x^{(m)}$ and $\hat{x}^{(m)}$, respectively, and $\langle\cdot\rangle$ denotes the statistical average operator. Since the frames $X^{(m)}$ and $\hat{X}^{(m)}$ are obtained by transforming the frames $x^{(m)}$ and $\hat{x}^{(m)}$, respectively, it is easy to show that P_W given by (2.16) can be expressed in terms of the difference

$$\varepsilon = X_k^{(m)} - \hat{X}_k^{(m)} \tag{2.17}$$

as

$$P_W = \frac{1}{L} \langle \varepsilon^2 \rangle = \frac{1}{L} \int_{\text{domain of } \varepsilon} \varepsilon^2 f_\varepsilon(\varepsilon) d\varepsilon \tag{2.18}$$

where $f_\varepsilon(\varepsilon)$ is the probability density function (PDF) of ε. The substitution of (2.12) in (2.17) yields

$$\varepsilon = X_k^{(m)} - \left\lfloor \frac{X_k^{(m)} + \beta + \gamma w_m}{\Delta} \right\rfloor \Delta - w_m \frac{\Delta}{2} - \alpha \tag{2.19}$$

By adding and subtracting the quantity $\beta + \gamma w_m$, (2.19) can be rearranged as

$$\varepsilon = \left(X_k^{(m)} + \beta + \gamma w_m\right) - \left\lfloor \frac{X_k^{(m)} + \beta + \gamma w_m}{\Delta} \right\rceil \Delta - w_m \frac{\Delta}{2} - \alpha - \beta - \gamma w_m \tag{2.20}$$

Thus, using the quantization given by (1.13), (2.20) can be expressed as

$$\varepsilon = z - \left(w_m \frac{\Delta}{2} + \alpha + \beta + \gamma w_m\right) \tag{2.21}$$

where

$$z = y - Q(y, \Delta) \tag{2.22}$$

and

$$y = X_k^{(m)} + \beta + \gamma w_m \tag{2.23}$$

The quantization step Δ in (2.22) must be chosen significantly smaller than the host sample y given by (2.23) in order to have a small quantization noise z. Therefore, according to Bennett's high rate model for quantization [77], the random variable z is uniformly distributed in the interval $[-\Delta/2\,, \Delta/2]$, i.e., its PDF is given by

$$f_z(z) = \begin{cases} \frac{1}{\Delta} & z \in \left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right] \\ 0 & \text{otherwise} \end{cases} \tag{2.24}$$

Let us assume that the watermark binary image constitutes of p white pixels (p 1s) and q black pixels (q 0s). The probability of the watermark bits can be expressed as

$$P(w_m) = \frac{1}{p+q}\left(q\ \delta(w_m) + p\ \delta(w_m - 1)\right) \tag{2.25}$$

where $\delta(l)$ is defined as

$$\delta(l) = \begin{cases} 1 & \text{if} \quad l = 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.26}$$

It is clear that ε given by (2.21) is a function of z and w_m, which are two independent random variables. Therefore, using (2.21), (2.24) and (2.25) in (2.18), the watermark power can then be expressed as

$$P_W = \frac{1}{\Delta L} \sum_{w_m=0}^{1} \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \left(z - w_m \frac{\Delta}{2} - \alpha - \beta - \gamma w_m\right)^2 \cdot \frac{\left(q\ \delta(w_m) + p\ \delta(w_m - 1)\right)}{p+q} dz \tag{2.27}$$

After some mathematical manipulations, (2.27) reduces to

$$P_W = \frac{1}{L}\left(\frac{\Delta^2}{12} + (\alpha + \beta)^2 + \frac{q(2\gamma + \Delta)}{4(p+q)}\left((\Delta + 2\gamma) + 4(\alpha + \beta)\right)\right) \tag{2.28}$$

By substituting (2.28) in (2.15), an expression for the SWR is obtained as

$$\mathrm{SWR} = \frac{P_S L}{\left(\frac{\Delta^2}{12} + (\alpha+\beta)^2 + \frac{q(2\gamma+\Delta)}{4(p+q)}\left((\Delta+2\gamma)+4(\alpha+\beta)\right)\right)} \tag{2.29}$$

2.3.2 Probability of error and the BER under an AWGN attack

In the case of transmission of the watermarked signal through an additive white Gaussian noise (AWGN) channel, the received noisy watermarked signal frame indexed by m is given by

$$\widehat{xn}^{(m)} = \hat{x}^{(m)} + n^{(m)} \tag{2.30}$$

where $n^{(m)}$ is the noise frame, which is a white Gaussian $\mathcal{N}(0, \sigma_n)$. By applying the linear discrete transformation T given by (2.11) on the frame $\widehat{xn}^{(m)}$ given by (2.30), we obtain

$$\widehat{XN}^{(m)} = \hat{X}^{(m)} + N^{(m)} \tag{2.31}$$

where $N^{(m)}$ is the transform of $n^{(m)}$ and it can be shown that it is also white Gaussian $\mathcal{N}(0, \sigma_n)$.

In the following, without loss of generality, we assume that the watermarked transformed frame $\hat{X}^{(m)}$ was obtained by embedding in the original host transformed frame $X^{(m)}$ the watermark bit $w_m = 0$ (the same analysis lead to the same results in the case of $w_m = 1$).

The extraction of the watermark bit from the received noisy frame $\widehat{XN}^{(m)}$ is performed according to (2.14) as

$$\begin{aligned}\hat{w}_m &= \begin{cases}1 & \text{if } \frac{\Delta}{4} \le \mathrm{mod}\left(\widehat{XN}_k^{(m)} - \alpha, \Delta\right) < \frac{3\Delta}{4} \\ 0 & \text{otherwise}\end{cases} \\ &= \begin{cases}1 & \text{if } \frac{\Delta}{4} \le \mathrm{mod}\left(\hat{X}_k^{(m)} + N_k^{(m)} - \alpha, \Delta\right) < \frac{3\Delta}{4} \\ 0 & \text{otherwise}\end{cases} \\ &= \begin{cases}1 & \text{if } \frac{\Delta}{4} \le \mathrm{mod}\left(N_k^{(m)}, \Delta\right) < \frac{3\Delta}{4} \\ 0 & \text{otherwise}\end{cases}\end{aligned} \tag{2.32}$$

To obtain (2.32), we have used the fact that $\left(\hat{X}_k^{(m)} - \alpha\right)$ is a multiple of Δ, which can clearly be seen from (2.12). It is clear from (2.32) that a transmission error occurs in the case of

$$\frac{\Delta}{4} \le \mathrm{mod}\left(N_k^{(m)}, \Delta\right) < \frac{3\Delta}{4} \tag{2.33}$$

which is satisfied for all values of $N_k^{(m)}$ in the range

$$r\,\Delta + \frac{\Delta}{4} \le N_k^{(m)} \le r\,\Delta + \frac{3\Delta}{4} \tag{2.34}$$

where r is any integer. Therefore, the probability of error is the sum over all possible values of r and is given by

$$P_e = P\left(\frac{\Delta}{4} \leq \text{mod}\left(N_k^{(m)}, \Delta\right) < \frac{3\Delta}{4}\right)$$

$$= \sum_{r=-\infty}^{\infty} P\left(r\,\Delta + \frac{\Delta}{4} \leq N_k^{(m)} \leq r\,\Delta + \frac{3\Delta}{4}\right) \tag{2.35}$$

Since $N^{(m)}$ is $\mathcal{N}(0, \sigma_n)$, the probability in (2.35) can be expressed as

$$P\left(r\,\Delta + \frac{\Delta}{4} \leq N_k^{(m)} \leq r\,\Delta + \frac{3\Delta}{4}\right) \quad = \int_{r\,\Delta+\frac{\Delta}{4}}^{r\,\Delta+\frac{3\Delta}{4}} \frac{e^{-\frac{N_k^{(m)^2}}{2\sigma_n^2}}}{\sqrt{2\pi\sigma_n^2}} dN_k^{(m)} \tag{2.36}$$

After computing (2.36) and using it in (2.35), the probability of error becomes

$$P_e = \frac{1}{2} \sum_{r=-\infty}^{\infty} \left(\text{erf}\left(\frac{\sqrt{2}\Delta\,(4r+3)}{8\sigma_n}\right) - \text{erf}\left(\frac{\sqrt{2}\,\Delta(4r+1)}{8\sigma_n}\right)\right) \tag{2.37}$$

Finally, using the fact that $\sigma_n^2 = P_n$, where P_n is the noise power, (2.37) can be expressed as

$$P_e = \frac{1}{2} \sum_{r=-\infty}^{\infty} \left(\text{erf}\left(\frac{\Delta\,(4r+3)}{\sqrt{32P_n}}\right) - \text{erf}\left(\frac{\Delta(4r+1)}{\sqrt{32P_n}}\right)\right) \tag{2.38}$$

and the bit error rate is defined in terms of the probability of error as

$$BER \overset{\text{def}}{=} 100P_e \tag{2.39}$$

2.4 Optimization technique

The main objective of this section is to find the values of the parameters α, β, γ and Δ for which the parametric QIM defined in (2.9) becomes optimal for the proposed watermarking technique. The criterion for the proposed optimization technique is to guarantee the imperceptibility while maximizing the robustness under an AWGN attack. This is equivalent to fix the SWR to a value SWR$_0$, which is according to IFPI must be higher than 20 dB, while minimizing the probability of error under an AWGN attack.

First, it is worth to notice that the probability of error derived in (2.38) depends only on the quantization step Δ and does not depend on the parameters α, β and γ, that is, $P_e = P_e(\Delta)$. However, the SWR derived in (2.29) depends on α, β, γ and Δ. Thus, in order to make the SWR constant, it is sufficient to fix the following function to a constant value as

$$g(\alpha, \beta, \gamma, \Delta) = \frac{\Delta^2}{12} + (\alpha+\beta)^2 + \frac{q(2\gamma+\Delta)}{4(p+q)}\big((\Delta+2\gamma) + 4(\alpha+\beta)\big) = \text{constant} \tag{2.40}$$

This optimization problem can easily be solved using the method of Lagrange multipliers [78], where the Lagrangian can be defined as

$$\mathcal{L}(\alpha, \beta, \gamma, \Delta) = g(\alpha, \beta, \gamma, \Delta) - \frac{1}{\kappa} P_e(\Delta) \tag{2.41}$$

with κ is the Lagrange multiplier. The optimal values of the parameters α, β and γ and the quantization step Δ can then be obtained by solving the following system of equations

$$\begin{cases} \vec{\nabla}_{\alpha,\beta,\gamma,\Delta}\mathcal{L}(\alpha,\beta,\gamma,\Delta) = \vec{0} \\ \mathrm{SWR}(\alpha,\beta,\gamma,\Delta) = \mathrm{SWR}_0 \end{cases} \tag{2.42}$$

where

$$\vec{\nabla}_{\alpha,\beta,\gamma,\Delta} = \frac{\partial}{\partial\alpha}\overrightarrow{e_\alpha} + \frac{\partial}{\partial\beta}\overrightarrow{e_\beta} + \frac{\partial}{\partial\gamma}\overrightarrow{e_\gamma} + \frac{\partial}{\partial\Delta}\overrightarrow{e_\Delta} \tag{2.43}$$

is the gradient operator. Finally, the solution of (2.42) is obtained as

$$\begin{cases} \alpha \in \mathbb{R} \\ \beta = -\alpha \\ \Delta = \sqrt{\dfrac{12P_sL}{\mathrm{SWR}_0}} \\ \gamma = -\dfrac{\Delta}{2} \end{cases} \tag{2.44}$$

It is worth to notice that the optimal quantization step Δ in (2.44) depends only on P_s, L and SWR_0. Consequently, the quantization step is the same for all the frames of the same audio signal, which means that the proposed quantization operation given by (2.12) is uniform. Using the optimal solution for the quantization step Δ given by (2.44) in (2.38), the probability of error reduces to

$$P_e = \frac{1}{2}\sum_{r=-\infty}^{\infty}\left(\mathrm{erf}\left(\sqrt{\frac{3L \times \mathrm{SNR}}{8\,\mathrm{SWR}_0}}(4r+3)\right) - \mathrm{erf}\left(\sqrt{\frac{3L \times \mathrm{SNR}}{8\,\mathrm{SWR}_0}}(4r+1)\right)\right) \tag{2.45}$$

where $\mathrm{SNR} = P_s/P_n$ is the signal to noise ratio. It is seen from (2.45) that the probability of error under an AWGN attack is independent of the considered discrete transform and the coefficient index k (the insertion position used to carry the watermark bit).

In order to provide a closed-form expression for the probability of error function obtained in (2.45) as an infinite summation, we propose an overall approximation of the form

$$\widetilde{P_e} = \mathrm{erfc}\left(\frac{1}{a - be^{-\frac{c}{x^d}}}\right) \tag{2.46}$$

which is parameterized by the parameters $a, b, c,$ and d, where

$$x = \sqrt{\frac{3\,L \times \mathrm{SNR}}{8\,\mathrm{SWR}_0}} \tag{2.47}$$

It should be mentioned that we have tried many forms, but the one given by (2.46) is the best.

The maximum error between the probability of error function P_e obtained in (2.45) and its approximation $\widetilde{P}_e$ provided in (2.46) is given in terms of the parameters a, b, c and d as

$$E(a,b,c,d) = \max_{x>0} \left| P_e - \widetilde{P}_e \right| \tag{2.48}$$

We use the particle swarm optimization (PSO) to find the numerical values $a_0, b_0, c_0,$ and d_0 of the parameters a, b, c and d, respectively, that minimize the maximum error in (2.48), and thus ensure that $\widetilde{P}_e$ is a good approximation of P_e. The PSO is an iterative numerical optimization technique, which is able to find a solution that globally minimizes a given objective function [73–75], and its concept is shown in Fig. 2.3.

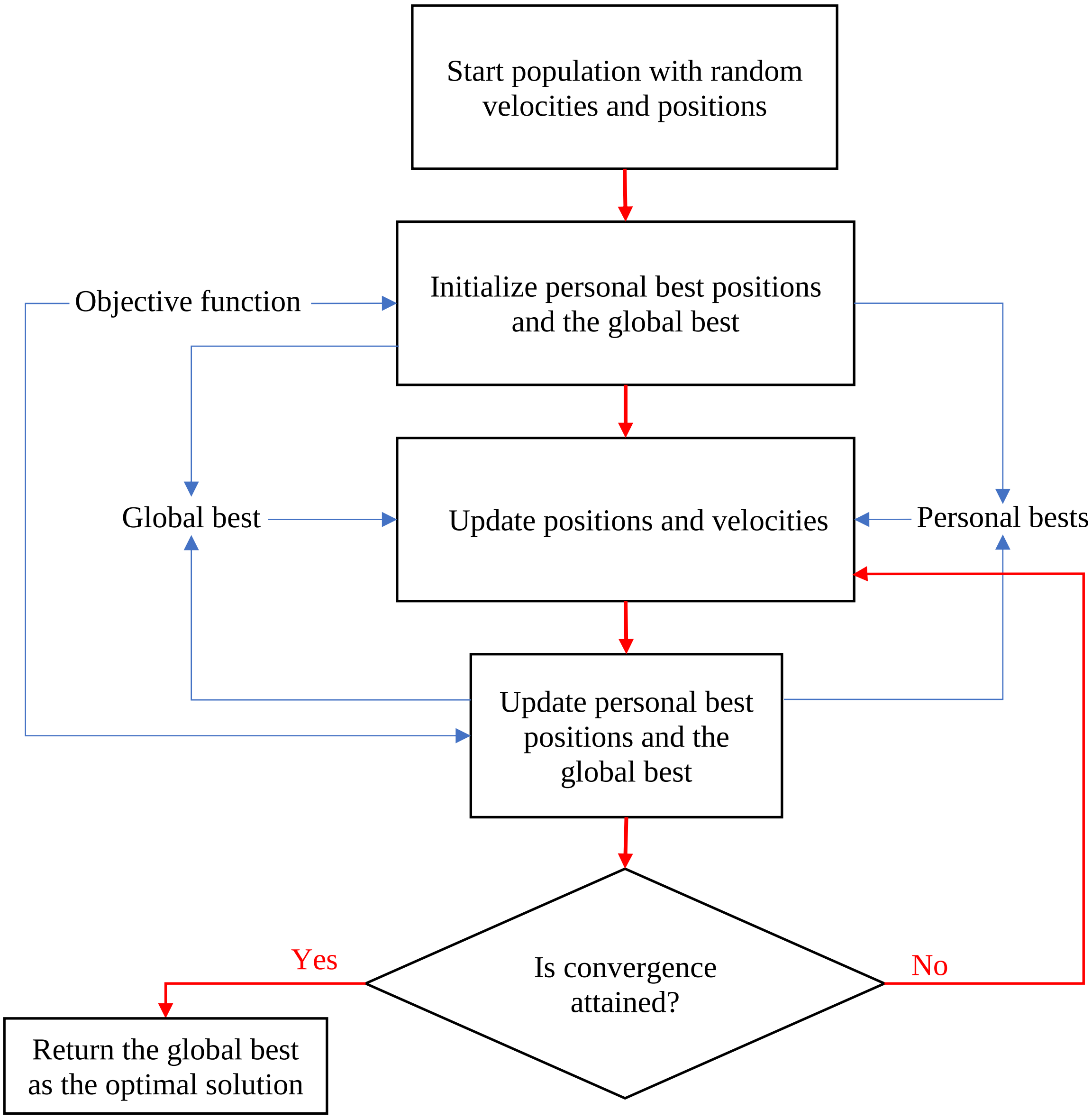

Start population with random velocities and positions
Initialize personal best positions and the global best
Objective function
Global best
Update positions and velocities
Personal bests
Update personal best positions and the global best
Yes
Is convergence attained?
No
Return the global best as the optimal solution

Fig. 2.3 Flowchart of PSO.

In our case, we use $E(a,b,c,d)$ as an objective function for PSO to find the solution (a_0,b_0,c_0,d_0) that satisfies

$$\forall\,(a,b,c,d)\in\mathbb{R}^4\colon\quad E(a_0,b_0,c_0,d_0)\leq\; E(a,b,c,d) \tag{2.49}$$

Let $\{S_1,S_2,S_3,\dots,S_N\}$ be a swarm of N particles, where the position and velocity of the k^{th} particle at the t^{th} iteration are denoted by $R_k(t)=[a_k(t),b_k(t),c_k(t),d_k(t)]$ and $v_k(t)=\left[v_k^{(a)}(t),v_k^{(b)}(t),v_k^{(c)}(t),v_k^{(d)}(t)\right]$, respectively, with $k=1,2,3,\dots,N$. The velocity and position of the k^{th} particle are calculated iteratively as

$$v_k(t+1)=\theta\cdot v_k(t)+\varsigma_1\,r_k^{(1)}(t)\,\left(\hat{R}_k(t)-R_k(t)\right)+\varsigma_2\,r_k^{(2)}(t)\,\left(G(t)-R_k(t)\right) \tag{2.50}$$

and

$$R_k(t+1)=R_k(t)+v_k(t+1) \tag{2.51}$$

respectively, where $\hat{R}_k(t)=\left[\hat{a}_k(t),\hat{b}_k(t),\hat{c}_k(t),\hat{d}_k(t)\right]$ is the personal best position in the position history of the k^{th} particle, i.e.,

$$\forall\,\tau\leq t\colon\;\; E\left(\hat{a}_k(t),\hat{b}_k(t),\hat{c}_k(t),\hat{d}_k(t)\right)\leq E\big(a_k(\tau),b_k(\tau),c_k(\tau),d_k(\tau)\big) \tag{2.52}$$

$G(t)=[a(t),b(t),c(t),d(t)]$ is the global best position in the position history of the swarm, i.e.,

$$\forall\,k\in\{1,2,3,\dots,N\}\colon\;\; E\big(a(t),b(t),c(t),d(t)\big)\leq E\left(\hat{a}_k(t),\hat{b}_k(t),\hat{c}_k(t),\hat{d}_k(t)\right), \tag{2.53}$$

$r_k^{(1)}(t)$ and $r_k^{(2)}(t)$ are random numbers uniformly distributed in the interval [0, 1], ς_1 and ς_2 are real-valued acceleration constants, which modulate the magnitude of the steps taken by the particle in the direction of its personal and global best positions, respectively, and θ is the inertia weight.

The implementation of PSO requires values for the initial conditions $R_k(0)$ and velocities $v_k(0)$, $k=1,2,3,\dots,N$, and parameters N, θ, ς_1, and ς_2. However, the choice of the required values has a large impact on the optimization performance. This problem has been the subject of many research and is still controversial [74]. In our implementation, we randomly and uniformly initialize the components of the particle positions and velocities from the intervals [0, 2] and [-2, 2], respectively, and use the values $\theta=0.9$, $\varsigma_1=\varsigma_2=0.5$, $N=1000$. The maximization over x in (2.48) should theoretically be taken over the infinite interval $]0,\infty[$. In our case, we perform the maximization over the interval $[0.001,10]$, which is sufficiently large due to the fact that P_e in (2.45) is a monotonically decreasing function and its values are less than 10^{-10} for $x>5$. We evaluate P_e for $|r|\leq 5000$ and confirm that larger values of $|r|$ lead to negligible terms having total contribution less than 10^{-7}. For the convergence criterion of PSO, we stop iterating (2.50) and (2.51) when the maximum speed of all particles is less than 10^{-8}, that is

$$\max_{k\in\{1,2,3,\dots,N\}}\left(\sqrt{\left(v_k^{(a)}(t)\right)^2+\left(v_k^{(b)}(t)\right)^2+\left(v_k^{(c)}(t)\right)^2+\left(v_k^{(d)}(t)\right)^2}\right)\le 10^{-8} \qquad (2.54)$$

Finally, the solution of the optimization problem is taken as the global best position $[a_0, b_0, c_0\ , d_0] = G(t_\mathrm{s}) = [a(t_\mathrm{s}), b(t_\mathrm{s}), c(t_\mathrm{s}), d(t_\mathrm{s})]$, where t_s is the iteration at which the convergence is attained, and the obtained numerical values are $a_0 = 2.1143$, $b_0 = 1.7831$, $c_0 = 0.4669$ and $d_0 = 2.0128$. By substituting this solution and (2.47) in (2.46), we obtain

$$\widetilde{P_e} = \operatorname{erfc}\left(\frac{1}{2.1143 - 1.7831\, e^{-\frac{0.4669}{\sqrt{\frac{3L\times\mathrm{SNR}}{8\,\mathrm{SWR}_0}}^{2.0128}}}}\right) \qquad (2.55)$$

The exact probability error P_e derived in (2.45) and its approximation $\widetilde{P_e}$ obtained in (2.55) are plotted in Fig. 2.4. In this figure, the evaluation of the exact probability function is done by taking the sum of the terms corresponding to $|r| \le 5000$, which implies a plotting precision of the order 10^{-7}. This figure shows clearly that $\widetilde{P_e}$, which has a closed-form expression, is a good overall approximation of P_e.

Using the optimal solutions obtained in (2.44) in the proposed parametric QIM defined in (2.9), a new one-parameter optimal QIM denoted by QIM$^{\alpha}$ can be defined as

$$\hat{X} = Q_\alpha(X, \Delta, w) = \left\lfloor \frac{X - w\frac{\Delta}{2} - \alpha}{\Delta} \right\rceil \Delta + w\frac{\Delta}{2} + \alpha \qquad (2.56)$$

and the corresponding watermark extraction process as

$$\hat{w} = Q_\alpha^{-1}\left(\hat{X}, \Delta\right) = \begin{cases} 1 & \text{if} \quad \frac{\Delta}{4} \le \operatorname{mod}\left(\hat{X} - \alpha, \Delta\right) < \frac{3\Delta}{4} \\ 0 & \quad\quad\quad \text{otherwise} \end{cases} \qquad (2.57)$$

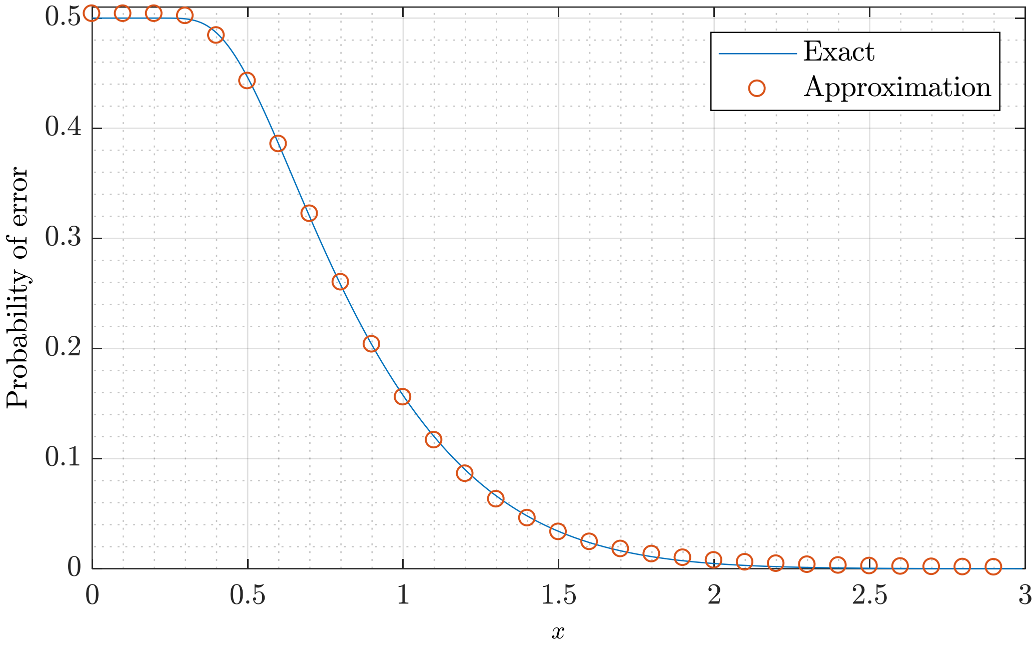


Fig. 2.4 Efficiency of the proposed probability approximation.

As mentioned earlier, the parameter α can take any real value. However, it is appropriate to choose it arbitrarily from the interval $[-\Delta/2, \Delta/2]$ and then use it as a secret key. To make this key more effective and significantly enhance the security of the proposed watermarking technique, we use for different frames different values of the parameter α, i.e., α_m for the m^{th} frame. Since the proposed method consists of inserting one bit per frame, the key becomes a vector of length equals to the number of the watermark bits. The values α_m of this vector can be selected value by value by the user, or simply by using chaotic maps, which have the ability to generate chaotic vectors of any desired length from a given initial condition known usually as the seed or simply as the key. Therefore, the coefficient $X_k^{(m)}$ of the transformed audio frame is obtained as

$$\hat{X}_k^{(m)} = \left[\frac{X_k^{(m)} - w_m \frac{\Delta}{2} - \alpha_m}{\Delta}\right]\Delta + w_m \frac{\Delta}{2} + \alpha_m \tag{2.58}$$

and the watermark bit is extracted as

$$\widehat{w}_m = \begin{cases} 1 & \text{if } \frac{\Delta}{4} \leq \operatorname{mod}\left(\hat{X}_k^{(m)} - \alpha_m, \Delta\right) < \frac{3\Delta}{4} \\ 0 & \text{otherwise} \end{cases} \tag{2.59}$$

2.5 Proposed procedure for best watermark embedding positions

We have shown in Section 2.4 that in the optimal case, the probability of error under an AWGN attack is independent of the considered orthogonal linear transform and watermark insertion position. However, it is desirable to select the transform and embedding positions that offer robustness against other signal processing manipulations. In this section, we propose a procedure to find the best interval for selecting the watermark embedding positions.

It is clear that the re-quantization operation additively introduces a uniform noise in the time domain, and for large frame length L, it converges to a Gaussian noise in the transform domain, hence its probability of error is independent of the used transform and embedding position. The resampling and lossy compression are operations that can be regarded as additive noise and filtering. Therefore, the essential operations that should be considered for choosing a transform and an embedding position are the high and low pass filtering operations.

Although the DFT gives a good interpretation of filtering attacks, the DCT can be used to give a similar interpretation and is preferable for its energy compaction capability and real-valued nature. It is clear that if the embedding position is close to the DC coefficient, the watermark survives in the case of the low pass filtering and fail to survive in the case of the high pass filtering, whereas if the embedding position is near high frequencies, then the contrary occurs. Therefore, an embedding position is selected to provide a good trade-off between the high and low pass filtering effects. Fig. 2.5 shows, for different embedding positions, the empirical maximum BER corresponding to high and low pass filtering attacks with cut-off frequencies of 100 Hz and 11025 Hz, respectively, where the sampling frequency of the used audio signals is 44100 Hz, and the frame length is 256. This empirical probability of error is obtained by taking the mean of about 100 experiments on different types of audio signals. The maximum BER (MBER) is calculated as

$$\mathrm{MBER} = \max(\mathrm{BERH}, \mathrm{BERL}) \tag{2.60}$$

where BERH and BERL denote the BERs resulted from the high and low pass filtering, respectively.

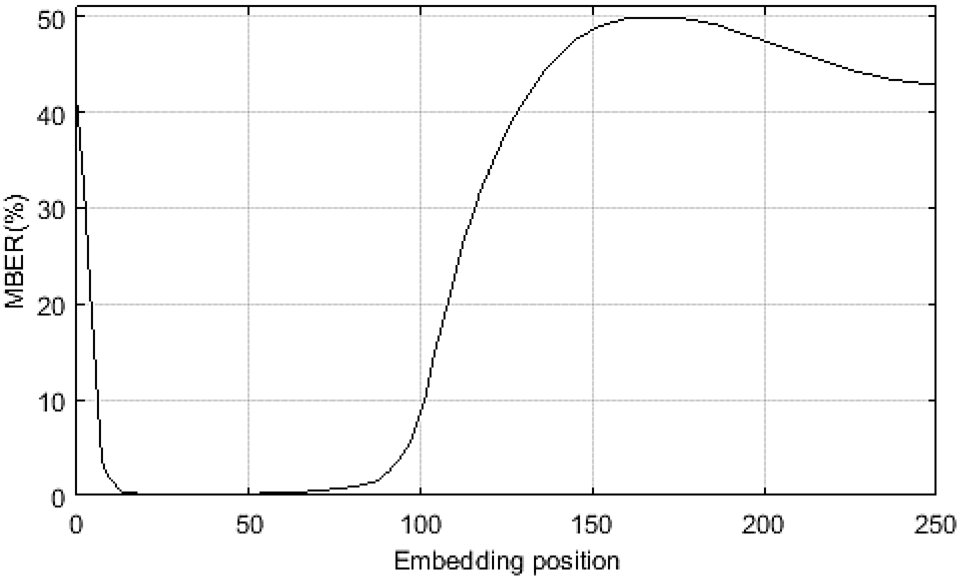


Fig. 2.5 Maximum BER resulted from low and high pass filtering attacks.

The results that are shown in Fig. 2.5 allow us to claim that the region where $k \in [12, 82]$ gives a good trade-off between the two filtering attacks. This is because the maximum BER is flatly approaching zero in this region. Further experiments that we have carried out for this purpose allow us to propose the region $k \in [0.05\ L, 0.32\ L]$ for the best selection of the insertion position k valid for any arbitrary frame length L, where the value of k must be an integer. The value of k can be chosen randomly from the proposed region and is not necessarily the same for all the frames. Therefore, the insertion position k_m is used for the frame indexed by m and the vector of length equals to the number of the watermark bits containing all the embedding positions can be coupled with the vector containing all the values of the parameter α_m to form a secret key for the proposed watermarking technique.

2.6 Fast implementation of the proposed watermarking technique

The computational complexity of a transform-based QIM watermarking technique is mainly due to the required forward and inverse transforms. Although this complexity can be reduced by using fast algorithms to compute the forward and inverse transforms, it is still a great challenge for modern systems requiring faster watermark embedding and extraction algorithms. Fortunately, the computational complexity can significantly be reduced in the proposed watermarking technique and in any other technique that modifies only one coefficient per frame in the transform domain. This can easily be achieved using the fast implementation scheme proposed below.

The main idea behind the proposed fast implementation scheme to pass from the time domain to the transform domain and vice versa is an appropriate exploitation of the fact that only one watermark bit w_m is embedded in the transformed frame $X^{(m)}$, specifically in the coefficient $X_k^{(m)}$, which is the only coefficient need to be computed using the forward transformation given by (2.11) for $r = k$, where k is the embedding position. It is clear that in this case at most L multiplications and $(L-1)$ additions are required. If the kernel $\phi_l(k)$ of the transform has some symmetry properties like in the case of the DFT or DCT, then the complexity can further be reduced. Specifically, in the case of the DCT of length that is an integral power of two, the required number of multiplications is $L/2$.

After embedding the bit w_m in the coefficient $X_k^{(m)}$ to get the coefficient $\hat{X}_k^{(m)}$ according to (2.12), the inverse transform given by (2.13) is then applied on the resulting watermarked frame $\hat{X}^{(m)}$ to obtain the watermarked frame $\hat{x}_l^{(m)}$, $l = 0, 1, \dots, L-1$, in the time domain. Since the coefficients of $\hat{X}^{(m)}$ are all identical to those of $X^{(m)}$ except the one situated in the insertion position k, the following holds

$$\hat{X}_r^{(m)} = X_r^{(m)} + \left(\hat{X}_k^{(m)} - X_r^{(m)}\right)\delta(k-r), \qquad r = 0, 1, \dots, L-1 \tag{2.61}$$

By substituting (2.61) in (2.13), the watermarked frame in the time domain can simply be computed as

$$\hat{x}_l^{(m)} = x_l^{(m)} + \left(\hat{X}_k^{(m)} - X_k^{(m)}\right)\phi_l(k), \qquad l = 0, 1, \dots, L-1 \tag{2.62}$$

It is clear that (2.62) requires at most L multiplications and $(L+1)$ additions. For the case of the DCT, the kernel is given by

$$\phi_l(r) = \Lambda(r)\cos\left(\frac{\pi\,(2l+1)\,r}{2L}\right) \tag{2.63}$$

with

$$\Lambda(r) = \begin{cases} \dfrac{1}{\sqrt{L}} & \text{if } l = 0 \\ \sqrt{\dfrac{2}{L}} & \text{otherwise} \end{cases} \tag{2.64}$$

In this case, and for values of L that are integral powers of two, the following symmetry

$$\phi_{L-1-u}(k) = (-1)^k \phi_u(k), \;\; u = 0, 1, \dots, \frac{L}{2} \tag{2.65}$$

can be exploited in (2.62) to further reduce the number of the required multiplications to only $L/2$.

Therefore, using the proposed approach described above, the total computational complexity required by the forward and inverse transformations in the proposed watermark embedding algorithm (or in any other technique that modifies only one coefficient per frame in the DCT domain) is only $L \times q$ multiplications and $2L \times q$ additions, where q is the number of watermark bits. The total computational complexity required by the forward transform in the proposed watermark extraction algorithm is only $(L/2) \times q$ multiplications and $(L-1) \times q$ additions. This is because the inverse transformation is not needed in the extraction process. Thus, the proposed approach reduces the complexity significantly compared to that required by the direct use of fast algorithms for the forward and inverse DCTs. For instance, if $L = 256$, then the proposed approach requires $256 \times q$ multiplications and $512 \times q$ additions in the embedding process, whereas the direct use of the fast algorithms reported in [79] for the forward and inverse DCTs requires $2048 \times q$ multiplications and $5634 \times q$ additions.

2.7 Experimental results

In this section, we evaluate the performance and robustness of the proposed audio watermarking technique and perform a comparison with the existing techniques reported in [13,14,20,65] that are based on QIM. For this purpose, we apply these and proposed techniques for watermarking audio signals that are stored in 16-bit signed mono waveform audio format files sampled at a frequency of 44.1 kHz. We present here only the experimental results obtained for three audio signals; namely music like signal, human speech like signal and mixture of music and human speech signal denoted by S1, S2, and S3, respectively. The evaluation in terms of bit error rate is carried out by considering the following known attacks:

- ***Re-sampling***: the watermarked audio signal is down-sampled from 44.1 kHz to 22.05, 11.025 or 5 kHz and the resulting signal is then up-sampled to 44.1 kHz.
- ***Common lossy audio compressions***: we consider the MPEG layer III compression (usually referred to as MP3), which is the most popular lossy compression in the music industry, and supported by many video file formats such as Audio Video Interleave (AVI), Matroska, MPEG-4 part 14 (MP4), MPEG-1 (MPG), Materiel eXchange Format (MXF), QuickTime, etc. We apply MP3 at several bit rates 192, 128, 96, 80 and 64 kbps. We also consider other common lossy audio compressions, namely, OPUS, advanced audio coding (AAC), and Vorbis. OPUS is used at a typical bit rate of 96 kbps. For the case of AAC, we apply two different versions, the low complexity AAC (LC-AAC), which is designed to reduce the complexity at the cost of quality, thus we apply it at a bit rate of 96 kbps, and the high-efficiency AAC (HE-AAC), which is optimized

to compress audio at low bit rates, thus we apply it at a bit rate of 56 kbps. The metric for Vorbis audio compression is quality, therefore, we perform Vorbis compression at a representative quality of 50%. The watermarked audio signal is compressed and then decompressed. Subsequently, the watermark is extracted in order to evaluate the robustness.

- ***Low pass filtering***: a low pass filter with a cut-off frequency of 11.025, 8 or 6 kHz is used to filter the watermarked audio signal.
- ***High pass filtering***: a high pass filter with a cut-off frequency of 20, 50, 100 or 4000 Hz is employed to filter the watermarked audio signal.
- ***Additive white Gaussian noise***: a white Gaussian noise is added to the watermarked audio signal with an SNR of 20, 15 or 10 dB.
- ***Re-quantization***: each of the watermarked audio samples is re-quantized from 16 bits to 8 or 4 bits.
- ***Amplitude scaling***: the watermarked audio signal is scaled by a factor of 0.8, 0.9, 1.1 or 1.2.

It is well known that there is a trade-off between the watermark payload (capacity), transparency measured by the SWR and robustness [38,67,80]. Therefore, for a fair comparison, we test the robustness for different attacks by fixing the SWR and capacity. The theoretical SWR that can be calculated using (2.29) is fixed to the desired level of 30 dB, which satisfies the IFPI recommendation and gives a good compromise between robustness and imperceptibility. As discussed earlier, the theoretical SWR can be calculated from (2.29) using $\alpha = \Delta/4\,, \beta = -\Delta/2$ and $\gamma = 0$ for the techniques reported in [13,20,65], and $\beta = -\alpha$ and $\gamma = -\Delta/2$ for the proposed technique. This desired theoretical SWR and the corresponding experimental SWRs obtained by different techniques are given in Table 2.1 for different audio signals. It is seen from this table that the theoretical and experimental values of SWR are similar. Furthermore, to confirm the validity of (2.29), we carried out many experiments with different quantization step sizes and two distinct frame lengths (i.e., $L = 256$ and $L = 512$). The theoretical and experimental results are compared in Fig. 2.6, which clearly shows the agreement of the theoretical and experimental SWR and confirms the reliability of the expression derived in (2.29) for the SWR. It should be noted that the desired theoretical SWR cannot be fixed for the technique reported in [14], since it uses a psychoacoustic model as a measure of transparency. Therefore, we use for [14] the obtained experimental SWRs that are less than 30 dB.

Table 2.1 Watermarking capacity, domain and SWR of different watermarking techniques for the signals S1, S2 and S3.

Technique	Capacity	Domain	Signal	SWR (dB)	
	Bits/sec			desired	Obtained
Proposed172	172	256-DCT	S1	30	29.79
			S2	30	29.80
			S3	30	29.92
Chen[65]		7 level DWT	S1	30	30.42
			S2	30	30.29
			3	30	30.29
Wu[20]		8 level DWT	S1	30	29.92
			S2	30	29.97
			S3	30	29.88
Proposed86	86	512-DCT	S1	30	29.97
			S2	30	29.72
			S3	30	29.98
Chen[13]		8 level DWT	S1	30	30.28
			S2	30	30.27
			S3	30	30.10
Hu[14]	85	4096-DCT	S1	/	28.58
			S2	/	25.74
			S3	/	23.50

The watermark payload (embedding capacity) is the number of watermark bits embedded per unit time (e.g. seconds) in the host digital audio signal. According to the IFPI [20], the payload should be greater than 20 bps (bits per second). Since the proposed technique embeds one bit per frame of length L, its watermark payload is obtained as

$$\mathcal{P} = \frac{f_s}{L} \tag{2.66}$$

where f_s is the sampling frequency of the host audio signal. The proposed technique can be applied for various desired watermark payloads. Since the payloads of the techniques in [20], [65] and [13] are 172, 172 and 86 bps, respectively, we fix the payload given by (2.66) of the proposed technique to 172 and then to 86. The proposed technique versions corresponding to $\mathcal{P} = 172$ and $\mathcal{P} = 86$ are denoted by Proposed172 and Proposed86, respectively. The embedding positions in these two versions are selected to be the sample numbers 47 and 94, respectively, which belong to the best region defined in Section 2.5. In [14], the authors proposed four variants of their technique, which differ in payload. Therefore, we select the variant with 85 bps of [14] and compare it to Proposed86 and the technique reported in [13].

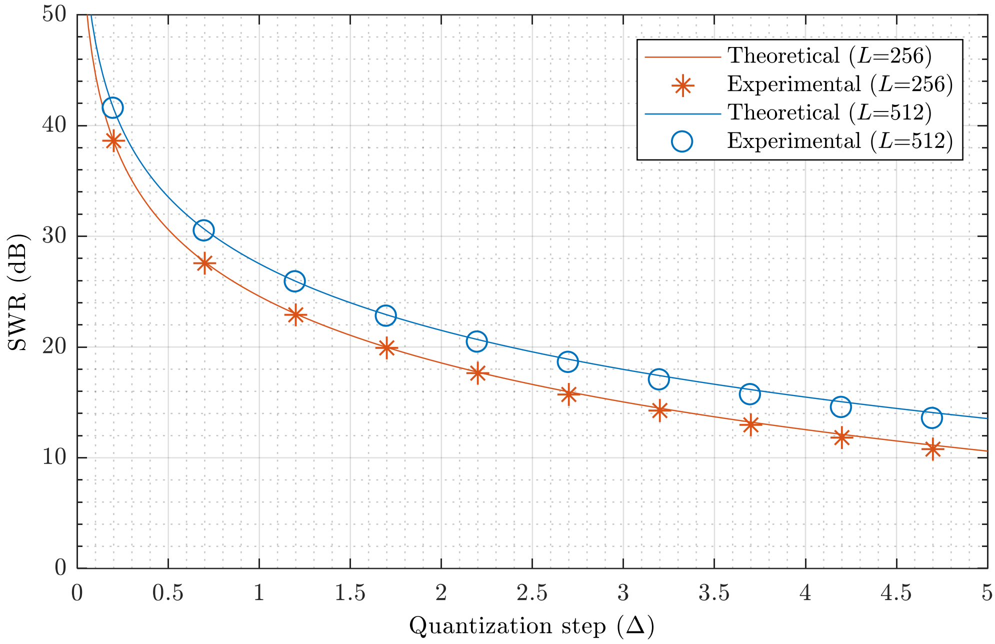


Fig. 2.6 Theoretical and experimental SWR.

The experimental results are compared in terms of the BER in Tables Table 2.2-Table 2.9. Table 2.2 and Table 2.3 show the BER obtained by different techniques in the cases of the resampling and low-pass filtering attacks. It is clear from these tables that all the techniques are equally robust to the two attacks. Table 2.4 shows clearly that the proposed technique versions are more robust against MPEG layer III compression attack than the corresponding existing techniques reported in [13,14,20,65]. Particularly, in the case of low bit rate compression, the robustness of the proposed technique becomes significantly remarkable. The same observation can be seen from Table 2.5, which illustrates the robustness against common lossy audio compressions. The reason behind the superiority of the proposed technique is the embedding region selection procedure introduced in Section 2.5, which ensures that the watermark embedding is in a significant region of the audio signal, and the optimization procedure developed in Section 2.4, which ensures robustness to the additive noise introduced by compression. The fragility of the methods reported in [13,20,65] against lossy compression attacks is due to the embedding in the lowest approximation band of the DWT that is fragile to lossy compression. The method in [14] performs better than the DWT-based techniques as it only embeds a small fraction of the watermark bits in the low frequency of the DCT coefficients, and it is worse than the proposed technique because of the use of the adaptive quantization step that is fragile to compression noise.

Table 2.2 Re-sampling BER.

Re-sampling frequency (Hz)	Signal	22050	11025	5000
Proposed172 BER (%)	S1	0	0	0
	S2	0	0	0
	S3	0	0	0
Chen[65] BER (%)	S1	0	0	0
	S2	0	0	0
	S3	0	0	0
Wu[20] BER (%)	S1	0	0	0
	S2	0	0	0
	S3	0	0	0
Proposed86 BER (%)	S1	0	0	0
	S2	0	0	0
	S3	0	0	0
Chen[13] BER (%)	S1	0	0	0
	S2	0	0	0
	S3	0	0	0
Hu[14] BER (%)	S1	0	0	0
	S2	0	0	0
	S3	0	0	0

Table 2.6 demonstrates that the proposed technique is strongly robust against the high-pass filtering attack even at exaggerated high cut-off frequencies, e.g., 4 kHz, and the existing techniques are significantly fragile to this attack even at low cut-off frequencies, e.g., 20 Hz, except the technique in [14], which is fragile only in the case of exaggerated high cut-off frequencies. The reason behind the strong robustness of the proposed technique in this attack is the proposed approach for the best watermark embedding positions introduced in Section 2.5, which is specifically designed to resist to high and low pass filtering. The weakness of the techniques in [13,20,65] is mainly due to the embedding in low frequencies of the host audio signal, which is very sensitive to the high-pass filtering. Table 2.7 shows that the proposed technique versions are more robust than the corresponding existing techniques of [13,14,20,65] in the case of AWGN attack. This can also be seen from Fig. 2.7, which illustrates the BER as a function of SNR resulting from an AWGN attack for the case of the signal S3. These results confirm the

effectiveness of the proposed optimization technique introduced in Section 2.4 to guarantee the imperceptibility while maximizing the robustness under an AWGN attack. Moreover, in order to confirm the usefulness of the theoretical analysis performed in Section 2.3, we show in Fig. 2.8 the theoretical BER obtained using (2.55) and the BER obtained experimentally by the proposed technique versions Proposed172 and Proposed86. Specifically, this figure demonstrates the correctness of (2.55) derived for the theoretical BER.

Table 2.3 Low-pass filtering BER.

Cut-off frequency (Hz)	Signal	11025	8000	6000
Proposed172 BER (%)	S1	0	0	0
	S2	0	0	0
	S3	0	0	0
Chen[65] BER (%)	S1	0	0	0
	S2	0	0	0
	S3	0	0	0
Wu[20] BER (%)	S1	0	0	0
	S2	0	0	0
	S3	0	0	0
Proposed86 BER (%)	S1	0	0	0
	S2	0	0	0
	S3	0	0	0
Chen[13] BER (%)	S1	0	0	0
	S2	0	0	0
	S3	0	0	0
Hu[14] BER (%)	S1	0	0	0
	S2	0	0	0
	S3	0	0	0

Table 2.8 shows that the proposed technique versions are more robust to re-quantization attack than the corresponding techniques reported in [13,14,20,65]. The re-quantization attack introduces a uniform noise in the time domain, which according to the central limit theorem converges to a Gaussian noise in the transform domain for sufficiently large frame length. This is confirmed by the experimental results shown in Fig. 2.9. Therefore, re-quantization in the time domain is essentially an AWGN in transform domain for which the robustness of the proposed technique is optimized. Table 2.9 confirms that the proposed technique outperforms the techniques reported in [13,14,20,65] in the case of the amplitude scaling attack, except that

the performance of the Proposed86 is almost equal to that of the technique in [14], still, the SWR obtained by [14] is less than the one obtained by the proposed technique. Moreover, the strong robustness of the technique of [14] is achieved in this attack due to the use of an adaptive quantization step that is invariant under scaling. However, this adaptive scheme has the drawback of low robustness against noise attacks and substantially increases the computational complexity of the audio watermarking technique of [14] compared to that of the proposed technique.

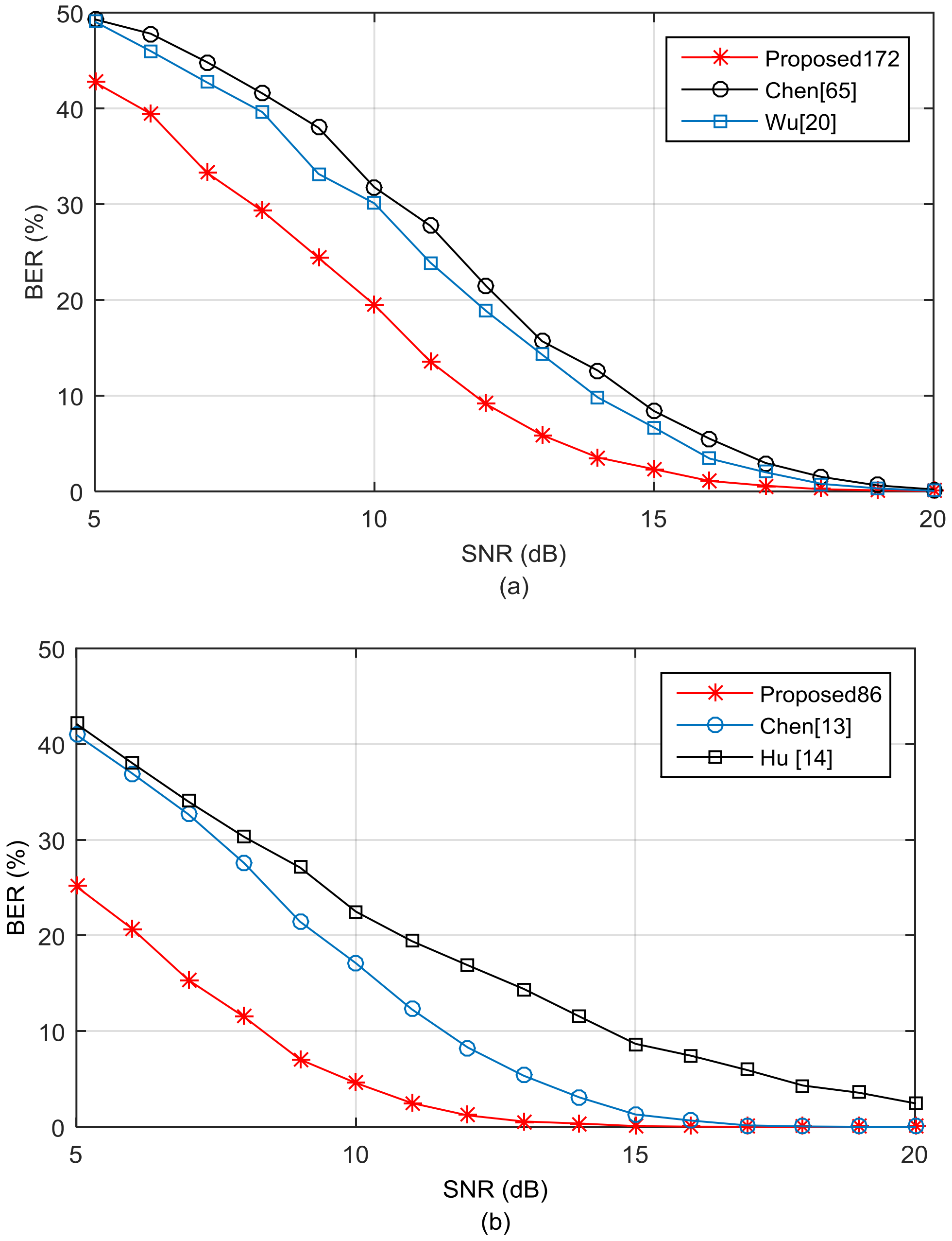


Fig. 2.7 BER of S3 under an AWGN attack, (a) techniques with a capacity of 172 bps, (b) techniques with a capacity of 86 bps.

Table 2.4 MPEG layer III compression BER.

Bit rate (kbps)	Signal	192	128	96	80	64
Proposed172 BER (%)	S1	0	0	0.5859	2.0508	4.8584
	S2	0	0	0	0	0.3418
	S3	0	0	0.1465	0.5127	2.3193
Chen[65] BER (%)	S1	0	1.1963	8.0078	17.6758	28.4668
	S2	0	0	0.8301	2.2949	7.7637
	S3	0	0.0977	2.4414	7.5195	15.7227
Wu[20] BER (%)	S1	0	1.0986	6.7139	15.0879	26.3916
	S2	0	0	0.6592	2.2461	6.4941
	S3	0	0.0488	2.0020	6.1523	13.2080
Proposed86 BER (%)	S1	0	0	0	0.1465	0.8301
	S2	0	0	0	0	0
	S3	0	0	0	0.1465	0.7080
Chen[13] BER (%)	S1	0	0	0.8057	3.4668	9.4971
	S2	0	0	0.4395	1.2695	3.4912
	S3	0	0	0.5127	2.0752	6.0059
Hu[14] BER (%)	S1	0	0	0.1953	1.2451	4.0283
	S2	0	0	0.0244	0.3662	0.6104
	S3	0	0	0	0.2197	0.8545

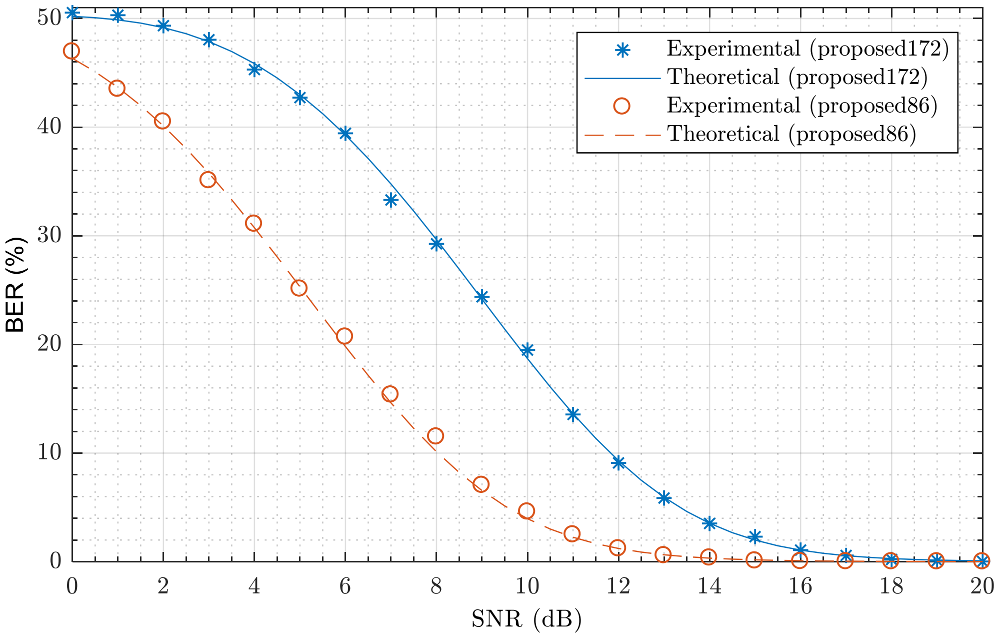


Fig. 2.8 Theoretical and experimental BER of the Proposed172 and Proposed86 versions under an AWGN attack.

Table 2.5 BER of common audio compressions at typical bitrate/quality.

Compression type	Signal	MP3 (96 kbps)	OPUS (96 kbps)	LC-AAC (96 kbps)	HE-AAC (56 kbps)	Vorbis (Quality 50/100)
proposed172 BER (%)	S1	0.5859	1.3428	0.2197	10.8398	3.1982
	S2	0	0.3906	2.6123	5.7617	1.6357
	S3	0.1465	0.5615	0.4395	9.4482	1.5381
Chen[65] BER (%)	S1	8.0078	41.7725	0.6104	35.0098	16.1621
	S2	0.8301	54.5654	5.9814	45.2637	18.0420
	S3	2.4414	39.5996	1.8066	36.7676	14.0381
Wu[20] BER (%)	S1	6.7139	50.7080	0.4639	32.5684	15.1123
	S2	0.6592	54.2236	5.5664	44.9951	16.7236
	S3	2.0020	53.4180	7.9102	35.9863	13.0615
Proposed86 BER (%)	S1	0	0.1953	0.0977	3.4668	0.5371
	S2	0	0.0488	0.9766	2.7100	0.7080
	S3	0	0.2197	0.1953	4.9316	0.4883
Chen[13] BER (%)	S1	0.8057	44.5068	0.7080	22.8516	8.9111
	S2	0.4395	52.8076	2.6367	40.4785	11.5723
	S3	0.5127	44.8242	0.4395	30.1270	6.2500
Hu[14] BER (%)	S1	0.1953	0.2441	0.7813	8.3496	0.5859
	S2	0.0244	0.3662	1.0010	8.5449	0.7080
	S3	0	0.6592	0.7080	7.3486	0.3418

In order to evaluate the security of the proposed technique, we used the logistic map to iteratively calculate α_m in (2.58) and (2.59) as

$$\alpha_{m+1} = \mu \left(\frac{1}{2} + \frac{\alpha_m}{\Delta}\right)\left(\frac{1}{2} - \frac{\alpha_m}{\Delta}\right)\Delta - \frac{\Delta}{2} \tag{2.67}$$

where the control parameter $\mu \in [3.57, 4]$ and the initial condition $\alpha_0 \in [-\Delta/2, \Delta/2]$ are employed as watermark embedding an extraction key. We confirmed by means of extensive experiments that, in the extraction process, adding a perturbation of 10^{-16} to the values of μ or α_0 used in the embedding process results in a failure of the watermark extraction process.

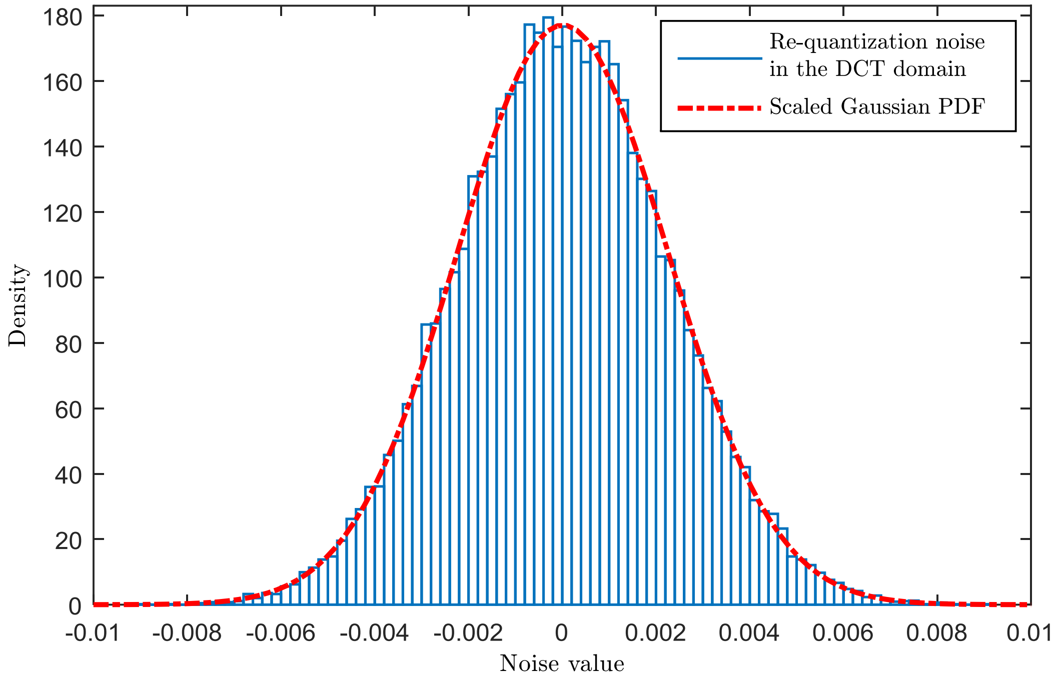


Fig. 2.9 Histogram of the re-quantization noise in the DCT domain fitted by a scaled Gaussian PDF for the frame size 256.

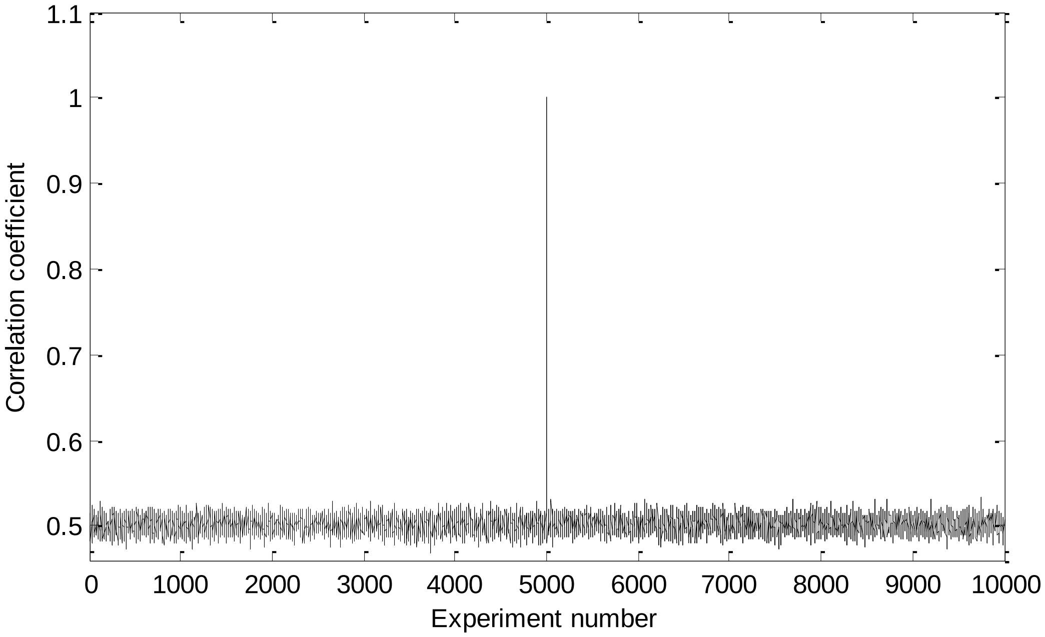


Fig. 2.10 Correlation between the extracted and embedded watermark for ten thousand experiments.

Moreover, we have embedded a watermark binary image of size 64 × 64 pixels in the signal S3 using values $\hat{\mu}$ and $\hat{\alpha}_0$ for the control parameter μ and initial condition α_0, respectively, and Fig. 2.10 shows the similarity between the embedded and extracted watermarks of ten thousand watermark extractions all performed with random values of μ and α_0 except the extraction number 5000, which is performed using identical values of μ and α_0 to those used in the embedding procedure. The similarity is measured in terms of the normalized correlation C_{NN}, which is calculated as

$$C_{NN} = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N} W_{mn}\widetilde{W}_{mn}}{\sqrt{(\sum_{m=1}^{M}\sum_{n=1}^{N} W_{mn}^2)(\sum_{m=1}^{M}\sum_{n=1}^{N} \widetilde{W}_{mn}^2)}} \tag{2.68}$$

where W_{mn} and $\widetilde{W}_{mn}, m = 1, \dots, M, n = 1, \dots, N$ are the embedded and extracted watermark bits, respectively.

Table 2.6 High-pass filtering BER.

Cut-off frequency (Hz)	Signal	20	50	100	4000
Proposed172 BER (%)	S1	0	0	0	0.9766
	S2	0	0	0	0.2930
	S3	0	0	0	0.5615
Chen[65] BER (%)	S1	49.8779	48.8037	50.1465	50.2197
	S2	50.5859	51.2451	50.1221	50.5371
	S3	49.4873	50.8057	50.1221	49.7314
Wu[20] BER (%)	S1	49.4629	49.3408	49.7559	49.4629
	S2	48.7061	50.7568	50.3174	50.6104
	S3	50.0000	49.8535	50.1953	49.7559
Proposed86 BER (%)	S1	0	0	0	0.0244
	S2	0	0	0	0
	S3	0	0	0	0
Chen[13] BER (%)	S1	49.0967	50.3662	50.0977	49.7070
	S2	51.0986	49.6094	51.4648	51.0742
	S3	49.1211	50.2197	50.7568	49.9756
Hu[14] BER (%)	S1	0	0	0	50.4832
	S2	0	0	0	49.9753
	S3	0	0	0	50.4437

Table 2.7 Additive white Gaussian noise BER.

SNR (dB)	Signal	20	15	10
Proposed172 BER (%)	S1	0	1.4648	16.8213
	S2	0	1.3916	17.0898
	S3	0	1.3672	16.0645
Chen[65] BER (%)	S1	0.1709	7.1777	33.1787
	S2	0.2197	8.4473	33.0322
	S3	0.2441	7.9102	32.7637
Wu[20] BER (%)	S1	0.1221	6.6650	30.1758
	S2	0.0732	6.5674	29.0283
	S3	0.1709	6.0303	29.5410
Proposed86 BER (%)	S1	0	0.0977	5.4199
	S2	0	0	5.2246
	S3	0	0.0488	4.5654
Chen[13] BER (%)	S1	0	1.3184	16.0889
	S2	0	1.2939	16.1289
	S3	0	1.2695	16.9189
Hu[14] BER (%)	S1	0.7324	5.1758	17.9932
	S2	2.0996	9.6191	22.9248
	S3	0.1465	1.4648	8.8867

Table 2.8 Re-quantization BER.

16 bits to	Signal	8 bits	4 bits
Proposed172 BER (%)	S1	0	0.3447
	S2	0	0.3906
	S3	0	0.1465
Chen[65] BER (%)	S1	0	14.0869
	S2	0	6.5430
	S3	0	1.7090
Wu[20] BER (%)	S1	0	11.0840
	S2	0	5.0537
	S3	0	1.2939
Proposed86 BER (%)	S1	0	0.0732
	S2	0	0.0732
	S3	0	0
Chen[13] BER (%)	S1	0	2.0508
	S2	0	4.1992
	S3	0	0.3174
Hu[14] BER (%)	S1	0	6.3721
	S2	0	9.7412
	S3	0	0.8057

Table 2.9 Amplitude scaling BER.

Scaling factor	Signal	0.8	0.9	1.1	1.2
Proposed172 BER (%)	S1	2.2949	0.2441	0.2441	2.2949
	S2	3.7354	0.0488	0.0488	3.7354
	S3	3.3203	0.2197	0.2197	3.3203
Chen[65] BER (%)	S1	44.3604	38.5986	38.5986	51.2939
	S2	47.6563	40.7227	40.7227	51.1963
	S3	39.5508	34.5215	34.5215	45.3613
Wu[20] BER (%)	S1	44.3848	37.6221	37.6221	51.1475
	S2	46.9727	39.6484	39.6484	51.2939
	S3	38.9893	33.5938	33.5938	44.5313
Proposed86 BER (%)	S1	0.4883	0	0	0.4883
	S2	1.5137	0	0	1.5137
	S3	1.4404	0.0244	0.0244	1.4404
Chen[13] BER (%)	S1	25.7813	16.9434	16.9434	32.2998
	S2	45.1172	37.3535	37.3535	45.1172
	S3	30.0293	15.3564	15.3564	38.9648
Hu[14] BER (%)	S1	0	0	0	0
	S2	0	0	0	0
	S3	0	0	0	0

2.8 Conclusion

In this chapter, a new blind audio watermarking technique has been proposed in the transform domain by introducing a parametric QIM. We have shown that most of the existing QIM realizations are special cases of the proposed parametric QIM. In order to find the optimal values for the parameters of the proposed parametric QIM that ensure the imperceptibility while maximizing the robustness of the proposed watermarking technique under an AWGN attack, we have derived theoretical expressions for the signal to watermark ratio and probability of error and used them in an optimization technique based on the method of Lagrange multipliers. This optimization technique has led to a parametric QIM that is optimal for the proposed watermarking technique. Furthermore, in order to provide a good trade-off between the effects of high and low pass filtering attacks for the proposed watermarking technique, we have devised an efficient procedure to find the interval for the best selection of the watermark embedding positions in the DCT domain. In addition, to make the proposed watermarking technique more attractive, we have developed a very efficient algorithm for its fast implementation. Another interesting property of the proposed technique is the fact that its watermarking secret key can be constituted of the parameters of the optimal parametric QIM coupled with the embedding positions. All the experiments that we have carried out confirm the usefulness of the theoretical analysis presented in this chapter and clearly show that the proposed technique outperforms the existing QIM-based audio watermarking techniques in most known attacks, specifically in AWGN, re-quantization, high pass filtering and common lossy compressions, and has performance similar to that of the existing QIM-based techniques in other attacks. Moreover, the proposed technique substantially reduces the computational complexity compared to the existing techniques. The proposed method does not provide excellent results in the case of amplitude scaling attack. Therefore, it is worth to investigate the maximization of the performance in this attack. This issue is considered in the next chapter.

Chapter 3

Proposed QIM-based technique for robust blind and semi-blind audio watermarking

3.1 Introduction

The conventional uniform QIM severely suffers from lack of robustness against gain attacks [15,81], which are caused by amplitude scaling of the watermarked signal samples. These attacks may not affect the signal perceptual quality while causing a high bit error rate in the extracted watermark. Many solutions have been proposed in the literature to solve the gain attacks problem of QIM, such as the rational dither modulation (RDM) [82], logarithmic QIM (LQIM) [77], adaptive quantization approach [14,49,71], and angular QIM (AQIM) [83]. The latter has been enhanced by the absolute AQIM (AAQIM) [84] and its quantization distortion has been optimized by the improved AQIM (IAQIM) [15]. However, the RDM has the drawback of high peak to average power ratio (PAPR), the LQIM only resists partially gain attacks, and the adaptive quantization approach has the deficiency of quantization steps recovery from an attacked signal that yields an additional error in the watermark extraction process. In addition, the adaptive quantization approach, LQIM, and AQIM have the disadvantage of low robustness to additive noise attacks compared to the conventional QIM. Due to the above interesting advantages of QIM, which is unfortunately fragile to gain attacks, and since the existing QIM-based solutions for gain attacks are vulnerable to additive noise, it is highly desirable to design a single QIM-based technique for blind and semi-blind audio watermarking that is robust to both additive noise and gain attacks.

In this chapter, we propose an efficient QIM-based technique for robust blind and semi-blind audio watermarking in the DWT domain [29]. It consists of embedding the watermark bits in the frames of the lowest 4-level DWT approximation band of the host audio signal. Moreover, we introduce a local minimum distortion criterion to select appropriate coefficients to carry the watermark bits. For each frame, the coefficient that carries the watermark bit is selected from a subset of the frame according to the local minimum distortion criterion. For the semi-blind operation mode of the proposed technique, the size of this selection subset is controlled by the desired size of side information, which can have many possible sizes. Furthermore, we devise an expression for the quantization step by suitably exploiting the characteristics of the DWT approximation and detail bands of the host audio signal. This quantization step is nearly invariant under an AWGN attack and renders the extraction process of the proposed technique invariant under an amplitude scaling attack. Thus, making the proposed technique robust to gain attacks while maintaining high robustness to additive noise attacks. For the embedding process, we use the dither modulation (DM) embedding method, which is a low complexity realization of QIM that can achieve good rate-distortion-robustness trade-offs [80]. In order to secure the proposed technique, we dither the quantization function using pseudo-random dithers and scramble the lowest 4-level DWT approximation band of the host audio signal using pseudo-random permutations (PRPs). The side information in semi-blind watermarking must usually be transmitted over a high-fidelity channel [18]. In order to avoid the use of an extra channel, we develop an efficient side information recovery procedure that can recover the side information entirely from a scaled or an attack-free watermarked signal, and partially from a watermarked signal that is exposed to other attacks. In addition, we derive the theoretical expressions for the watermarking distortion, probability of error under AWGN attack, and probability of an error recovery of the proposed side information recovery procedure. Moreover, we theoretically demonstrate that the proposed technique is highly robust to gain attacks while maintaining the known high robustness to additive noise attacks of QIM. Finally, we conduct various experiments to verify the consistency between the theoretical and empirical results, evaluate the security of the proposed technique and assess its robustness against content preserving attacks and manipulations such as resampling, low-pass filtering, re-quantization, echo addition, amplitude scaling, uniform multiplicative noise, AWGN, MPEG layer III lossy compression, etc. The experimental results show that the proposed technique outperforms the conventional QIM-based techniques reported in [13,18], the QIM-based solutions to gain attacks introduced in [14,15], the blind techniques proposed in [13–15], and the semi-blind techniques presented in [12,18].

3.2 A brief review of the DWT

The DWT is a transform that is capable of giving time-frequency representation of any given signal [85]. It decomposes a given audio signal into an approximation and detail bands denoted by A1 and D1, respectively, whereas the inverse DWT (IDWT) aims at perfectly reconstructing the given audio signal from the bands A1 and D1. In order to achieve higher decomposing levels, the DWT can be applied successively on any resulting approximation band as shown in Fig. 3.1, which illustrates the 4-level DWT decomposition and reconstruction concepts that are exploited in the subsequent sections.

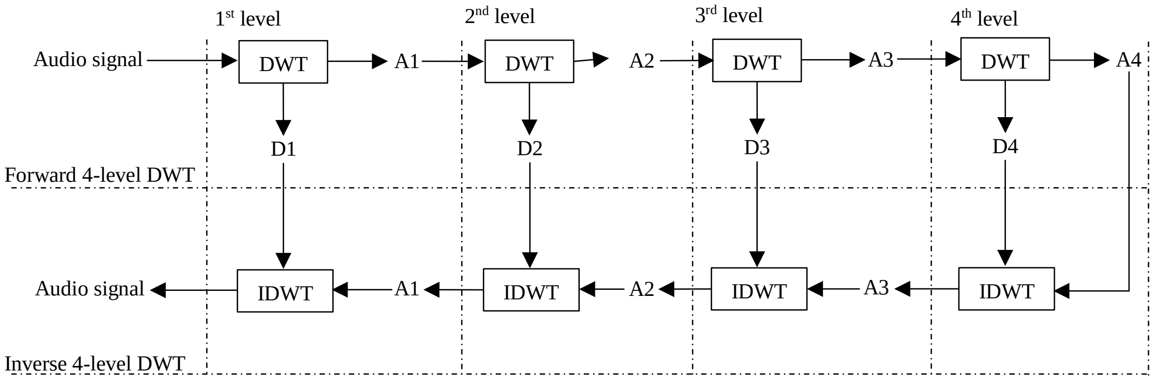


Fig. 3.1 Forward 4-level DWT and its inverse.

3.3 Proposed technique

In this section, we propose a quantization-based technique for robust blind and semi-blind audio watermarking in the DWT domain. Without loss of generality, we use the Haar DWT for its low computational complexity. The embedding and extraction procedures of the proposed audio watermarking technique are described in detail in the following subsections.

3.3.1 Watermark embedding

The proposed watermark embedding procedure is depicted in Fig. 3.2, where the inputs of the embedder are the host audio signal, a scrambling key K_S, a dithering key K_D, a non-negative integer q that controls the size of the side information, and the watermark bits to be embedded denoted by w_m, $m = 1, 2, 3, \dots, L_w$, with L_w being the total number of the watermark bits. The embedding procedure consists of the following steps:

Step 1: Apply the 4-level Haar DWT on the host audio signal as shown in Fig. 3.1 to obtain the 4th level approximation band A4 and detail bands denoted by $\mathrm{D1}, \mathrm{D2}, \mathrm{D3}$, and $\mathrm{D4}$.

Step 2: Scramble the approximation band A4 using pseudo-random permutations (PRPs) and the scrambling key K_S, and then split the resulting scrambled band into two equally-sized sequences A and B, and calculate the quantization step Δ as

$$\Delta = \delta \cdot \sqrt{|\sigma_B^2 - \sigma_D^2|} \tag{3.1}$$

where δ is a constant that controls the watermark transparency and σ_B^2 and σ_D^2 are the variances of the sequence B and the detail band D4, respectively. It should be mentioned that the variance σ_Y^2 of any zero-mean sequence $Y = [Y_1, Y_2, , \dots, Y_L]$ is calculated as

$$\sigma_Y^2 = \frac{1}{L}\sum_{k=1}^{L} Y_k^2 \tag{3.2}$$

Step 3: Segment the sequence A into non-overlapping frames $F^{(m)}, m = 1, 2, 3, \dots, L_w$, of length L_F, and generate the dither coefficients $D^{(m)} \in [0, 1]$, $m = 1, 2, 3, \dots, L_w$ using a pseudo-random noise generator (PRNG) and the dithering key K_D.

Step 4: For all $m \in \{1, 2, 3, \dots, L_w\}$, embed the watermark bit w_m in the frame $F^{(m)}$ using the quantization step Δ and the dither coefficient $D^{(m)}$ according to a local minimum distortion QIM (LMD-QIM) targeting the embedding of the watermark bit w_m in the coefficient of $F^{(m)}$ that gives the lowest embedding distortion among the first $N = 2^q$ coefficients of $F^{(m)}$ as follows:

1. Form the quantized sequence $u^{(m)} = \left[u_1^{(m)}, u_2^{(m)}, u_3^{(m)}, \dots \dots \dots, u_N^{(m)}\right]$ of length N whose coefficients are given by

$$u_k^{(m)} = Q_\Delta\left(F_k^{(m)}, w_m, D^{(m)}\right), \; k = 1, 2, 3, \dots, N \tag{3.3}$$

where $F_k^{(m)}$ is the kth coefficient of the frame $F^{(m)}$, and $Q_\Delta\left(F_k^{(m)}, w_m, D^{(m)}\right)$ is a DM embedding function given by

$$Q_\Delta\left(F_k^{(m)}, w_m, D^{(m)}\right) = \left(\left[\frac{F_k^{(m)}}{\Delta} - \frac{w_m}{2} + D^{(m)}\right] + \frac{w_m}{2} - D^{(m)}\right)\Delta \tag{3.4}$$

with $[\cdot]$ denoting the rounding operation. It should be noted that we use a DM embedding function due to its good rate-distortion-robustness trade-offs. Moreover, it has been shown in the previous chapter that the form given in (3.4) minimizes the probability of error under AWGN attack at any transparency, regardless of the embedding domain.

2. Calculate the absolute error sequence $\varepsilon^{(m)} = \left[\varepsilon_1^{(m)}, \varepsilon_2^{(m)}, \varepsilon_3^{(m)}, \dots, \varepsilon_N^{(m)}\right]$ as

$$\varepsilon_k^{(m)} = \left|F_k^{(m)} - u_k^{(m)}\right|, \qquad k = 1, 2, 3, \dots, N \tag{3.5}$$

and find the index S_m of the smallest component of the absolute error sequence $\varepsilon^{(m)}$ as

$$S_m = \underset{k \in \{1,2,3,\dots,N\}}{\operatorname{argmin}} \left\{\varepsilon_k^{(m)}\right\} \tag{3.6}$$

and then construct the watermarked frame $F'^{(m)}$, whose coefficients are given by

$$F'^{(m)}_r = \begin{cases} F_r^{(m)} & \text{if } \ r \neq S_m \\ u_r^{(m)} & \text{if } \ r = S_m \end{cases} \qquad r = 1, 2, 3 \dots, L_F \tag{3.7}$$

Step 5: (i) Concatenate the watermarked frames $F'^{(m)},\ m = \ 1, 2, 3, \dots, L_W$ to form the watermarked sequence A', and (ii) descramble the result of concatenating A' and B using the scrambling key K_S to form the watermarked 4th level approximation band, which is then used together with the detail bands D1, D2, D3 and D4 by the 4-level inverse Haar DWT to obtain the watermarked audio signal.

It is worth to notice that the value of q is restricted to be in the interval $[0, \log_2(L_F)]$. For $q = 0$, the indices $S_m = 1, \forall\, m \in \{1, 2, 3, \dots, L_W\}$, which do not need to be considered as a side information and thus the watermark extraction can be done blindly, whereas for $q \geq 1$, the sequence $S = \left[S_1, S_2, \dots, S_{L_W}\right]$ must be transmitted to the extractor as a side information and thus the watermark extraction is performed semi-blindly. In the latter, each index S_m requires exactly q bits, and hence the total size of the side information is $q \times L_W$ bits.

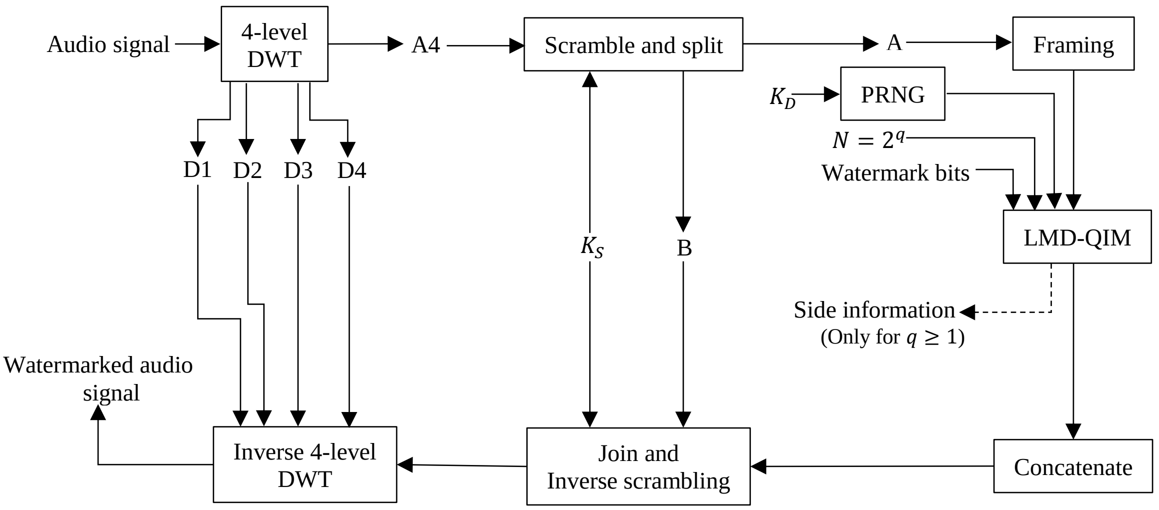


Fig. 3.2 Block diagram of the proposed watermark embedding procedure.

3.3.2 Watermark extraction

The proposed extraction procedure is shown in Fig. 3.3, where the inputs of the extractor are the watermarked audio signal, scrambling key K_S, dithering key K_D, and side information $S = [S_1, S_2, \dots, S_{L_w}]$ when $q \geq 1$. The extraction is performed according to the following steps:

Step 1: Apply the 4-level Haar DWT on the watermarked audio signal in order to obtain the 4th level detail and approximation bands denoted by D4′ and A4′, respectively.

Step 2: Scramble the approximation band A4′ using the PRPs and scrambling key K_S employed in the embedding procedure, and then split the resulting scrambled band into two equally-sized sequences A' and B', and calculate the quantization step Δ' using the variances σ'^2_B and σ'^2_D of the sequence B' and detail band D4′, respectively, as

$$\Delta' = \delta \cdot \sqrt{|\sigma'^2_B - \sigma'^2_D|} \tag{3.8}$$

Step 3: Segment the sequence A' into non-overlapping frames $F'^{(m)}, m = 1, 2, 3, \dots, L_w$, of length L_F, and generate the dither coefficients $D^{(m)} \in [0, 1]$, $m = 1, 2, 3, \dots, L_w$, using the PRNG and dithering key K_D employed in the embedding procedure.

Step 4: $\forall\, m \in \{1, 2, 3, \dots, L_w\}$, extract the watermark bit w'_m from the coefficient $F'^{(m)}_{S_m}$ of the frame $F'^{(m)}$ using the dither coefficient $D^{(m)}$ and index S_m according to the minimum-distance decoder as

$$w'_m = \underset{w=\{0,1\}}{\operatorname{argmin}} \left| F'^{(m)}_{S_m} - Q_{\Delta'}\left(F'^{(m)}_{S_m}, w, D^{(m)}\right) \right| \tag{3.9}$$

The value of the index S_m in (9) depends on the operation mode of the proposed technique. It is unity in the blind mode, whereas in the semi-blind mode it must be provided to the extractor as side information.

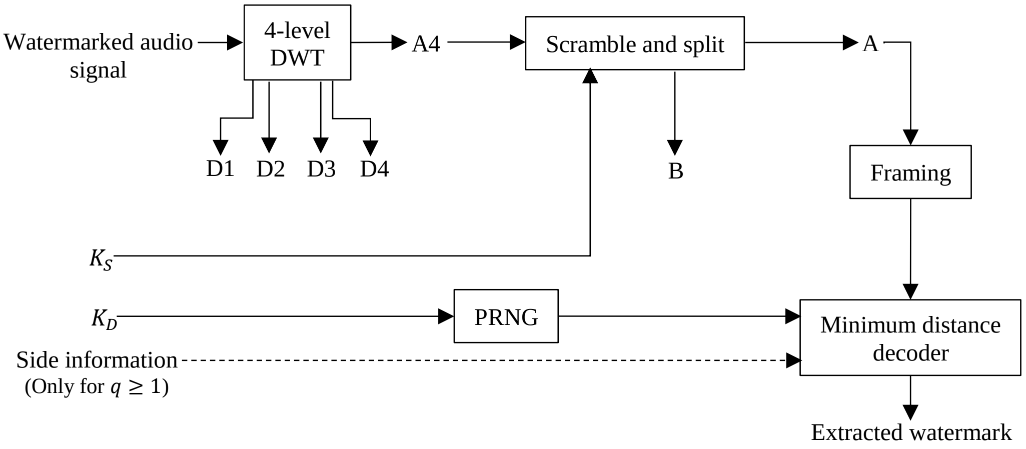


Fig. 3.3 Block diagram of the proposed watermark extraction procedure.

3.4 Performance analysis of the proposed technique

In this section, we derive theoretical expressions of the embedding distortion and probability of error for the proposed audio watermarking technique under an AWGN attack. Furthermore, we theoretically demonstrate that the proposed technique is robust to gain attacks while maintaining high robustness to additive noise attacks.

3.4.1 Embedding distortion

It can be seen from (3.5), (3.6) and (3.7) that the absolute embedding error caused by the proposed embedding procedure in the m^{th} frame is

$$\xi^{(m)} = \min_{k \in \{1,2,3,\ldots,N\}} \left\{ \varepsilon_k^{(m)} \right\} \tag{3.10}$$

Furthermore, the frame coefficients $F_k^{(m)}$, $k = 1, 2, 3, \ldots, N$, in the embedding procedure are obtained from the approximation band of the host audio signal, and thus are independent and identically distributed random variables [12,15,86]. Therefore, the absolute errors $\varepsilon_k^{(m)}, k = 1, \ldots, N$, which are obtained in (3.5) using the frame coefficients and their corresponding quantized coefficients are independent and identically distributed random variables. Consequently, the absolute embedding error $\xi^{(m)}$ is the minimum of N independent and identically distributed. random variables, and its cumulative density function (CDF) is given by [87]

$$F_{\xi}\left(\xi^{(m)}\right) = 1 - \left(\int_{\xi^{(m)}}^{\infty} f_{\varepsilon}(\varepsilon)\, d\varepsilon\right)^{N} \tag{3.11}$$

where $f_{\varepsilon}(\cdot)$ is the probability density function (PDF) of $\varepsilon_k^{(m)}, k = 1,2,3,\dots,N$. The PDF of the absolute error $\xi^{(m)}$ can be obtained as

$$f_{\xi}\left(\xi^{(m)}\right) \overset{\text{def}}{=} \frac{d}{d\xi^{(m)}} F_{\xi}\left(\xi^{(m)}\right) = N\, f_{\varepsilon}\left(\xi^{(m)}\right) \left(\int_{\xi^{(m)}}^{\infty} f_{\varepsilon}(\varepsilon)\, d\varepsilon\right)^{N-1} \tag{3.12}$$

and the average watermarking distortion σ_W^2 is given by [22]

$$\sigma_W^2 \overset{\text{def}}{=} \frac{1}{2^5 L_{\mathrm{F}}} E\left\{\left(\xi^{(m)}\right)^2\right\} = \frac{1}{2^5 L_{\mathrm{F}}} \int_0^{\infty} \left(\xi^{(m)}\right)^2 f_{\xi}\left(\xi^{(m)}\right) d\xi^{(m)} \tag{3.13}$$

where $E\{\cdot\}$ stands for the expectation operator, and the factor 2^5 is introduced by the five divisions performed on the host audio signal in the embedding process, i.e., four divisions performed by the 4-level DWT, and one division performed by splitting the 4th level approximation band A4. Replacing (3.12) in (3.13), gives

$$\sigma_W^2 = \frac{N}{2^5 L_{\mathrm{F}}} \int_0^{\infty} \left(\xi^{(m)}\right)^2 f_{\varepsilon}\left(\xi^{(m)}\right) \left(\int_{\xi^{(m)}}^{\infty} f_{\varepsilon}(\varepsilon)\, d\varepsilon\right)^{N-1} d\xi^{(m)} \tag{3.14}$$

The imperceptibility requirement imposes the usage of a small quantization step, which implies that $\varepsilon_k^{(m)}, k = 1,2,3,\dots,N$ are uniformly distributed in the interval $[0, \Delta/2]$ [22], i.e.,

$$f_{\varepsilon}(\varepsilon) = \begin{cases} \dfrac{2}{\Delta} & 0 \le \varepsilon \le \dfrac{\Delta}{2} \\ 0 & \text{otherwise} \end{cases} \tag{3.15}$$

By substituting $f_{\varepsilon}(\varepsilon)$ from (3.15) in (3.14), we obtain

$$\sigma_W^2 = \frac{N}{2^5 L_{\mathrm{F}}} \left(\frac{2}{\Delta}\right)^{N} \int_0^{\frac{\Delta}{2}} \left(\xi^{(m)}\right)^2 \left(\frac{\Delta}{2} - \xi^{(m)}\right)^{N-1} d\xi^{(m)} \tag{3.16}$$

By calculating (3.16), we obtain a closed-form expression of the watermarking distortion for the proposed audio watermarking technique as

$$\sigma_W^2 = \frac{\Delta^2}{2^6\,(N+1)\,(N+2)\,L_F} \tag{3.17}$$

3.4.2 *Probability of error under an AWGN attack*

In the extraction procedure, the watermark bits are extracted using the minimum distance decoder. Therefore, under an additive white Gaussian noise attack with a noise n of zero mean and variance σ_n^2 (i.e., $n \sim \mathcal{N}(0, \sigma_n^2)$), the probability of error extraction is given by [20]

$$P_e \approx \Pr\left(|n| > \frac{\Delta}{4}\right) = 1 - \frac{2}{\sqrt{2\pi\sigma_n^2}} \int_0^{\frac{\Delta}{4}} e^{-\frac{n^2}{2\sigma_n^2}}\, dn \tag{3.18}$$

where $\Pr(A)$ stands for the probability of occurrence of a given event A. By simplifying (3.18), we obtain

$$P_e \approx \operatorname{erfc}\left(\frac{\Delta}{4\sqrt{2\sigma_n^2}}\right) \tag{3.19}$$

where $\operatorname{erfc}(\cdot)$ refers to the complementary error function defined as

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2}\, dt$$

By obtaining the quantization step Δ from (3.17) and replacing it in (3.19), the probability of error can be expressed in terms of the watermarking distortion as

$$P_e \approx \operatorname{erfc}\left(\sqrt{2\,(N+1)\,(N+2)\,L_F\,\frac{\sigma_W^2}{\sigma_n^2}}\right) \tag{3.20}$$

The watermark to noise ratio (WNR) is defined as the ratio between the watermarking distortion σ_W^2 and the noise power σ_n^2 as

$$\text{WNR} \overset{\text{def}}{=} \frac{\sigma_W^2}{\sigma_n^2} \tag{3.21}$$

Therefore, by using (3.21) in (3.20), we provide a closed-form expression of the probability of error for the proposed audio watermarking technique as

$$P_e \approx \operatorname{erfc}\left(\sqrt{2\,(N+1)\,(N+2)\,L_F\,\text{WNR}}\right) \tag{3.22}$$

3.4.3 *Robustness against amplitude scaling and AWGN*

On one hand, if the watermarked signal is scaled by a factor β, then its DWT coefficients are scaled by the same factor due to the linearity of the Haar DWT [88]. On the other hand, any scaling of the DWT coefficients when used in (3.2) leads to a similar scaling of the quantization step given by (3.8) used in the extraction process. Due to this scaling property of the proposed expression for the quantization step and since

$$\underset{w=\{0,1\}}{\operatorname{argmin}}\left|\beta\, {F'}_{S_m}^{(m)}-Q_{\beta\Delta'}\left(\beta\, {F'}_{S_m}^{(m)},w,D^{(m)}\right)\right|=\underset{w=\{0,1\}}{\operatorname{argmin}}\left|\beta\, {F'}_{S_m}^{(m)}-\beta Q_{\Delta'}\left({F'}_{S_m}^{(m)},w,D^{(m)}\right)\right|$$
$$=\underset{w=\{0,1\}}{\operatorname{argmin}}\left|{F'}_{S_m}^{(m)}-Q_{\Delta'}\left({F'}_{S_m}^{(m)},w,D^{(m)}\right)\right| \qquad (3.23)$$

the watermark extraction performed by the minimum distance decoder in (3.9) is invariant under amplitude scaling of the watermarked signal. Furthermore, if the watermarked signal undergoes an AWGN attack $n\sim\mathcal{N}(0,\sigma_n^2)$, then the variances ${\sigma'}_B^2$ and ${\sigma'}_D^2$ of the noisy sequence B' and the noisy 4th level detail band $D4'$, respectively, can be approximated in terms of σ_B^2, σ_D^2 and σ_n^2 as ${\sigma'}_B^2\approx\sigma_B^2+\sigma_n^2$ and ${\sigma'}_D^2\approx\sigma_D^2+\sigma_n^2$, and hence ${\sigma'}_B^2-{\sigma'}_D^2\approx\sigma_B^2-\sigma_D^2$, which implies that the quantization step of the proposed technique is near invariant under an AWGN attack, and hence the quantization step used in the extraction process is very close to that used in the embedding process ($\Delta'\approx\Delta$). Therefore, the proposed quantization step does not introduce additional errors in the extraction process in the case of the AWGN attack, in contrast, the adaptive quantization step used in [14,49,71] introduces significant errors due to the fact that its value changes under AWGN attack.

To show more rigorously the near invariance property of the proposed quantization step, we define the auxiliary variables $\zeta=\sigma_B^2-\sigma_D^2$ and $\zeta'={\sigma'}_B^2-{\sigma'}_D^2$. We denote the coefficients of the sequence B by $b_k, k=1,2,3,\dots,L_B$, and those of the 4th level detail band $D4$ by $d_l, l=1,2,3,\dots,L_D$, where L_B and L_D are the lengths of B and $D4$, respectively. In a similar way, we denote the coefficients of the noisy sequence B' by ${b'}_k, k=1,2,3,\dots,L_B$ and those of the noisy 4th level detail band $D4'$ by ${d'}_l; l=1,2,3,\dots,L_D$. Then, ζ and ζ' can be expressed as

$$\zeta\overset{\text{def}}{=}\frac{1}{L_B}\sum_{k=1}^{L_B}b_k^2-\frac{1}{L_D}\sum_{l=1}^{L_D}d_l^2 \qquad (3.24)$$

and

$$\zeta'\overset{\text{def}}{=}\frac{1}{L_B}\sum_{k=1}^{L_B}{b'}_k^2-\frac{1}{L_D}\sum_{l=1}^{L_D}{d'}_l^2=\frac{1}{L_B}\sum_{k=1}^{L_B}\left(b_k+n_k^{(b)}\right)^2-\frac{1}{L_D}\sum_{l=1}^{L_D}\left(d_l+n_l^{(d)}\right)^2 \qquad (3.25)$$

respectively, where $n_k^{(b)}$ and $n_l^{(d)}$ are the AWGN samples. By using (3.24) in (3.25), we obtain

$$\zeta' = \zeta + \frac{1}{L_B}\sum_{k=1}^{L_B}\left(n_k^{(b)} + 2\, b_k\right) n_k^{(b)} - \frac{1}{L_D}\sum_{l=1}^{L_D}\left(n_l^{(d)} + 2\, d_l\right) n_l^{(d)} \tag{3.26}$$

The mean μ'_ζ of the auxiliary variable ζ' is given by

$$\mu'_\zeta \overset{\text{def}}{=} E\left\{\zeta + \frac{1}{L_B}\sum_{k=1}^{L_B}\left(n_k^{(b)} + 2\, b_k\right) n_k^{(b)} - \frac{1}{L_D}\sum_{l=1}^{L_D}\left(n_l^{(d)} + 2\, d_l\right) n_l^{(d)}\right\} \tag{3.27}$$

By expanding (3.27) and using the fact that the considered noises are zero-mean AWGN, i.e., $E\left\{n_k^{(b)}\right\} = E\left\{n_l^{(d)}\right\} = 0$, we obtain $\mu'_\zeta = \zeta$, which we use to find the variance ${\sigma'_\zeta}^2$ of ζ' as

$${\sigma'_\zeta}^2 \overset{\text{def}}{=} E\left\{\left(\zeta' - \mu'_\zeta\right)^2\right\} = E\{(\zeta' - \zeta)^2\} \tag{3.28}$$

The substitution of (3.26) in (3.28) yields

$$\begin{aligned} {\sigma'_\zeta}^2 = E\Bigg\{&\left(\frac{1}{L_B}\sum_{k=1}^{L_B}\left(n_k^{(b)} + 2\, b_k\right) n_k^{(b)} - \frac{1}{L_D}\sum_{l=1}^{L_D}\left(n_l^{(d)} + 2\, d_l\right) n_l^{(d)}\right) \\ &\times \left(\frac{1}{L_B}\sum_{r=1}^{L_B}\left(n_r^{(b)} + 2\, b_r\right) n_r^{(b)} - \frac{1}{L_D}\sum_{s=1}^{L_D}\left(n_s^{(d)} + 2\, d_s\right) n_s^{(d)}\right)\Bigg\} \end{aligned} \tag{3.29}$$

By using the following properties of the zero-mean AWGN

- $E\left\{\left(n_k^{(b)}\right)^p \left(n_l^{(d)}\right)^m\right\} = E\left\{\left(n_k^{(b)}\right)^p\right\} \cdot E\left\{\left(n_l^{(d)}\right)^m\right\}$
- $E\left\{\left(n_k^{(b)}\right)^p \left(n_r^{(b)}\right)^m\right\} = E\left\{\left(n_k^{(b)}\right)^{p+m}\right\} \cdot \delta_{k,r} + E\left\{\left(n_k^{(b)}\right)^p\right\} E\left\{\left(n_r^{(b)}\right)^m\right\} \cdot (1 - \delta_{k,r})$
- $E\left\{\left(n_l^{(d)}\right)^p \left(n_s^{(d)}\right)^m\right\} = E\left\{\left(n_l^{(d)}\right)^{p+m}\right\} \cdot \delta_{l,s} + E\left\{\left(n_l^{(d)}\right)^p\right\} E\left\{\left(n_s^{(d)}\right)^m\right\} \cdot (1 - \delta_{l,s})$
- $E\left\{\left(n_k^{(b)}\right)^{2p+1}\right\} = E\left\{\left(n_l^{(d)}\right)^{2p+1}\right\} = 0$
- $E\left\{\left(n_k^{(b)}\right)^2\right\} = E\left\{\left(n_l^{(d)}\right)^2\right\} = \sigma_n^2$
- $E\left\{\left(n_k^{(b)}\right)^4\right\} = E\left\{\left(n_l^{(d)}\right)^4\right\} = 3\sigma_n^4$

where p and m can be any positive integers, and $\delta_{k,r}$ is the Kronecker-delta defined as

$$\delta_{k,r} = \begin{cases} 1; & k = r \\ 0; & k \neq r \end{cases}$$

the expression of ${\sigma'_\zeta}^2$ given in (3.29) can be simplified as

$$
{\sigma'_\zeta}^2 = 2\sigma_n^2 \left(\left(\frac{\sigma_n^2}{L_B} + \frac{\sigma_n^2}{L_D} \right) + 2 \left(\frac{\sigma_B^2}{L_B} + \frac{\sigma_D^2}{L_D} \right) \right) \tag{3.30}
$$

It can be seen from (3.30) that for sufficiently large L_B and L_D, and relatively small noise variance σ_n^2, the variance ${\sigma'_\zeta}^2$ of the auxiliary variable ζ' approaches zero, which implies that auxiliary variable ζ' is approximately a sure event and its value equals to its mean i.e., $\zeta' \approx \mu'_\zeta = \zeta$, that is, ${\sigma'}_B^2 - {\sigma'}_D^2 \approx \sigma_B^2 - \sigma_D^2$.

To illustrate that the assumption of the largeness of L_B and L_D is practical, we consider the case of embedding a binary image of size 64x64 pixels at a bit rate of 172 bits per second (bps) using the proposed technique in an audio signal whose variance is $\sigma_X^2 = 0.05$, which is sampled at 44.1 kHz. In this case, $L_D = 65536$ and $L_B = 32768$, and Fig. 3.4 demonstrates the practical variations of the proposed quantization step given by (3.8) under an AWGN attack compared with those of the average quantization step of the adaptive approach [14,49,71]. From this figure, the near invariance property of the proposed quantization step under an AWGN attack is evident even for an extremely noisy signal, i.e., when $\sigma_n^2 \gg \sigma_X^2$.

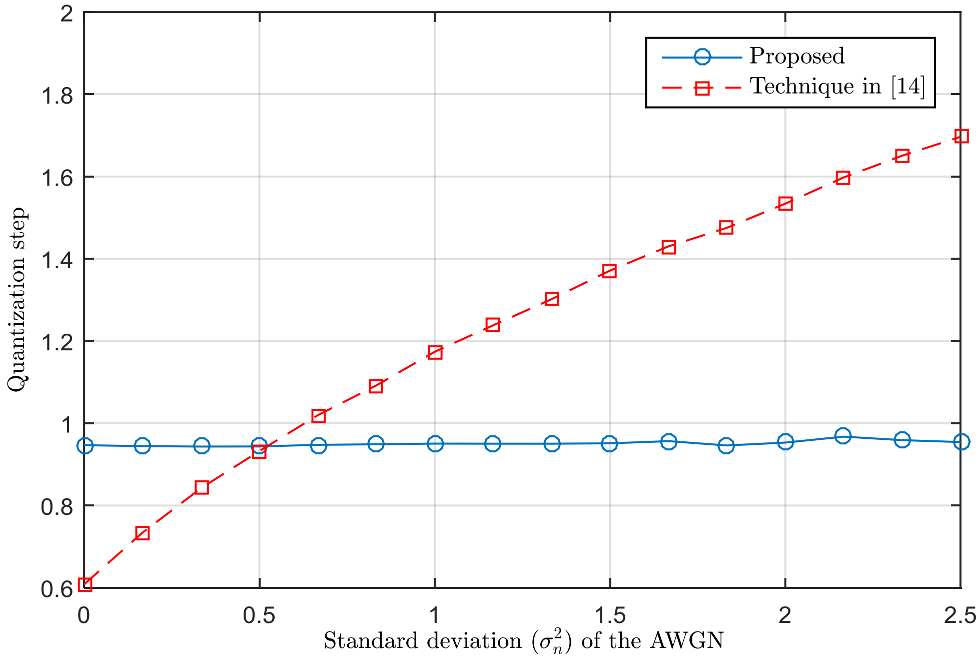


Fig. 3.4 Practical variations of the proposed quantization step compared with those of the average of the adaptive quantization step [14] under an AWGN attack.

3.5 Side information recovery procedure for the semi-blind operation mode

Due to the importance of the side information in semi-blind watermarking, we propose in this section a side information recovery procedure for the proposed technique when operating semi-blindly. This procedure uses steps similar to those used in the watermark extraction procedure (presented in Subsection 3.3.2) with slight modifications. Specifically, we apply the steps 1, 2 and 3 in the same manner as in the watermark extraction procedure, then for each frame $F'^{(m)}, m = 1, 2, 3, \ldots, L_w$, we recover the index $\tilde{S}_m$ using the quantization step Δ' and the corresponding dither coefficient $D^{(m)}$ as

$$\tilde{S}_m = \underset{\varsigma\in\{1,2,3,\ldots,N\}}{\operatorname{argmin}} \left\{ \min_{w\in\{0,1\}} \left\{ \left| F'^{(m)}_{\varsigma} - Q_{\Delta'}\left(F'^{(m)}_{\varsigma}, w, D^{(m)}\right) \right| \right\} \right\} \tag{3.31}$$

and finally, form the recovered side information sequence $\tilde{S} = \left[\tilde{S}_1, \tilde{S}_2, \tilde{S}_3, \ldots, \tilde{S}_{L_W}\right]$. By using this procedure, the side information can be recovered accurately from an attack-free or scaled watermarked signal, and can also be recovered from an attacked watermarked signal with a probability of error recovery that is given by

$$\begin{aligned} P_{er} &= \Pr\left(\tilde{S}_m \neq S_m\right) \\ &= \Pr\left(\min_{\substack{\varsigma\in\{1,2,3,\ldots,N\} \\ \varsigma\neq S_m}} \left\{ \min_{w\in\{0,1\}} \left\{ \left| F'^{(m)}_{\varsigma} - Q_{\Delta'}\left(F'^{(m)}_{\varsigma}, w, K_2^{(m)}\right) \right| \right\} \right\} \right. \\ &\qquad\qquad \left. \leq \min_{w\in\{0,1\}} \left\{ \left| F'^{(m)}_{S_m} - Q_{\Delta'}\left(F'^{(m)}_{S_m}, w, K_2^{(m)}\right) \right| \right\} \right) \end{aligned} \tag{3.32}$$

where S_m is the correct index.

To evaluate the proposed side-information recovery procedure, we investigate the probability of error recovery when the watermarked signal is attacked by a low-power AWGN $n \sim \mathcal{N}(0, \sigma_n^2)$, specifically, for the case when $\sigma_n \ll \Delta'$. To simplify the analysis, we first rewrite (3.32) as

$$P_{er} = \Pr\left(\min_{\substack{\varsigma=\{1,2,3,\ldots,N\} \\ \varsigma\neq S_m}} \left\{ G^{(m)}_{\varsigma} \right\} \leq G^{(m)}_{S_m} \right) \tag{3.33}$$

where

$$G^{(m)}_{\varsigma} = \min_{w\in\{0,1\}} \left\{ \left| F'^{(m)}_{\varsigma} - Q_{\Delta'}\left(F'^{(m)}_{\varsigma}, w, K_2^{(m)}\right) \right| \right\} \tag{3.34}$$

Under an AWGN attack, the coefficients of the m^{th} received frame $F'^{(m)}_{\varsigma}$ are related to the coefficients of the m^{th} watermarked frame $F^{(m)}_{\varsigma}$ and the noise coefficients n_{ς} by $F'^{(m)}_{\varsigma} = F^{(m)}_{\varsigma} + n_{\varsigma}$. Using the low-power noise assumption and the fact that $F^{(m)}_{S_m}$ has already been quantized in the embedding process, it is seen that $G^{(m)}_{S_m} \approx |n_{S_m}|$, and hence (3.33) can be expressed as

$$P_{er} = \Pr\left(\min_{\substack{\varsigma \in \{1,2,3,\dots,N\} \\ \varsigma \neq S_m}} \left\{ G^{(m)}_{\varsigma} \right\} \leq |n_{S_m}| \right)$$
$$= 1 - \Pr\left(\min_{\substack{\varsigma \in \{1,2,3,\dots,N\} \\ \varsigma \neq S_m}} \left\{ G^{(m)}_{\varsigma} \right\} > |n_{S_m}| \right) \tag{3.35}$$

Moreover, the received coefficients $F'^{(m)}_{\varsigma}$ are obtained by decomposing the noisy signal using the DWT, and thus independent and identically distributed [12,15,86], which implies the independence and identicalness of $G^{(m)}_{\varsigma}$ given in (3.34). Therefore, the probability in (3.35) can be expressed as

$$\Pr\left(\min_{\substack{\varsigma \in \{1,2,3,\dots,N\} \\ \varsigma \neq S_m}} \left\{ G^{(m)}_{\varsigma} \right\} > |n_{S_m}| \right) = \prod_{\substack{\varsigma=1 \\ \varsigma \neq S_m}}^{N} \Pr\left(G^{(m)}_{\varsigma} > |n_{S_m}| \right)$$
$$= \left(\Pr\left(G^{(m)}_{\varsigma} > |n_{S_m}| \right) \right)^{N-1} \tag{3.36}$$

Using (3.36) in (3.35), the expression of the probability of error recovery reduces to

$$P_{er} = 1 - \left(\Pr\left(G^{(m)}_{\varsigma} > |n_{S_m}| \right) \right)^{N-1} \tag{3.37}$$

Furthermore, since the random variables $\left| F'^{(m)}_{\varsigma} - Q_{\Delta'}\left(F'^{(m)}_{\varsigma}, w, K^{(m)}_{2} \right) \right|$ are uniformly distributed in the interval $[0, \Delta'/2]$ for all $\varsigma \neq S_m$ [22], the random variables $G^{(m)}_{\varsigma}$ in (3.34) are uniformly distributed in the interval $[0, \Delta'/4]$ for all $\varsigma \neq S_m$, which implies that

$$\Pr\left(G^{(m)}_{\varsigma} > |n_{S_m}| \right) = \int_{0}^{\frac{\Delta'}{4}} \frac{4}{\Delta'} \int_{-G^{(m)}_{\varsigma}}^{G^{(m)}_{\varsigma}} \frac{e^{-\frac{n^2_{S_m}}{2\sigma^2_n}}}{\sqrt{2\pi\sigma^2_n}} \, dn_{S_m} dG^{(m)}_{\varsigma}$$
$$= \operatorname{erf}\left(\frac{\sqrt{2}\,\Delta'}{8\sigma_n} \right) + \frac{4\sqrt{2}\sigma_n}{\sqrt{\pi}\,\Delta'} \left(e^{\frac{-\Delta'^2}{32\sigma^2_n}} - 1 \right) \tag{3.38}$$

where $\mathrm{erf}(\cdot) = 1 - \mathrm{erfc}(\cdot)$ is the error function. Using the near invariance of the quantization step under an AWGN attack established in the previous section, (3.17), and (3.21), (3.38) can be rewritten as

$$\Pr\left(G_{\varsigma}^{(m)} > \left|n_{S_m}\right|\right) = \mathrm{erf}\left(\sqrt{2(N+1)(N+2)L_F\,\mathrm{WNR}}\right) + \frac{\left(e^{-2\,(N+1)(N+2)L_F\,\mathrm{WNR}} - 1\right)}{\sqrt{2\,\pi\,(N+1)(N+2)L_F\,\mathrm{WNR}}} \quad (3.39)$$

Finally, by substituting (3.39) in (3.37), a closed-form expression for the probability of error recovery under a low-power AWGN attack is given by

$$P_{er} = 1 - \left(\mathrm{erf}\left(\sqrt{2(N+1)(N+2)L_F\,\mathrm{WNR}}\right) + \frac{\left(e^{-2\,(N+1)(N+2)L_F\,\mathrm{WNR}} - 1\right)}{\sqrt{2\,\pi\,(N+1)(N+2)L_F\,\mathrm{WNR}}}\right)^{N-1} \quad (3.40)$$

3.6 Experimental results

To evaluate the performance of the proposed technique, we use a binary image of size 64×64 bits as a watermark to be embedded in audio signals sampled at 44.1 kHz and stored in WAVE format (16-bit, mono). We carried out many experiments and present here the results for three distinct signals, namely speech signal, music signal, and mixture of music and speech denoted by S1, S2, and S3, respectively. Since different audio types have different characteristics, we present also the average results of twenty signals including jazz, piano, classical, pop, country, and rock music. We implement the proposed technique for different values of $q = 0, 1, 2,$ and 3, and denote the corresponding implementations by Proposed-0, Proposed-1, Proposed-2, and Proposed-3, respectively. The Proposed-0 is blind and the Proposed-1, Proposed-2, and Proposed-3 are semi-blind. For the purpose of comparison, we also implemented the techniques reported in [12–15,18] and compare the Proposed-0 with the blind techniques reported in [13–15], and the Proposed-1, Proposed-2 and Proposed-3 with the semi-blind techniques reported in [12,18]. It has been shown by Wang et al. in [15] that the performance of IAQIM against gain attacks and AWGN is independent of the used transform domain and the type of the host media. For a fair comparison, we implement IAQIM in a way similar to that of the blind operation mode of the proposed technique, specifically the implementation is achieved by replacing the embedding function in (3.4) by the IAQIM embedding function and using the IAQIM decoder instead of the QIM minimum distance decoder given by (3.9).

It should be noted that there is a trade-off between transparency, capacity, and robustness [20]. The signal to watermark ratio (SWR) defined as the ratio between the variance of the host signal σ_X^2 and the watermarking distortion i.e., $\mathrm{SWR} = \sigma_X^2/\sigma_W^2$, is usually used as an objective measure of transparency [5]. Capacity (the watermarking payload) is the number of the embedded watermark bits per unit of time. The bit error rate (BER) defined as the number of unsuccessfully extracted watermark bits divided by the total number of watermark bits, is extensively used to assess the robustness [32]. The SWR and capacity of the proposed technique can be adjusted by varying, respectively, δ in (3.1) and the frame length L_F.

For a fair comparison, we use an SWR of approximately 25 dB for all the considered techniques except the technique in [12], which employs a psychoacoustic model in a way that prevents any adjustment of the SWR, and its SWR is much less than that used for the proposed technique as shown in Table 3.1. The value 25 dB of the SWR satisfies the IFPI requirement [20,21] and gives a good compromise between transparency and robustness. The capacity of the blind techniques in [13,14] is roughly 86 bps, and thus we use a capacity of 86 bps for the techniques in [13–15] and the blind implementation of the proposed technique. The capacity of the semi-blind technique in [12] is 172 bps, and since the capacity of the technique in [18] can be adjusted to 172 bps as well, we use for techniques in [12,18] and the semi-blind implementations of the proposed technique a capacity of 172 bps.

Table 3.1 SWR of different techniques.

Signal	Proposed-0	Proposed-1	Proposed-2	Proposed-3	[13]	[14]	[15]	[18]	[12]
S1	25.04	24.87	25.39	26.34	24.96	27.20	24.96	24.90	18.23
S2	24.97	24.90	24.87	24.92	25.10	25.74	24.92	24.99	18.63
S3	25.00	24.82	24.97	25.05	24.82	23.50	25.66	25.15	19.20
Average	25.00	24.86	25.07	25.43	24.96	24.48	25.18	25.01	18.69
Average over 20 signals				≈ 25					19.5

We then compare the considered techniques in terms of their robustness against content preserving attacks and manipulations listed in Table 3.2, and the results are presented in tables Table 3.3-Table 3.5. It is clear from the obtained results that all the considered techniques give excellent robustness to low-pass filtering except the technique in [18], whereas, the other attacks can be regarded as gain attacks (i.e., amplitude scaling, multiplicative uniform noise) or noise addition attacks (i.e., resampling, re-quantization, echo addition, AWGN, lossy compression). It can be seen from Tables Table 3.3-Table 3.5 that:

1. The proposed technique in its blind mode of operation outperforms the blind techniques reported in [13–15] in terms of robustness against content preserving attacks and manipulations.
2. Compared to QIM-based solutions to gain attacks presented in [14,15], the proposed technique offers slightly better robustness to gain attacks and much higher robustness to noise addition attacks.
3. The proposed technique offers much higher robustness against content preserving attacks and manipulations than the conventional QIM-based techniques reported in [13,18].
4. In its semi-blind mode of operation, the proposed technique is much more robust than the semi-blind techniques in [12,18] against gain attacks. However, the proposed-1 is slightly less robust than the techniques in [12,18] against noise addition attacks, whereas the Proposed-2 is much more robust than techniques in [12,18] against all considered attacks, and the Proposed-3 is significantly more robust against all considered attacks than all considered techniques. Furthermore, the side information sizes required by the Proposed-1, Proposed-2, and Proposed-3 without compression are 0.02%, 0.05%, and 0.07% of the host audio signal size, respectively, which are far less than the 0.3% with compression reported in [12] and the 0.2% without compression of the technique in [18]. Hence the proposed technique in its semi-blind mode of operation outperforms the semi-blind techniques in [12,18] while demanding less storage requirements.

The robustness of the proposed technique to the filtering attack is a result of embedding of the watermark bits in the approximation band of the host signal, whereas the strong robustness of the proposed technique against gain attacks is due to the proposed quantization step that renders the extraction process invariant under amplitude scaling of the watermarked signal, and the robustness of the proposed technique against noise addition attacks is due to (i) the near invariance property of the proposed quantization step under AWGN attack, (ii) the selection criterion for the coefficient that carries the watermark, and (iii) the utilized quantization function given in (3.4), which can achieve favorable rate-distortion-robustness trade-offs.

Table 3.2 List of considered attacks and their abbreviations and descriptions.

Attack	Abbreviation	Description
Resampling	Res.	The watermarked signal is resampled from 44.1 kHz to 8 kHz and back to 44.1 kHz
Re-quantization	Req. (16→8)	The watermarked signal is re-quantized from 16-bits resolution to 8-bits resolution
	Req. (16→4)	The watermarked signal is re-quantized from 16-bits resolution to 4-bits resolution
Low-pass filtering	LPF	A second order Butterworth low-pass filter of 8 kHz cut-off frequency is applied to the watermarked signal.
Echo addition	Echo	Add an echo of amplitude 10% and a delay of 50 ms to the watermarked signal
Amplitude scaling	Scale-150	The watermarked signal is a scaled by a factor of 150%
	Scale-50	The watermarked signal is a scaled by a factor of 50%
Multiplicative uniform noise	MUN-0.1	The watermarked signal samples are multiplied by a noise uniformly distributed in the interval [0.9, 1.1]
	MUN-0.5	The watermarked signal samples are multiplied by a noise uniformly distributed in the interval [0.5, 1.5]
Additive white Gaussian noise (AWGN)	AWGN-20	The watermarked signal is attacked by an AWGN with an SNR=20 dB.
	AWGN-10	The watermarked signal is attacked by an AWGN with an SNR=10 dB.
	AWGN-5	The watermarked signal is attacked by an AWGN with an SNR=5 dB.
MPEG layer III compression	MP3-96	MPEG layer III compression at 96 bps bit rate is applied to the watermarked signal
	MP3-64	MPEG layer III compression at 64 bps bit rate is applied to the watermarked signal

Table 3.3 Comparison of the proposed blind technique with existing blind techniques in terms of BER (%) under the considered attacks.

Attack	Proposed-0			[13]			[14]			[15]		
	S1	S2	S3	S1	S2	S3	S1	S2	S3	S1	S2	S3
Res.	0	0	0	0	0	0	0	0	0	0	0	0
Req. (16→8)	0	0	0	0	0	0	0	0	0	0	0	0
Req. (16→4)	0	0	0	0.293	0.293	0	4.81	4.48	0.5615	2.08	4.91	2.61
LPF	0	0	0	0	0	0	0	0	0	0	0	0
Echo	0.024	1.587	0.708	21.48	28.66	7.959	0.293	1.587	1.294	6.98	10.67	11.62
Scale-150	0	0	0	50.75	50.87	49.12	0	0	0	0	0	0
Scale-50	0	0	0	50.07	49.82	49.88	0	0	0	0	0	0
UMN-0.1	0	0	0	0	0	0	0.756	0.805	0.438	0	0.073	0.122
UMN-0.5	0.683	1.416	1.196	1.172	2.051	1.294	13.25	12.32	6.811	0.415	0.928	2.832
AWGN-20	0	0	0	0	0	0	2.002	1.831	0.524	0.537	1.318	1.367
AWGN-10	0.073	0.049	0.024	1.050	0.952	0.977	26.66	21.19	7.813	4.688	11.45	12.01
MP3-96	0	0	0	0	0	0.049	0.171	0.024	0	0.195	0.464	1.685
MP3-64	0	0.073	0.049	0.122	0.220	0.513	1.464	0.708	0.854	0.854	1.685	5.054

Table 3.4 Comparison of the proposed semi-blind technique with existing semi-blind techniques in terms of BER (%) under the considered attacks.

Attack	Proposed-1			Proposed-2			Proposed-3			[18]			[12]		
	S1	S2	S3	S1	S2	S3	S3	S2	S3	S1	S2	S3	S1	S2	S3
Res.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Req. (16→8)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Req. (16→4)	0	0	0	0	0	0	0	0	0	0	0	0	0.15	0.10	0
LPF	0	0	0	0	0	0	0	0	0	0.01	0.02	0.04	0	0	0
Echo	0.02	0.85	0.59	0	0	0	0	0	0	0	0	0	0.46	1.42	1.68
Scale-150	0	0	0	0	0	0	0	0	0	49.5	49.5	49.5	17.2	28.6	30.1
Scale-50	0	0	0	0	0	0	0	0	0	8.08	10.9	7.74	50.4	49.2	50.9
UMN-0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
UMN-0.5	0.68	0.93	0.78	0.02	0.20	0.05	0	0	0	6.59	13.4	6.88	0.83	3.81	6.67
AWGN-20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AWGN-10	0.07	0.05	0.02	0	0	0	0	0	0	0	0	0	0.88	2.66	0.88
AWGN-5	5.37	4.57	5.40	0.27	0.20	0.12	0	0	0	2.03	1.21	1.37	0.85	1.83	3.81
MP3-96	0	0	0	0	0	0	0	0	0	0	0	0	0.01	0.01	0
MP3-64	0.024	0	0.05	0	0	0	0	0	0	0.02	0.05	0.10	0.13	0.11	0.09

Table 3.5 Average BER (%) over twenty signals for different techniques under the considered attacks.

Category	Blind techniques				Semi-blind techniques				
Capacity	≈ 86 bps				≈ 172 bps				
Attack	Proposed-0	[13]	[14]	[15]	Proposed-1	Proposed-2	Proposed-3	[18]	[12]
Res.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Req. (16→8)	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Req. (16→4)	0.73	3.18	5.35	5.69	0.55	0.00	0.00	0.24	0.41
LPF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00
Echo	0.53	4.61	1.54	4.28	0.49	0.03	0.00	0.10	0.33
Scale-150	0.00	50.1	0.00	0.00	0.00	0.00	0.00	49.8	18.98
Scale-50	0.00	50.0	0.00	0.00	0.00	0.00	0.00	11.3	49.95
UMN-0.1	0.00	0.00	0.08	0.09	0.00	0.00	0.00	0.13	0.00
UMN-0.5	0.87	1.43	5.36	1.53	0.70	0.11	0.01	7.98	2.02
AWGN-20	0.00	0.00	3.25	1.06	0.00	0.00	0.00	0.00	0.00
AWGN-10	0.05	1.02	13.27	8.09	0.05	0.00	0.00	0.00	0.82
AWGN-5	4.92	15.37	24.7	18.8	5.09	0.27	0.00	2.47	2.76
MP3-96	0.00	0.02	0.15	1.17	0.00	0.00	0.00	0.01	0.01
MP3-64	0.18	0.38	1.34	3.83	0.19	0.02	0.00	0.18	0.11

In order to verify the theoretical analysis carried out in this work, we use the proposed technique to embed a binary image of size 64x64 pixels at a bit rates of 172 and 86 bps in the signal S3. Fig. 3.5 shows the theoretical BER given by (3.22) and the empirical BER under AWGN attack for a capacity of 172 bps and values of $q = 0, 1$ and 2, whereas Fig. 3.6 shows the analytical watermarking distortion given by (3.17) and its empirical values for different values of δ and for capacities of 172 and 86 bps. These figures clearly validate the correctness of the theoretical expressions given by (3.17) and (3.22).

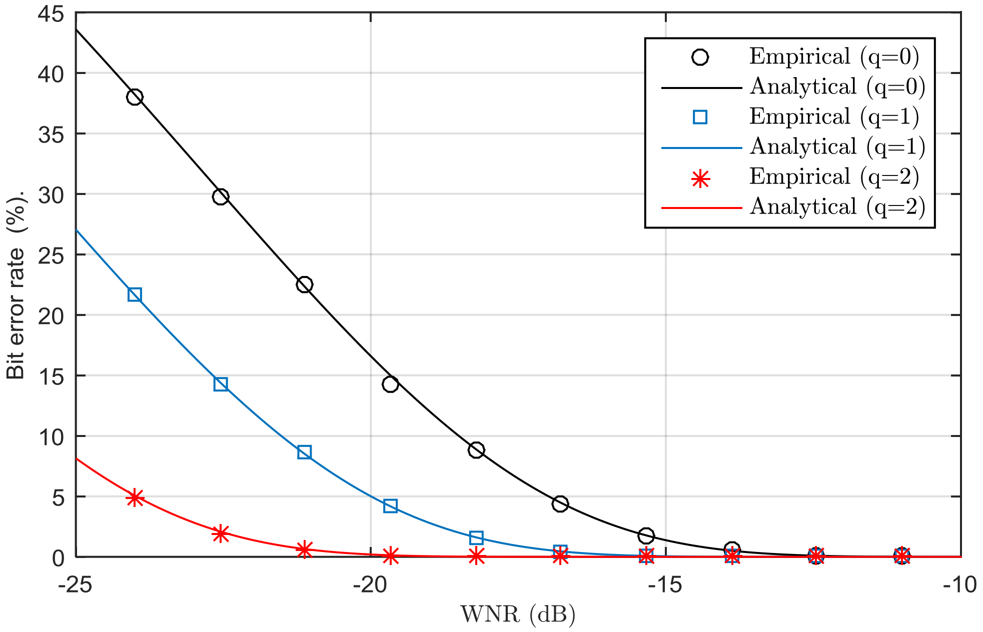


Fig. 3.5 Analytical and empirical BER for the proposed watermarking technique under an AWGN attack.

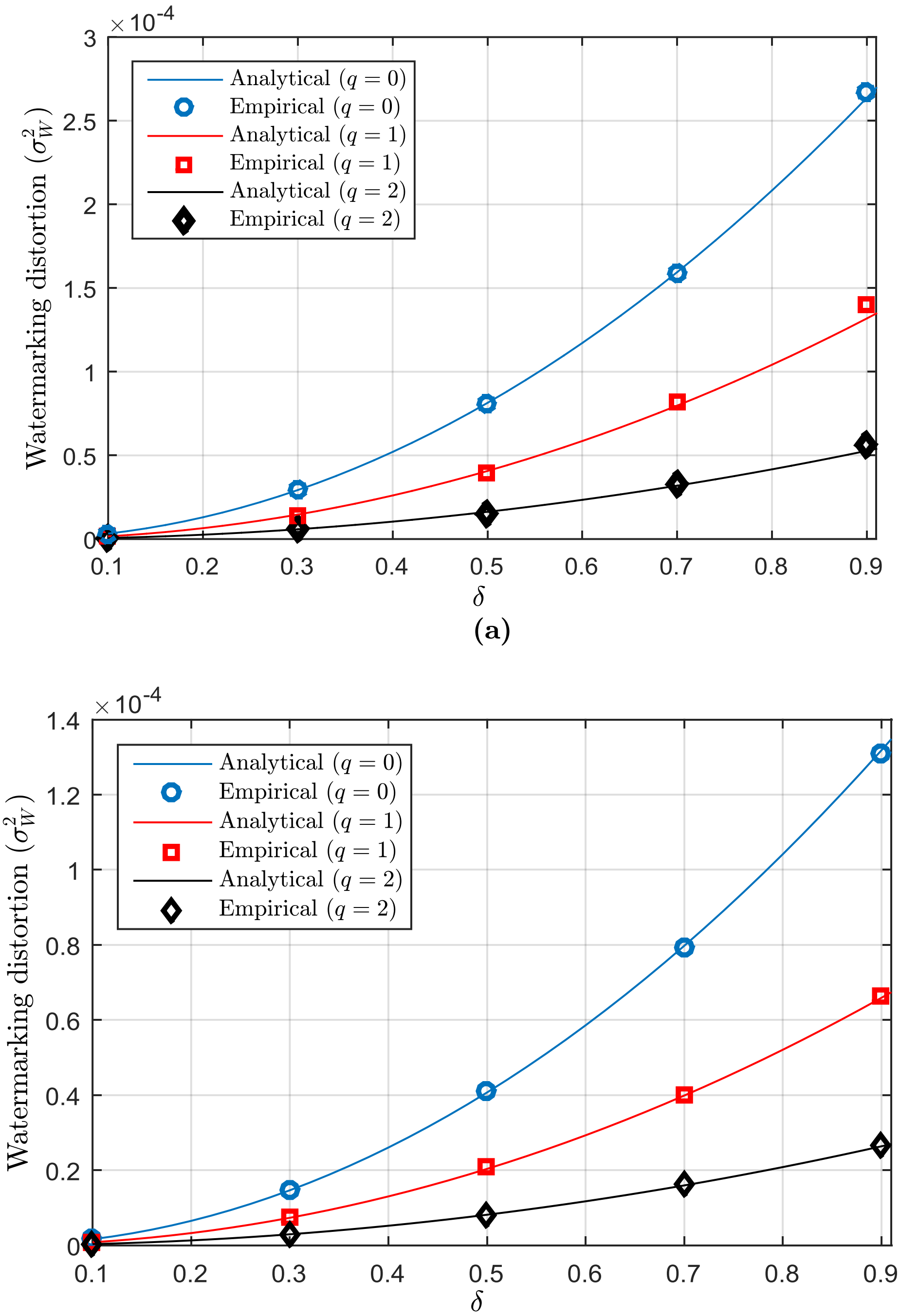
Fig. 3.7 shows the evaluation of the side information recovery procedure developed in Section 3.5 against AWGN attack in terms of the probability of error recovery given by (3.40) and the empirical normalized Hamming distance $H(S, \tilde{S})$ between the original side information S and the recovered side information $\tilde{S}$, which is calculated as

$$P_{er} = H(S, \widetilde{S}) = \frac{1}{L_S} \sum_{k=1}^{L_S} [\widetilde{S}_k \neq S_k] \qquad (3.41)$$

where L_s is the length of the side information, and $[\tilde{S}_k \neq S_k]$ is the Iverson bracket notation

$$[\tilde{S}_k \neq S_k] = \begin{cases} 1 & \text{if } \tilde{S}_k \neq S_k \\ 0 & \text{if } \tilde{S}_k = S_k \end{cases}$$

It is clear from Fig. 3.7 that the probability of error side information recovery given by (3.40) exhibits excellent agreement with the empirical normalized Hamming distance.



(a)

(b)

Fig. 3.6 Theoretical and empirical watermarking distortions for different values of $\boldsymbol{\delta}$: (a) capacity of 172 bps, (b) capacity of 86 bps.

Finally, to evaluate the security of the proposed technique, and without loss of generality, we use Zhou chaotic map [89,90] to calculate the dither coefficients $D^{(m)}$ iteratively as

$$D^{(m+1)} = \psi\left(D^{(m)}, \lambda_D\right) = \begin{cases} \dfrac{D^{(m)}}{\lambda_D}, & 0 \leq D^{(m)} < \lambda_D \\ \dfrac{D^{(m)} - \lambda_D}{0.5 - \lambda_D}, & \lambda_D \leq D^{(m)} < 0.5 \\ \psi\left(1 - D^{(m)}, \lambda_D\right), & 0.5 \leq D^{(m)} < 1 \end{cases} \tag{3.42}$$

where the initial value $D^{(0)} \in [0,1]$ and the control parameter $\lambda_D \in [0, 0.5]$ constitute the dithering key, i.e., $K_D = \{D^{(0)}, \lambda_D\}$. Similarly, we generate another pseudo-random sequence from Zhou map using another initial value $R^{(0)} \in [0,1]$ and control parameter $\lambda_R \in [0, 0.5]$ to perform the PRPs in the proposed technique, and thus the scrambling key is $K_S = \{R^{(0)}, \lambda_R\}$. Given that the values of $D^{(0)}$, λ_D, $R^{(0)}$, and λ_R used in the watermark embedding procedure of the proposed technique are respectively $\widehat{D}^{(0)}$, $\hat{\lambda}_D$, $\hat{R}^{(0)}$, and $\hat{\lambda}_R$, extensive experiments confirm that adding a perturbation of 10^{-16} in the watermark extraction procedure to $\widehat{D}^{(0)}$, $\hat{\lambda}_D$, $\hat{R}^{(0)}$, or $\hat{\lambda}_R$ results in a failure of the watermark extraction. Hence the key sensitivity is of order $10^{-4\times16}$ and the key space is of order $0.5^2 \times 10^{4\times16} = 2.5 \times 10^{63}$, which is largely sufficient from cryptographic viewpoint [89].

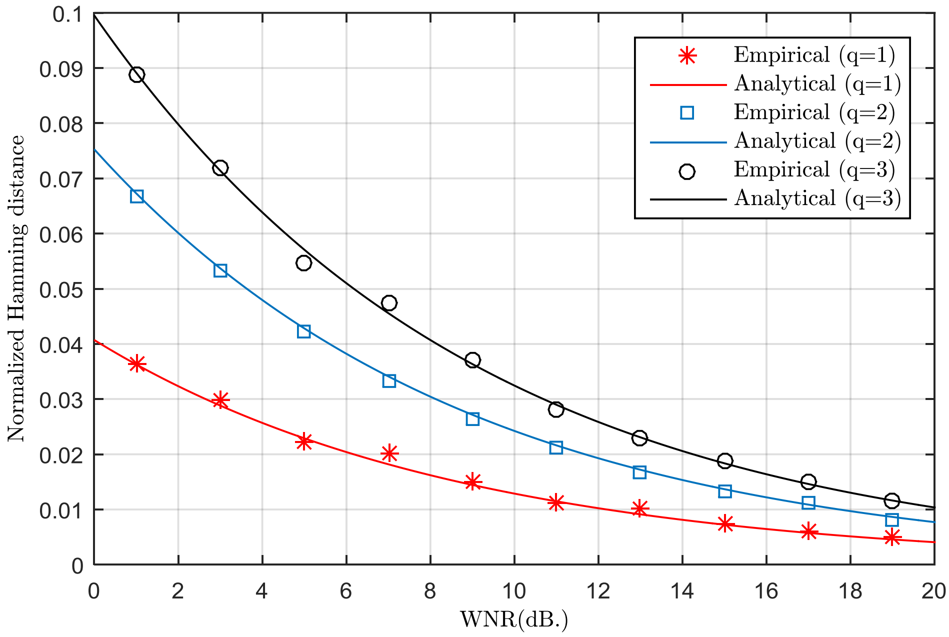


Fig. 3.7 Empirical and analytical probability of side information error recovery for the proposed side information recovery procedure under an AWGN attack.

Fig. 3.8 shows the similarity between the embedded watermark and 1000 extracted watermarks where the 500th extraction utilizes the same values of $D^{(0)}$, λ_D, $R^{(0)}$ and $\lambda_R$used in the embedding procedure, and the remaining 999 extractions use (a) the same values λ_D, $R^{(0)}$, λ_R used in the embedding procedure and random values of $D^{(0)}$, (b) the same values $D^{(0)}$, $R^{(0)}$, λ_R used in the embedding procedure and random values of λ_D, (c) the same values $D^{(0)}$, λ_D, λ_R used in the embedding procedure and random values of $R^{(0)}$, and (d) the same values $D^{(0)}$, λ_D, $R^{(0)}$ used in the embedding procedure and random values of λ_R. It can be seen from this figure that the proposed technique is sensitive to each component of the embedding key (the values of $D^{(0)}$, λ_D, $R^{(0)}$, and λ_R).

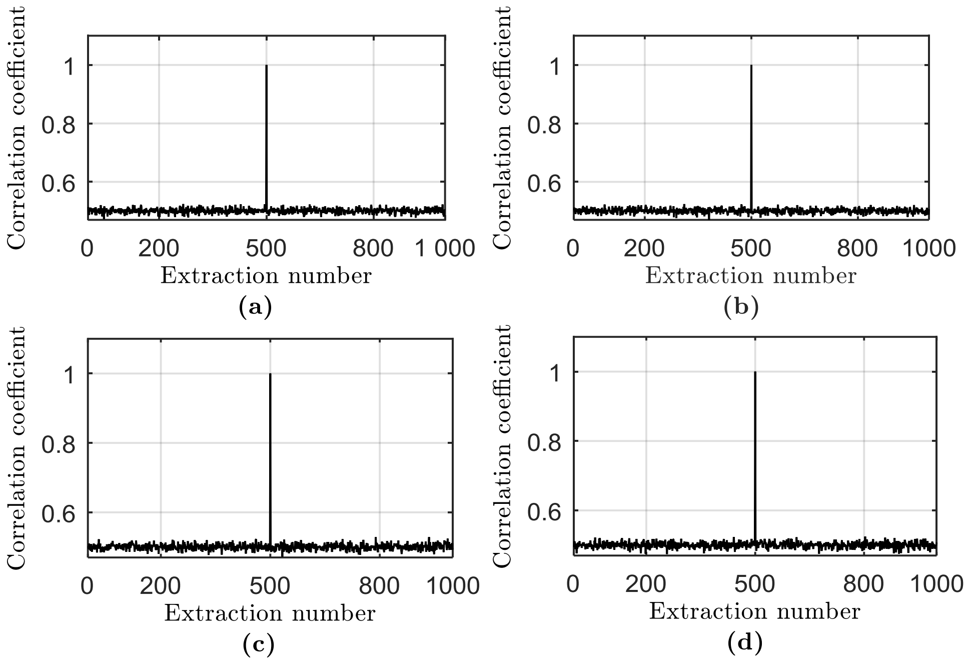


Fig. 3.8 Similarity between the embedded and a thousand extracted watermarks where the 500th extraction uses the same values of the embedding key $\boldsymbol{D^{(0)}}$, $\boldsymbol{\lambda_D}$, $\boldsymbol{R^{(0)}}$, and $\boldsymbol{\lambda_R}$ used in the embedding procedure, and the 999 remaining extractions use the same embedding key with random values of: (a) $\boldsymbol{D^{(0)}}$, (b) $\boldsymbol{\lambda_D}$, (c) $\boldsymbol{R^{(0)}}$, and (d) $\boldsymbol{\lambda_R}$.

3.7 Conclusion

In this chapter, a novel QIM-based watermarking technique has been proposed for robust blind and semi-blind audio watermarking. The DWT has been exploited in the proposed technique for its multiresolution property. Moreover, we have introduced an expression for the quantization step by a suitable exploitation of the DWT bands of the host audio signal. The resulting quantization step is near invariant under an AWGN attack and makes the watermark extraction process invariant to gain attack. For the QIM realization, we have adopted the dither modulation owing to its good rate-distortion-robustness trade-offs. In order to secure the proposed audio watermarking technique, pseudo-random dithering of the quantization function and pseudo-random permutations of the DWT approximation band have been employed together.

Moreover, an efficient side information recovery procedure has been proposed for the semi-blind operation mode of the proposed technique. Theoretical performance of this side information recovery procedure and that of the proposed audio watermarking technique have been investigated and verified experimentally. In addition, several experiments have been carried out to evaluate the security of the proposed technique and assess its robustness against content preserving attacks and manipulations. Experimental results clearly show that the proposed technique outperforms the existing QIM-based watermarking techniques, QIM-based solutions to gain attacks and existing blind and semi-blind watermarking techniques.

The proposed technique has many other pertinent advantages such as the ability to operate in both blind and semi-blind modes and accept different side information sizes for its semi-blind mode, which has an efficient procedure for side information recovery. This and the above advantages would make the proposed audio watermarking technique more attractive than the existing techniques. In the next chapter, we apply watermarking and robust-hashing in audio fingerprinting.

Chapter 4

Joint hashing-watermarking for audio fingerprinting

4.1 Introduction

Audio fingerprinting is a technology that allows signal identification by summarizing big audio signals into compact digests that are robust to content preserving manipulations [26,91,92]. It can be achieved by watermarking/robust-hashing methods, i.e., embedding/extracting a/an watermark/robust-hash in/from the host signal. This watermark/robust-hash serves as a fingerprint that uniquely identifies the signal. Fig. 4.1 shows a typical fingerprinting system. First, the fingerprint of the audio signal is embedded/extracted and stored with the metadata of the signal in the database. The watermarked/original signal can then be distributed over multimedia systems and/or communication networks. When the received signal is subject to identification, its fingerprint is extracted and then compared with the fingerprints in the database. In case of matching, the metadata of the received signal is obtained to serve the identification. Note that robustness is required to face attacks and manipulations that can happen during transmission, e.g., noise, filtering, lossy compression, etc.

In contrast to watermarking, which reduces the perceptual quality of the signal by altering its samples, robust-hashing requires no modification. On the other hand, watermarking has less storage requirements and can achieve stronger robustness. Therefore, it is natural to exploit robust-hashing and watermarking together to devise a technique for the fingerprinting application that is robust and has low storage requirements and better transparency than conventional watermarking techniques. Unfortunately, such a task is arduous due to the fact that the concepts of robust-hashing and watermarking techniques are completely different although their application in fingerprinting is very similar. Hence, to design a joint technique, one must either develop a robust hashing technique that has a functioning principle similar to that of watermarking or vice versa.

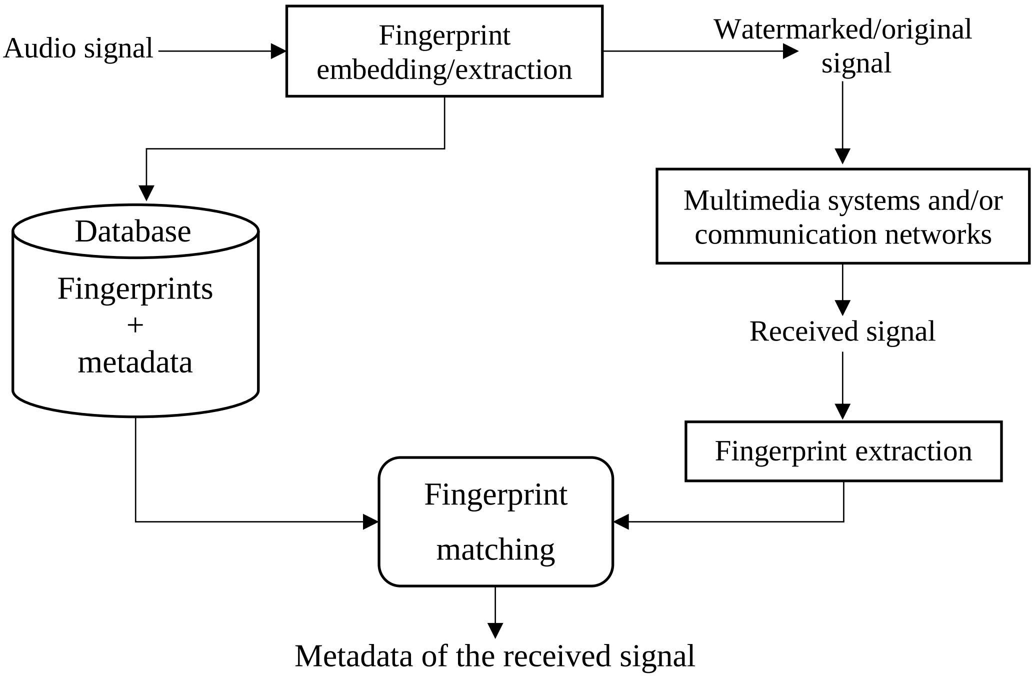

Audio signal
Fingerprint embedding/extraction
Watermarked/original signal
Database
Fingerprints + metadata
Multimedia systems and/or communication networks
Received signal
Fingerprint extraction
Fingerprint matching
Metadata of the received signal

Fig. 4.1 A typical fingerprinting system.

In this chapter, we propose a new key-dependent audio fingerprinting technique by exploiting the multiresolution decomposition property of the DWT in order to summarize the audio signal samples [30]. Specifically, we use the Haar DWT for its low computational complexity. In addition, inspired by the success of the quantization minimum distance in watermarking as a decoder, we introduce it in audio fingerprinting in the DWT domain as a hash extractor. Furthermore, we dither the quantization function of the quantization minimum distance using coefficients that are obtained from a chaotic map whose initial value is employed as a fingerprint extraction secret key. As mentioned earlier, the proposed hashing technique utilizes the quantization minimum distance as a hash extractor, which is used in QIM-based watermarking techniques as a watermark extractor. This allows us to develop and propose a joint hashing-watermarking technique [31] by substituting the LMD-QIM of the embedding procedure of the technique proposed in the previous chapter by a quantization minimum distance hash extractor and a QIM embedder.

We conduct various experiments to assess the proposed robust-hashing and joint hashing-watermarking techniques. To evaluate the performance of the proposed robust-hashing technique, we compare it with Haitsma's Philips robust hash (PRH) algorithm [91,93], which is considered as one of the standard approaches [92] in robust-hashing. Specifically, we compare it with the baseline PRH [91] as well as with two of its recent enhancements, namely, the multiple hashing (MLH) reported in [26] and the asymmetric matching method PRH (AMM-PRH) developed in [92]. As for the proposed joint hashing-watermarking technique, we perform the comparison with the blind implementation of the watermarking technique proposed in the previous chapter.

4.2 Proposed key-dependent audio fingerprinting technique based on a quantization minimum-distance hash extractor in the DWT domain

4.2.1 Proposed fingerprinting technique

The block diagram of the proposed fingerprinting technique is depicted in Fig. 4.2. Firstly, the audio signal is segmented into non-overlapping frames $f^{(m)}, m = 1, 2, 3, \dots, N_f$, each of length L, with N_f being the total number of frames. Secondly, an n-level Haar DWT is performed on each of these frames to obtain their corresponding approximation bands $F^{(m)}, m = 1, 2, 3, \dots, N_f$. Thirdly, we extract a hash from each of these bands. The m^{th} hash denoted by $H^{(m)} = \left\{h_0^{(m)},\ h_1^{(m)}, \dots, h_{N-1}^{(m)}\right\}$ is obtained from the N mid coefficients of $F^{(m)}$, where $h_r^{(m)}, r = 0, 1, 2, \dots, N-1$, are bits and calculated using a quantization minimum distance hash extractor as

$$h_r^{(m)} = \underset{\eta \in \{0,1\}}{\operatorname{argmin}} \left(\left| F_{r+\frac{L-N}{2}}^{(m)} - Q_{\Delta^{(m)}} \left(F_{r+\frac{L-N}{2}}^{(m)}, \eta, C_r^{(m)} \right) \right| \right) \tag{4.1}$$

The quantization function $Q_{\Delta^{(m)}} \left(F_{r+\frac{L-N}{2}}^{(m)}, \eta, C_r^{(m)} \right)$ in (4.1) is defined as

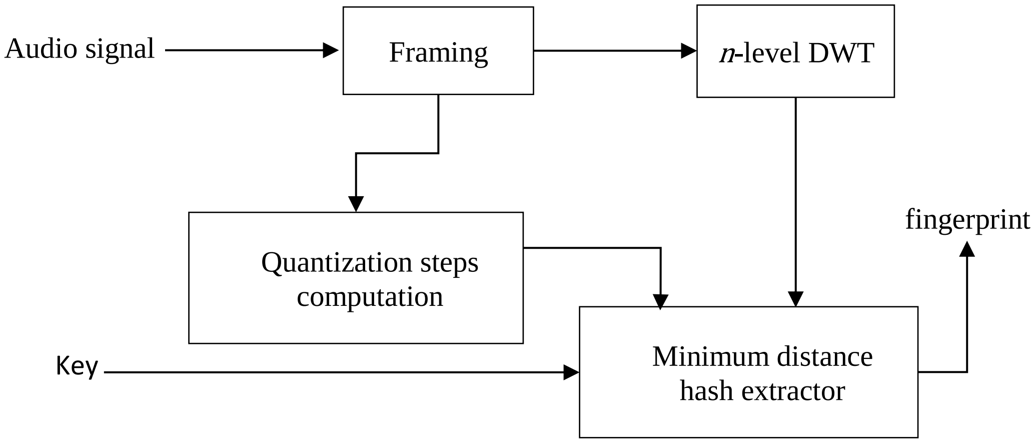

Audio signal
Framing
n-level DWT
Quantization steps computation
fingerprint
Key
Minimum distance hash extractor

Fig. 4.2 Block diagram of the proposed fingerprinting technique.

$$Q_{\Delta^{(m)}}\left(F^{(m)}_{r+\frac{L-N}{2}},\eta,C^{(m)}_r\right)=\left(\left[\frac{F^{(m)}_{r+\frac{L-N}{2}}}{\Delta^{(m)}}+\frac{\eta}{2}+C^{(m)}_r\right]-\frac{\eta}{2}-C^{(m)}_r\right)\Delta^{(m)} \tag{4.2}$$

where $[\cdot]$ denotes the rounding operation, the values of the coefficients $C_r^{(m)}$are obtained using a chaotic map with the initial value $C_0^{(1)}$, which is exploited as a fingerprint extraction key, and the quantization step $\Delta^{(m)}$ is defined as

$$\Delta^{(m)}=\alpha\sqrt{\frac{1}{L}\sum_{k=1}^{L}\left(f_k^{(m)}\right)^2} \tag{4.3}$$

with α is a proportionality constant. Finally, the hashes $H^{(m)}, m=1,2,3\ldots N_f$ are concatenated to form a fingerprint $P=\left\{H^{(1)},H^{(2)},\ldots,H^{(N_f)}\right\}$ of length $N_P=N_f\mathrm{x}N$.

4.2.2 Discrimination procedure

Let S and $\tilde{S}$ be two audio signals with P and $\tilde{P}$ their corresponding fingerprints. The aim of the discrimination procedure is to accept or reject the hypothesis $\mathcal{H}_0$ stating that $\tilde{S}$ is an identical or a manipulated version of S. For this purpose, we use the normalized Hamming distance $D(P,\tilde{P})$ between P and $\tilde{P}$ given by

$$x=D(P,\tilde{P})=\frac{1}{N_P}\sum_{k=0}^{N_P-1}\left(P_k-\tilde{P}_k\right)^2 \tag{4.4}$$

and accept the hypothesis $\mathcal{H}_0$ if $x<\tau$ and reject it otherwise, where τ is a predefined threshold, and P_k and $\tilde{P}_k$ are the k^{th} elements of P and $\tilde{P}$, respectively. This discrimination procedure may lead to two different types of errors. One type occurs when $\mathcal{H}_0$ is valid and the discrimination procedure rejects $\mathcal{H}_0$, i.e., $x\geq\tau$. In this case, the error is called a false rejection error. The other type is called a false acceptance error, which occurs when $\mathcal{H}_0$ is not valid and the discrimination procedure accepts $\mathcal{H}_0$, i.e., $x<\tau$. Note that for a large value of N_P, x in (4.4) is according to the central limit theorem a Gaussian random variable $\mathcal{N}(\mu,\sigma)$ and hence the conditional events $(x|\mathcal{H}_0)$ and $(x|\overline{\mathcal{H}_0})$ are $\mathcal{N}(\mu_0,\sigma_0)$ and $\mathcal{N}(\mu_1,\sigma_1)$, respectively, where $\overline{\mathcal{H}_0}$ denotes the complementary (alternative) hypothesis of $\mathcal{H}_0$. Therefore, the false acceptance rate (FAR) and the false rejection rate (FRR) denoted by R_{FA} and R_{FR} can, respectively, be derived as

$$
\begin{aligned}
R_{\mathrm{FA}} &= \Pr(x < \tau|\overline{\mathcal{H}_0}) = \int_{-\infty}^{\tau} f(x|\overline{\mathcal{H}_0})\, dx \\
&= \frac{1}{2}\operatorname{erfc}\left(\frac{\mu_1 - \tau}{\sqrt{2}\sigma_1}\right)
\end{aligned} \tag{4.5}
$$

and

$$
\begin{aligned}
R_{\mathrm{FR}} &= \Pr(x \geq \tau|\mathcal{H}_0) = \int_{\tau}^{\infty} f(x|\mathcal{H}_0)\, dx \\
&= \frac{1}{2}\operatorname{erfc}\left(\frac{\tau - \mu_0}{\sqrt{2}\sigma_0}\right)
\end{aligned} \tag{4.6}
$$

where $\Pr(\cdot)$, $f(\cdot)$ and $\operatorname{erfc}(\cdot)$ denote the probability, probability density function and complimentary error function, respectively. It is clear from (4.5) and (4.6) that in order to complete the theoretical characterization of the R_{FA} and R_{FR}, it is worth to find theoretical expressions for μ_0, σ_0, μ_1 and σ_1. By assuming in the case when $\mathcal{H}_0$ is not valid that P_k and $\tilde{P}_k$ are independent and uniformly distributed, it is easy to show that $\mu_1 = 0.5$ and $\sigma_1 = 1/\sqrt{4N_p}$. However, in the case when $\mathcal{H}_0$ is valid, μ_0 and σ_0 are manipulation dependent.

4.2.3 *Experimental results*

For the implementation of the proposed technique, we use $N_f = 256$, $L = 1024$, $n = 6$ and $N = 16$. Thus, the size of the fingerprint in the case of the proposed technique is $N_p = 4096$ bits, which is significantly less than the sizes 8192 bits in [91,92] and 32768 bits in [26]. The proportionality constant α in (4.3) controls the quantization step sizes and thus has an important impact on the performance of the proposed technique. To find an appropriate value of α, we have carried out extensive experiments and observed that the use of $\alpha = 3$ provides a good compromise between FRR and FAR.

To evaluate the performance of the proposed technique, we use 1000 audio signals belonging to various genres, such as classical, jazz, pop, rock, and hip-hop. Then, we extract 1000 fingerprints from these signals using the proposed technique and calculate the Hamming distance between two different fingerprints for all possible combinations (inter-distances). This yields a total of $1000 \times 999/2 = 499500$ Hamming distances whose empirical mean and variance are $\hat{\mu}_1 = 0.4996$ and $\hat{\sigma}_1^2 = 6.14 \times 10^{-5}$, which are close to the theoretical mean and variance $\mu_1 = 0.5$ and $\sigma_1^2 = 6.10 \times 10^{-5}$, respectively. Fig. 4.3 shows the empirical histogram of the Hamming distances fitted by a scaled Gaussian PDF and demonstrates the validity of the assumption that these distances are Gaussian made in the previous section. Fig. 4.4 shows the empirical and theoretical FAR for different values of the threshold τ and confirms the validity of (4.5). The FAR for a threshold $\tau = 0.25$ is 5.45×10^{-225}, which demonstrates that the proposed technique achieves excellent discrimination between audio signals of different contents.

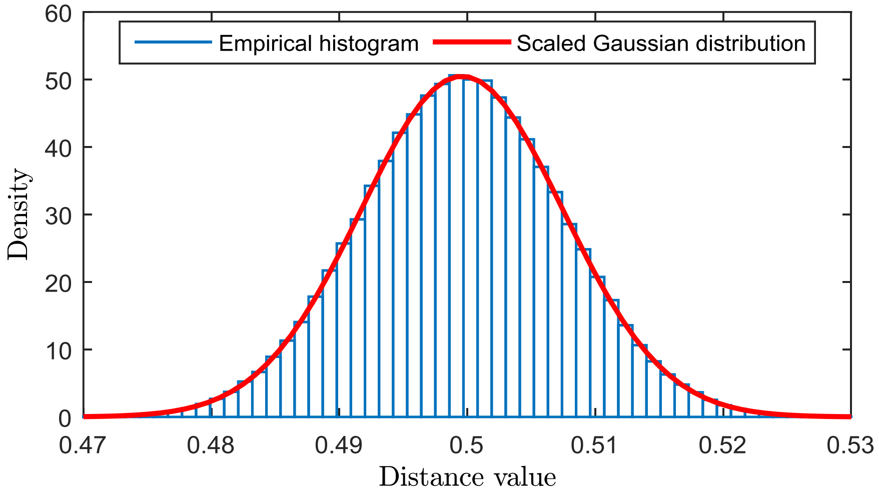


Fig. 4.3 Empirical histogram of the hamming distances fitted by a scaled Gaussian PDF.

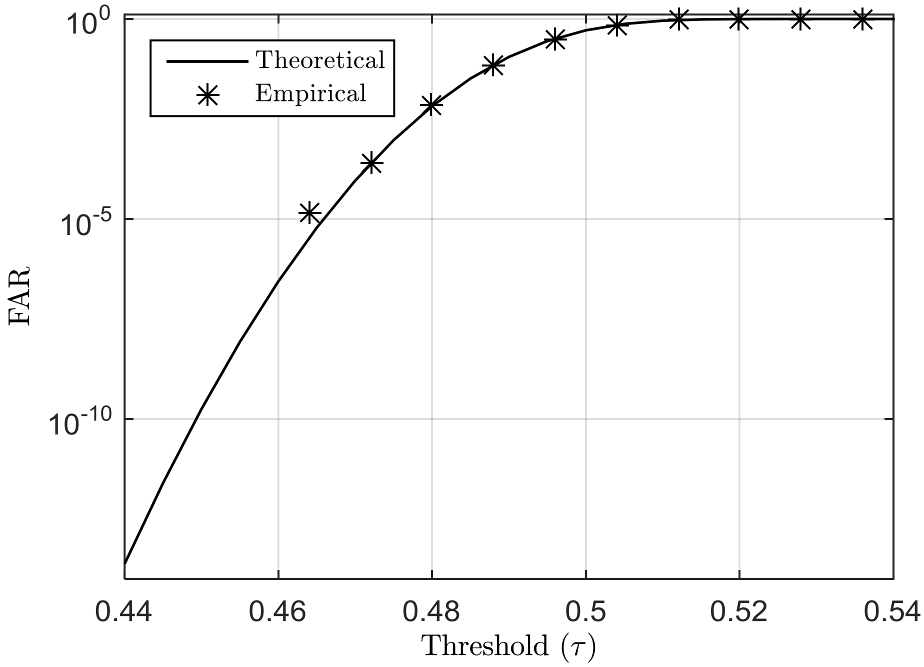


Fig. 4.4 Comparison between theoretical and empirical FAR values.

To assess the robustness of the proposed technique to noise addition, we attack the considered signals by a 10 dB AWGN and then calculate the Hamming distances between the fingerprints obtained from the original signals and those obtained from the corresponding noisy signals (intra-distances). The experimental mean and variance of the Hamming distances obtained in the case of the proposed technique are $\hat{\mu}_0 = 0.0442$ and $\hat{\sigma}_0^2 = 2.28 \times 10^{-4}$, respectively. For the purpose of comparison, we also implement the PRH-based fingerprinting techniques reported in [26,91,92]. Fig. 4.5 and Fig. 4.6 show, respectively, the empirical and theoretical FRR obtained by different techniques under a 10 dB AWGN attack as a function of the threshold value. It is clear from these figures that the proposed technique is more robust to AWGN than the existing PRH-based techniques. For instance, for a threshold $\tau = 0.25$, the theoretical FRR of the proposed technique is 1.78×10^{-42}, whereas those of the techniques in [26], [91] and [92] are 6.51×10^{-26}, 7.57×10^{-5} and 1.75×10^{-12}, respectively.

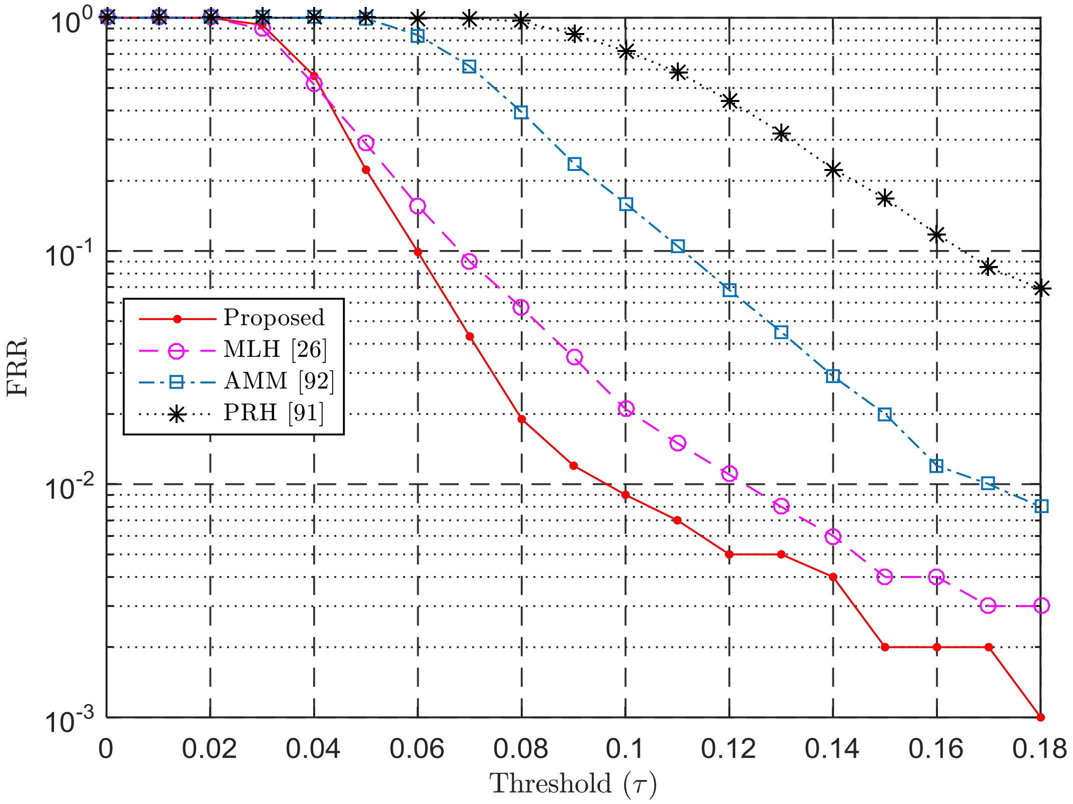


Fig. 4.5 Empirical FRR of different methods under a 10 dB AWGN attack.

Fig. 4.7 shows the inter- and intra-distance distributions for the techniques in [26,91,92] and the proposed technique. It can be seen from this figure that the proposed technique has a discrimination gap of 0.3188, whereas the gaps of [91] and [26] are 0.1819 and 0.2319, respectively, and are less than that of the proposed fingerprinting technique.

Table 4.1 shows the results of identification under some content preserving attacks and manipulations given in Table 3.2. It can be clearly seen from Table 4.1 that the proposed fingerprinting technique achieves higher identification rates than the techniques in [26,91,92].

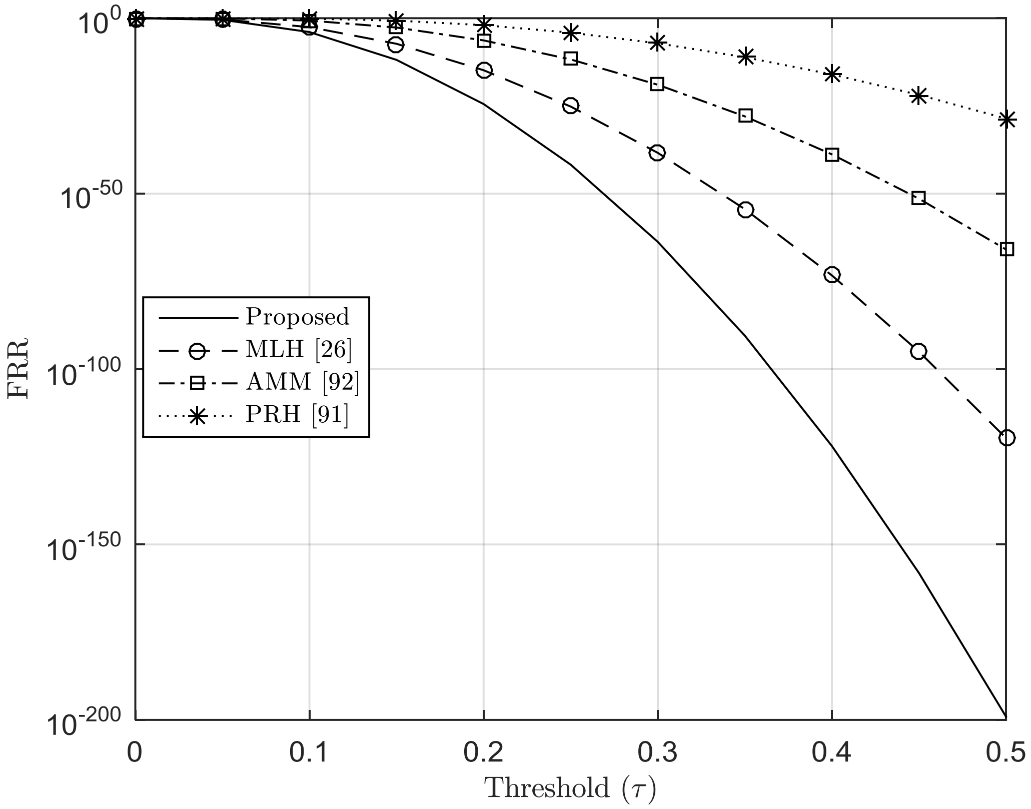
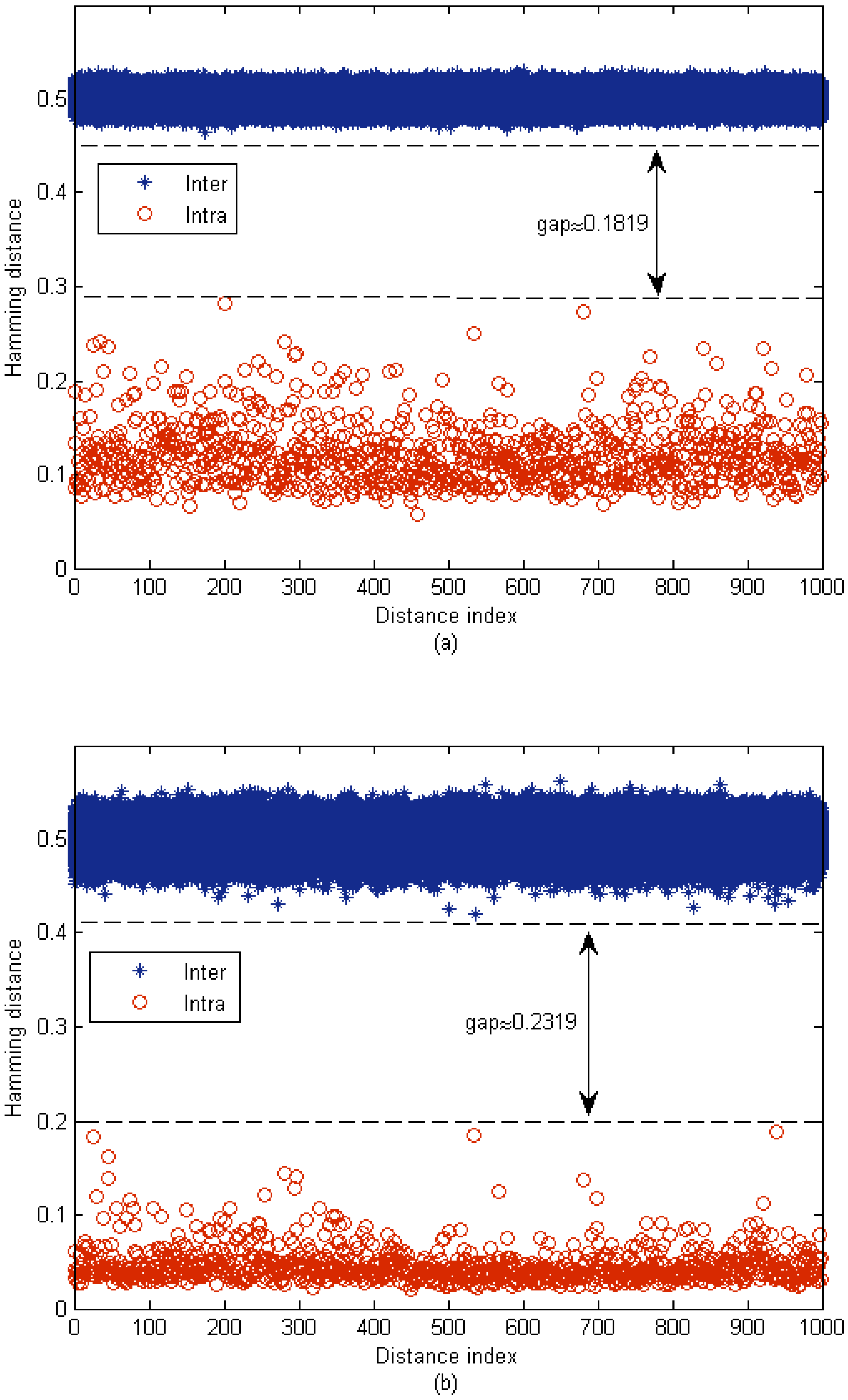


Fig. 4.6 Theoretical FRR of different methods under a 10 dB AWGN attack.

Table 4.1 Identification rates (%) of different techniques under various attacks (threshold $\boldsymbol{\tau = 0.25}$).

Attack	Proposed	PRH [91]	AMM [92]	MLH [26]
Req. (16→4)	99.9	98.1	98.57	91.12
Res.	100	98.91	99.45	99.82
LPF	100	100	100	99.65
MP3-96	99.89	99.77	99.81	99.85
MP3-64	99.05	95.73	95.98	96.50
Scale-150	100	100	100	100
Echo	99.97	99.18	99.48	99.78


Inter
Intra
gap≈0.1819
Hamming distance
0
0.1
0.2
0.3
0.4
0.5
0
100
200
300
400
500
600
700
800
900
1000
Distance index
Inter
Intra
gap≈0.2319
Hamming distance
Distance index

(a)

(b)

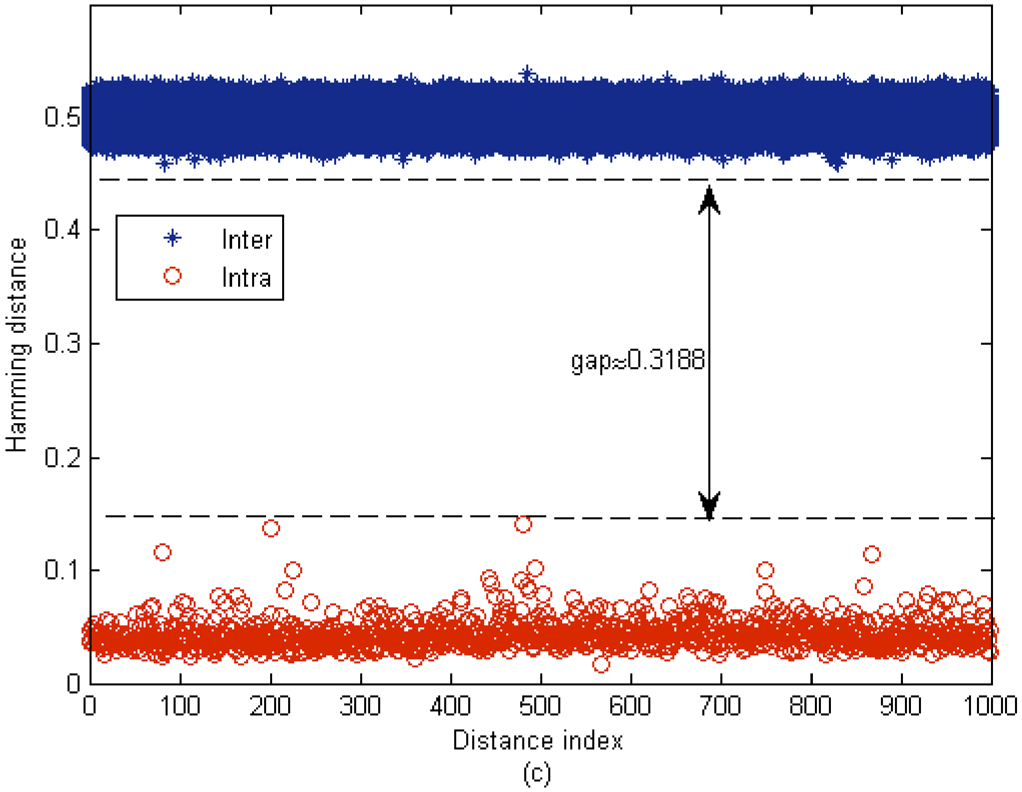


Fig. 4.7 Distribution of inter- and intra-distances: (a) PRH [91], (b) MLH [26], and (c) the proposed fingerprinting technique.

4.3 Proposed joint hashing-watermarking technique for audio fingerprinting

4.3.1 Proposed technique

The proposed joint hashing-watermarking technique is similar to the blind watermarking technique proposed in the previous chapter. However, instead of embedding an arbitrary watermark, we (i) use an approach similar to that devised for the proposed fingerprinting technique presented in the previous section to extract binary hashes from the designated samples that will carry the watermark, and (ii) instead of embedding a user-specified watermark, we use the extracted binary hashes as a watermark to be embedded using the technique given in the previous chapter.

Therefore, it is sufficient to replace the LMD-QIM in Step 4 of the embedding procedure given in the previous chapter by the one that we refer to as the hash QIM (H-QIM) as shown in Fig. 4.8. This H-QIM is depicted in Fig. 4.9 for the case of the k^{th} coefficient of the m^{th}

frame $F_k^{(m)}$. First, a binary hash $h^{(m)}$ is extracted using the dither coefficient $D^{(m)}$ and the quantization step Δ as

$$h^{(m)} = \underset{\eta\in\{0,1\}}{\operatorname{argmin}} \left(\left|F_k^{(m)} - Q_\Delta\left(F_k^{(m)}, \eta, D^{(m)}\right)\right|\right) \tag{4.7}$$

then, the extracted hash is embedded in the coefficient $F_k^{(m)}$ to obtain the watermarked coefficient $F'^{(m)}_k$ as

$$F'^{(m)}_k = Q_\Delta\left(F_k^{(m)}, h^{(m)}, D^{(m)}\right) \tag{4.8}$$

Note that Δ is calculated using (3.1), and the extraction procedure is identical to that of the technique given in the previous chapter.

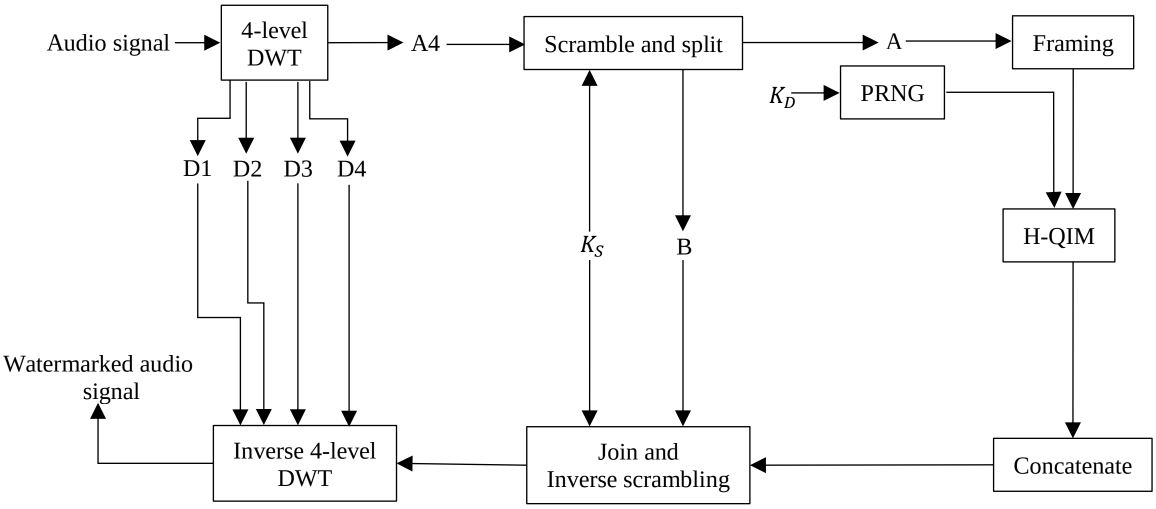


Fig. 4.8 Block diagram of the proposed joint hashing-watermarking procedure.

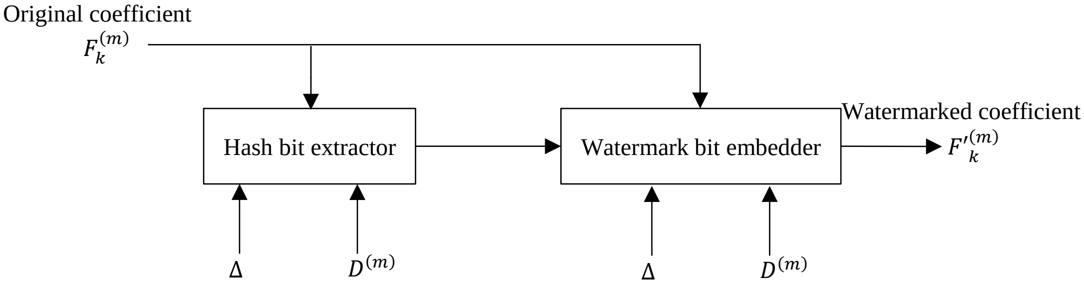


Fig. 4.9 Block diagram of the proposed H-QIM.

4.3.2 Experimental results

In order to evaluate the proposed joint hashing-watermarking technique, we use the same evaluation procedure used in the previous chapter, i.e., $SWR = 25$, capacities of 172 and 86 bps, and the attacks listed in Table 3.2. We compare the proposed joint hashing-watermarking technique (denoted by Proposed-H) with the blind watermarking technique proposed in the previous chapter (denoted by Proposed-0), and the results are given in Table 4.2 From this table, it can be seen that the Proposed-H is much more robust than Proposed-0 against all considered signal processing attacks and manipulations. Fig. 4.10 shows BER of the Proposed-0 and Proposed-H under an AWGN for different values of the WNR and SWRs of 55, 50, and 45 dB (embedding capacity of 172 bps), whereas Fig. 4.11 shows the BER of the Proposed-0 and Proposed-H under an AWGN for different values of the WNR and for capacities of 354 and 172 bps (SWR=45 dB). These figures clearly demonstrate that the Proposed-H is much more robust against AWGN than the Proposed-0. Furthermore, It can be seen from Fig. 4.12, which shows the SWR of the Proposed-H and Proposed-0 techniques for different values of the quantization step, that the Proposed-H achieves better transparency than the Proposed-0 even when operating at higher capacity, that is, it is seen from Fig. 4.12 that the Proposed-H at a capacity of 172 bps has higher SWR than the Proposed-0 at a capacity of 86 bps.

Since it has been shown in section 3.6 that the Proposed-0 is more robust than the techniques given in [13–15], and it is confirmed in this section that the Proposed-H is much more robust than the Proposed-0, it is clear that the Proposed-0 is significantly more robust than the techniques given in [13–15].

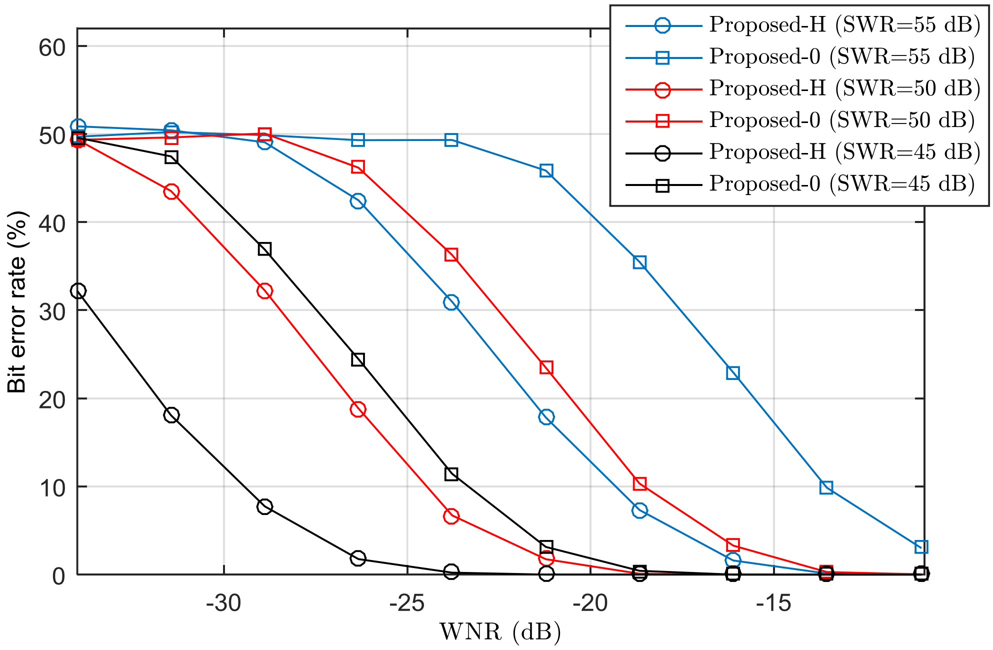


Fig. 4.10 BER of the Proposed-0 and Proposed-H under an AWGN for different values of the WNR and SWRs of 55, 50, and 45 dB (embedding capacity of 172 bps).

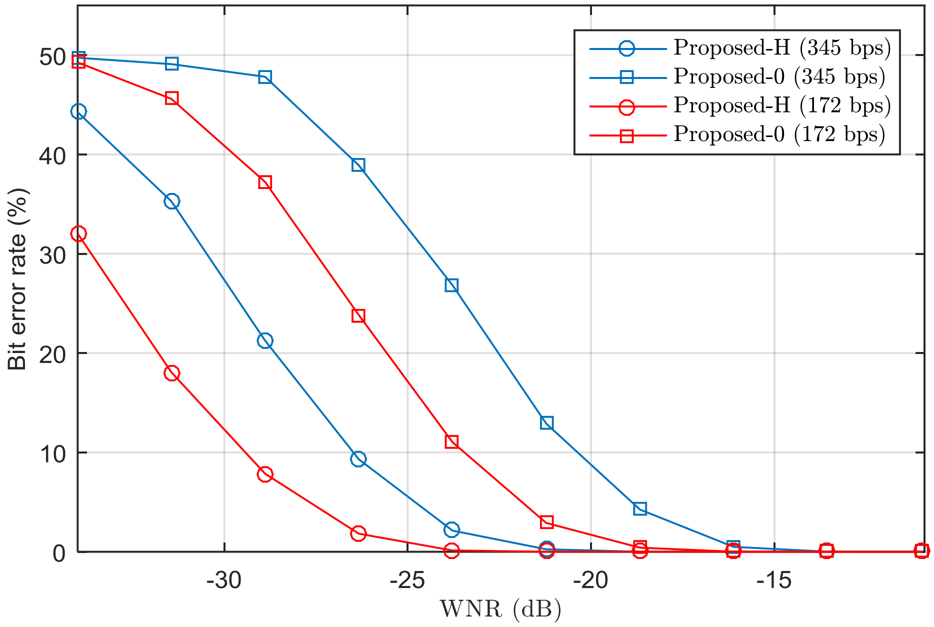


Fig. 4.11 BER of the Proposed-0 and Proposed-H under an AWGN for different values of the WNR and for capacities of 354 and 172 bps (SWR=45 dB).

Table 4.2 BER (%) of the Proposed-H and Proposed-0 under various attacks.

	Proposed-H			Proposed-P			Proposed-H	Proposed-P
Attack	S1	S2	S3	S1	S2	S3	Average over 20 signals	Average over 20 signals
Res.	0	0	0	0	0	0	0	0.00
Req. (16→8)	0	0	0	0	0	0	0	0.00
Req. (16→4)	0	0	0	0	0	0	0	0.73
LPF-8	0	0	0	0	0	0	0	0.00
Echo	0	0	0	0.024	1.587	0.708	0	0.53
Scale-150	0	0	0	0	0	0	0	0.00
Scale-50	0	0	0	0	0	0	0	0.00
UMN-0.1	0	0	0	0	0	0	0	0.00
UMN-0.5	0	0.05	0.02	0.683	1.416	1.196	0.03	0.87
AWGN-20	0	0	0	0	0	0	0	0.00
AWGN-10	0	0	0	0.073	0.049	0.024	0	0.05
AWGN-5	0.02	0	0.05	4.39	5.10	5.05	0.01	4.92
MP3-96	0	0	0	0	0	0	0	0.00
MP3-64	0	0	0	0	0.073	0.049	0	0.18

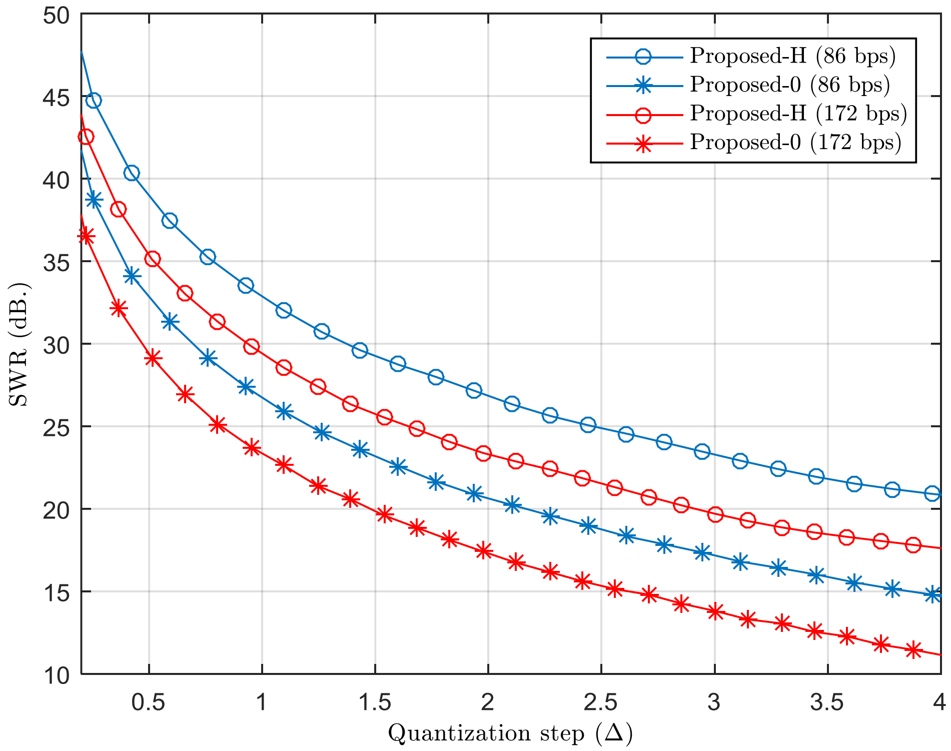


Fig. 4.12 SWR of the Proposed-H and Proposed-0 techniques for different values of the quantization step.

Finally, we use the Proposed-H to extract and embed fingerprints of sizes 2048 bits from each signal of the 1000 signal that has been used in subsection 4.2.3. Table 4.3 shows the identification rates of the Proposed-H technique under various attacks using the same discrimination procedure described in subsection 4.2.2 and a threshold value $\tau = 0.25$. This table together with Table 4.1 show that the Proposed-H technique has a much higher identification rates than robust hashing techniques while using smaller fingerprints.

Table 4.3 Identification rates (%) of the Proposed-H technique under various attacks (threshold $\boldsymbol{\tau = 0.25}$).

Attack	Identification rate
Req. (16→4)	100
Res.	100
LPF	100
MP3-96	100
MP3-64	100
Scale-150	100
Echo	100

4.4 Conclusion

An efficient key-dependent audio fingerprinting technique has been proposed in this chapter by exploiting the quantization minimum distance as a hash extractor in the DWT domain. Moreover, the initial condition of the chaotic map used to generate the dithers of the quantizer can be exploited as a secret key for fingerprint extraction. The quantization minimum distance hash extractor has been joined with one of our techniques to develop an efficient joint hashing-watermarking technique for audio fingerprinting. We have experimentally shown that the proposed fingerprinting technique achieves an excellent discrimination and outperforms the existing PRH-based techniques in terms of storage requirement, security, and robustness to AWGN and various content preserving manipulations and attacks. Moreover, the experimental results have shown that the proposed joint hashing-watermarking technique is highly robust against content preserving manipulations and attacks than the existing watermarking techniques while having a much higher transparency. Furthermore, it outperforms the conventional robust-hashing techniques while requiring a lower fingerprint size.

General conclusion

An extensive literature review has shown that QIM is a good class of watermarking methods that is widely popular in the development of new audio watermarking techniques. However, QIM suffers the drawback of the diversity of its realizations and lack of robustness against gain attacks, and although there exist many solutions to the gain attacks problem of QIM, the existing solutions are vulnerable to noise addition attacks. Consequently, our objective has been to develop new audio watermarking techniques in the frequency domain by exploiting the important characteristics of QIM while finding efficient solutions to eliminate its drawbacks.

In our first contribution, we have proposed a blind audio watermarking technique in the DCT domain by introducing a parametric QIM, then formulating watermarking as a mathematical optimization problem, i.e., maximizing robustness at any level of transparency. The parametric QIM has been proposed to solve the problem of the diversity of QIM realizations. Moreover, the solution of the optimization problem has served as a way to find values of the parameters for which the proposed parametric QIM is optimal. Furthermore, we have developed an approach to select the coefficients that carry the watermark in a manner that the watermark bits survive high- and low-pass filtering attacks. In addition, a fast implementation of the proposed technique has been introduced to reduce the computational complexity. Besides, theoretical expressions for the watermarking distortion and probability of error have been derived and experimentally verified.

In our second contribution, we have proposed a QIM-based technique that is robust against both gain and additive noise attacks, and unifies blind and semi-blind audio watermarking in a single framework. The robustness against noise addition and gain attacks has been achieved by proposing an expression for obtaining a quantization step that is near invariant under noise addition and renders the minimum distance decoder of QIM invariant under amplitude scaling. Moreover, to avoid the usage of an extra channel, we have proposed a side-information recovery procedure for the semi-blind mode of operation of the proposed technique. In addition, closed-form expressions of the watermarking distortion, probability of error under AWGN attack, and probability of error recovery of the proposed recovery procedure have been derived. These theoretical expressions have been validated by subjecting them to verification through comparison with empirical values.

Our third contribution has been on the application of watermarking in audio fingerprinting and consists of two parts: (i) we have proposed a quantization based robust-hashing technique for audio fingerprinting in the DWT domain, and (ii) we have proposed a joint hashing-watermarking technique for audio fingerprinting. The main idea of the hashing technique is to summarize the audio samples using the DWT and exploit the minimum distance decoder of QIM as a hash extractor. This idea has allowed us to successfully combine robust-hashing and watermarking to develop an efficient joint hashing-watermarking technique for audio fingerprinting, and hence to extremely reduce the embedding distortion and storage requirement while enhancing the robustness. Moreover, the theoretical FAR and FRR have also been derived.

Finally, the four techniques proposed in this thesis as well as various relevant audio watermarking robust-hashing techniques have been implemented and applied on different types of audio signals including human speech, songs, and several kinds of music such as jazz, classical, rock, etc. The experimental results and comparison with the existing techniques have shown the efficiency of: (i) of the parametrization of QIM, the optimization through Lagrange multipliers method and the approach of best embedding position proposed in the second chapter, (ii) the proposed approach for solving the gain and additive noise attacks problem of QIM, the unification of blind and semi-blind audio watermarking, and the side information recovery procedure proposed in the third chapter, and (iii) the combination of the DWT and the quantization minimum distance for robust audio hashing, and the joining of robust-hashing with watermarking for audio fingerprinting proposed in the fourth chapter. Moreover, various experiments have been conducted to verify and validate the theoretical expression developed in this thesis.

Future research directions may include:

- AQIM watermarking methods are of a great interest and it is required to find a solution to their vulnerability to additive noise attacks. Moreover, the conventional AQIM methods operate in two dimensions of spaces and thus deal only with one angle. Therefore, a generalization of AQIM methods to higher dimensions might be interesting.
- Until now, the most successful method against synchronization attacks (i.e., time scaling attacks) is the embedding of synchronization codes. However, this solution

is not very efficient, and it is desirable to find more efficient solutions against synchronization attacks for QIM-based techniques.

- On one hand, quantization is an essential part of lossy compression methods such as MPEG layer II and III, AAC, Vorbis, etc. On the other hand, quantization-based watermarking by QIM is highly efficient. Therefore, it is suitable to develop a QIM-based audio watermarking technique to be integrated into lossy compression systems. Such systems would significantly reduce the computational complexity compared with applying watermarking and compression separately.

Bibliography

[1] N. Cvejic, T. Seppänen, Digital Audio Watermarking Techniques and Technologies, IGI Global, Hershey, 2008. doi:10.4018/978-1-59904-513-9.

[2] I.J. Cox, M.L. Miller, The First 50 Years of Electronic Watermarking, EURASIP J. Adv. Signal Process. 2002 (2002) 820936. doi:10.1155/S1110865702000525.

[3] W.-T. Huang, S.-Y. Tan, Y.-J. Chang, C.-H. Chen, A robust watermarking technique for copyright protection using discrete wavelet transform, WSEAS Trans. Comput. 9 (2010) 485–495.

[4] T. Rabie, I. Kamel, High-capacity steganography: a global-adaptive-region discrete cosine transform approach, Multimed. Tools Appl. (2016). doi:10.1007/s11042-016-3301-x.

[5] Y. and W.H.A. Lin, Audio Watermark, A Comprehensive Foundation Using Matlab, 2015. doi:10.1007/978-3-319-07974-5_1.

[6] M.M. Soliman, A.E. Hassanien, H.M. Onsi, An adaptive watermarking approach based on weighted quantum particle swarm optimization, Neural Comput. Appl. 27 (2015) 469–481. doi:10.1007/s00521-015-1868-1.

[7] M.E. Farfoura, S.J. Horng, J.M. Guo, A. Al-Haj, Low complexity semi-fragile watermarking scheme for H.264/AVC authentication, Multimed. Tools Appl. 75 (2016) 7465–7493. doi:10.1007/s11042-015-2672-8.

[8] P.V. S., C.M. P. V. S. S. R., A robust semi-blind watermarking for color images based on multiple decompositions, Multimed. Tools Appl. (2017) 1–34. doi:10.1007/s11042-017-4355-0.

[9] X. Wang, P. Wang, P. Zhang, S. Xu, H. Yang, A norm-space, adaptive, and blind audio watermarking algorithm by discrete wavelet transform, Signal Processing. 93 (2013) 913–922. doi:10.1016/j.sigpro.2012.11.003.

[10] M. Zareian, H.R. Tohidypour, Robust quantisation index modulation-based approach for image watermarking, Image Process. IET. 7 (2013) 432–441. doi:10.1049/iet-ipr.2013.0048.

[11] T.M. Thanh, P.T. Hiep, T.M. Tam, K. Tanaka, Robust semi-blind video watermarking based on frame-patch matching, AEU - Int. J. Electron. Commun. 68 (2014) 1007–1015. doi:10.1016/j.aeue.2014.05.004.

[12] M.A. Akhaee, N. Khademi Kalantari, F. Marvasti, Robust audio and speech watermarking using Gaussian and Laplacian modeling, Signal Processing. 90 (2010) 2487–2497. doi:10.1016/j.sigpro.2010.02.013.

[13] S.-T. Chen, H.-N. Huang, Optimization-based audio watermarking with integrated quantization embedding, Multimed. Tools Appl. 75 (2016) 4735–4751. doi:10.1007/s11042-015-2500-1.

[14] H.T. Hu, L.Y. Hsu, Robust, transparent and high-capacity audio watermarking in DCT

domain, Signal Processing. 109 (2015) 226–235. doi:10.1016/j.sigpro.2014.11.011.

[15] Y.-G. Wang, G. Zhu, An improved AQIM watermarking method with minimum-distortion angle quantization and amplitude projection strategy, Inf. Sci. (Ny). 316 (2015) 40–53. doi:10.1016/j.ins.2015.04.029.

[16] G. Kasana, S.S. Kasana, Reference based semi blind image watermarking scheme in wavelet domain, Opt. - Int. J. Light Electron Opt. 142 (2017) 191–204. doi:10.1016/j.ijleo.2017.05.027.

[17] D.K. Thind, S. Jindal, A semi blind DWT-SVD video watermarking, in: Procedia Comput. Sci., Elsevier Masson SAS, 2015: pp. 1661–1667. doi:10.1016/j.procs.2015.02.104.

[18] M. Hemis, B. Boudraa, D. Megías, T. Merazi-Meksen, Adjustable audio watermarking algorithm based on DWPT and psychoacoustic modeling, Multimed. Tools Appl. 77 (2018) 11693–11725. doi:10.1007/s11042-017-4813-8.

[19] Z. Liu, A. Inoue, Audio Watermarking Techniques Using Sinusoidal Patterns Based on Pseudorandom Sequences, IEEE Trans. Circuits Syst. Video Technol. 13 (2003) 801–812. doi:10.1109/TCSVT.2003.815960.

[20] S. Wu, J. Huang, D. Huang, Y.Q. Shi, Efficiently self-synchronized audio watermarking for assured audio data transmission, IEEE Trans. Broadcast. 51 (2005) 69–76. doi:10.1109/TBC.2004.838265.

[21] X. Wang, P. Wang, P. Zhang, S. Xu, H. Yang, A blind audio watermarking algorithm by logarithmic quantization index modulation, Multimed. Tools Appl. 71 (2014) 1157–1177. doi:10.1007/s11042-012-1259-x.

[22] B. Chen, G.W. Wornell, Quantization index modulation: a class of provably good methods for digital watermarking and information embedding, IEEE Trans. Inf. Theory. 47 (2001) 1423–1443. doi:10.1109/18.923725.

[23] Y. Jiang, Y. Zhang, W. Pei, K. Wang, Adaptive spread transform QIM watermarking algorithm based on improved perceptual models, AEU - Int. J. Electron. Commun. 67 (2013) 690–696. doi:10.1016/j.aeue.2013.02.005.

[24] J.P. Boyer, P. Duhamel, J. Blanc-Talon, Scalar DC-QIM for semifragile authentication, IEEE Trans. Inf. Forensics Secur. 3 (2008) 776–782. doi:10.1109/TIFS.2008.2004285.

[25] J.P. Boyer, P. Duhamel, J. Blanc-Talon, Performance analysis of scalar DC-QIM for zero-bit watermarking, IEEE Trans. Inf. Forensics Secur. 2 (2007) 283–289. doi:10.1109/TIFS.2007.897279.

[26] Y. Liu, H.S. Yun, N.S. Kim, Audio fingerprinting based on multiple hashing in DCT domain, IEEE Signal Process. Lett. 16 (2009) 525–528. doi:10.1109/LSP.2009.2016837.

[27] F. Balado, N.J. Hurley, E.P. McCarthy, G.C.M. Silvestre, Performance analysis of robust audio hashing, IEEE Trans. Inf. Forensics Secur. 2 (2007) 254–266. doi:10.1109/TIFS.2007.897258.

[28] Y. Terchi, S. Bouguezel, A blind audio watermarking technique based on a parametric quantization index modulation, Multimed. Tools Appl. 77 (2018) 25681–25708. doi:10.1007/s11042-018-5813-z.

[29] Y. Terchi, S. Bouguezel, A QIM-based technique for robust blind and semi-blind audio watermarking, “Submitted to” Digit. Signal Process. (2018).

[30] Y. Terchi, S. Bouguezel, Key-dependent audio fingerprinting technique based on a quantisation minimum-distance hash extractor in the DWT domain, Electron. Lett. 54 (2018) 720–722. doi:10.1049/el.2018.0045.

[31] Y. Terchi, S. Bouguezel, Joint hashing-watermaring technique for audio fingerprinting, “Submitted to” IEEE Signal Process. Lett. (2018).

[32] G. Hua, J. Huang, Y.Q. Shi, J. Goh, V.L.L. Thing, Twenty years of digital audio watermarking—a comprehensive review, Signal Processing. 128 (2016) 222–242. doi:10.1016/j.sigpro.2016.04.005.

[33] A. Cheddad, J. Condell, K. Curran, P. Mc Kevitt, Digital image steganography: Survey and analysis of current methods, Signal Processing. 90 (2010) 727–752. doi:10.1016/j.sigpro.2009.08.010.

[34] T. Jahnke, J. Seitz, Digital Watermarking and Its Impact on Intellectual Property Limitation for the Digital Age, J. Electron. Commer. Organ. 3 (2005) 72–82. doi:10.4018/jeco.2005010105.

[35] N. Liu, P. Amin, A. Ambalavanan, K.P. Subbalakshmi, An Overview of Digital Watermarking, in: Multimed. Secur. Technol. Digit. Rights Manag., Elsevier, 2006: pp. 167–195. doi:10.1016/B978-012369476-8/50009-9.

[36] O.T.C. Chen, W.C. Wu, Highly robust, secure, and perceptual-quality echo hiding scheme, IEEE Trans. Audio, Speech Lang. Process. 16 (2008) 629–638. doi:10.1109/TASL.2007.913022.

[37] W.N. Lie, L.C. Chang, Robust and high-quality time-domain audio watermarking based on low-frequency amplitude modification, IEEE Trans. Multimed. 8 (2006) 46–59. doi:10.1109/TMM.2005.861292.

[38] Y. Xiang, I. Natgunanathan, D. Peng, W. Zhou, S. Yu, A dual-channel time-spread echo method for audio watermarking, IEEE Trans. Inf. Forensics Secur. 7 (2012) 383–392. doi:10.1109/TIFS.2011.2173678.

[39] X. Zhu, A.T.S. Ho, P. Marziliano, A new semi-fragile image watermarking with robust tampering restoration using irregular sampling, Signal Process. Image Commun. 22 (2007) 515–528. doi:10.1016/j.image.2007.03.004.

[40] N. Chen, H. Xiao, Perceptual audio hashing algorithm based on Zernike moment and maximum-likelihood watermark detection, Digit. Signal Process. 23 (2013) 1216–1227. doi:10.1016/j.dsp.2013.01.012.

[41] M. Arnold, M. Schmucker, S.D. Wolthusen, Techniques and Applications of Digital Watermarking and Content Protection, Artech House, Boston, 2003.

[42] X. He, Signal Processing, Perceptual Coding and Watermarking of Digital Audio, IGI Global, Hershey, 2012. doi:10.4018/978-1-61520-925-5.

[43] I.F. Jafar, K.A. Darabkh, R.T. Al-Zubi, R.R. Saifan, An efficient reversible data hiding algorithm using two steganographic images, Signal Processing. 128 (2016) 98–109. doi:10.1016/j.sigpro.2016.03.023.

[44] J. Zhou, W. Sun, L. Dong, X. Liu, O.C. Au, Y.Y. Tang, Secure Reversible Image Data Hiding Over Encrypted Domain via Key Modulation, IEEE Trans. Circuits Syst. Video Technol. 26 (2016) 441–452. doi:10.1109/TCSVT.2015.2416591.

[45] F. Hartung, M. Kutter, Multimedia watermarking techniques, Proc. IEEE. 87 (1999) 1079–1107. doi:10.1109/5.771066.

[46] Sang-Kwang Lee, Yo-Sung Ho, Digital audio watermarking in the cepstrum domain, IEEE Trans. Consum. Electron. 46 (2000) 744–750. doi:10.1109/30.883441.

[47] C. Baras, N. Moreau, P. Dymarski, Controlling the inaudibility and maximizing the robustness in an audio annotation watermarking system, IEEE Trans. Audio, Speech Lang. Process. 14 (2006) 1772–1782. doi:10.1109/TASL.2006.879808.

[48] Y. Liu, S. Tang, R. Liu, L. Zhang, Z. Ma, Secure and robust digital image watermarking scheme using logistic and RSA encryption, Expert Syst. Appl. 97 (2018) 95–105. doi:10.1016/j.eswa.2017.12.003.

[49] H.-T. Hu, L.-Y. Hsu, H.-H. Chou, Variable-dimensional vector modulation for perceptual-based DWT blind audio watermarking with adjustable payload capacity, Digit. Signal Process. 31 (2014) 115–123. doi:10.1016/j.dsp.2014.04.014.

[50] I.J. Cox, M.L. Miller, J. a Bloom, T. Kalker, J. Fridrich, Digital Watermarking and Steganography Second Edition, 2008. doi:10.1017/CBO9781107415324.004.

[51] A.G. Acevedo, Audio watermarking quality evaluation, E-Bus. Telecommun. Networks. (2006) 272–283.

[52] G. Stoll, F. Kozamernik, EBU listening tests on Internet audio codecs, EBU Tech. Rev. (2000) 24. doi:10.1049/cp:19971266.

[53] ITU-R, Recommendation BS.1116-1: Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems, Int. Telecommun. Union, Geneva. (1997) 1–26.

[54] Itu, BS.1534-1: Method for the subjective assessment of intermediate quality level of coding systems, Recomm. R. (2003) 1–18.

[55] Y. Lin, W. Abdulla, Objective quality measures for perceptual evaluation in digital audio watermarking, Signal Process. IET. 5 (2011) 623–631. doi:10.1049/iet-spr.2010.0069.

[56] S. Xiang, J. Huang, Histogram-based audio watermarking against Time-Scale Modification and cropping attacks, IEEE Trans. Multimed. 9 (2007) 1357–1372. doi:10.1109/TMM.2007.906580.

[57] W. Bender, D. Gruhl, N. Morimoto, A. Lu, Techniques for data hiding, IBM Syst. J. 35 (1996) 313–336. doi:10.1147/sj.353.0313.

[58] N.K. Kalantari, M.A. Akhaee, S.M. Ahadi, H. Amindavar, Robust multiplicative patchwork method for audio watermarking, IEEE Trans. Audio, Speech Lang. Process. 17 (2009) 1133–1141. doi:10.1109/TASL.2009.2019259.

[59] I. Natgunanathan, Y. Xiang, Y. Rong, W. Zhou, S. Guo, Robust patchwork-based embedding and decoding scheme for digital audio watermarking, IEEE Trans. Audio, Speech Lang. Process. 20 (2012) 2232–2239. doi:10.1109/TASL.2012.2199111.

[60] I.J. Cox, J. Kilian, F.T. Leighton, T. Shamoon, Secure spread spectrum watermarking for multimedia, IEEE Trans. Image Process. 6 (1997) 1673–1687. doi:10.1109/83.650120.

[61] H.S. Malvar, D.A.F. Florêncio, Improved spread spectrum: A new modulation technique for robust watermarkimg, IEEE Trans. Signal Process. 51 (2003) 898–905. doi:10.1109/TSP.2003.809385.

[62] R. Li, S. Xu, H. Yang, Spread spectrum audio watermarking based on perceptual characteristic aware extraction, IET Signal Process. 10 (2016) 266–273. doi:10.1049/iet-spr.2014.0388.

[63] H.J. Kim, Y.H. Choi, J. Seok, J. Hong, Audio Watermarking Techniques, in: Intell. Watermarking Tech., WORLD SCIENTIFIC, 2004: pp. 185–217. doi:10.1142/9789812562524_0008.

[64] M. Zareian, H. Hasani, Robust image watermarking based on quantization index modulation, IEEE Int. Conf. Commun. (2013) 2106–2110. doi:10.1109/ICC.2013.6654837.

[65] S.-T. Chen, G.-D. Wu, H.-N. Huang, Wavelet-domain audio watermarking scheme using optimisation-based quantisation, IET Signal Process. 4 (2010) 720. doi:10.1049/iet-spr.2009.0187.

[66] M. Hemis, B. Boudraa, T. Merazi-meksen, Intelligent Audio Watermarking Algorithm using Multi-objective Particle Swarm Optimization, in: 2015 4th Int. Conf. Electr. Eng., IEEE, 2015: pp. 0–4. doi:10.1109/INTEE.2015.7416777.

[67] Y. Xiang, D. Peng, I. Natgunanathan, W. Zhou, Effective pseudonoise sequence and decoding function for imperceptibility and robustness enhancement in time-spread echo-based audio watermarking, IEEE Trans. Multimed. 13 (2011) 2–13. doi:10.1109/TMM.2010.2080668.

[68] B.Y. Lei, I.Y. Soon, Z. Li, Blind and robust audio watermarking scheme based on SVDDCT, Signal Processing. 91 (2011) 1973–1984. doi:10.1016/j.sigpro.2011.03.001.

[69] D. Renza, D.M. Ballesteros, H.D. Ortiz, Text Hiding in Images Based on QIM and OVSF, IEEE Lat. Am. Trans. 14 (2016) 1206–1212. doi:10.1109/TLA.2016.7459600.

[70] Y. Yubao Bai, S. Sen Bai, G. Guibin Zhu, C. Chunyan You, B. Bowen Liu, A blind audio watermarking algorithm based on FFT coefficients quantization, in: 2010 Int. Conf. Artif. Intell. Educ., IEEE, 2010: pp. 529–533. doi:10.1109/ICAIE.2010.5640958.

[71] Q. Li, I.J. Cox, Using perceptual models to improve fidelity and provide resistance to valumetric scaling for quantization index modulation watermarking, IEEE Trans. Inf. Forensics Secur. 2 (2007) 127–138. doi:10.1109/TIFS.2007.897266.

[72] A. Al-Haj, A dual transform audio watermarking algorithm, Multimed. Tools Appl. 73 (2014) 1897–1912. doi:10.1007/s11042-013-1645-z.

[73] M. Clerc, J. Kennedy, The particle swarm-explosion, stability, and convergence in a multidimensional complex space, IEEE Trans. Evol. Comput. 6 (2002) 58–73. doi:10.1109/4235.985692.

[74] F. Marini, B. Walczak, Particle swarm optimization (PSO). A tutorial, Chemom. Intell. Lab. Syst. 149 (2015) 153–165. doi:10.1016/j.chemolab.2015.08.020.

[75] R.C. Eberhart, Yuhui Shi, Particle swarm optimization: developments, applications and resources, Proc. 2001 Congr. Evol. Comput. (IEEE Cat. No.01TH8546). 1 (2001) 81–86. doi:10.1109/CEC.2001.934374.

[76] B. Chen, G.W. Wornell, Quantization index modulation: A class of provably good methods for digital watermarking and information embedding, IEEE Trans. Inf. Theory. 47 (2001) 1423–1443. doi:10.1109/18.923725.

[77] N.K. Kalantari, S.M. Ahadi, A Logarithmic Quantization Index Modulation for Perceptually Better Data Hiding, IEEE Trans. Image Process. 19 (2010) 1504–1517. doi:10.1109/TIP.2010.2042646.

[78] B. Wah, T. Wang, Efficient and Adaptive Lagrange-Multiplier Optimization, J. Glob. Optim. (1999) 1–25. doi:10.1023/A.1008203422124.

[79] M. Puschel, Cooley-Tukey FFT like algorithms for the DCT, in: 2003 IEEE Int. Conf. Acoust. Speech, Signal Process. 2003. Proceedings. (ICASSP ’03)., IEEE, 2003: p. II-501-4. doi:10.1109/ICASSP.2003.1202413.

[80] B. Chen, G.W. Wornell, Preprocessed and postprocessed quantization index modulation methods for digital watermarking, in: P.W. Wong, E.J. Delp III (Eds.), SPIE - Secur. Watermarking Multimed. Contents II, 2000: pp. 48–59. doi:10.1117/12.384995.

[81] M.A. Akhaee, S.M.E. Sahraeian, C. Jin, Blind Image Watermarking Using a Sample Projection Approach, IEEE Trans. Inf. Forensics Secur. 6 (2011) 883–893. doi:10.1109/TIFS.2011.2146250.

[82] F. Pérez-González, C. Mosquera, M. Barni, A. Abrardo, Rational Dither Modulation: A high-rate data-hiding method invariant to gain attacks, IEEE Trans. Signal Process. 53 (2005) 3960–3975. doi:10.1109/TSP.2005.855407.

[83] F. Ourique, V. Licks, R. Jordan, F. Perez-Gonzalez, Angle QIM: A Novel Watermark Embedding Scheme Robust Against Amplitude Scaling Distortions, in: Proceedings. (ICASSP ’05). IEEE Int. Conf. Acoust. Speech, Signal Process. 2005., IEEE, 2005: pp. 797–800. doi:10.1109/ICASSP.2005.1415525.

[84] E. Nezhadarya, Z.J. Wang, R.K. Ward, Robust Image Watermarking Based on Multiscale Gradient Direction Quantization, IEEE Trans. Inf. Forensics Secur. 6 (2011) 1200–1213. doi:10.1109/TIFS.2011.2163627.

[85] A. Al-Haj, An imperceptible and robust audio watermarking algorithm, EURASIP J. Audio, Speech, Music Process. 2014 (2014) 37. doi:10.1186/s13636-014-0037-2.

[86] M.A. Akhaee, N.K. Kalantari, F. Marvasti, Robust Multiplicative Audio and Speech Watermarking Using Statistical Modeling, in: 2009 IEEE Int. Conf. Commun., IEEE, 2009: pp. 1–5. doi:10.1109/ICC.2009.5199424.

[87] M. Shaked, On the Distribution of the Minimum and of the Maximum of a Random Number of I.I.D. Random Variables, in: A Mod. Course Stat. Distrib. Sci. Work, Springer Netherlands, Dordrecht, 1975: pp. 363–380. doi:10.1007/978-94-010-1842-5_29.

[88] R.S. Stanković, B.J. Falkowski, The Haar wavelet transform: its status and achievements, Comput. Electr. Eng. 29 (2003) 25–44. doi:10.1016/S0045-7906(01)00011-8.

[89] S.E. Azoug, S. Bouguezel, A non-linear preprocessing for opto-digital image encryption using multiple-parameter discrete fractional Fourier transform, Opt. Commun. 359 (2016) 85–94. doi:10.1016/j.optcom.2015.09.054.

[90] H. Zhou, X.T. Ling, Problems with the chaotic inverse system encryption approach, IEEE Trans. Circuits Syst. I Fundam. Theory Appl. 44 (1997) 268–271. doi:10.1109/81.557386.

[91] J. (Philips) Haitsma, A Highly Robust Audio Fingerprinting System, Ircam. (2002) 2742648.

[92] J.S. Seo, An asymmetric matching method for a robust binary audio fingerprinting, IEEE Signal Process. Lett. 21 (2014) 844–847. doi:10.1109/LSP.2014.2310237.

[93] J. Haitsma, T. Kalker, A Highly Robust Audio Fingerprinting System, Proc. Int. Conf. Music Inf. Retr. (2002).

Abstract

In this thesis, we present three contributions in the field of digital audio watermarking in the frequency domain. In the first contribution, we propose a parametric quantization index modulation (QIM) for which we find the optimal values of the parameters using the Lagrange multipliers method. Moreover, we present an approach for selecting the embedding positions in the discrete cosine transform (DCT) domain that gives the immunity against low- and high-pass filtering attacks. Furthermore, a fast algorithm for the proposed technique is developed. In the second contribution, we introduce a QIM-based technique that unifies blind and semi-blind audio watermarking in the discrete wavelet transform (DWT) domain. We also propose an expression that gives a quantization step, which is invariant under additive white Gaussian noise and renders the minimum distance decoder of QIM invariant under amplitude scaling of the host signal. Moreover, to avoid the usage of an extra channel, we propose an efficient side-information recovery procedure for the semi-blind mode of operation of the proposed technique. In the third contribution, we propose a robust-hashing method by using QIM's minimum distance decoder in the DWT domain, then we propose a joint hashing-watermarking technique for audio fingerprinting to enhance the perceptual quality and robustness while reducing the storage requirement compared with other audio watermarking techniques. Finally, we conduct various experiments to validate the correctness of the theoretical expressions derived in this thesis and also to evaluate the proposed techniques and compare them with existing relevant techniques.

Keywords: Audio watermarking, Quantization index modulation, Robust-hashing, fingerprinting.

Résumé

Dans cette thèse, nous présentons trois contributions dans le domaine du tatouage audio numérique dans le domaine fréquentiel. Dans la première contribution, nous proposons une modulation d'indice de quantification paramétrique (QIM) pour laquelle nous trouvons les valeurs optimales des paramètres en utilisant la méthode des multiplicateurs de Lagrange. De plus, nous présentons une approche pour sélectionner les positions d'insertion dans le domaine de transformée en cosinus discrète (DCT) qui confère l'immunité contre les attaques de filtrage passe-bas et passe-haut. De plus, un algorithme rapide pour la technique proposée est développé. Dans la deuxième contribution, nous introduisons une technique basée sur QIM pour unifie le tatouage audio aveugle et semi-aveugle dans le domaine de la transformation discrète en ondelettes (DWT). Nous proposons également une expression qui donne un pas de quantification, qui est invariant sous le bruit gaussien blanc additif et rend le décodeur par distance minimale de QIM invariant au changement d'échelle du signal hôte. De plus, pour éviter l'utilisation d'un canal supplémentaire, nous proposons une procédure efficace de récupération d'informations latérales pour le mode de fonctionnement semi-aveugle de la technique proposée. Dans la troisième contribution, nous proposons une méthode de hachage robuste en utilisant le décodeur par distance minimale de QIM dans le domaine DWT, puis nous proposons une technique de hachage-tatouage conjoint pour améliorer la qualité perceptive et la robustesse tout en réduisant les besoins de stockage par rapport aux autres techniques de tatouage audio. Enfin, nous menons diverses expériences pour valider l'exactitude des expressions théoriques dérivées dans cette thèse et aussi pour évaluer les techniques proposées et les comparer avec les techniques pertinentes existantes.

Mots-clés: Tatouage audio, modulation par indice de quantification, hachage robuste, empreintes digitales.

ملخص

في هذه الأطروحة ، نقدم ثلاث مساهمات في ميدان وسم الصوت الرقمي في مجال التردد. في أول مساهمة ، نقترح تمييزا لمعادلة التضمين بمؤشر التكميم (QIM) والذي نجد من أجله القيم المثلى للمعلمات باستخدام طريقة مضاعفات لاجرانج. علاوة على ذلك ، نقدم طريقة لاختيار مواقع التضمين في مجال تحويل جيب التمام المنفصل (DCT) و هو ما يمنح المناعة ضد هجمات الترشيح المنخفضة والعالية. علاوة على ذلك، تم تطوير خوارزمية سريعة للتقنية المقترحة. في الإسهام الثاني ، نقدم تقنية تعتمد على QIM والتي توحد وسم الصوت الأعمى وشبه الأعمى في مجال تحويل المويجات المنفصلة (DWT). كما نقترح تعبيرًا يعطي خطوة تكميم ثابتة في ظل الضوضاء الغاوسية المضافة وتجعل وحدة فك ترميز المسافة الدنيا الخاص بـ QIM لا يتأثر بتأثر سعة الإشارة المضيفة. علاوة على ذلك ، لتجنب استخدام قناة إضافية ، فإننا نقترح إجراءًا فعالًا لاستعادة المعلومات الجانبية للطريقة التشغيل شبه العمياء الخاصة بالتقنية المقترحة. في المساهمات الثالثة ، نقترح طريقة تجزئة قوية باستخدام مفكك تشفيرالحد الأدنى الخاص بـ QIM في نطاق DWT ، ثم نقترح تقنية مشتركة بين الوسم و التجزئة و ذلك لغرض تعزيز الجودة المدركة والمتانة مع تقليل متطلبات التخزين مقارنة مع تقنيات وسم الصوت التقليدية. وأخيرًا ، نجري تجارب مختلفة للتحقق من صحة التعبيرات النظرية المشتقة في هذه الرسالة ، وكذلك لتقييم التقنيات المقترحة ومقارنتها بالتقنيات الحالية ذات الصلة.

كلمات مفتاحية: وسم الصوت، التضمين بمؤشر التكميم ، تجزئة قوية ، وبصمات الكترونية.