Title	動的位相変調に基づいたロバスト音響データハイディングとその応用
Author(s)	Ngo, Minh Nhut
Citation	
Issue Date	2015-09
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/12963
Rights	
Description	Supervisor:鵜木　祐史，情報科学研究科，博士

Robust Audio Data Hiding Based on Dynamic Phase Manipulation and Its Applications

Nhut Minh Ngo

Japan Advanced Institute of Science and Technology

Doctoral Dissertation

Robust Audio Data Hiding Based on Dynamic Phase Manipulation and Its Applications

Nhut Minh Ngo

Supervisor: **Associate Professor Masashi Unoki**

School of Information Science
Japan Advanced Institute of Science and Technology

September 2015

Abstract

Recent years have seen a rapid development of multimedia communication technologies which facilitate our life, but at the same time put security of digital audio at many risks, such as copyright infringement, malicious tampering. Audio data hiding techniques, in which special codes are embedded into actual audio content without any affection to its normal use, have been proposed as a potential solution for these issues. In general, audio data hiding methods must satisfy five requirements: (i) inaudibility—keeping watermark imperceptible to users, (ii) blindness—avoiding using double storage and communication channels, (iii) robustness—preventing intentional attacks from illegal users, (iv) high capacity—conveying a large amount of data, and (v) high reliability—precisely detecting the watermark.

Although most reported methods could partly satisfy the requirements, the trade-off among the requirements is still challenging. To keep the watermark inaudible, it is straightforward that perceptually insensitive features of audio signals should be exploited for embedding. The resistance against modifications such as lossy compression becomes weak since the hidden data could be easily destroyed without degrading the sound quality. Finding out suitable acoustic features to ensure both inaudibility and robustness simultaneously is one of the most important tasks of audio data hiding design.

The aim of this research is to propose an audio data hiding method that achieves a reasonable trade-off among the requirements and is applicable in practical problems. This research introduces a concept of dynamic phase manipulation for audio watermarking in which the human sound-perception mechanism and sophisticated embedding rule are utilized to solve the conflict among the requirements. First, the dynamic phase manipulation scheme finds out frequency components that are insensitive to human ears and resistant against signal processing operations to keep hidden data inaudible and robust. Second, an appropriate embedding rule is employed to account for blindness and high embedding capacity. Accordingly, the proposed method of audio data hiding could obtain the inaudibility and the robustness simultaneously. The proposed method is then applied to three typical applications: copy prevention, annotated audio, and information carrier over AM radio broadcast.

The phase manipulation technique is used to embed a bit of data into an audio signal by changing the phase according to two phase patterns. The amount of phase modification is an important factor which directly decides the performance of the data hiding system on the inaudibility and the robustness. The smaller amount keeps the hidden data less audible but more weak against processing and vice versa. The main goal of this research is to find out a region of frequency components suitable for embedding and corresponding amount of phase-modification for the embedding region based on the characteristics of human auditory system (HAS) and the variability of audio signals.

According to these considerations, the dynamic phase manipulation scheme is constructed as follows. Original audio is firstly analyzed to find out suitable frequency region for embedding. The phase modification of a frequency component cause distortion in a manner that is directly proportional to the magnitude of that component. Therefore, the amount of phase-modification should also be adapted to the magnitude. The amount of phase-modification is determined based on the energy of the embedding region. The modified phase spectrum and original magnitude spectrum are processed to yield a watermarked signal. In data extraction process, the same analysis steps are performed to identify the embedding frequency components and the amount of phase-modification. Watermark decoder is then performed on the phase of these components to extract embedded data. Experimental results have shown that variant amount of phase-modification improves performance on inaudibility and robustness remarkably compared with the case that a fixed amount of phase-modification is used. It suggested that the dynamic phase manipulation scheme is effective for audio data hiding.

The proposed framework ensures the inaudibility and the robustness by exploiting the advantages human perception mechanism and the variability of audio signals. The blindness and the embedding capacity are achieved by the nature of the embedding rules. To combat against cropping and shifting attacks, the proposed framework is built with a frame synchronization scheme. The frame synchronization is performed by searching the starting point around the size of one frame. A correct starting point is detected when the confidence of extracting a bit is the highest among all the points in that frame. Confidentiality is ensured by incorporating security parameters. Watermark is encrypted with a secret key before being embedded into audio signal.

The proposed framework is evaluated with respect to inaudibility, robustness, blindness, and embedding capacity. The inaudibility is confirmed by objective difference grade and subjective listening test. The robustness is confirmed by the accuracy of extracted data against signal processing operations and attacks. Subjective test and robustness test are also carried out to confirm the effectiveness of the dynamic phase manipulation scheme. Bit rate is varied to investigate embedding capacity as well. The proposed audio data hiding method is then applied to protecting digital audio, audio entertainment, and information carrier over AM radio broadcast.

Keywords: audio data hiding, robustness, reliability, adaptive phase modulation, quantization index modulation.

Acknowledgments

"If you want to go fast, go alone. If you want to go far, go together."—African Proverb. I could not sit down here to complete the final part of my PhD course without guidance, support, help, and encouragement of many people. I would like to express my deep gratitude to all those who gave me the possibility to complete this dissertation.

First, I would like to give my special thanks to my supervisor, Associate Professor Masashi Unoki whose guidance and support have helped me in all the time of my study and research. I sincerely appreciate discussion, suggestion, and encouragement that Associate Professor Unoki has given me in my PhD course. His good advice has always opened a way for me to learn many skills as well as knowledge during my research. His dedicated help has also made my daily life in JAIST much easier from the first day to the present.

I am also grateful to my vice supervisor, Professor Masato Akagi who always welcomes any question of mine. I am thankful Professor Akagi for always spending his precious time giving me many helpful comments and invaluable suggestions when I presented my progress in the Lab. meetings. I am deeply impressed with Professor Akagi's enthusiasm and willingness when I want to discuss with him about background knowledge related to my research or freely talk to him about any daily life topic.

I greatly appreciate Associate Professor Brian Michael Kurkoski for his supervision of my minor research. I would like to take this opportunity to thank Associate Professor Kurkoski for frequently spending time on discussion with me and giving me helpful suggestions during the time I did my minor research. I acknowledge Associate Professor Kurkoski for his valuable comments on the manuscript of my paper and my dissertation. I would like to thank Professor Akinori Ito of Tohoku University for his precious comments on my dissertation.

I would also like to express my thanks to all kind colleagues and friends in my laboratory who have made my research life in JAIST colorful. Especially, I appreciate huge support with kindness and willingness of my tutor, Mr. Yasuaki Kanai in first days I came to JAIST. I am also thankful to Ms. Rieko Kubo for her kind help with my troubles in daily life. I would like to thank Assistant Professor Ryota Miyauchi for his welcoming answers to my research questions. I am grateful my lab mates who have paid a lot of efforts to organize many enjoyable events for everyone in the Lab.

Life without finance would be much more difficult. I pay my acknowledgment to the supporting grants: Doctoral Research Fellow (DRF)—giving me support for tuition fee and daily expenditures during my PhD course, JASSO grant for student exchange—opening an opportunity for me to improve my English skill abroad, A3 Foresight program—giving me many chances to attend workshops and meetings, ICT Global Leader protect—providing me a travel fund to attend the IWDW 2014 conference in Taiwan, JAIST Research Grant—offering me a travel fund to attend the ICASSP 2015 conference in Australia, and Telecommunications Advancement Foundation—giving me a travel grant to attend the EUSIPCO 2015 conference in France. All those grants gave me an amazing opportunity to learn, study, and do research in an academic and professional environment.

I owe it to Associate Professor Thuc Dinh Nguyen (HCMC University of Science, Vietnam) sincere gratitude for giving me the chance to open all other chances. I am really impressed with heartfelt assistance and care that Associate Professor Nguyen has paid for me and other students of his.

I would like to thank all my dear friends in JAIST who have shared joyful and memorable moments in daily life with me. I also thank all my distant friends in Vietnam who have always listened to my stories and had unforgettable night conversation with me. My stresses were gone quickly with their timely appearance.

Last but not least, I would like to give many thanks to my family. I would like thank my parents for their unconditional love and the lifetime efforts that they have been paying for me, my brother, and my sister. I deeply appreciate my sister who has always taken care of me with endless love. I appreciate great assistance that my brother have given me whenever I need. I also appreciate all other members in my family who always support and stand beside me.

Contents

List of Figures

List of Tables

Glossary

Technical Terms

BCH	a class of error-correcting codes, invented by Alexis Hocquenghem, Raj Bose, and D. K. Ray-Chaudhuri.
blind detection	the ability of detecting the watermark without using of the host signal.
direct current	used to refer to the zero-frequency component. This term is borrowed from electronics where it is used to refer to the constant, zero-frequency, or slowly-variant local mean value of a current.
embedding capacity	the watermark should convey a high enough amount of information for the intended applications. Embedding capacity is often measured by the number of bits per second (bps).
fingerprint	a short numeric sequence associated with a digital content and is used to identify the digital content.
fingerprinting	a process that embeds a unique identity of digital data as watermark into digital content. In the literature of information security, it may be referred to as a technique that associates a digital content to a much shorter numeric sequence, namely fingerprint and use this sequence to identify the digital content.
frame synchronization	the ability to detect positions of frames automatically in order to ensure that watermark could be correctly detected.
host sigal	a signal that is used to convey a watermark. The watermark is embedded into the host signal by a watermark embedder. It may also be referred to as original signal or unwatermarked signal.

inaudibility — the ability of keeping the watermark inaudible to listeners during normal use. It requires that the perceived sound quality of watermarked signals should be as good as that of host signals. The presence of watermarks should not lower sound quality of host signals.

MP3 — MPEG-1 or MPEG-2 Audio Layer III, a standard of digital-audio lossy-compression, designed by the Moving Picture Experts Group (MPEG).

MP4 — MPEG-4 HE-AAC (High-Efficiency Advanced Audio Coding), an audio coding format for digital-audio lossy-compression, defined as an MPEG-4 Audio profile.

pirate — a person who illegally copies multimedia content and sells them or a person or intentionally performs signal processing operations on a watermarked signal in order to destroy the embedded watermark.

reliability — the watermark should be extracted with acceptable error rates. This requirement is usually accompanied with or implicitly included in robustness and high embedding capacity, which means that improving robustness and increasing embedding capacity must keep the error rates in reasonable limit.

robustness — the watermark should be withstanding against signal processing operations and intentional attacks from illegal users. Watermarked audio signals may be processed before, during, and after distribution over the Internet. The processing includes operations to improve the quality of the audio signal or lossy compression to reduce the size. Intentional attacks are the processing that attempts to invalidate or remove the embedded watermark.

watermarked sigal	a signal that is conveying a watermark. A host signal after embedded with a watermark becomes a watermarked signal. It may also be referred to as embedded signal or stego signal.
zero-contour	refer to the circle going over a zero of the transfer function of an IIR filter in z-plane. This circle's radius is equal to the modulus of the complex zero of the transfer function.

Notation

β	complex pole of IIR APF
β^*	complex conjugate of β
f_{s}	sampling frequency
$x(n)$	host audio signal
$x_i(n)$	host-signal frame
$H(z)$	transfer function
r	modulus of IIR APF's pole
N_{bpf}	number of bits in per frame
N_{bps}	number of bits per second (bps)
θ	angle of IIR APF's pole-zero
s	watermark, a string of bits
$\hat{s}$	detected watermark
$y(n)$	watermarked audio signal
$y_i(n)$	watermarked frame

Acronym and Abbreviation

AM	amplitude modulation
APF	all-pass filter
APM	adaptive phase modulation
AWGN	additive white Gaussian noise
BDR	bit detection rate
BER	bit-error rate
bps	bit per second
CD	cochlear delay
CDb	cochlear delay watermarking method with blind detection
CZT	chirp-z transformation
D/A–A/D	digital-to-analogue and analogue-to-digital
dB	decibel
DC	direct current
DCT	discrete cosine transform
DFT	discrete Fourier transform
DPC	dynamic phase coding
DRM	digital right management
DSB-SC	double-sideband suppressed carrier
DSB-WC	double-sideband with carrier
DSS	direct spread spectrum
DSSS	direct sequence spread spectrum
DWT	discrete wavelet transform
ECC	error control coding
FFT	fast Fourier transform
FM	frequency modulation
HAS	human auditory system

IFFT	inverse fast Fourier transform
IFPI	International Federation of the Phonographic Industry
IIR	infinite impulse-response
ISS	improved spread spectrum
LS-bit	least significant bit
LSB	lower sideband
LSD	log-spectral distortion
MCLT	modulated complex lapped transform
non-OLA	non-overlap and add
ODG	objective difference grade
OLA	overlap and add
PEAQ	perceptual evaluation of audio quality
PSK	phase-shift keying
QIM	quantization index modulation
RWC	Real World Computing
SER	signal-to-error ratio
SNR	signal-to-noise ratio
SNS	social networking service
SPC	static phase coding
SPL	sound pressure level
SPM	static phase modulation
SS	spread spectrum
SSB	single-sideband modulation
SVD	singular value decomposition
SVR	support vector regression
SWR	signal-to-watermark ratio

TV	television
USB	upper sideband

Chapter 1

Introduction

1.1 Importance of Research

Recent development in digitalization and the Internet technologies has made life much easier. Most multimedia content such as text, photo, audio, and video are newly created in digital form. Even old content that was born in analog form can be easily converted into machine-readable formats [1]. In the digital revolution, a major landmark was the vast transition of music from analog to digital records thanks to the invention of optical compact discs [2, 3]. In combination with explosive growth of social networking services (SNSs), multimedia content can be shared over the global network in second and nearly at zero cost. However, the ease of digital world has also been a challenge for traditional ways of managing intellectual property [4, 5].

Enforcing evil users to obey copyright law in digital world has proven incredibly difficult. The nature of digital content, which is inherently easily copyable, mutable, and transferable, exposes many ways of illegally using and maliciously manipulating copyrighted digital content. For instance, there is a widespread social concern about piracy of popular music and film. The International Federation of the Phonographic Industry (IFPI) pointed out that 95 % of the worldwide music downloaded via the Internet were pirated in 2010. Global sales of recorded music dropped about 10 % because compensation of revenue from digital services was insufficient for a successive fall in sales of compact discs in 2009 [6]. Zentner found that there has been a positive correlation between progress in the broadband Internet and reductions in music sales recently [7]. Consequently, combating against copyright infringement of digital property has become urgent demand of the world industry.

Solutions such as copy restriction, encryption, and addition of copyright proof into the file header have been proposed for copyright protection of digital multimedia over recent years. Digital right management (DRM) utilizes techniques of copy restriction that are able to restrict playback or copy of digital content to protect its copyright. DRM software is installed on users' computer to prevent widespread copy of digital files in a manner that users are able to view or listen to content on their computer but cannot copy or distribute it. However, DRM is unable to stop sharing digital files across computers or playing music on certain MP3 players that lack DRM software because digital files were born to be easily copied [8]. Copyright protection based on encryption is inconvenient for legal users. Although encryption can securely conceal data from unauthorized users, it requires authorized users to decrypt data before playing or viewing. The technique of adding copyright proof as meta-data into file header has also been considered as well. However, it is insecure because the codes can easily be removed once the file header is analyzed. These techniques have limitation in many aspects, hence have not become a good method of copyright protection for digital multimedia data.

Contrary to these techniques, digital data hiding, a technique that embeds watermark into actual content of digital data without any effect to its normal use, has been proposed as a promising solution for copyright protection [9]. Watermark could be ownership information of digital content so that it can be used to verify the ownership in a copyright dispute. The embedding procedure is implemented in a manner that watermark is kept transparent to users, i.e., digital data after embedded can be viewed or played without any difference compared with unprocessed digital data. Since the watermark is embedded into the actual content of digital data, but not its bit representation, this technique assures that watermark is secure against removal. Any attempt to destroy the watermark may cause the content severely destroyed too.

Originally proposed for copyright protection, but digital data hiding has also been considered as potential solutions for other issues such as usage tracking, copy protection, tampering detection, covert communication, and added information for entertainment [10, 11]. Usage tracking is enabled when creators sell a copy of their digital data embedded with a unique identity. Based on embedded identity creators can determine a specific customer who may be making unauthorized copies and selling them to others for profit. Watermark can be used as a fingerprint of digital audio for examination of copyrighted data [12]. Fragile watermark can be used to detect malicious tampering in digital data [13]. Covert communication involves secure information transmission, e.g., transactions between military and state offices, banks and other financial organizations [14]. Watermark can also be used to enable program broadcaster to collect information about customers' interest on television (TV) programs or used as complementary information for services such as advertisement, weather broadcasting [15].

Digital data hiding is, therefore, an indispensable technology nowadays. As audio and music have been playing an important role in human daily lives from the most ancient cultures, this thesis focuses on digital data hiding for audio signals.

1.2 Challenges with Audio Data Hiding

As a potential and promising solution for many issues arising in the digital era, audio data hiding has been studied for decades but there are still many challenges that need to be addressed. Literature has seen a variety of audio data hiding algorithms in recent years [9, 10, 16–24], but most of them are not fully satisfied necessary requirements to be applied in practice.

To be effectively applied in practice, a method of audio data hiding should satisfy following general requirements [9]:

Inaudibility: keeping the watermark inaudible to listeners during normal use. It requires that the perceived sound quality of watermarked signals should be as good as that of host signals. The presence of watermarks should not lower sound quality of host signals. This requirement is important because high sound quality keeps the commercial value of host signals. It also helps conceal watermarks in host signals securely. If the presence and precise location of watermarks are revealed, they can be exploited for malicious attacks to distort the embedded watermarks by pirate.

Robustness: the watermark should be withstanding against signal processing operations and intentional attacks from illegal users. Watermarked audio signals may be processed before, during, and after distribution over the Internet. The processing includes operations to improve the quality of the audio signal or lossy compression to reduce the size. A pirate may attempt to apply signal processing operations to destroy the watermark without severely distorting the quality of watermarked signal. Common signal processing operations that could be applied to watermarked signals are discussed by Cox and Linnartz [25]. This requirement ensures that the watermarking method is invulnerable to deliberate attempts to forge, remove, or invalidate watermarks.

It is reasonable to assume that such processing, regardless of intention, should not severely distort the sound quality of watermarked signals. Digital multimedia data were created to represent and transmit information at an high level of resolution. Low quality of audio signals greatly reduces their commercial value [26].

Blind detection: detecting the watermark without the need of original signals. In general, there are two types of detection scheme: blind detection and non-blind detection. As its name would suggest, a blind watermarking scheme can extract watermarks without using the original signal. In contrast, a non-blind watermarking scheme needs to use the original signal to extract watermarks in detection process.

Non-blind watermarking may not be so practical because it requires double storage and double transmission for watermark detection. However, using original signals in detection process could offer a possibility of enhancing performance of watermarking system, e.g., increasing detection accuracy. Non-blind watermarking scheme may be suitable for applications such as: verifying the ownership of digital content in a copyright dispute [27], usage tracking where sellers embed customers' information as watermarks into host signals and extract watermarks to determine the customer who made unauthorized duplications.

Frame synchronization: automatically detecting positions of frames. Most methods of audio data hiding is implemented on a frame-basis. In practice, frame position may be unavailable in scenarios when we need to detect watermark from live streaming sources or watermarked signals that are subjected to time cropping attack. This requirement enables frame positions to be synchronized so that watermarks could be correctly detected. Frame synchronization should work properly even when the watermarked signal is affected by innocent signal-processing operations and harmful attacks.

Embedding capacity: the watermark should convey a high enough amount of information for the intended applications. Embedding capacity is often measured by the number of bps. Given that other requirements have been satisfied, the higher the bit rate, the better the watermarking system. It is noteworthy that higher bit rate should not degrade the sound quality of watermarked signals. This requirement is especially important for applications such as covert communication that need large-sized watermarks. Even image or other audio might need to be embedded in such applications.

Reliability: the watermark should be extracted with acceptable error rates. This requirement is usually accompanied with or implicitly included in robustness and high embedding capacity, which means that improving robustness and increasing embedding capacity must keep the error rates in reasonable limit.

Besides the main requirements, there are other subsidiary requirements such as confidentiality—securely concealing of the watermark, low computational complexity—useful for real-time processing and implementation on low-power devices, reversibility—ability to recover the host signal from the watermarked signal, or fragility—applicable for integrity-verification applications, requiring watermark to be robust against innocent processing but fragile against malicious attacks. These requirements help increase applicability of methods of audio watermarking in practice.

All these requirements are respected to a certain extent, depending on the application. Different classes of application demand different types of watermarking schemes with different sets of requirement. For example, resistance to signal processing operations such as filtering, resampling, or compression is mostly necessary. Robustness against malicious attacks that try to invalidate watermark detection must be the first priority for copyright protection. For integrity-verification applications, the watermark should, however, be no longer detectable when audio content is intentionally modified. While some applications (such as information carrier) need watermark with high bit rate, watermark like numeric sequence is adequate for others (such as fingerprinting).

The solution is very hard because there is a trade-off in these requirements, especially inaudibility, robustness, and embedding capacity. It is straightforward that perceptually insensitive features should be exploited for embedding watermarks. But this is a challenge for robustness, since processing can distort the watermark without degrading the sound quality. Increasing embedding capacity adds more data into the host signal, causing more distortion in the watermarked signal and making the watermark likely less recoverable.

Sensitivity of human auditory system

The human auditory system (HAS) has been proven incredibly sensitive to sound in real environment [28]. The frequency range of human hearing is generally considered to be from 20 Hz to 20 kHz, and it is much more sensitive to sounds between 1 kHz and 4 kHz. The ear can perceive sounds as low as sound pressure level (SPL) of 0 dB at 3 kHz, but require SPL of 40 dB at 100 Hz with an amplitude increase of 100 times. The ear can discriminate two tones which differ in frequency by more than about 0.3% at 3 kHz. Human ear has an amazingly wide dynamic range of pressure sensitivity. It is capable of hearing the sounds with the lowest amplitude and the highest amplitude different in order of one-million. It adapts to environment very well; the ear drum vibrates even less than the diameter of a single molecule, when listening to very quiet sounds. Due to the sensitivity of the HAS, embedding inaudible watermarks into an audio signal is a challenging task.

Variety of common signal-processing operations

Digital audio can be subjected to a variety of signal-processing operations such as analysis-by-synthesis, domain transformation, and filtering that aim to both refine its quality and compress the size. These processing operations modify insignificant components of the signal in such a way that does not induce change in perception to human ear. However, they may accidentally destroy the watermark because it is usually embedded in perceptual insignificant portion of audio signal or they may even be exploited by attackers to invalidate the watermark. Therefore, finding out a suitable portion of audio signals for embedding robust watermark is very important for the design of watermarking algorithm.

Trade-off in requirements

Applications of audio data hiding could be categorized into two classes: copyright-related applications and those related to added-values. For the first one, robustness against malicious attacks is essentially important. To preserve the high quality, perceptual insignificant features are chosen for embedding. However, the features can also be changed by attackers without degrading the sound quality. The watermark must be securely concealed in the host signal so that it is very difficult for attackers to find out a way for impairing it.

For applications related to added-values, harmful attack is not the problem because it just adds the complementary value. However, high capacity of watermark is necessary to convey data as much as possible, which leads to a trade-off between inaudibility and high embedding capacity. It is obvious that the larger the watermark is embedded, the more distortion the audio signal suffers. This trade-off must also be tackled in building a new data hiding method for digital audio.

1.3 Motivation and Research Aims

As the promising advantages and potential applications of audio data hiding, it is urgent to conduct research on it to develop a new technology for protecting copyright, integrity, and value of digital audio. For years, researcher have been working on audio data hiding, but there is still not an efficient method applicable for practical problems. We usually meet a trade-off in the requirements when implementing a watermarking system for digital audio. Selecting suitable acoustic features for watermarking that satisfy both inaudibility and robustness is an important task for the design of watermarking algorithms.

Phase has intensively been explored for audio data hiding in recently due to its advantages in robustness and inaudibility. Although human ear is relatively sensitive to phase of audio signals, perceptible distortion would not occur if phase relations between frequency components are not drastically changed [29–31]. Furthermore, the distortion caused by phase-modification of a frequency component is directly related to its magnitude. According, the amount of phase-modification should be adjusted to the magnitude. More precisely, higher-magnitude frequency components would have slight phase-modification whereas, smaller-magnitude frequency components would have relatively higher phase-modification. As long as the modification of phase is kept sufficiently small, distortion due to watermark is under reasonable level.

On the other hand, most audio signals in realistic environment are non-stationary, i.e., their characteristics are time-variant. However, most of the reported methods of audio data hiding have not taken the variability of audio signals into consideration, which remarkably influence the performance of the whole system. Depending on local characteristics of the audio signal, a suitable phase-manipulation scheme should be figured out in order to ensure the best performance of the embedding system.

Based on these considerations, this dissertation aims to propose an efficient and practicable method of audio data hiding. The proposed method would be responding enough necessary requirements for the demanding applications in practice. In order to reach the ultimate goal, the proposed method is expected to achieve the following objectives:

(i) Obtaining a reasonable trade-off among the requirements. The highest-priority property that the proposed method would have is inaudibility. For all the applications, the existence of embedded watermark must be transparent to normal use of digital audio content. The watermark algorithm must be withstanding common signal-processing operations to ensure watermark correctly extracted. The robustness against malicious attacks and the high capacity are supposed to trade together. Depending on a specific class of application, i.e., protection related or add-values related, these properties is balanced for satisfying necessary application requirements.

In order to attain the reasonable properties, acoustic features is thoroughly investigated and chosen the suitable one. The implicit property, blind-detection ability, is remarked by choosing a sophisticated embedding rule that is able to detect watermark with the absence of the host signal. Knowledge from psychoacoustic study is exploited to better assure high quality of the watermarked signal. Furthermore, watermark is embedded into resistant portion of audio signals to attain robustness. The amount of modification of embedding components is adaptively adjust to local characteristics of audio signals to ensure both inaudibility and robustness.Finally, the watermarking algorithm is expected to be flexible in a manner that is able to adjust its properties to meet the demanded application.

(ii) Wide applicability to practical problems. The proposed method would be capable of being a solution for resolving the urgent social problems. The proposed method is expected to obtain a reasonable trade-off in the requirements to have enough necessary properties for the applications of protecting digital audio and adding complementary values toward enriching audio content.

The first application would be a scheme for prevention of uploading copyrighted digital audio to sharing services like YouTube. Nowadays, the viral broadcast of copyrighted audio content is done mainly thanks to the popularity of SNS and cloud data services. To solve this problem, copyright mark is embedded as watermark in the digital audio content. When a user upload an audio file with copyright mark, the sharing service can examine the copyright mark and stops the uploading and give certain further actions on this user. The watermark is expected to be very robust against attacks to secure the copyright mark. It is also expected to not cause audible distortion in the audio signal in order to preserve its commercial value. The main usefulness of this application is helping prevent widespread of copyrighted digital audio via sharing services in the Internet. Its limitation is not able to stop sharing directly between users.

The second application would use watermark as a channel carrying complementary information for digital audio content. Content information about an audio signal is interesting to users, especially in music applications, helping enrich users' experience on listening to music. Another usage scenario is to enable broadcasters to enhance entertainment experiences for their consumers. The watermark is expected to have high capacity in order to convey data as much as possible. Inaudibility is also desired to maintain the high sound quality.

1.4 Thesis Outline

This dissertation is composed of six chapters. A schematic overview of this dissertation is shown in Fig. 1.1. Chapters 2–6 are organized as follows.

Chapter 2 presents background knowledge related to this study and key principles of the proposed methods. This chapter starts with detailed descriptions about audio data hiding and literature review of this research field. State-of-the-art methods of audio data hiding are surveyed and their remaining issues are pointed out as well. Next, important background knowledge related to the proposed methods including phase characteristics and human perception on phase of audio signals is presented. Then, general scheme of audio data hiding based on phase manipulation is described and key principles for efficient methods of audio data hiding based on dynamic phase manipulation are proposed.

Chapter 3 describes the concept of watermarking based on adaptive phase modulation and the implementation. This chapter proposes a method of audio data hiding based on adaptive phase modulation. Audio signals are usually non-stationary, i.e., their own characteristics are time-variant. The features for watermarking are usually not selected by combining the principle of variability, which affects the performance of the whole watermarking system. The proposed method embeds watermark into an audio signal by adaptively modulating its phase with the watermark using IIR all-pass filters. The frequency location of the pole-zero of IIR all-pass filter which characterizes the transfer function of the filter is adapted based on signal power distribution on sub-bands in magnitude spectrum domain. The pole-zero locations are adapted in a manner that the phase modulation produces slight distortion in watermarked signals to achieve the best sound quality. Reasonable trade-off between inaudibility and robustness could be obtained by balancing the phase modulation scheme.

Chapter 4 provides description on the concept of audio data hiding based on dynamic phase coding and its implementation. This chapter proposes an audio watermarking method based on dynamic phase coding and error control coding. The technique of quantization index modulation is employed for embedding watermarks into the phase spectrum of audio signals. Since most of audio information is distributed in somewhat low frequencies, to increase robustness, this frequency region is chosen for embedding watermarks. Phase modification causes sound distortion in a manner that is proportional to the magnitude. Therefore, the amount of phase modification is adjusted according to the magnitude to balance inaudibility and robustness. Error control coding is incorporated to further increase reliability of watermark detection.

Chapter 5 introduces three applications of the proposed methods. The applicability of the proposed methods is further verified by practical applications. The first application is a copy prevention scheme that enables sharing services to prevent users from uploading copyright audio content. The second application is capable of adding complementary information for audio content toward enriching digital audio. The last application uses watermarks as complementary information for radio programs in radio broadcasting systems, e.g., advertisement, weather forecasting, and emergency alert warnings.

Finally, **Chapter 6** summarizes this study, gives some concluding remarks, highlights our contributions to the research field, and opens a number of directions for future research.

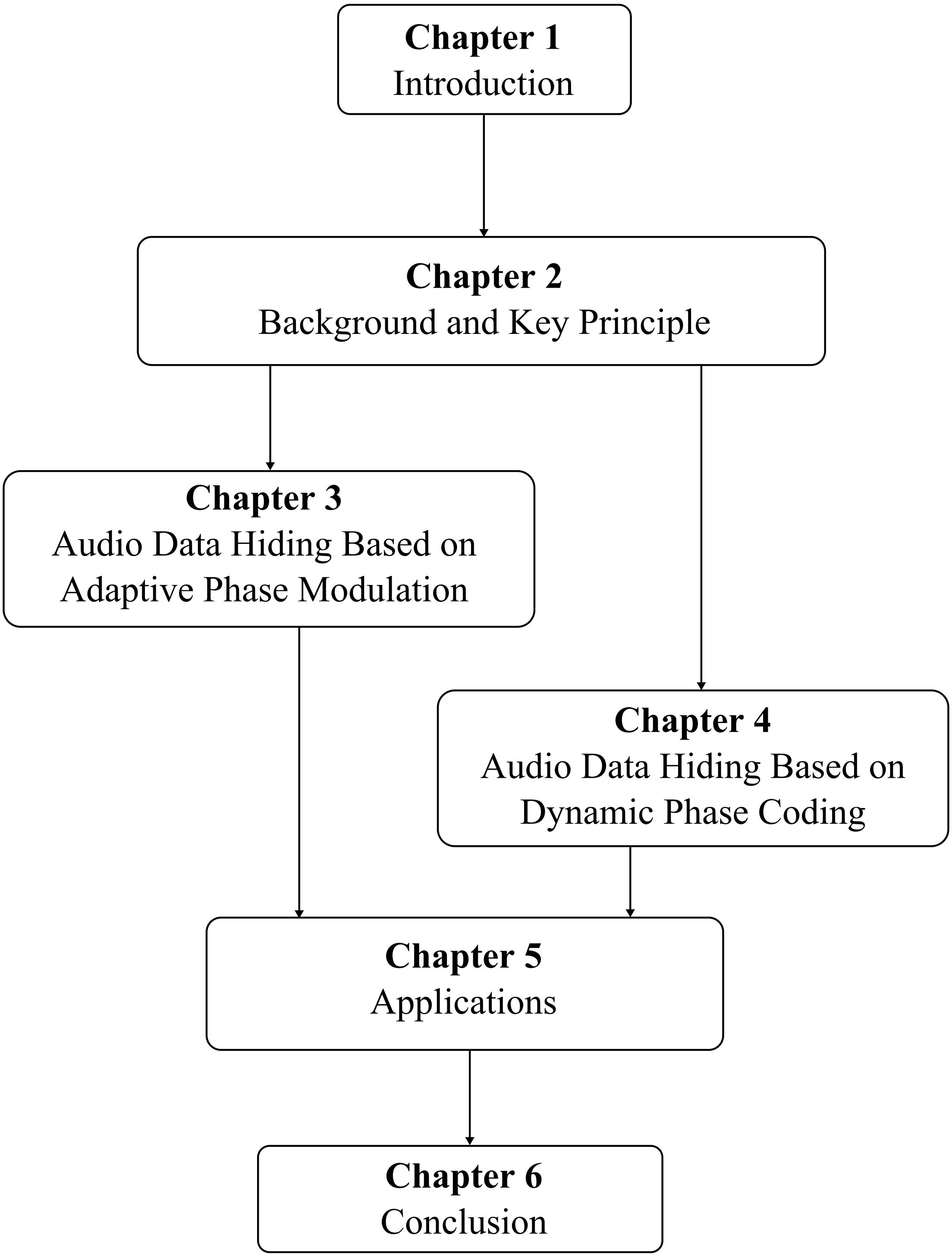


Figure 1.1: Schematic outline of the dissertation

Chapter 2

Background and Key Principle

2.1 Overview of Audio Data Hiding

2.1.1 General scheme

Figure 2.1 depicts a general scheme of audio data hiding which consists of two processes: watermark embedding and watermark detection. At embedding side, a watermark, s, is embedded into a host audio signal, $x(n)$, by a watermark embedder which produces an output, namely watermarked audio signal, $y(n)$. The watermarked signal is sent to others and may suffer from some kinds of processing or attacks. At detection side, receivers extract the watermark, $\hat{s}$, by using a watermark detector in which the input is the watermarked signal or both the host signal and the watermarked signal.

Usually, the host signal is split into frames, in which a watermark-bit or a few watermark-bits are embedded into each frame. All watermarked frames are then combined together by non-overlap and add (non-OLA) or certain frame-transition technique. In that manner, the watermarked signal is also split into frames to extract the bits in every frames. And accordingly, frame positions need to be detected before watermark-detection process. Automatic frame-synchronization is thus an important issue should be taken into account for the design of watermarking algorithm.

2.1.2 Performance evaluation

Research of audio data hiding has been conducted for about twenty years, resulting in several approaches with numerous methods. To have a better view on effectiveness and performance of the reported methods of audio data hiding in the literature, this section starts with an overview on performance evaluation of audio data hiding. Objective evaluations such as the signal-to-watermark ratio (SWR), log-spectral distortion (LSD) [32], and perceptual evaluation of audio quality (PEAQ) [33] are usually used to measure the

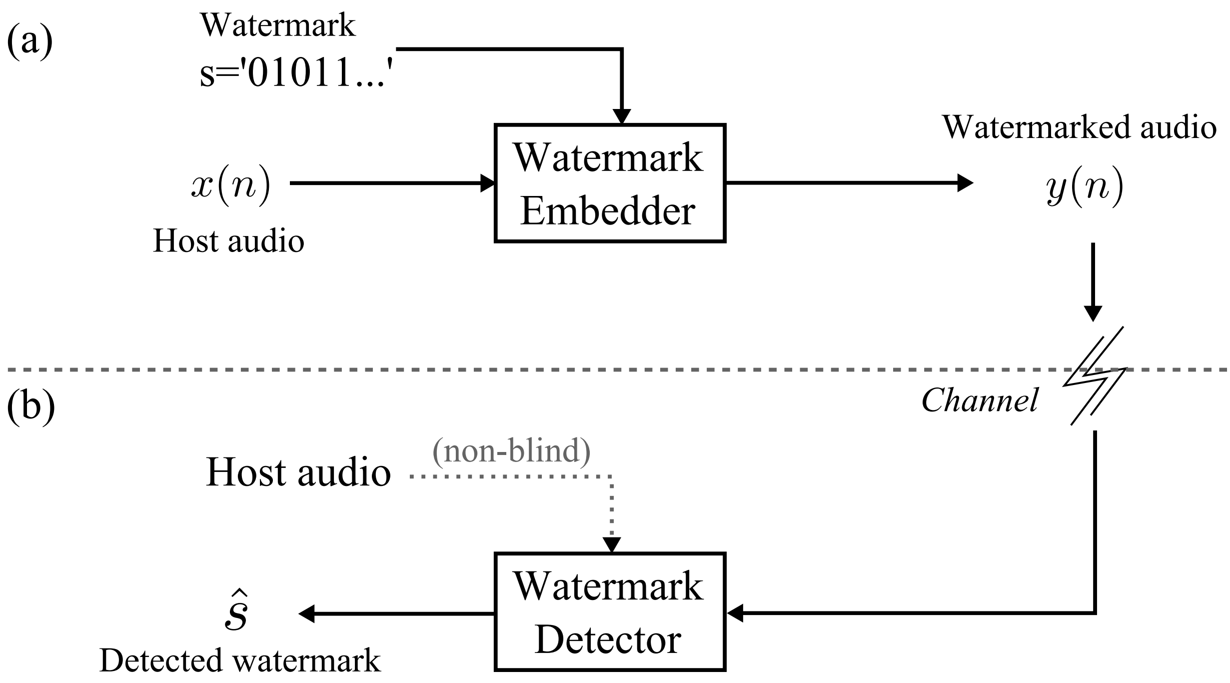


Figure 2.1: Overview of an audio data hiding system: (a) process at embedding side and (b) process at detection side.

sound quality of the watermarked signals. Bit detection rate (BDR) is used to measure the accuracy of watermark detection process.

Signal-to-watermark ratio

SWR was used to compare the level of a host signal to the level of watermark, which may be referred to as signal-to-noise ratio (SNR) or signal-to-error ratio (SER). A higher SWR signal indicated better sound quality. SWR is defined in dB by:

$$\mathrm{SWR}(x, y) = 10 \log_{10} \left(\frac{\sum_{n=1}^{N} x^2(n)}{\sum_{n=1}^{N} (y(n) - x(n))^2} \right) \quad \text{(dB)}, \tag{2.1}$$

where $x(n)$ is the host signal, $y(n)$ is the watermarked signal, and N is the number of samples.

Log-spectral distortion

LSD was used to measure the distance or distortion between two spectra. A lower LSD value indicates a better result. LSD is defined by:

$$\mathrm{LSD}(X, Y) = \sqrt{\frac{1}{K} \sum_{k=1}^{K} \left(10 \log_{10} \frac{||Y(k)||^2}{||X(k)||^2} \right)^2} \quad \text{(dB)}, \tag{2.2}$$

Table 2.1: Quality degradation of sound and PEAQ (ODG).

Quality degradation	ODG
Imperceptible	0
Perceptible, but not annoying	-1
Slightly annoying	-2
Annoying	-3
Very annoying	-4

where $X(k)$ and $Y(k)$ are the short-time Fourier transform of the host and watermarked signals, respectively, in which an overlap rate of 0.6 was used for this evaluation. k is the frame index and K is the number of frames.

Perceptual evaluation of audio quality

PEAQ [33] is used to measure quality degradation in audio according to the objective difference grade (ODG) which ranges from -4 to 0. ODG indicates the sound quality of target signals as shown in Table 2.1.

Detection accuracy

Detection accuracy was measured by BDR, the ratio between the numbers of correct bits and total bits as follows.

$$\mathrm{BDR} = \frac{\text{number of correctly detected bits}}{\text{total number of bits}} = \frac{\sum \overline{\mathrm{xor}(s, \hat{s})}}{\sum \overline{\mathrm{xor}(s, s)}}, \tag{2.3}$$

where xor is the exclusive$-$or function.

2.1.3 Watermark attack

Watermark attacks are defined as processing that aims to enhance the quality of the watermarked signals or perceptually compress its size or as attempts that try to remove or invalidate the embedded watermark without affecting the sound quality. The term attack here requires further clarification. To simplify the notation of the general scheme of watermarking, both common signal-processing operations and intentional attempts are regarded attacks. Furthermore, any processing damaging the embedded watermark can be a potential method for pirate to remove the watermark. A good watermark should resist the attacks unless the sound quality of watermarked signals is severely degraded. The following attacks have been widely used individually or in a combination [34, 35].

Additive white Gaussian noise (AWGN). White Gaussian noise is usually added to the watermarked signal with the aim of masking the watermark and make it not fully recoverable. The sound quality of the resulting signal will be reduced to a certain extent.

Digital-to-analogue and analogue-to-digital (D/A–A/D) conversion. This attack consists in converting the digital audio samples into analog form. Some kinds of distortion such as amplitude scaling and phase shifting may be applied. The resulting analog-signal is then converted back into the digital form.

Re-sampling. Re-sampling attacks involve down-sampling the audio signal and then up-sampling the resulting signal to the original sampling-frequency. Aliasing distortion may occur if the re-sampling rate is low.

Perceptual compression. Knowledge from psychoacoustics study is exploited to remove inaudible portions of the signal, reducing its size. Frequency components falling outside the human audible range and the components masked by the stronger ones are removed. The state-of-the-art technique includes MP3 and MP4 compression.

Re-quantization. This attack re-maps the value of audio samples to a smaller set by reducing the number of bits for each sample. It affects the audio signal in a similar manner to that of additive white noise. The number of quantization bits is reduce in such a way that do not make the noise audible.

Filtering. Filters are used to remove unwanted frequency components or change the phase of audio signals. This attack includes low-pass filtering, bandpass filtering, and all-pass filtering. The filter coefficients are suitably designed so that only slight distortion occurs.

Echo addition. Delayed and attenuated copies (i.e., echoes) of the audio signal are created and added back to the signal in a manner that human ears cannot easily differentiate. Echoes with short delay-time actually emphasize the signal, making it louder, rather than cause distortion.

Time scaling. The audio signal is played faster or more slowly, resulting in a change in pitch. Related attacks include time-stretching, which preserves pitch (speed change) and pitch stretching, which preserves tempo (frequency scaling). These attacks are also used as de-synchronization attacks.

2.2 General Applications of Audio Data Hiding

Although the main purposes of audio watermarking is to combat against copyright infringement, it has shown advantages in a vast majority of other applications as well [36]. While copyright-related applications of data hiding require a certain degree of robustness against intentional attacks which greatly increases the complexity of the system, malicious attacks do not influence applications related to information carrier since the information actually adds complementary value to the signal. The watermark should still be robust to common operations, such as filtering, compression, D/A–A/D conversion.

This subsection presents some typical applications of audio data hiding that have been reported in the literature so far. In each application, we emphasize essential properties that the watermarking method must have in order to be efficiently applicable.

2.2.1 Proof of ownership

Audio watermarking was originally proposed as a solution for protecting copyright of digital music. It is necessary to have a mechanism to determine who is the owner of a public music track. In this application, a proof of ownership is embedded as watermark into the music track. When there is a dispute on ownership right, the audio watermarking system enables the actual author to demonstrate the presence of this watermark to claim his ownership.

The watermark is expected to be totally robust and secure, i.e., to survive common signal processing operations and especially intentional attacks that can invalidate the watermark. Such a watermark is ideal to carry copyright proof in audio signals since they can be both inaudible and inseparable from the watermarked signal that contains it. However, watermarking without extremely strong robustness may still be useful in embedding copyright notice into audio signals in order to generate a warning for considerate users.

2.2.2 Copy protection

As an effort to prevent copyright infringement, copy protection prevents publication of copyrighted digital data to file sharing services. Unlike conventional copy prevention that restricts duplication of digital audio on users' computer that has been proven impossible [8], this technique is made feasible by enforcing file sharing services to comply with the copyright law. Users cannot upload digital audio to file sharing services without rightful ownership.

In copy protection, unique identity or digital fingerprint of copyrighted audio is embedded as watermark into the audio signal. The digital fingerprint should be unchanged with imperceptible modification in the audio signal. When the audio signal is uploaded to a file sharing website, the embedded watermark is extracted and used to verify if it is legal to be uploaded. Media sites like YouTube can analyze files and compare their fingerprints with a database of copyrighted material and stop users from uploading these files [12].

The watermark should be very robust and secure to withstand against attempts to distort it. This application offers a possibility to limit widespread sharing of copyrighted digital audio. However it cannot stop peer-to-peer file sharing, which transfers digital audio directly between users.

2.2.3 Tampering detection

Digital technologies allows normal users to modify digital audio without leaving any perceptible marks. In digital forensics, digital audio is used as an evidence in court where its authenticity and integrity is crucial. Without authentication, we cannot verify whether the audio signal is really the original one or has been modified. It is therefore demanded to detect such tampering to verify integrity of audio signals.

In contrast to robust watermarking which aims at copyright protection, fragile watermarking has been proposed for dealing with tampering detection [37–39]. The challenge is that it is difficult to differentiate malicious tampering from common signal processing operations. Fragile watermark is embedded in the original audio signal and can be used to check if the original signal is tampered and detect the location of tampering.

To achieve that goal, fragile watermarking must be robust against common signal processing operations but fragile against malicious attacks that try to remove watermark.

2.2.4 Usage tracking

Another technique in digital right management (DRM) is tracing the source of illegal duplication [40, 41]. This technique enables the owner to monitor the customers who have infringed the license agreement on sold digital audio. Certain punishment can be made to compensate for the owner's commercial loss.

Usage tracking is done by firstly embedding the digital fingerprint of each copy of digital audio as watermark into audio content. This process is also referred to as fingerprinting. Unlike copy protection, this application requires that the fingerprint must be unique for each copy. This fingerprint could be the unique identity of the customer who buy digital audio. Based on embedded watermark, the owner can identify the unauthorized copy and track back the customer that broke the license.

The main challenge for this technique is assuring watermark robust against collusion attack where several users can combine their copies together to invalidate the watermark, e.g., averaging all the copies [42].

2.2.5 Broadcast monitoring

Growing with digital development has notably been expansion of television and radio channels over the last few years. Management of media assets sold for broadcasters has become increasingly important for the owner of media content. For example, the owners might want to check if their advertisement is actually airing, if their program is edited without permission, or it is broadcast multiple times without consent [43]. All those need to be used according to contract to ensure fair compensation for the owner. The straightforward solution is having human observers record what they see or hear on a broadcast. This method, however, becomes costly and error prone. Therefore, it is desirable to replace it with a more sophisticated version, and digital watermarks can be an ideal solution.

By embedding a watermark into audio or video content at the time of production and broadcasting, the owner can easily identify when and where the content is broadcast, who is the broadcaster, and how long it lasts. Being a part of the content, the watermark is strongly withstanding against attempts to destroy or remove it. Digital watermarks can be referred to a database with complete meta-data associated with the content, e.g., title, type, author, broadcast schedule, and so on. The broadcast monitor uses a computer with a pre-installed watermark decoder to examine and track inappropriate television channels or radio stations precisely and quickly. Watermarking has the advantage that the broadcast program is compatible with available TV or radio devices since the watermark does not affect to normal use of audio or video and that it does not requires extra resources such as other frequencies or header files.

The watermark has to be robust against transmission over the air channel which is analogue and noisy to be compatible with analog radios and TVs. Inaudibility is highly required to remain the high quality of digital audio. High embedding capacity is of interest to carry more information so that the system can have enough data without using a database which needs frequent update.

2.2.6 Information carrier

Watermark can be employed as a channel for carrying any kind of information [44–47]. For example, it could be the content information which describes for a music track or the information that adds value for radio or TV programs.

Content information is defined as information about an audio signal that is interesting for users, especially in music applications. Depending on specific application, several levels of content information could be defined. Some examples of interest we can mention are: content information describing a song (timbre, rhythmic, or harmonic description), meta-data describing a music track (title, composer, created date, performer, performance date, lyrics, album cover art, copyright notice).

Broadcasters can embed text to their radio or TV programs to enhance entertainment experiences for consumers [15, 48, 49]. The text is then displayed on a small screen of the radio receiver or TV. Here are some examples that we can think about: station identification, current time, advertisement, breaking news, weather forecast, stock prices.

Another interesting application of using watermark as information channel is identification of the aircraft in voice communication between aircraft pilots and controllers [50]. Aircraft pilots communicate with ground controllers mainly by voice via analog amplitude modulation (AM) radios [51]. The identification of the aircraft, which is very important for the controller to give appropriate instructions, is until now transmitted verbally. The identification step is inherently threatened by quality of the speech signal and human error, which might cause incidents. The aircraft number could be embedded as a watermark in the speech signal to prevent safety-critical misunderstandings.

These applications requires to transfer an as-much-as-possible amount of information, thus embedding capacity is highly necessary. Inaudibility should also be satisfied to maintain the high sound quality of audio signals.

2.2.7 Covert communication

Similar with information carrier application, watermark can also be employed as a channel for transmission of secret information [14, 52–54]. Covert means hidden or secret, implying that the information should be invisible to everyone. The term "covert communication" may be referred to as steganography. Watermarking has an advantage that the transmission of secret data is transparent and fuzzy. People are not aware that the important message is being transmitted because it is hidden inside the host audio that is being used for general purposes. In contrast, in conventional techniques of securing data by employing encryption, encrypted data, which have no meaning, implicitly say that the message is important and thus becomes a target of attacks.

This application is a combination between encryption and watermarking. At the sender side, the message is encrypted using a public key and then the encrypted message is embedded as watermark into the audio signal by watermarking. At the receiver side, the watermark is detected and then decrypted using a private key to extract the message. Secret message can be embedded as watermark in transactions between military and state offices, banks and other financial organizations.

Since security and large amount of information are very important, this application requires almost all the requirements: robustness, blind detection, high reliability, frame synchronization, and high capacity.

Summary

Above are some classes of application of data hiding in practice. Each application in each class has some specific requirements. However, there are some requirements that all applications in each class should be satisfied. Table 2.2 lists some necessary requirements recommended for these classes of application. In this table, strong robustness means that the watermark should resist intentional attacks and it is assumed that the resistance against innocent processing is always needed. It should be mentioned that the degree of requirement also varies with different applications. For instance, embedding capacity is

Table 2.2: Recommended requirements for typical classes of application.

	(1)	(2)	(3)	(4)	(5)	(6)
Proof of ownership	✓	✓	✓	✓		✓
Copy protection	✓	✓	✓	✓		✓
Tampering detection	✓		✓	✓		✓
Usage tracking	✓	✓	✓	✓		✓
Broadcast monitoring	✓		✓	✓	✓	✓
Information carrier	✓		✓	✓	✓	✓
Covert communication	✓	✓	✓	✓	✓	✓

(1) Inaudibility (2) Strong robustness (3) Blind detection
(4) Frame synchronization (5) Embedding capacity (6) Reliability

needed in many degrees or robustness is desired for certain attacks. Requirements for watermarking should be figured out in details when it is designed for realistic applications.

2.3 History of Audio Data Hiding

2.3.1 Methods based on least significant bits

The most conventional and simplest technique of audio data hiding works based on least significant bit (LS-bit) [55]. Watermark is embedded into audio signals by substituting or manipulating the LS-bits of audio samples according to watermark bits. Since LS-bits do not have much effect to the quality of audio signal, the distortion caused by this method is not severe, hence the watermark is inaudible to users. The other advantage is the high bit rate, e.g., 44100 bps for audio signals with a sampling frequency of 44.1 kHz. However, this method has a serious limitation with robustness, even against simple signal-processing manipulation such as addition of white noise.

A few smarter versions of this technique [56–58] attempts to improve the robustness by wisely modifying bits in higher bit layers, i.e., second LS-bit, third LS-bit, etc., for embedding watermark. However, the improvement is effective for additive white noise only; the watermark still do not resist other signal-processing operations such as resampling or MP3 compression.

2.3.2 Methods based on spread spectrum

Inspired from the concept used in spread spectrum (SS) communication, a number of methods of audio data hiding based on spread spectrum have been proposed. The principle is that the watermark is spread over the large bandwidth spectrum so that the energy in a single frequency is very small and it can be undetectable. This also enables the watermark to be easily detected even if there is interference on some frequencies.

The simplest technique based on spread spectrum is direct sequence spread spectrum (DSSS) [17, 59]. In this case, the host signal is firstly split into frames in which a bit is embedded into each frame. Supposed that the watermark bits is represented by $\{-1, 1\}$. A pseudo random sequence is generated with the length as same as that of the frame using a secret key. Then, this sequence, after multiplied with the watermark bit, is added into the host frame to produce the watermarked frame. To keep the watermark bit inaudible, the random sequence is scaled by a factor beforehand resulting in a sequence with amplitude is about 0.5 % of the dynamic range of the host frame [60]. The secret key which is used to control the sequence generator offers the possibility of protecting the watermark privacy.

In order to extract the watermark, the watermarked frame is correlated with the pseudo random sequence. If the result is positive, bit '1' is detected, otherwise, bit '–1' is detected. The secret key needs to be available at the detection side. As the watermark is spread and added into the host signal, the interference caused by the host signal may increase the error rate of watermark detection even in the absence of any attack.

In DSSS, the watermark is directly embedded into the host signal and the information of host signal is not used. Malvar and Florencio [61] have proposed a method, namely improved spread spectrum (ISS), in order to reduce the effect of the host signal interference. The ISS method takes advantage of the known information of the host signal to compensate for the host signal interference, reducing the error rate of watermark detection. Using this approach, the performance can improved in term of robustness against white noise or error rate of detected watermark and the capacity of watermark channel can also increase as well [62].

Although the watermark can be securely protected and robust against noise in the methods based on spread spectrum, the watermark represented by the pseudo random sequence is audible even its power is low. Some other improvements on traditional spread spectrum technique have also been made, e.g., embedding watermarks in modulated complex lapped transform (MCLT) domain [63, 64], incorporating perceptual entropy psychoacoustic model into spread spectrum and adaptively embedding watermark [65], and spreading watermark over selective frequencies [66, 67].

In spread spectrum, high frequencies are more accounted for the inaudibility of the watermark but are less resistant against robustness whereas low frequencies have the opposite characteristics. The watermark is spread over the whole audible spectrum in order to solve this conflict. This is the appeal of spread spectrum technique for audio data hiding. However, a common limitation in spread spectrum approach is that it is vulnerable to multiple watermark, i.e., only one bit is embedded into one frame. While we can think of reducing the frame length to increase the bit rate, it may also lead to more interference of the host signal, resulting in higher bit error rate. Hence, the bit rate of watermark is relatively low.

2.3.3 Methods based on echo hiding

The simplest method based on echo hiding embeds information into the host signal by introducing an echo in the host signal, which is an annutated delayed version of the host signal [16]. The watermark is encoded by the delay time of the echo, e.g., two different values of delay time are used to encode one watermark bit. The delay times are carefully chosen so that the watermark is inaudible and recoverable. By optimzing the delay time, adding the echo to the host signal essentially emphasizes the host signal, making it perceived louder, hence the watermark is basically inaudible to human auditory system (HAS).

To extract the embedded watermark bit, a technique based on examining complex cepstrum of watermarked signal is used [16]. Accordingly, a magnitude peak occurs at the delay time in the watermarked signal's cepstrum. The magnitudes at two locations (delay times) of the autocorrelation of the watermarked signal's cepstrum are compared and the watermark bit is chosen as the one corresponding to a higher magnitude.

In order to extract the watermark correctly, the magnitude of the echo needs to be considerably large, but, on the other hand, large magnitude echoes degrad the quality of watermarked signals. Toward overcoming this drawback, multiple small magnitude echoes with different delay times were used to embed watermark [68]. This approach, however, has higher error rate of watermark detection. Oh et al. [69] proposed echo kernels with positive and negative magnitude to compensate for distortion. Kim and Choi [70] proposed backward and forward echo kernels for audio data hiding. These methods use multiple echo kernel with relatively smaller magnitude achieving better performance on detection accuracy and inaudibility.

Since the watermark can be simply detected by autocorrelation of the watermarked signal's cepstrum, echo hiding methods face with a serious issue in security. The watermark is easily detected by using cepstrum analysis or destroyed by echo cancellation.

Ko et al. [71] proposed a time-spread echo which acts as reverberation in real environment consisting of many small echoes. By spreading the echoes, the energy of each echo can be reduced, thus the distortion induced by echo addition is inaudible to HAS while the error rate of watermark detection is maintained as small as conventional echo hiding methods. Spreading echo signals is done by using a pseudo random sequence which acts as a secret key. The secret key needs to be used in watermark detection, resulting in increasing security of watermark. This method, however, meets a drawback with host signal interference, which may reduce robustness against noise, when sound quality of watermarked signals is kept at an acceptable level.

In [72], Chen and Wu presented an echo hiding method to increase security, robustness, and perceptual quality by introducing interlaced kernels, frequency hopping, and analysis-by-synthesis approach. The magnitude of echo signals are suitably adapted to the host signal's characteristics and its resistance against attacks. In the interlaced kernels, delay times for embedding bit '0' and '1' are alternately exchanged in order to minimize the host signal's influence and the attacks' effect. Frequency hopping, incorporated with pseudo-noise sequence as a secret key, is aimed at increasing robustness and security. As for interlaced kernels, the author assumed that the audio frame has similar properties in the first half part and the last half pass, for correct detection of watermarks. This assumption does not hold for all the audio signals in practice, this method may not always robust.

Several other methods have improved inaudibility, robustness, and security of echo hiding method. Xiang et al. [73] proposed the echo kernel generated with pseudo-noise sequence has frequency characteristics that the echo signals have smaller magnitudes in perceptually significant region, toward higher perceptual quality of watermarked signals. Xiang et al. [74] also proposed a dual-channel time-spread method which adds two sets of echo signals with opposite polarity into two sub-signals of an audio frame, i.e., one consisting of even samples sequence and the other consisting of odd samples. This embedding mechanism yields an inherent cepstral feature in the two watermarked frame, reducing error rate of watermark detection.

In [75], Hua et al. proposed an optimal echo filter coefficients in which power spectrum of the echo filter is shaped by the absolute threshold of hearing of HAS to ensure the optimal inaudibility. Additionally, the autocorrelation function of the echo filter coefficients is optimally quantitatively specified. Based on this echo kernel, the authors have demonstrated the outperforming results in term of inaudibility, robustness, and security over other echo hiding methods.

The challenging point in echo hiding is that it is hard to determine exactly delay time of low-magnitude echoes by autocorrelation when the watermarked signals are subjected to attacks. Robustness and inaudibility are conflicted each other, with regard to echo magnitude and delay time. Furthermore, at the boundary between two successive frames containing different bits, there may be disrupt change in the watermarked signal caused by fast phase-shift. This disruption may become audible to human. A transition period can be incorporated to smoothly connect the frames in order to minimize the distortion.

2.3.4 Methods in transformed domain

Methods in transformed domain exploit advantages of simultaneous masking characteristics of HAS to embed inaudible watermarks. It is easier to incorporate perceptual knowledge into the embedding algorithm in transformed domain than in time domain. A simple way of exploiting perceptual knowledge is to embed watermark into the mid-range frequency component, since the low frequency components are more sensitive to noise and the high frequency components can be remove without degrading the audio quality. Also, many of the state-of-the-art compression techniques such as MP3 work in the same framework. Watermark can be adapted with the models to resist perceptual compression.

In general, quantization index modulation (QIM) technique is used for embedding watermark in transformed domain because of its good robustness and blind nature [19]. The embedding rule is quite simple. Suppose that a watermark bit needs to be embedded into a variable, QIM quantizes the value of this variable according to the corresponding scale. The watermarked variable is transmitted and may suffer from noise. To detect, the distance from the watermarked variable to its closet point in each scale are calculated and the watermark bit is decided by choosing the bit corresponding to the scale with the lower distance. When applied to audio watermarking, QIM needs to specific acoustic features, e.g., magnitude spectrum or phase spectrum, for embedding and reasonable quantization step size (the distance between two point in the scales).

Wu et al. [44] proposed an algorithm that embeds watermark into low frequency coefficients in discrete wavelet transform (DWT) domain with self-synchronization of frame positions. Due to taking the advantage of the time-frequency localization capability of DWT, this method can reduce computational load occurring in searching synchronization codes and detect watermark. In [76], Chen et al. presented an approach based on group-amplitude quantization. To enhance the robustness, the watermark is embedded in the low-frequency components in DWT domain. Chen et al. solved the trade of between inaudibility and robustness, by choosing an optimal quantization equation in order to reduce the distortion in the host signal and the detection error introduced by quantization.

Huang et al. [77] employed QIM technique and exploited perceptually significant frequency components in discrete cosine transform (DCT) main for embedding watermark toward robustness against noise and MP3 compression. Wang and Zhao [78], aside from the employment of adaptive QIM, exploited the multiresolution analysis of the DWT and the energy-compression characteristics of the DCT to account for the inaudibility of the watermark and explored the resistance of low frequency component to attain the robustness against attacks.

Later in [79], Wang et al. proposed a method based on QIM in DCT domain, in which the quantization step sizes are adapted to the local acoustic features and human auditory masking and the well-trained support vector regression (SVR) are incorporated to precisely recover the watermark. This scheme has been reported to achieve effective audio watermarking in robustness, imperceptibility, and capacity. However, it suffered from a minor shortcoming that the watermark is not 100 % recovered when the embedding bit rate becomes very high.

Hu et al. [80, 81] have proposed an improvement of this approach by introducing perceptual QIM in DCT domain. This method has been reported to be capable of achieving inaudibility, blind detection, robustness, and high embedding capacity, becoming the state-of-the-art in the literature.

It has been reported that the HAS is relatively insensitive to the phase components of sound [82, 83]. Phase distortion is perceptible to HAS only when the phase relation between each frequency component of the signal is dramatically changed [31]. This property of HAS is potential and has been exploited for inaudible audio watermarking for years [16, 31, 60, 84, 85]. These methods work by modifying the phase of the host audio signal according to a phase pattern using all-pass filter (APF) [86–88] or by quantization of the phase spectrum using QIM [22, 89, 90]. The reported results suggest that audio watermarking based on phase modification well performs with regard to inaudibility, robustness, blind detection, and high embedding capacity.

Apart from DWT and DCT domains, audio watermarking can also be implemented in other domains such as discrete Fourier transform (DFT) [91], cepstrum [92, 93], and singular value decomposition (SVD) [23, 94–96]. According to the reported methods, theory analysis and simulation results have shown that transform-domain techniques are generally more robust than time-domain techniques because they can better exploit advantage of signal characteristics and auditory properties.

2.3.5 Other methods

Psychoacoustics is the study of the mechanism of sound perception by human. Human ears have several limitations in distinguishing frequency components; in particular, when two frequency components are close together, the stronger one masking the other, makes it inaudible. Psychoacoustic models generalize the frequency masking effect by calculating a masking-threshold curve for a specific signal, called a masker. If another signal has power spectral density below the masking-threshold curve, it will be masked by the masker. The watermark is often shaped in frequency so that its power spectrum is under the masking-threshold curve of the host signal.

From the early days of audio data hiding, psychoacoustic model has been incorporated into designing watermark embedding algorithm to maintain high sound quality of the host signal. Swanson et al. [97] proposed a method that directly embeds watermark by modifying audio samples in incorporation with temporal and perceptual frequency masking effect. Methods based on the presented approach such as spread spectrum and echo hiding have also incorporated masking effects into watermark embedding [65, 73, 75, 98, 99].

Besides methods based on the above approaches, there are other methods that can be mentioned: a patchwork-based method [24], a histogram-based method [100], a method based on neural network [101], a method based on SVR [79], and methods based on QIM [19, 21, 102, 103].

2.4 Audio Data Hiding Based on Phase Modification

2.4.1 Phase characteristic of audio signal

To thoroughly understand phase characteristic of sound, we should first be clear about phase of a pure tone which consists of a single frequency. Let take a look at a sinusoid function which is a mathematic model of pure tone. This function is represented by the following formula:

$$s(t) = \cos(2\pi f t + \phi), \tag{2.4}$$

where f is frequency, ϕ is initial phase, and t is time variable. We may easily observe that $s(t)$ changes with the increase or decrease of ϕ, but the change pattern of $s(t)$ is hard to imagine. To see the point more clearly, we slightly modify the function of $s(t)$ as follows.

$$s(t) = \cos(2\pi f(t + t_0)), \tag{2.5}$$

where $t_0 = \frac{\phi}{2\pi f}$. Now, it is obvious that with the increase of ϕ, $s(t)$ shifts to the left hand-side, whereas $s(t)$ shifts to the right hand-side with the decrease of ϕ along the time axis. Thus, modifying the phase of a pure-tone signal actually makes it come earlier or later in time domain. By extending this principle for a complex tone or a realistic audio-signal which is composed of multiple frequency components, it turns out that modification of the phase spectrum is the process that delays or advances the waveforms of corresponding frequency-components.

Human ears perceive sound by separating into group of frequencies; each of group is perceived as a single frequency. This phenomenon is called frequency selectivity of HAS. Moderately modifying phase of frequencies in a group in synchronization does not induce perceptible distortion but when phase relation between components in one group is changed, the timbre changes drastically, leading to perceptible distortion. Moore et al. [30] demonstrated that human ears appear insensitive to relative phase when the frequency components are resolved, but with the unresolved frequency components, human ears can easily detect distortion due to changing relative phase.

In another study, Ozawa et al. [29] revealed that human ears is hardly sensitive to the difference in phase between two complex tones in high frequency range whereas with low frequency components, the ears can perceive the difference as timbre change. It is generally believed that if the phase-modification is kept sufficiently small, it is difficult for human ears to distinguish between the original sound and the modified sound [104]. Phase is more related to spatial information of sound, hence more important for real or 3D sound and less important for recorded sound [105, 106].

On the other hand, lossy compression technique such as MP3 is based on psychoacoustic models to estimate the masking-threshold curve. The spectral components are quantized and coded in such a way that the noise introduced by quantization is below the masking threshold [107]. Accordingly, phase is relatively resistant to perceptual compression due to the following reasons. First, perceptual compressors do not modify the phase of audio signal, just the magnitude. Second, they are based on the frequency masking-effect, i.e., masked components are coded with less number of bits. Finally, as long as the magnitude is not two small, changing the magnitude does not affect the phase.

Since the human ears are less sensitive to the phase of sound than to the noise and the phase is less affected by processing, these properties have been exploited for inaudible and robust watermarking over the recent years [31, 60, 84–86]. The methods work by modifying the phase of the original audio signal according one of two amounts of modification, each one encoding a bit of information. That is, the watermark data is represented by a phase shift in the phase of the host signal.

The original signal is firstly split into a series of short frames. Next, fast Fourier transform (FFT) is applied to each frame, generating a phase spectrum and a magnitude spectrum. Then, the phase spectrum is added with the phase response of an infinite impulse-response (IIR) APF or is quantized by QIM; two APFs or two quantization step sizes are pre-defined to represent for each bit. The amount of modification is specified in a manner that the watermark does not induce severe distortion and resist signal-processing operations. There are two types of watermarking methods developed using the phase of host signal, namely phase-modulation based watermarking and phase-coding based watermarking.

2.4.2 Phase modulation for watermarking

In phase-modulation approach, an APF is used to embed watermarks into the host audio signal, $x(n)$, by introducing imperceptible phase-modification. The technique of phase-shift keying (PSK) which is popularly used in data-transmission systems is adopted to realize this scheme. To embed a watermark-bit ('0' or '1'), a corresponding APF with a specific pole-zero is used to modify the phase spectrum of the host signal. Two APFs are pre-

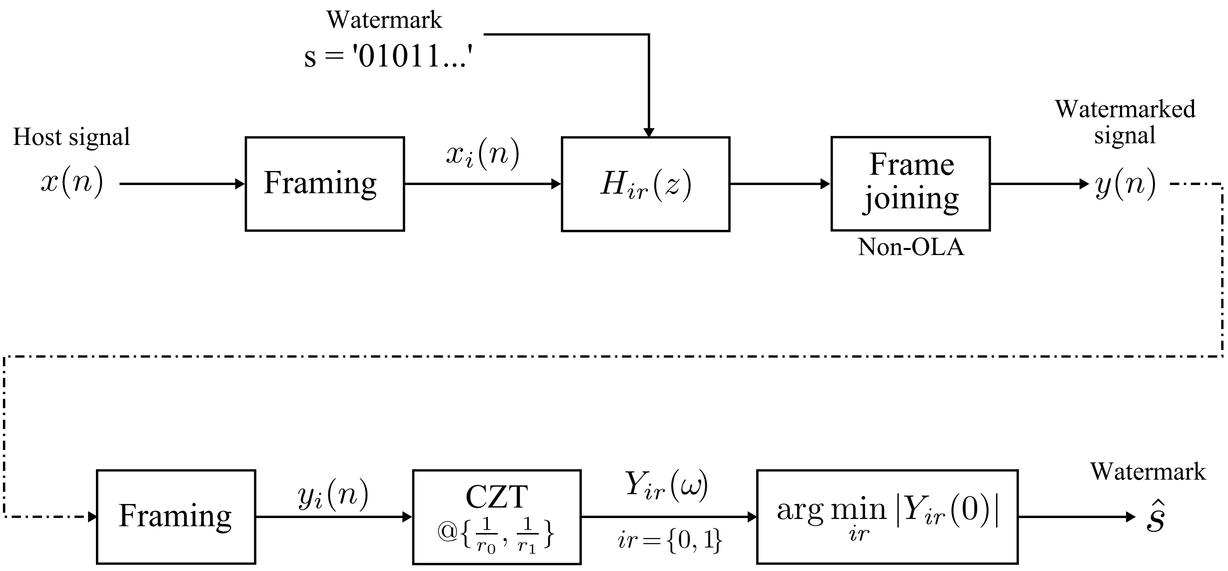


Figure 2.2: A scheme of audio data hiding based on phase modulation

designed and each one is picked up to use according to the bit. The transfer function of a first-order IIR APF is represented by:

$$H(z) = \frac{-r + z^{-1}}{1 - rz^{-1}} \tag{2.6}$$

where r is the filter's pole and $\frac{1}{r}$ is the filter's zero. The parameter r of the APF is modulated with the embedded bit, i.e., r_0 and r_1 are used for the corresponding bits.

Watermark embedding

Figure 2.2 depicts a block diagram of the watermark algorithm. Watermark $s(i)$ is embedded into the host signal, $x(n)$, as follows.

Step 1. Host signal $x(n)$ is first split into frames $x_i(n)$. Each frame is filtered by an APF which is designed by the next two steps.

Step 2. The watermark bit to be embedded is are represented by the filter parameter r. The values of parameter r, r_0 and r_1 are earlier determined by experimental analysis. According to the watermark bit '0' or '1', r is set to r_0 or r_1, respectively. For example, if the bit is '0', then $r = r_0$.

Step 3. Filter $H(z)$ is designed with the parameter r by Eq. (2.6). Then, frame $x_i(n)$ is filtered by $H(z)$.

Steps 2–3 are repeated until we reach the final frame.

Step 4. Finally, the filtered frames are joined together by non-OLA technique to yield watermarked signal $y(n)$.

Watermark detection

To detect embedded watermark-bit in a frame, each watermarked frame is analyzed to determine the filter that was used to process that frame in embedding process. It should be reminded that the filtering process can be performed in frequency domain as follows:

$$Y(z) = X(z)H(z), \tag{2.7}$$

where $X(z)$ and $Y(z)$ are z-transform of the original frame and the watermarked frame, respectively. Assume z_0 be a zero of the filter $H(z)$. As the nature of a zero of an IIR filter, $H(z_0) = 0$. Hence, $Y(z_0) = 0$. This is a hint for blind watermark detection. We can use chirp-z transformation (CZT) [108] to calculate $Y(z_0)$ as follows:

$$Y(z) = \sum_{n=0}^{N-1} y(n)z^{-n} \tag{2.8}$$

CZT over the zero-contour of the IIR-APF exposes a minimum at the pole-zero frequency in the log-magnitude spectrum. Based on the zero location over which the minimum occurs, we can determine which APF was used and then detect the embedded bit. The watermark is extracted as follows.

Step 1. Watermarked signal $y(n)$ is first split into the frames $y_i(n)$.

Step 2. Frame $y_i(n)$ is then analyzed by two types of CZT over two contours with radii of $1/r_0$ and $1/r_1$. The outputs are $Y_{ir}(z)$, $ir = \{0, 1\}$.

Step 3. The log-magnitudes at direct current (DC), $|Y_{ir}(0)|$, $ir = \{0, 1\}$, are compared together and the watermark is chosen as the bit corresponding to the lower magnitude.

Steps 2–3 are repeated for the total number of frames to extract all the embedded watermark-bits.

2.4.3 Phase coding for watermarking

In phase-coding approach, watermark is embedded into the phase spectrum of the host signal by QIM technique. The phase is quantized to two pre-defined scales in which each one represents for a watermark-bit ('0' or '1'). The QIM step-size which quantifies for the amount of phase-modification is pre-determined by experimental analysis. The higher QIM step-size results in higher robustness but less degradation of watermarked-signal quality and vice versa.

Watermark embedding

The embedding process starts with frame segmentation of the host signal, $x(n)$ into frames $x_i(n)$. Each watermark-bit from watermark s is embedded into one frame. Figure 2.3(a)

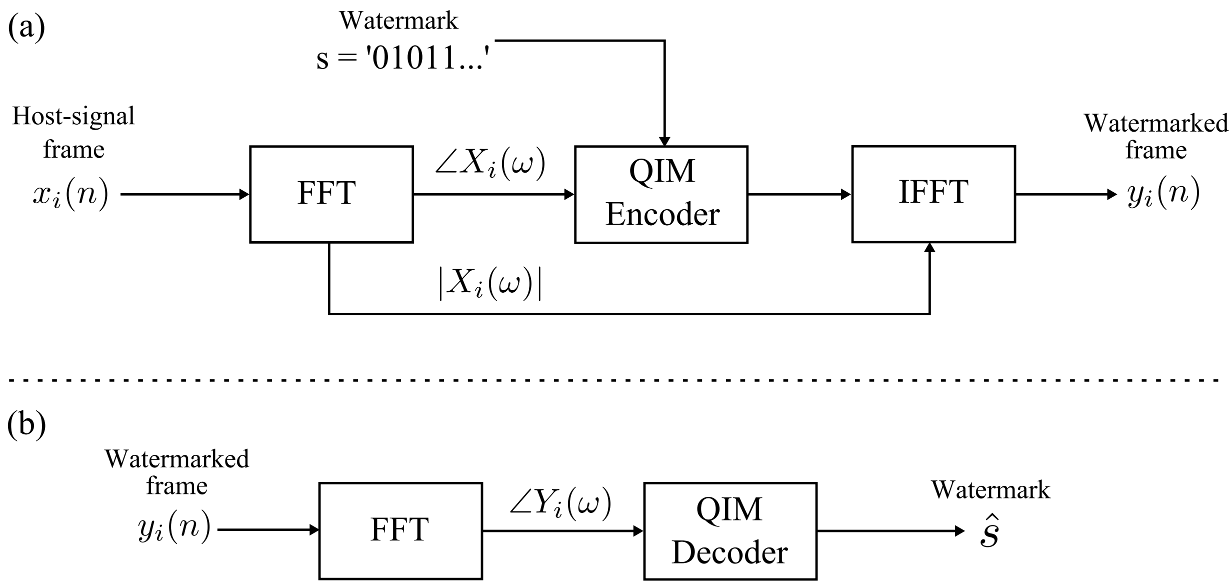


Figure 2.3: A scheme of audio data hiding based on phase coding: (a) watermark-embedding process and (b) watermark-detection process

depicts a block diagram of the watermark-embedding process. A watermark-bit is embedded into an audio frame as follows.

Step 1. Original frame $x_i(n)$ is transformed into the Fourier spectrum $X_i(\omega)$ by FFT. Magnitude spectrum $|X_i(\omega)|$ and phase spectrum $\angle X_i(\omega)$ are calculated.

Step 2. The watermark-bit are encoded into the phase of frequency-components by QIM encoder and a quantized phase spectrum $\angle Y_i(\omega)$ is obtained. Although each bit can be embedded in only one component, it is embedded in all components to increase robustness.

Step 3. The magnitude spectrum, $|X_i(\omega)|$ and the quantized phase spectrum, $\angle Y_i(\omega)$, are combined into Fourier spectrum $Y_i(\omega)$ and is then transformed into time domain signal $y_i(n)$ by inverse fast Fourier transform (IFFT).

Finally, all the processed frames are combined together to yield a watermarked signal $y(n)$.

Watermark detection

The detection process also starts with frame segmentation of the watermarked signal, $y(n)$ into frames $y_i(n)$. Figure 2.3(b) shows a block diagram of the watermark-detection process. A Watermark-bit is detected from a watermarked frame as follows.

Step 1. Watermarked frame $y_i(n)$ is firstly transformed into $Y_i(\omega)$ by FFT. Phase spectrum $\angle Y_i(\omega)$ is calculated.

Step 2. The phase of all frequency-components is decoded by QIM decoder all the bits. The output bit is determined by majority decision, e.g., if the number of '0' are

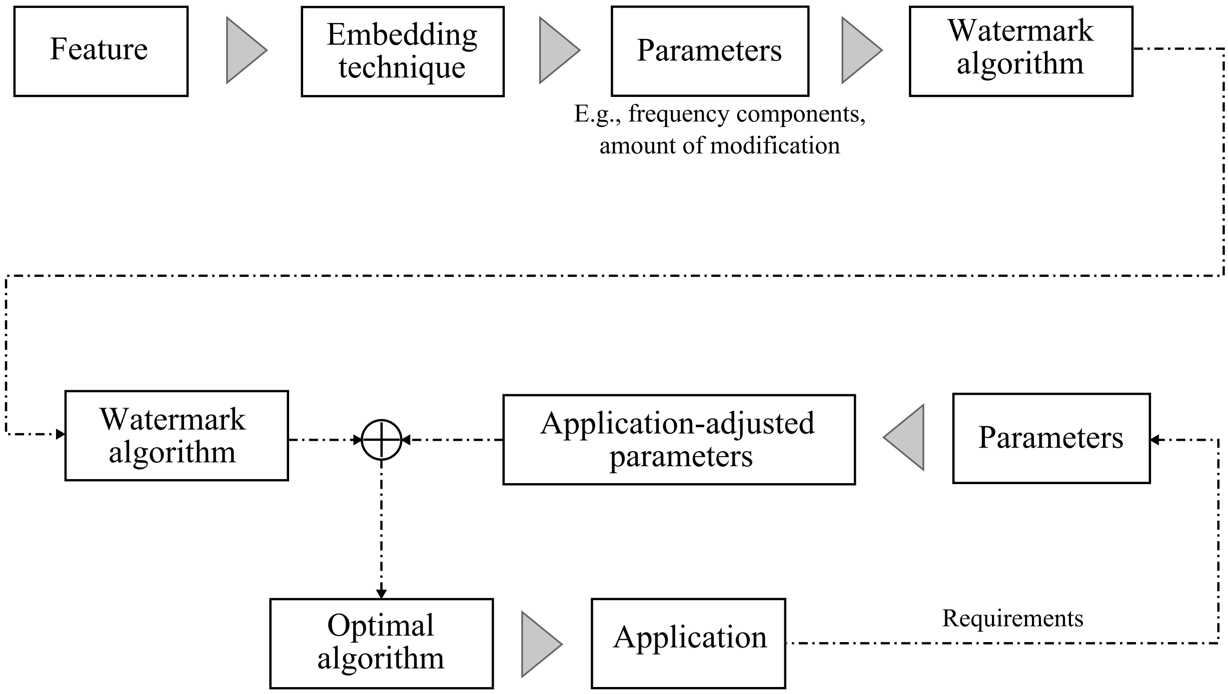


Figure 2.4: Schematic flowchart of research methodology for proposing a new method of audio data hiding

greater the number of ‘1’, the output is ‘0’.

These steps are repeated until we reach the final frame and all the watermark-bits are detected.

Remark

These are basic schemes of audio data hiding based on phase manipulation. With phase-modulation approach, discontinuity at frame boundary should be taken into account. Meaningfulness of frequency-components, i.e., whether they turn to a garbage component after processed, should be put into consideration with phase-coding approach. To achieve a good performance, these schemes need to be improved. This study relies on these scheme and exploited the variability of audio signals to develop wiser and more effective methods of audio data hiding. Key principles and detailed implementation are presented in later parts of this dissertation.

2.5 Research Methodology and Key principle

For years, many researchers have tried to proposed an efficient method of audio watermarking by employing mathematical models to account for inaudibility and robustness. Mathematical models were used to establish a set of constraints for watermarking algorithm in order to diminish influence of watermark to HAS and maintain the resistance of watermark. Nevertheless, the way human perceive sound is complicated and does not always fit with mathematical models. This study has conducted a better methodology for exploring, investigating, and proposing a new method of audio data hiding based on phase manipulation. Figure 2.4 shows a schematic flowchart of research methodology used in this study.

One should think about acoustic features for watermarking first before going to other steps. Choosing suitable features benefits a lot for performance of the method to be created. A good acoustic feature should be not only relatively inaudible to HAS, but also resistant against signal-processing operations. Secondly, embedding technique is important in many aspects such as blind-detection ability, not easily prone to distortion, streaming processing, and low-computational cost. Then, parameters set-up is strongly decisive in optimizing the performance of the proposed method. By figuring out these factors, a generic watermark algorithm is created with most of the requirements satisfied.

However, it has been proved that audio watermarking always meets a trade-off among all the requirements. Therefore, the constraint of all the requirements should be loosened depending on a specific application. The necessary requirements then need to be determined for the demanded application. These requirements are later used to adjust the parameters in order to bring out an optimal watermark algorithm applicable for the practical problem.

Based on the above methodology, this dissertation aims to propose two applicable methods of audio data hiding with the following key principles:

Acoustic features: Phase is used as an acoustic features for embedding as it is likely more resistant to other features such as magnitude. In well-known signal-processing systems, it has generally been assumed that auditory system is less sensitive to changes in relative phase than changes in magnitude [30]. Most signal-processing operations such as quality enhancement, have treated with the magnitude of frequency components rather than the phase. Perceptual compressors such as MP3 which are based on psychoacoustic models compress the magnitude of frequency components masked by the others to reduce the size of audio data. Holding watermarks in phase could help attain the robustness for the watermarking system.

Embedding technique: PSK and QIM are used as embedding technique as for its advantages in simple implementation, robustness, and blind-detection ability. PSK is employed to realize phase modulation. QIM has been popularly applied to multimedia watermark and obtained promising results. The nature of QIM is blind-detection, hence it helps create widely-applicable watermark algorithms for various practical problems. It has also been demonstrated in the literature that QIM produces watermarks robust to noise. In this research, two approaches of audio watermarking based on phase manipulation are studied: the phase-modulation-based approach adopting PSK and the the phase-coding-based approach adopting QIM.

Parameters: Parameters are chosen by incorporating the principle of invariability. Since realistic audio signals are time-variant, i.e., the properties of acoustic features which are potentially suitable for embedding watermarks also vary with time, for a fixed set of parameters, the algorithm may perform well for some signal but bad for others. Therefore, the variability properties of audio signals should be taken into account for designing a watermarking algorithm to achieve the best performance. The parameters are adaptively adjusted to local characteristics of audio signals to assure the algorithm works well for all kinds of audio signals.

Adjustable algorithm: It is easy to optimize the algorithm for the demanded application by adjusting the parameters. The algorithm has ability to change its properties to meet the requirements for different applications. In general, the amount of phase modification is directly proportional to robustness but inversely proportional to inaudibility. Therefore, the parameters that qualify for the amount of phase modification can be adjusted to balance inaudibility and robustness for difference applications with different requirements.

The sole philosophy in this research is dynamic phase-manipulation scheme. As to achieve the robustness against signal-processing operations for the proposed methods, audio phase is chosen as the embedding feature. It may conflict to the inaudibility of watermark. Study from psychoacoustics has shown that human ears are relatively sensitive to phase change. For instance, Ozawa et al. [29] pointed out that human beings have difficulty in discriminating two complex tones by the phase difference in high frequency components while in the low frequency region, phase differences between frequency components certainly affect the timbre of complex tones. In [30], Moore et al. have shown that when the frequency components are resolved, i.e., separately perceived as different tones, human ears appear relatively insensitive to the phase relation of the frequency components. For these reasons, the amount of phase modification is strictly controlled in order to not produce dramatically distortion in phase relation, resulting in hardly perceptible distortion in watermarked audio signals.

On the other hands, the phase-modification of a frequency component causes the distortion in watermarked audio signals in a manner that is directly proportional to the magnitude of that component. Therefore, the level of phase-modification of a frequency component is adaptively adjusted according to its magnitude. Larger-magnitude frequency components have small phase modification, whereas smaller-magnitude frequency components have somewhat higher phase modification. It is expected that the proposed methods could obtain reasonable trade-off between inaudibility and robustness with the dynamic phase-manipulation scheme.

Chapter 3

Audio Data Hiding Based on Adaptive Phase Modulation

This chapter proposes a method of watermarking for digital audio signals based on adaptive phase modulation. Audio signals are usually non-stationary, i.e., their own characteristics are time-variant. The features for watermarking are usually not selected by combining the principle of variability, which affects the performance of the whole watermarking system. The proposed method embeds watermark into an audio signal by adaptively modulating its phase with the watermark using IIR APFs. The frequency location of the pole-zero of IIR APF which characterizes the transfer function of the filter is adapted based on signal power distribution on sub-bands in magnitude spectrum domain. The pole-zero locations are adapted in a manner that the phase modulation produces slight distortion in watermarked signals to achieve the best sound quality. The experimental results show that the proposed method could embed inaudible watermark into various kinds of audio signals and correctly detect watermark without the aid of original signals. Reasonable trade-off between inaudibility and robustness could be obtained by balancing the phase modulation scheme. The proposed method has the ability to embed watermark into audio signals up to 100 bits per second with 99 % accuracy and 6 bits per second with 94 % accuracy in the cases of no attack and attacks, respectively.

3.1 Introduction

In general, audio watermarking methods embed watermarks directly into audio samples in time domain or acoustic features in transformed domain. Typical methods in time domain embed watermarks by modifying LS-bits or insert watermarks which are perceptually shaped according to the HAS [19, 109]. Although these methods could embed inaudible watermarks into audio signals, they are not robust against processing (e.g., LS-bit) or

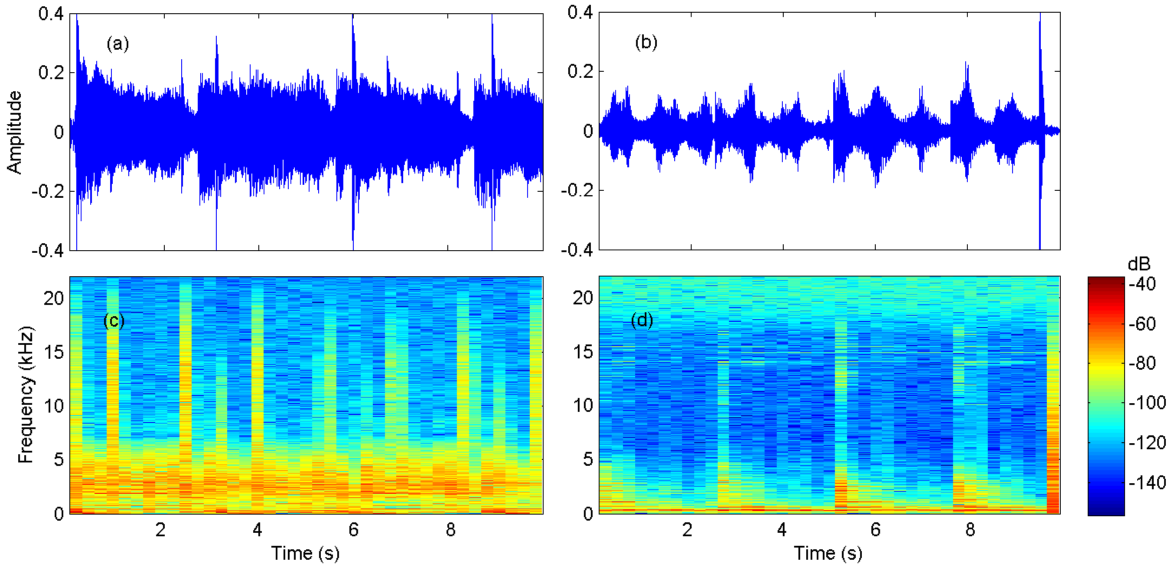


Figure 3.1: Waveform and power spectrogram of two pieces of music: (a) waveform of a pop song, (b) waveform of a dance song, (c) spectrogram of (a), and (d) spectrogram of (b).

have low embedding capacity [109]. Methods in transformed domain exploit advantages of simultaneous masking characteristics of HAS [17, 97] to embed inaudible watermarks. It has been reported that controlled phase alteration results in inaudible change in sound to HAS [82, 83]. This characteristic of HAS is potential and has been exploited for inaudible audio watermarking [16, 31, 60, 84, 85].

Audio signals in practice are usually non-stationary or, in other words, their perceptually insignificant features which are potentially suitable for embedding inaudible watermarks also vary with time. As a result, a watermark is inaudible for some signals but may be audible for others. This property has not been taken into consideration for designing a watermark embedding algorithm. In a nutshell, these proposed methods have a shortcoming that the embedding scheme is not adjusted according to the audio's own characteristics, which drastically influences the performance of the whole watermarking system.

Based on the above consideration, we studied feasibility of an audio watermarking method based on adaptive phase modulation. Depending on a local characteristic of each signal frame, a suitable phase modulator is chosen to ensure inaudibility or robustness on demand. More precisely, portion of modification for watermark embedding in phase spectral domain is adapted to local signal energy distribution in a manner that human perceives watermarks as slightest distortion or watermarks could be most withstanding against attacks. The experimental results show that the method can produce watermarked sound with better sound quality in comparison with other static watermarking methods. However, the method works with an assumption that frame positions were available in

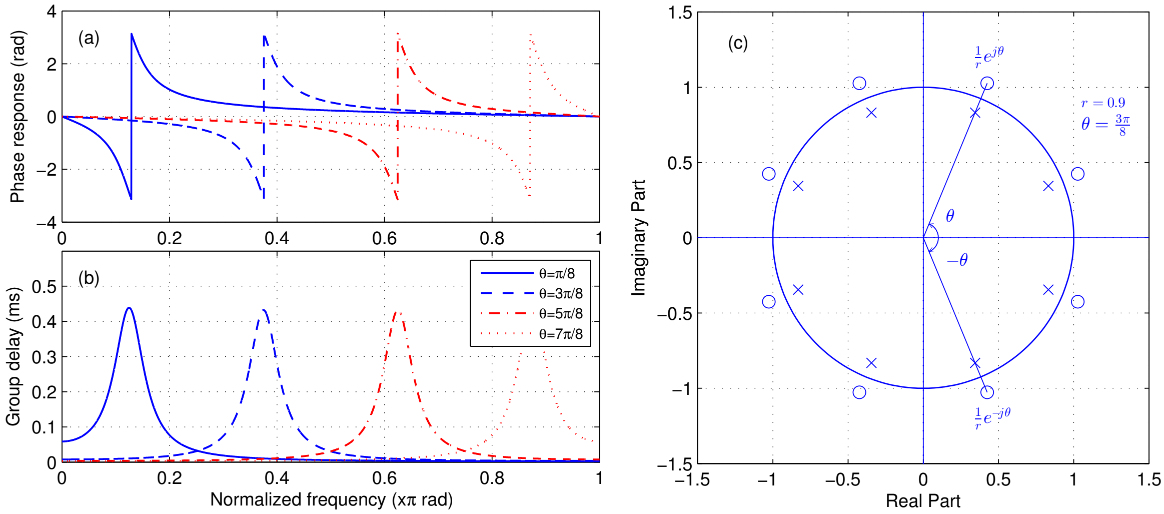


Figure 3.2: Characteristics of second-order APFs with respect to θ: (a) phase response, (b) group delay, and (c) pole-zero locations in z-plane.

watermark detection process, which may be unrealistic in some scenarios. Besides, the method meets a trade-off between inaudibility and robustness.

In this chapter, we propose an audio watermarking method based on adaptive phase modulation with automatic frame synchronization and good balance between inaudibility and robustness. Unlike conventional techniques for frame synchronization that embed synchronization codes along with watermarks, the proposed algorithm uses cues for detecting watermarks as a hint for detecting frame positions. Since the proposed method is implemented on a frame basis, the artifacts at frame boundaries deeply affect the sound quality, i.e., shorter frames cause larger degradation in watermarked signals. Instead of embedding a bit into a frame, we increase the number of bits in one frame to increase frame size in order to further improve sound quality while robustness is still ensured.

The rest of this chapter is organized as follows. Section 3.2 describes the concept of the proposed method, the details of watermark embedding and detection processes, frame synchronization, and illustrative examples for these processes. We present the results of evaluations with regard to effectiveness of the proposed adaptive scheme, inaudibility, robustness, and comparison with conventional watermarking methods in Sect. 3.3. Discussions are given in Sect. 3.4. Section 3.5 gives some concluding remarks and suggests a direction for future work.

3.2 Proposed Method

3.2.1 Principles of adaptive phase modulation

Real world audio signals are non-stationary, i.e., their characteristics vary with time. For this reason, acoustic signal processing applications usually divide signals into frames for non-stationary signal analysis. An obvious example of the variability characteristic of audio signals is that their power distribution over frequency components is time-variant. Figure 3.1 shows two pieces of music with their power distribution in time-frequency domain. The top panels show the waveforms and the bottoms show their power spectrogram. For the left signal, there are some periods having power distributed over the whole range of frequencies and the others having power distributed in only low frequencies. For the right signal, most of the periods have power distributed on low and high region of frequencies. The variability characteristic of audio signals should be taken into account for designing a watermarking algorithm to achieve the best performance.

Audio watermarking methods based on phase modulation usually take the advantage of controlled phase alteration resulting in relatively inaudible change to human ears [82, 83]. The phase of an audio signal can be manipulated by an APF. An APF passes all the frequency components equally but changes the phase according to the filter phase response. The principle of embedding a watermark into an audio signal is as follows. The audio signal is firstly split into frames. According to the watermark in each frame, a corresponding APF is designed. The watermark is then embedded into a frame by using the designed filter to enhance the phase spectrum of the audio frame. Since enhancing the audio phase by an APF results in relatively inaudible change, the watermarked frame is perceived without any severe difference. However, an issue occurring is that sound quality is affected by the abrupt change at frame boundaries due to phase shift. Shorter frame size may produce less artifacts. These issues should be thoroughly taken into consideration when implementing a watermarking method based on phase modulation.

In practice, a second-order IIR APF can be used to flexibly control phase change in audio signals by adjusting the modulus and angle of the filter's poles and zeros. This filter is represented as follows:

$$H(z) = \left(\frac{-\beta^* + z^{-1}}{1 - z^{-1}\beta}\right)\left(\frac{-\beta + z^{-1}}{1 - z^{-1}\beta^*}\right), \tag{3.1}$$

where $\beta = re^{j\theta}$ and β^* is complex conjugate of β. It has a pair of zeros ($\frac{1}{\beta}$ and $\frac{1}{\beta^*}$) and a pair of poles (β and β^*) that are symmetric on the real axis in z-plane. A second-order IIR APF can be characterized by the modulus and angle of its poles and zeros, i.e., r and θ, respectively. Figure 3.2 shows the phase characteristics and pole-zero locations of second-order APFs with respect to the pole-zero frequency θ. Because of the nature of APFs,

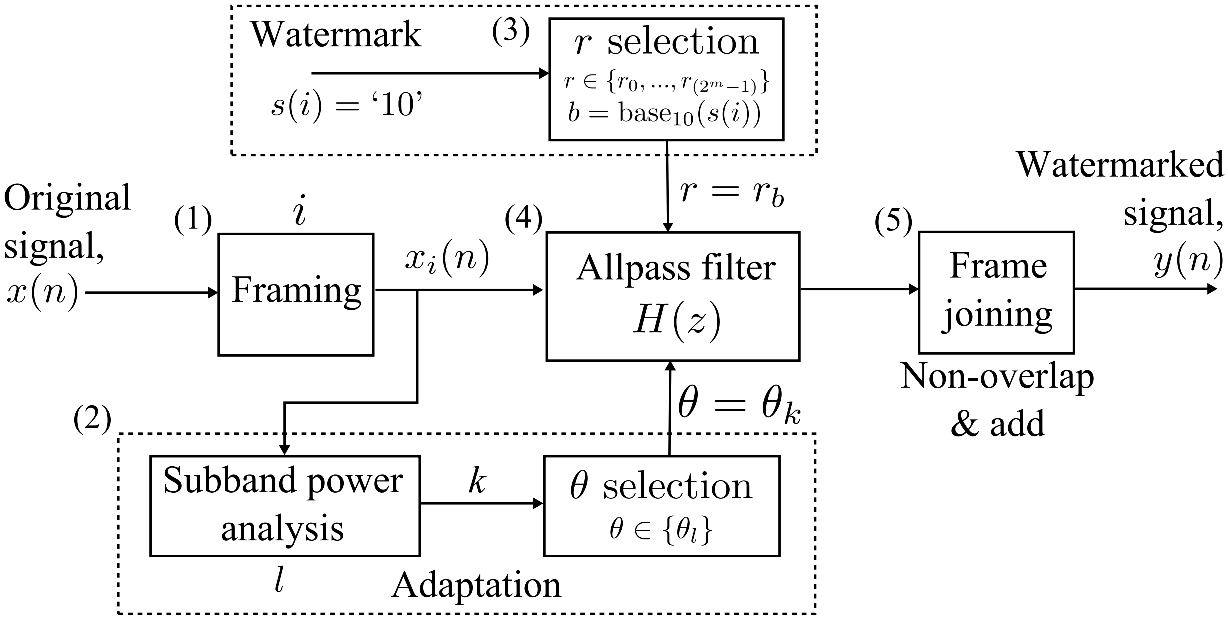


Figure 3.3: Watermarking based on phase modulation: embedding process.

the magnitude spectrum of a filtered signal is not manipulated but its phase spectrum is changed as follows:

$$\angle Y(e^{j\omega}) = \angle X(e^{j\omega}) + \angle H(e^{j\omega}), \tag{3.2}$$

where $\angle X(e^{j\omega})$, $\angle Y(e^{j\omega})$, and $\angle H(e^{j\omega})$ are phase spectra of an input signal and an output signal and phase response of the APF, respectively.

The absolute phase response and the group delay have a peak at the frequency that corresponds to the angle of the filter pole-zero θ (referred to hereafter as pole-zero frequency). According to Eq. (3.2), this frequency can be interpreted as the location where the signal is modified the most in phase spectrum. We have observed that the region around the pole-zero frequency is more affected than the outside region. In other words, if a signal has its energy concentrating at the pole-zero frequency then the filtered signal has faster disruption at frame boundaries than the signal having its energy outside of the pole-zero frequency. From these observations, we found that we can design an IIR APF with its pole-zero frequency lying on low energy region in spectrum domain to achieve the best sound quality. On the other hand, artifacts at frame boundaries could be reduced by increasing frame size. To further improve sound quality while keeping the same bit rate, we consider to embed more bits in a frame. These are the key principle for our proposed watermarking method based on frame-adaptive phase modulation.

Table 3.1: Representation of the watermark bits by the parameters r

	r_0	r_1	r_2	r_3	r_4	r_5	r_6	r_7
2-ary	0	1	–	–	–	–	–	–
4-ary	00	01	10	11	–	–	–	–
8-ary	000	001	010	011	100	101	110	111

3.2.2 Watermark embedding

Second-order IIR APFs are used to realize the adaptive phase modulation scheme. N_{bpf} watermark bits are embedded into a frame. The modulus r of the APF is m-ary modulated with the embedded bits, i.e., r_0, r_1, ..., and $r_{(m-1)}$ are used ($m = 2^{N_{\text{bpf}}}$) for the corresponding bits. The angle θ is adapted to power distribution in sub-bands of audio frames. In each frame, the pole-zero angle θ is determined by the center frequency of the lowest power sub-band.

Figure 3.3 depicts a block diagram of the proposed embedding process. Watermark $s(i)$ is embedded into an original signal, $x(n)$, as follows.

Step 1. Original signal $x(n)$ is first split into frames $x_i(n)$, in which the frame length $N = f_s/N_{\text{bps}}$; f_{s} is the sampling frequency and N_{bps} is the number of watermark bit per second (bps). Each frame is filtered by the APF which is designed by the next two steps.

Step 2. Frame $x_i(n)$ is then decomposed into K sub-bands with equivalent bandwidth by directly modifying the FFT of $x_i(n)$. Power of each sub-band is calculated and the lowest power sub-band index, k, is identified by Eq. (3.3). The output of this step is the parameter θ, i.e, the center frequency of the k-th sub-band which is calculated by Eq. (3.4).

$$k = \arg\min_{\ell} \left\{ 10 \log_{10} \left(\sum_{\omega} |X_\ell(e^{j\omega})|^2 \right) \right\}, \tag{3.3}$$

where $X_\ell(e^{j\omega})$ is the Fourier transform of the ℓ-th sub-band of the frame $x_i(n)$.

$$\theta_k = \frac{(2k-1)\pi}{2K} \quad \text{(rad)} \tag{3.4}$$

Step 3. The watermark bits which are represented by the filter parameter r are embedded into the frame in this step. The representation of the watermark bits by the parameters r is shown in Tab. 3.1. The values of parameter r, r_0, r_1, ..., and $r_{(m-1)}$, are earlier determined by experimental analysis as shown in Tab. 3.2. According to watermark bits $s(i)$, r is set to r_b, where $b = \text{base}_{10}(s(i))$ ($\text{base}_{10}(\cdot)$ is the function that converts a binary number to a decimal number).

For example, in the case 2 bits are embedded into each frame, we have 4 predefined

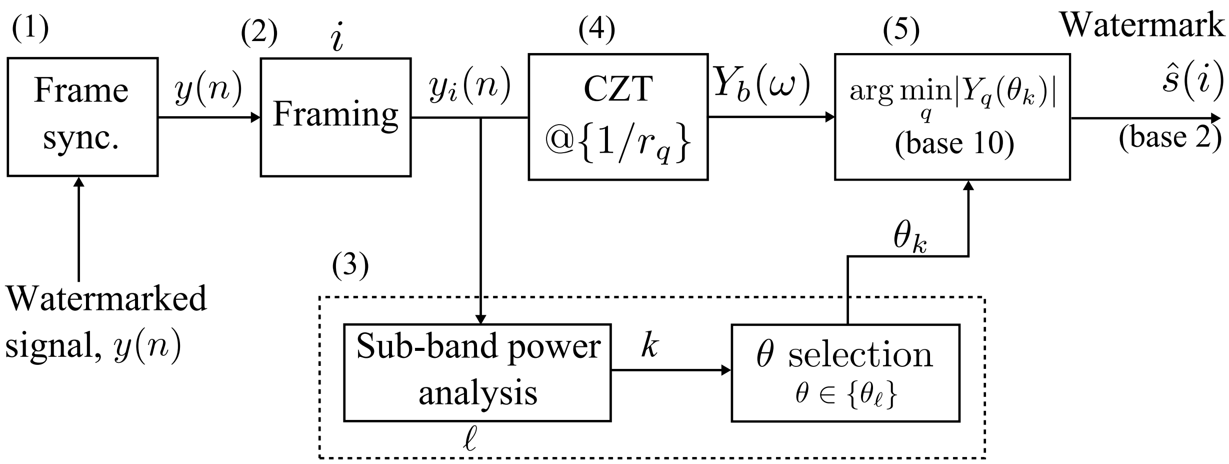


Figure 3.4: Watermarking based on phase modulation: detection process.

values of the parameter r: r_0, r_1, r_2, and r_3. If the watermark bits $s(i) =$ '10', then $r = r_2$.

Step 4. Filter $H(z)$ is designed with the two parameters, r and θ, by Eq. (3.1). Then, frame $x_i(n)$ is filtered by $H(z)$.

Steps 2–4 are repeated until we reach the final frame.

Step 5. Finally, the filtered frames are joined together by non-OLA technique to yield watermarked signal $y(n)$.

3.2.3 Watermark detection

To detect embedded watermark bits in a frame, each watermarked frame is analyzed to determine the filter that was used to process that frame in embedding process. Before that, frame synchronization is performed to identify frame positions. It should be noted that the filtering process can be performed in frequency domain as follows:

$$Y(z) = X(z)H(z), \tag{3.5}$$

where $X(z)$ and $Y(z)$ are z-transform of the original frame and the watermarked frame, respectively. Let z_0 be a zero of the filter $H(z)$. As the definition of a zero of an IIR filter, $H(z_0) = 0$. Hence, $Y(z_0) = 0$. This is a hint for blind watermark detection. We can use CZT [108] to calculate $Y(z_0)$ by (2.8).

CZT over the zero location of the APF exposes a minimum at the pole-zero frequency which corresponds to the pole-zero angle θ in the log-magnitude spectrum. Based on the zero location over which the minimum occurs, we can determine which APF was used and then detect the embedded bits.

Figure 3.4 depicts a block diagram of the proposed detection process. Watermark is detected in five steps as follows.

Step 1. Frame synchronization is performed. The details of the algorithm for frame

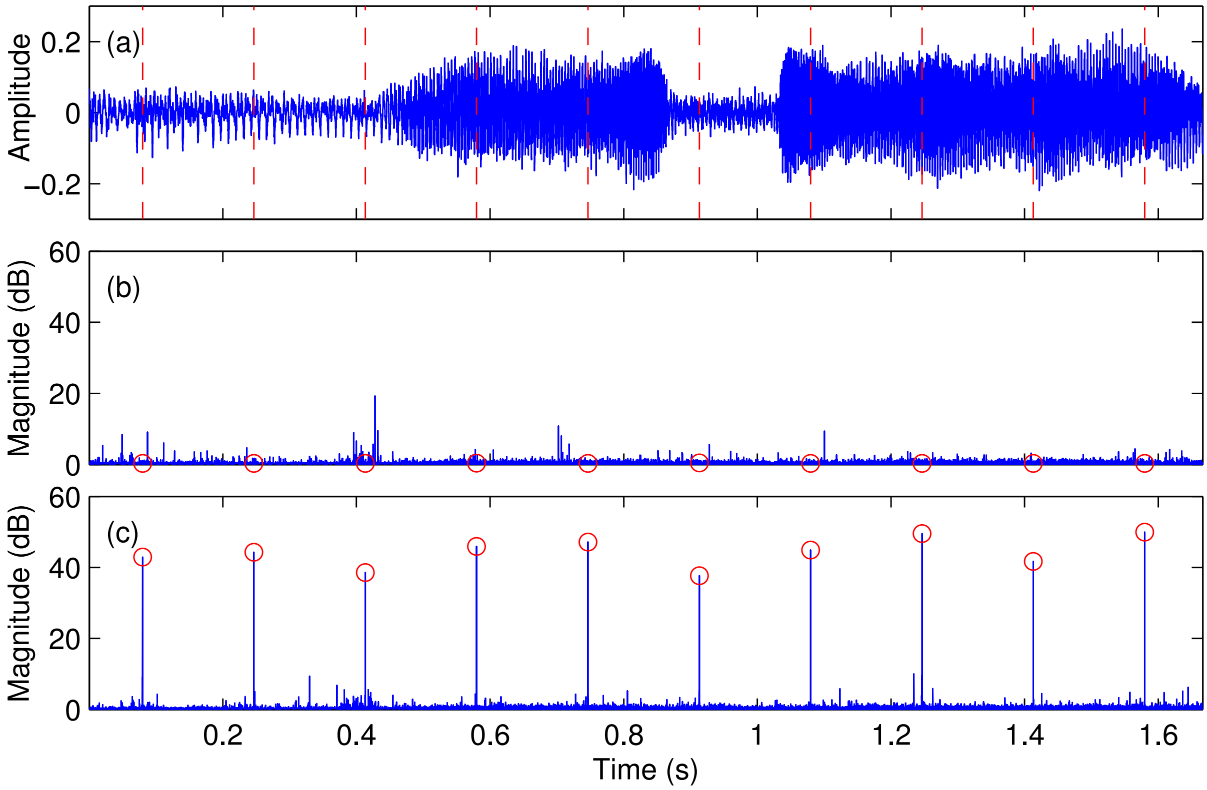


Figure 3.5: Frame synchronization in the case frame length = 167 ms: (a) watermarked signal, (b) synchronization on original signal, and (c) synchronization on watermarked signal.

synchronization is described in the next subsection.

Step 2. Watermarked signal $y(n)$ is first split into the frames $y_i(n)$ in which the frame length is as same as that in embedding process.

Step 3. Parameter θ_k is determined as Step 2 in the embedding process. Since the APF does not change the magnitude spectrum of the output signal, the power distribution of the watermarked frames in sub-bands could be obtained as same as that of the original frames.

Step 4. Frame $y_i(n)$ is then analyzed by m types of CZT over m contours with radii of $1/r_q$, $q = \{0, 1, ..., m-1\}$. The outputs are $Y_q(z)$, $q = \{0, 1, ..., m-1\}$.

Step 5. The log-magnitudes at θ_k, $|Y_q(\theta_k)|$, $q = \{0, 1, ..., m-1\}$, are compared together to detect watermark $\hat{s}(i)$ as follows.

$$\hat{s}(i) = \text{base}_2(\arg\min_q |Y_q(\theta_k)|), \tag{3.6}$$

where $Y_q(\theta_k) = Y(z)|_{z=\frac{1}{r_q}e^{j\theta_k}}$ and $\text{base}_2(\cdot)$ is the function that converts a decimal number to a binary number.

Steps 3–5 are repeated for the total number of frames.

3.2.4 Frame synchronization

The detection process works with an assumption that the frame positions are synchronized. In practice, this assumption may be invalid, so we have to identify a correct starting point before detecting watermarks. Based on the mechanism of watermark detection, we can see that CZT of a correct watermarked frame over the zero location of the IIR APF theoretically exposes a negative infinite at the pole-zero frequency in the log-magnitude spectrum. The negative infinite can be used as a hint for realizing frame synchronization.

We can search for a correct frame position over a frame length of signal. It is obvious that if we select a correct frame $y_i(n)$, the log-magnitude at θ_k of CZT of this frame over one of the contours is approximate to a negative infinite. This log-magnitude can be obtained by:

$$|Y_p(\theta_k)| = \min |Y_q(\theta_k)|, \qquad q \in \{0, 1, ..., m-1\}, \tag{3.7}$$

Then the average differences between $|Y_p(\theta_k)|$ and $|Y_q(\theta_k)|$, $ir \neq p$, which is calculated by:

$$d = \frac{1}{m-1} \sum_{b=0}^{m-1} \big| |Y_p(\theta_k)| - |Y_b(\theta_k)| \big|, \tag{3.8}$$

is very large. By contrast, an incorrectly selected frame has a small d. We can shift the frame sample by sample and calculate d for each frame. The frame with d which is the maximum is a correct frame. Once the correct starting point is identified, all the frame positions can be synchronized.

Figure 3.5 shows an illustration of frame position detection. We computed d for a ten-frame period of signal with the initial point randomly selected in two cases: (a) without watermark embedded and (b) with watermark embedded. There is no obvious maximum in Case (a) while high peaks occur at the correct frame positions in Case (b).

3.2.5 Example

Figure 3.6 shows an example of watermark embedding at the bit rate of 4 bps and $N_{\text{bpf}} = 1$. Watermark bits s='1001' were embedded into four frames. Parameter r was set to either r_0 or r_1 based on $s(i)$. The total number of sub-bands, K, is five. Parameter θ was set differently for each frame based on its power distribution in sub-bands. For example, frame #1 has $\theta_{\#1} = \frac{3\pi}{10}$ and frame #2 has $\theta_{\#2} = \frac{\pi}{2}$ as shown in Figs. 3.6(b) and Figs. 3.6(c).

Figure 3.7 shows an example of detecting the above embedded bits in the first two frames. The pole-zero frequency, θ, was determined by the same way as in the embedding process. The CZT of $y_1(n)$ over $\frac{1}{r_1}$ exposes a minimum at $\theta_{\#1}$ which decodes a bit '1' while the CZT of $y_2(n)$ over $\frac{1}{r_0}$ exposes a minimum at $\theta_{\#2}$ which decodes a bit '0'. The two bits

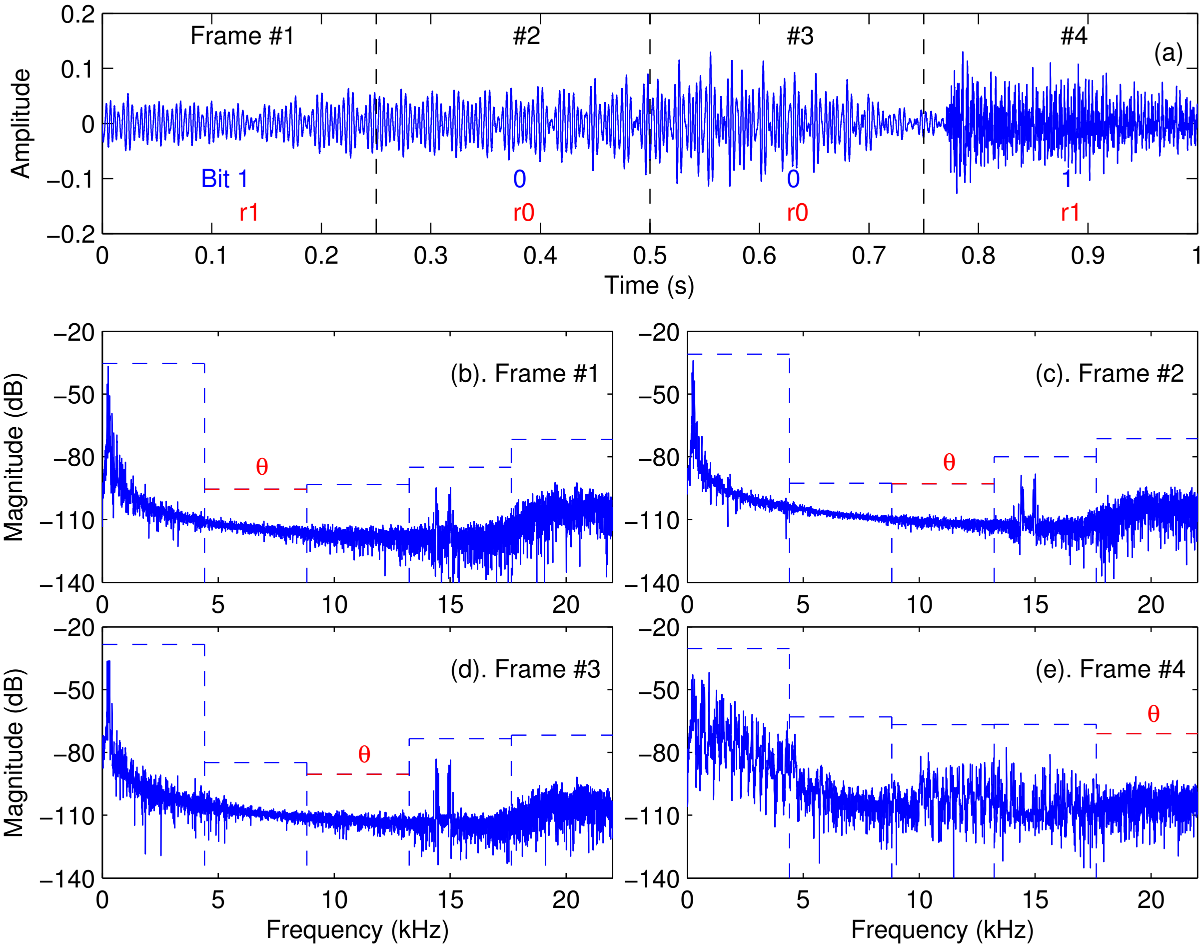


Figure 3.6: Example on modulation of r with embedded bits and adaptively adjusting θ: (a) four audio frames and the value of r in each frame; (b), (c), (d), and (e) power spectra of frame #1, frame #2, frame #3, and frame #4 and determination of θ in each frame.

were correctly detected.

3.3 Evaluations

3.3.1 Database and condition

We carried out several experiments to evaluate the proposed method with the Real World Computing (RWC) music database [110] which consists of 102 music tracks of various music genres. Each music track has a sampling frequency of 44.1 kHz, 16-bit quantization and is cut to be 60-second long. The watermark was set as random binary data. The bit rate, N_{bps}, was chosen from 6 to 400 bps. We investigated the m-ary modulation scheme with $m = 2, 4,$ and 8 (1, 2, and 3-bit encoding in a frame). The total number of sub-bands, K, was set to 5. The parameters r were determined by experimental analysis and set to the values as shown in Tab. 3.2.

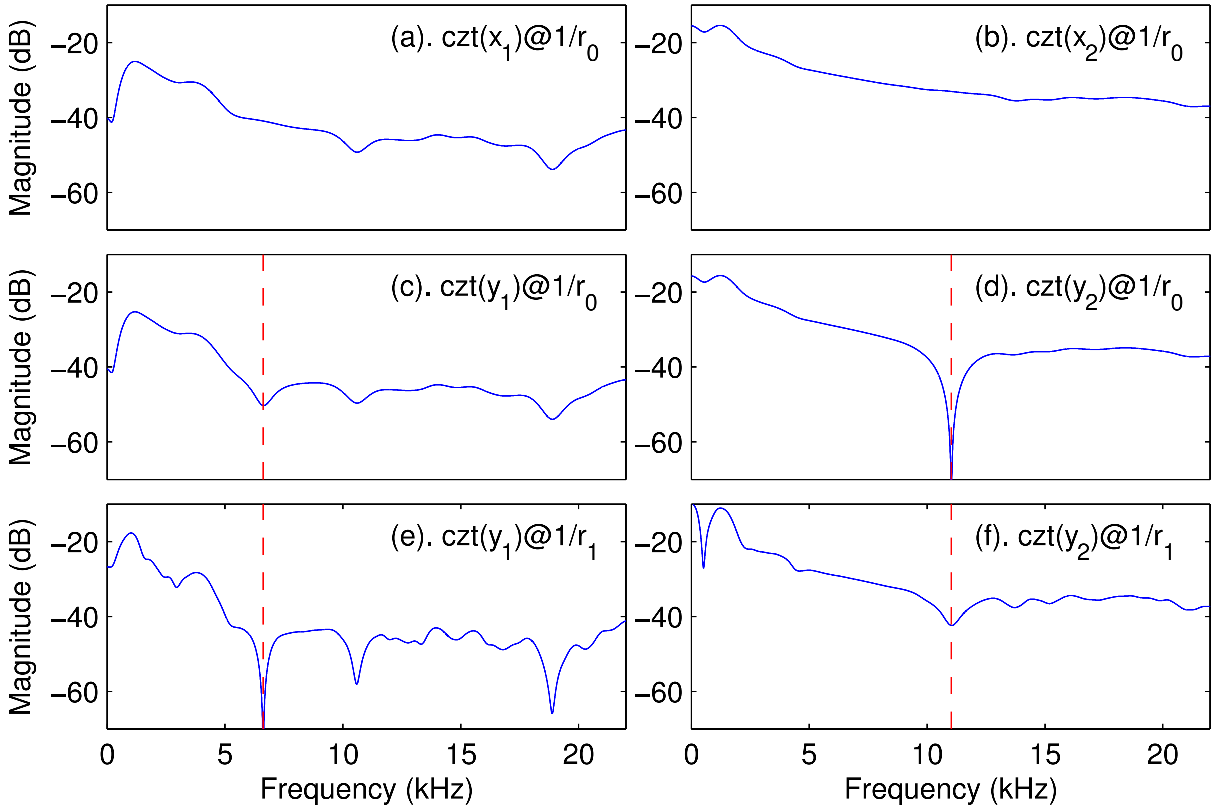


Figure 3.7: Minima in CZT spectra of watermarked signals: (a) and (b) show CZT of original frames #1 and #2 over $\frac{1}{r_0}$, (c) and (e) show CZT of watermarked frame #1 over $\frac{1}{r_0}$ and $\frac{1}{r_1}$, respectively, and (d) and (f) show CZT of watermarked frame #2 over $\frac{1}{r_0}$ and $\frac{1}{r_1}$, respectively. Bits of '1' and '0' were embedded into $y_1(n)$ and $y_2(n)$, respectively. Minima at $\theta_{\#1}$ in (d) and at $\theta_{\#2}$ in (e) indicate that the watermark bits were correctly detected.

Inaudibility of the proposed method is objectively tested by PEAQ [33]. PEAQ is used to measure degradation in audio according to the ODG which ranges from -4 to 0 indicating from very annoying to imperceptible. The evaluation criterion for PEAQ was set to an average ODG of -1 (not annoying). The watermark detection accuracy is measured by BDR which is defined as the ratio between the number of correctly detected bits and the total number of embedded bits.

3.3.2 Effectiveness of proposed method

In order to confirm the adaptation effectiveness of the proposed method, we comparatively evaluated the proposed scheme based on adaptive phase modulation (adaptive phase modulation (APM)) with schemes based on static phase modulation (static phase modulation (SPM)). Static schemes were implemented by setting a fixed value of APF parameter θ instead of frame-variant parameter θ in the adaptive scheme. Parameters θ of $\pi/32$, $\pi/16$, $\pi/8$, and $\pi/4$ were chosen in the non-adaptive schemes.

Table 3.2: Configuration on the parameter r

	r_0	r_1	r_2	r_3	r_4	r_5	r_6	r_7
2-ary	0.920	0.990	–	–	–	–	–	–
4-ary	0.885	0.920	0.955	0.990	–	–	–	–
8-ary	0.745	0.780	0.815	0.850	0.885	0.920	0.955	0.990

Figure 3.8 shows the average ODG in different music genres for the proposed method in comparison with the non-adaptive schemes. The bit rate was set to 6 bps in this experiment. The proposed method has good sound quality (ODG≥ -1) for all the genres except for the classical genre while some of the non-adaptive schemes have a few genres with bad quality. For all the genres, the proposed method has better sound quality than the non-adaptive schemes. The proposed method has total average ODGs greater than -0.47 and more than 89 % of all the music tracks have ODG greater than -1. The non-adaptive schemes seem to have a trend that when θ increases, the ODGs also increase. However, for the jazz, march, and vocal genres, the best case of sound quality is when θ is equal to $\pi/8$.

Figure 3.9 depicts the average ODGs and the average BDR with respect to embedding bit rate. The parameter m is set to 4 in this experiment. For the bit rates less than 100 bps, the adaptive scheme satisfies the evaluation criterion and has better sound quality performance in comparison with the non-adaptive schemes. When the bit rate increases, the sound quality of both the adaptive scheme and non-adaptive schemes decrease. This is because the more bits are embedded into audio signals, the more quality degradation the watermarked signals suffer. The BDR of the adaptive scheme is relatively lower than that of the static schemes. When the bit rate is greater or equal to 300 bps, the adaptive scheme has higher BDR than the static schemes with θ of $\pi/16$ and $\pi/32$.

The proposed APM scheme has much higher sound quality than the APM schemes while the BDR is comparable when the bit rate less that 100 bps. These results suggest that the proposed APM is more effective than the SPM scheme for audio data hiding.

3.3.3 Subjective listening test

The subjective listening test was carried out in a soundproof room. The audio was played via an amplifier of "Audio-Technica Stereo Headphone Amplifier AT-HA5000" and a headphone of "Sennheiser HDA 200". Twenty 10-second-long tracks (no. 5, 9, 10, 12, 14, 21, 22, 23, 26, 27, 29, 58, 63, 86, 88, 90, 95, 97, 98, and 99) in the RWC music database were used. The watermark is the same random binary data. The bit rate was set to 4 bps. These experimental configurations were referred to those in [111].

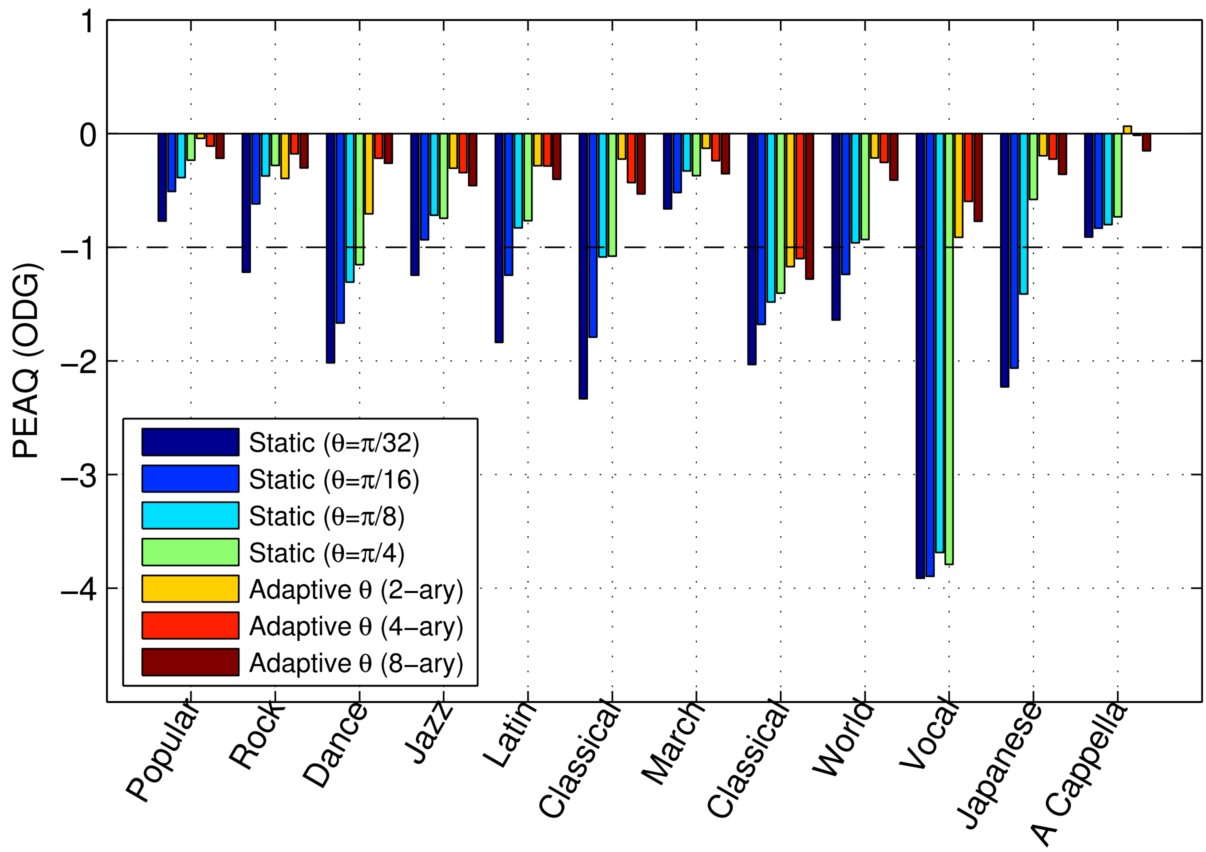


Figure 3.8: Objective evaluation on PEAQ of the proposed method in different music genres.

Eight voluntary subjects participated in the experiment. The subjects are asked to listen to a pair of signals: a reference signal and a test signal and then give a score of the similarity between the two signals according to Tab. 3.3. For each track, there are four pairs in a random order, in which the first one in the pair is the original track (Org) and the other one is the original track or an watermarked track created by the proposed method (APM), the cochlear delay watermarking method with blind detection (CDb) [112], or the direct spread spectrum (DSS) method [17]. Each subject evaluated the similarity of 80 test pairs in total, which is 20 tracks multiplied by 4 pairs (Org-Org, APM-Org, CDb-Org, and DSS-Org).

Figure 3.10 shows the mean scores of four test pairs for each track. The mean scores of the pairs APM-Org are approximate to those of the pairs Org-Org, suggesting that the distortion caused by the watermark embedding of the proposed method is hardly perceptible. In contrast, the distortion caused by the CDb method and the DSS method is slightly perceptible and perceptible for most of the tracks.

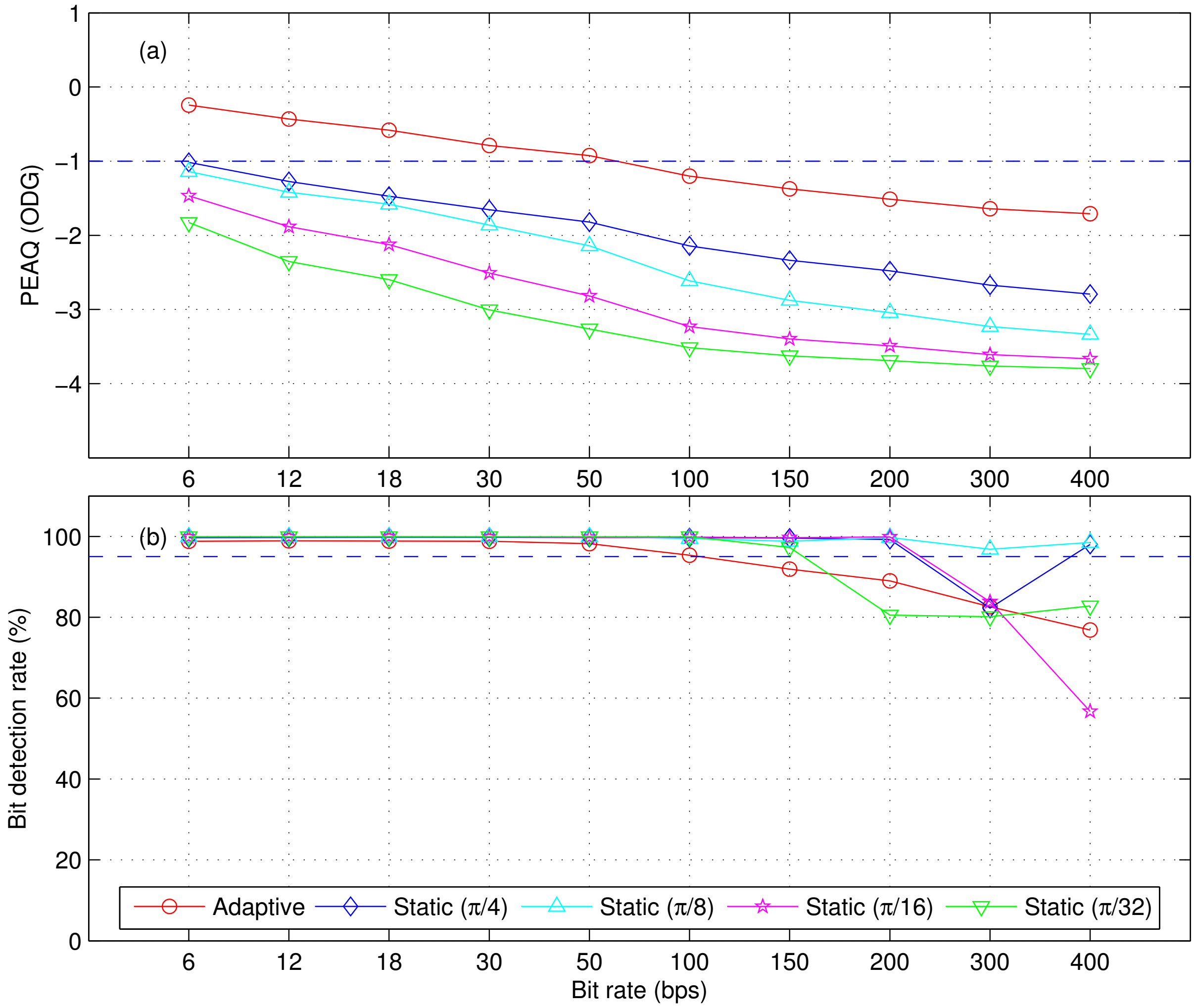


Figure 3.9: Evaluation on PEAQ and BDR of the proposed method in comparison with static phase modulation schemes.

3.3.4 Robustness evaluation

We carried out experiments to evaluate robustness of the proposed method against the following processing: addition of AWGN with an SNR of 36 dB, MP3 compression at 128 kbps, MP4 compression at 96 kbps, re-quantization to 8 bits, re-sampling to 22 kHz and 16 kHz, and bandpass filtering with the passband between 100 Hz and 6 kHz and the stopband of −12 dB per octave. The embedding bit rate was set to 6 bps in this experiment.

We set up the robustness test with two different configurations of adaptive phase modulation schemes to investigate the trade-off between the inaudibility and robustness of the proposed method. With default configuration (Config. 1), parameter θ is adjusted adaptively to the lowest power sub-band. Since the least sensitive portion of audio signal is modified, the sound quality is the best but the robustness may be quite low. We inves-

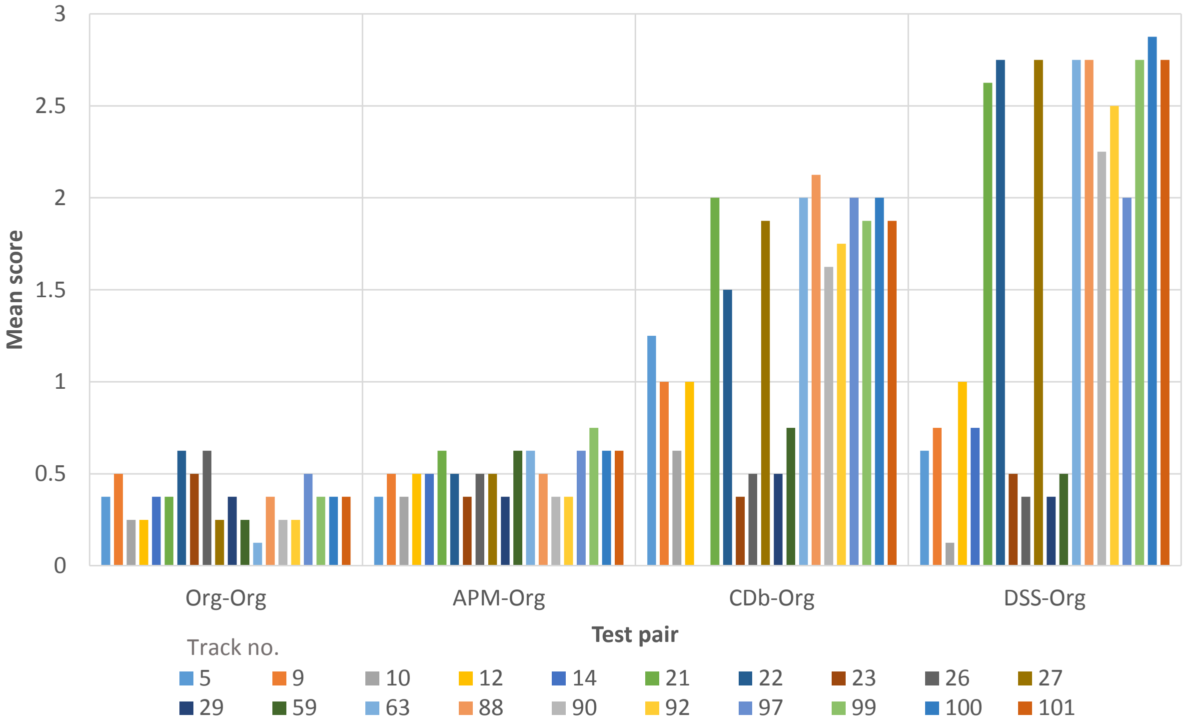


Figure 3.10: Results of subjective listening test.

Table 3.3: Score of similarity between the two tracks in subjective listening test.

Similarity	Score
Completely the same	0
Probably the same	1
Probably different	2
Completely different	3

tigated another scheme of adaptation in which parameter θ is adjusted adaptively to the highest power sub-band (Config. 2).

Table 3.4 shows the average BDR of the watermark detector in two configurations. With Config. 1 (2-ary, 4-ary, and 8-ary), the BDRs are all lower than 90 % while the PEAQs are close to imperceptible grade. By contrast, with Config. 2 (2-ary and 4-ary), the BDRs are greater than 90 % for all the attacks except for bandpass filtering while the PEAQs are less than those with Config. 1. The PEAQ with 4-ary is greater than that with 2-ary. On the other hand, 8-ary seems to be inferior to 2-ary and 4-ary. It is because 8-ary using various values of parameter r for phase modulation causes the change at frame boundaries more abrupt and tracking of the parameter r in detection process incorrect. Encoding 4-ary is the most suitable for the proposed method to ensure high sound quality and high robustness. These results suggest that the proposed method could achieve a

Table 3.4: Inaudibility and robustness evaluation with two configurations.

Measure	Config. 1			Config. 2		
	2-ary	4-ary	8-ary	2-ary	4-ary	8-ary
PEAQ (ODG)	-0.42	-0.35	-0.47	-1.23	-0.69	-0.77
BDR(%)						
No processing	99.08	99.56	99.51	99.75	99.96	99.76
Resample 22k	44.35	49.71	49.89	99.22	99.11	93.19
Resample 16k	43.99	49.52	49.78	98.27	98.21	89.08
AWGN 36 dB	83.07	81.23	79.91	97.43	97.98	96.60
Requantize	73.97	74.59	72.30	92.06	95.33	91.61
MP3 128 kbps	58.17	55.65	54.99	94.98	95.14	89.20
MP4 96 kbps	44.08	50.13	49.91	94.51	94.30	81.88
Bandpass filtering	46.29	49.91	50.18	61.26	64.06	55.49

reasonable balance between inaudibility and robustness with Config. 2 (4-ary).

We investigated the trade-off among inaudibility, robustness, and capacity by carrying out an experiment to evaluate the sound quality and the BDR with respect to the bit rate. The Config. 2 with 4-ary was set up in this experiment. The bit rate varies from 6 to 100 bps. The BDR was computed in both the cases of no attack and under attacks. The results of PEAQ and BDR are plotted in Fig. 3.11. When the bit rate increases, the PEAQ decreases while the BDR is relatively steady. The sound distortion is relatively high when the bit rate is greater than 12 bps. These results suggest that the proposed method has a good performance on inaudibility and robustness at low bit rates but the sound quality is bad at higher bit rates.

3.3.5 Comparative evaluation

We compared the proposed APM method with three other conventional methods: the CDb method [112], the DSS method [17], and the echo hiding method (ECHO) [16]. All these methods detect watermark using a blind detection scheme. In this experiment, the bit rate was set to 6 bps and the Config. 2 and 4-ary was set up for the proposed method. The watermarked signals is processed various kinds of attacks.

Figure 3.12 plots the average ODG and BDR of the proposed method and the compared methods. The PEAQs of the proposed method are higher than all the compared methods. The DSS method has all the bit detection rates greater than 95 %. The proposed method robust against all the processing except for bandpass filtering. The DSS method is the most

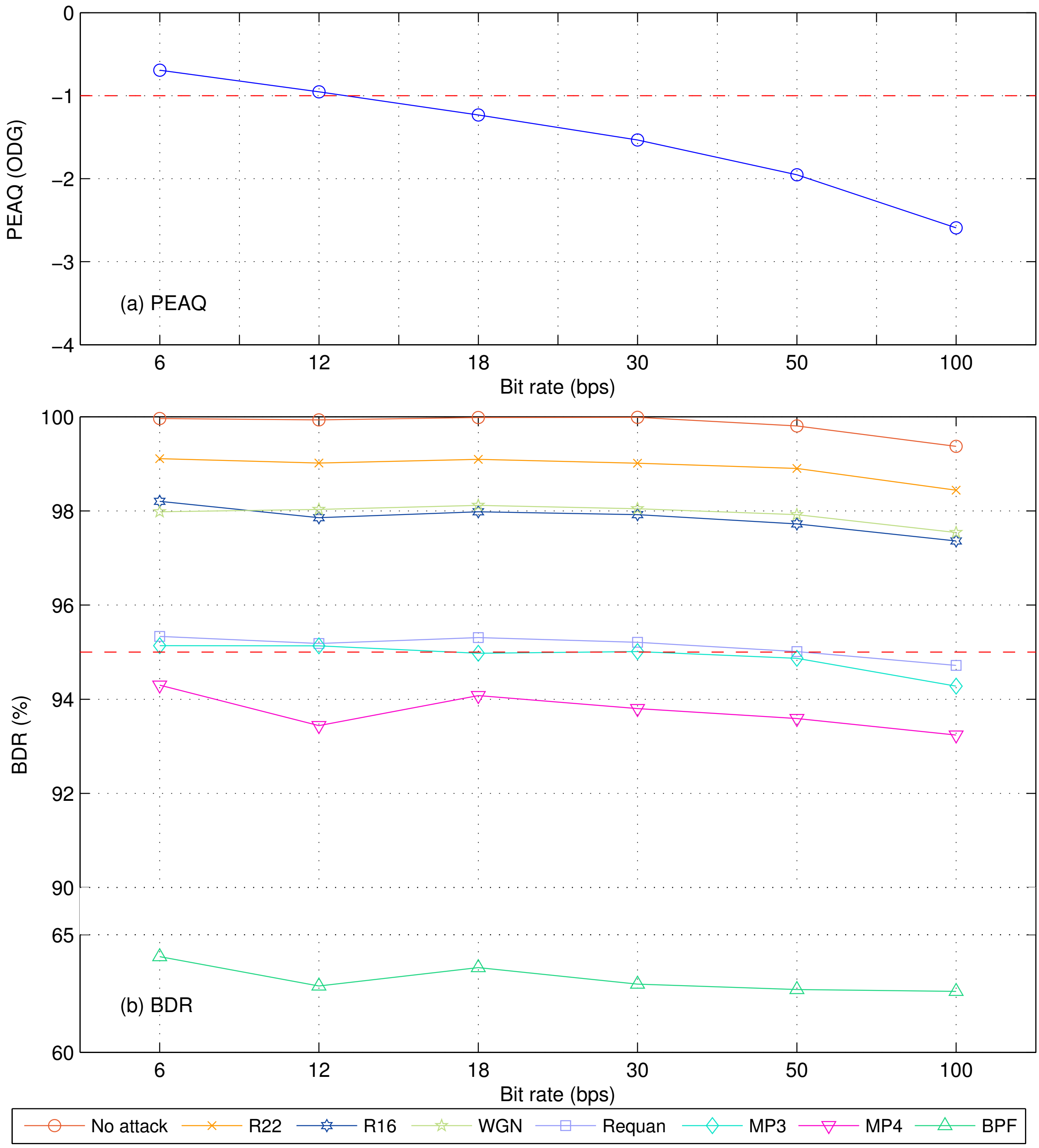


Figure 3.11: Evaluation results of (a) sound quality (PEAQ) and (b) BDR with respect to bit rate.

robust while the proposed method has the best inaudibility and is the second most robust. The ECHO method is good in robustness and the CDb method is good in inaudibility. These results reveal that the proposed method has the best trade-off between inaudibility and robustness.

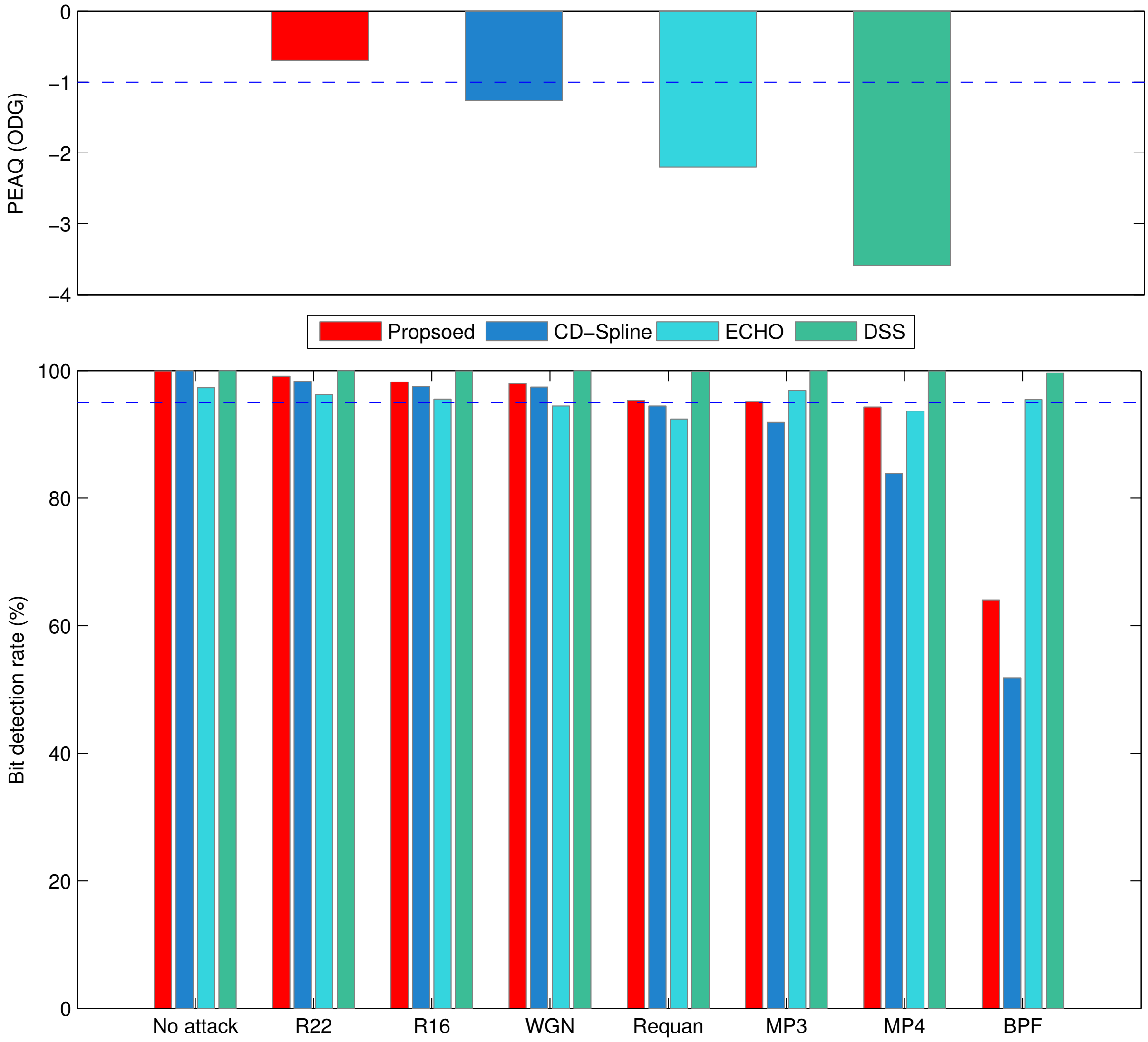


Figure 3.12: Comparative evaluation on PEAQ and BDR with respect to types of attacks.

3.4 Discussion

Phase modulation itself does not actually make sound quality of each frame degraded but introduces discontinuity at frame boundaries in the whole signal. Discontinuity creates spread of spectrum like pulsive sound's spectrum in watermarked signals which could be perceived as annoying distortion. Overlap and add (OLA) technique is usually employed to eliminate discontinuity in conventional signal processing applications. However, OLA does not work with the proposed method because it destroys the audio samples around the frame boundaries which are a key factor for blind detection by CZT. For that reason, the proposed method modifies the least significant portion to embed watermark instead of using OLA technique for high sound quality.

The proposed method meets a trade-off between inaudibility and robustness. To achieve the best sound quality, the proposed method chooses the least perceptually insensitive portion of audio signals for embedding watermark. Robustness could be achieved by choosing resistant portion of audio signals for embedding but at the expense of reducing sound quality. Toward a reasonable trade-off between robustness and inaudibility, we consider to reduce artifacts at frame boundaries by reducing frame size or, in other words, embedding more bits into a frame. However, larger-ary phase modulation produces the more abrupt change at frame boundaries and makes determination of the parameter r in detection process harder. Therefore, 4-ary scheme is the most suitable for the proposed method.

3.5 Summary

In this chapter, we proposed an audio watermarking method with reasonable trade-off between inaudibility and robustness based on the concept of adaptive phase modulation. The parameter of phase modulator was adaptively adjusted to the center frequency of the lowest or highest power sub-band of an audio frame to achieve high sound quality or high robustness. The results of PEAQ and BDR suggested that the proposed method can embed watermarks into various kinds of audio signals with good inaudibility and blindly detect embedded watermarks under various kinds of attacks. Compared with conventional methods, the proposed adaptive watermarking method provides better trade-off between inaudibility and robustness.

An embedding bit rate of 100 bps with 99 % accuracy and 6 bps with 94 % accuracy could be obtained in the case of no processing and the case of attacks, respectively. It was revealed that the proposed method could satisfy three requirements: *inaudibility*, *blindness*, and *robustness*.

Chapter 4

Audio Data Hiding Based on Dynamic Phase Coding

This chapter proposes an audio watermarking method based on dynamic phase coding and error control coding. The technique of quantization index modulation is employed for embedding watermarks into the phase spectrum of audio signals. Since most of audio information is distributed in somewhat low frequencies, to increase robustness, this frequency region is chosen for embedding watermarks. Phase modification causes sound distortion in a manner that is proportional to the magnitude. Therefore, the amount of phase modification is adjusted according to the magnitude to balance inaudibility and robustness. Error control coding is incorporated to further increase reliability of watermark detection. The experimental results show that the watermarks could be kept inaudible in watermarked signals and robust against various attacks. Error control coding is effective in increasing watermark detection accuracy remarkably.

4.1 Introduction

In general, audio watermarking methods should satisfy four requirements: *inaudibility*, *blindness*, *robustness*, and *high capacity*. The solution is very hard because there is a trade-off among these requirements. It is straightforward that perceptually insensitive features should be exploited for embedding watermarks. But this is a challenge for robustness, since processing can distort the watermark without degrading the sound quality. Selecting suitable acoustic features for watermarking that satisfy both inaudibility and robustness is an important task for the design of watermarking algorithms.

Audio watermarking methods typically embed a watermark directly into audio samples in the time domain or acoustic features in a transformed domain. Some methods replace LS-bits with watermark bits or insert a watermark which are perceptually shaped

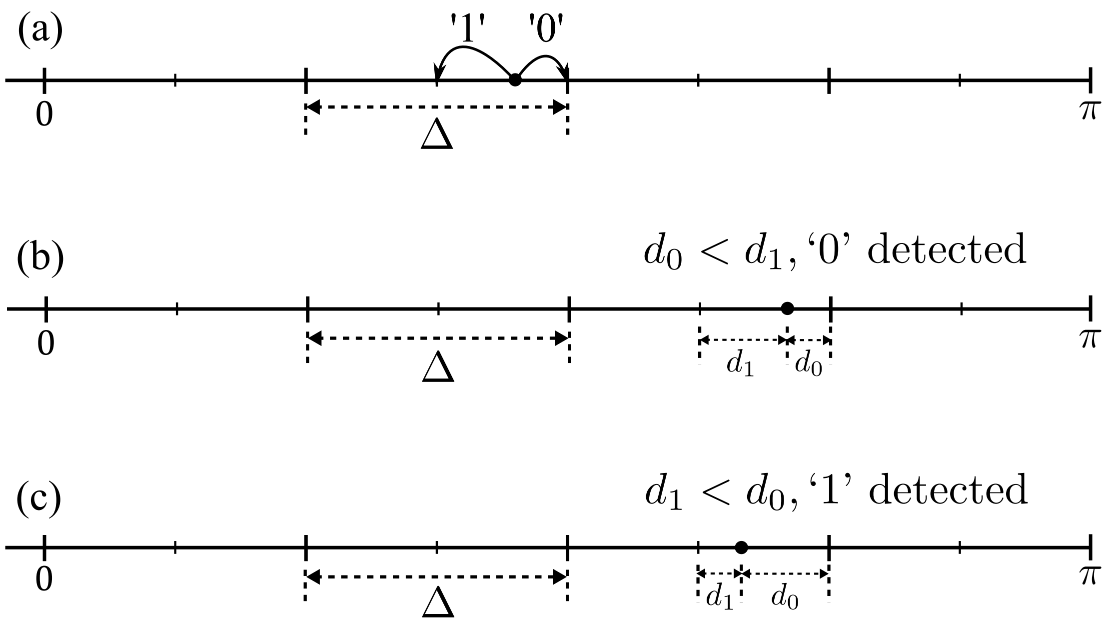


Figure 4.1: Illustration of watermarking based on QIM: (a) embedding, (b) detection of '0', and (c) detection of '1'

according to the HAS [109]. Other methods take the advantages of simultaneous masking characteristics of HAS [97] or relative insensitivity of phase change [20] to embed inaudible watermarks. Phase has been exploited for inaudible audio watermarking since controlled phase alteration results in inaudible change in sound to HAS [82]. Several audio watermarking methods have been proposed based on QIM [19, 22, 89, 90] and showed that QIM is a promising technique for robust watermarking.

We preliminarily study an audio watermarking method based on phase coding that applies QIM to the phase of low frequency components. The experimental results show that the watermark is robust but the sound quality decreases when the bit rate increases. In this method, to embed watermarks, the phase of frequency components is statically modified, regardless of how resistant each frequency component is. However, strong frequency components could be less modified to reduce sound distortion while the resistance of watermarks is still ensured.

We then extend the preliminary study to obtain reasonable trade-off between inaudibility and robustness. We replace the static modification with a dynamic phase coding (DPC) scheme for watermarking, in which the amount of phase modification is adjusted according the frequency component's magnitude. Large-magnitude frequency components are more sensitive to the modification of the phase of that component, with respect to sound distortion. Accordingly, to ensure low sound distortion and high robustness, larger-magnitude frequency components will have small phase modification, whereas smaller-magnitude frequency components will have somewhat higher phase modification.

In some applications, such as fingerprinting and authentication, very high capacity is not required but watermarks need to be perfectly extracted. We further reduce watermark detection error by incorporating error control coding (ECC) into the watermarking system. The experimental results show that the embedded watermarks are inaudible and robust against various attacks. The incorporation of ECC is effective in a manner that watermarks could be extracted without any detection error at a bit rate of 102 bps.

4.2 Proposed Method

4.2.1 Quantization index modulation

QIM has been considered as a class of provably good methods for digital watermarking [19]. The procedure of embedding and detecting watermarks is quite simple. Figure 4.1 shows an illustration of embedding and detection processes. To embed a bit m, '0' or '1' into a scalar variable x, we quantize x to the nearest point that is an even or odd multiple of $\frac{\Delta}{2}$, respectively as (4.1). The obtained variable, y, is sent to receivers and might be affected by channel noise, hence becomes $\hat{y}$. To decode the embedded bit from $\hat{y}$, we calculate the distances between $\hat{y}$ and the nearest even multiple of $\frac{\Delta}{2}$, d_0 and the nearest odd multiple of $\frac{\Delta}{2}$, d_1 and then compare d_0 and d_1 to decode the bit as (4.2) and (4.3).

$$y = Q(x, m) = \begin{cases} \Delta \left\lfloor \dfrac{x}{\Delta} + \dfrac{1}{2} \right\rfloor & \text{if } m = \text{`0'} \\ \Delta \left\lfloor \dfrac{x}{\Delta} \right\rfloor + \dfrac{\Delta}{2} & \text{if } m = \text{`1'} \end{cases} \tag{4.1}$$

where $\lfloor . \rfloor$ is the floor function and Δ is the QIM step size.

$$d_j = \hat{y} - Q(\hat{y}, j), \quad j = \{\text{`0'}, \text{`1'}\} \tag{4.2}$$

$$\hat{m} = \arg\min_j d_j \tag{4.3}$$

4.2.2 Principle of dynamic phase coding

We apply QIM to the phase spectrum of audio signals to construct an inaudible, robust, and reliable audio watermarking system with the following considerations. (i) Phase alteration is relatively inaudible [82], hence slightly modifying the phase keeps watermarks inaudibly embedded. (ii) Most of audio information is distributed in somewhat low frequencies [28], thus this frequency region is more robust against attacks and should be chosen for embedding. (iii) Since modifying the phase of a frequency component causes sound distortion in a manner that is proportional to magnitude of that component, the amount of phase modification should be adjusted to the magnitude. (iv) To increase the reliability, non-meaningful frequency components, i.e., very low magnitude components,

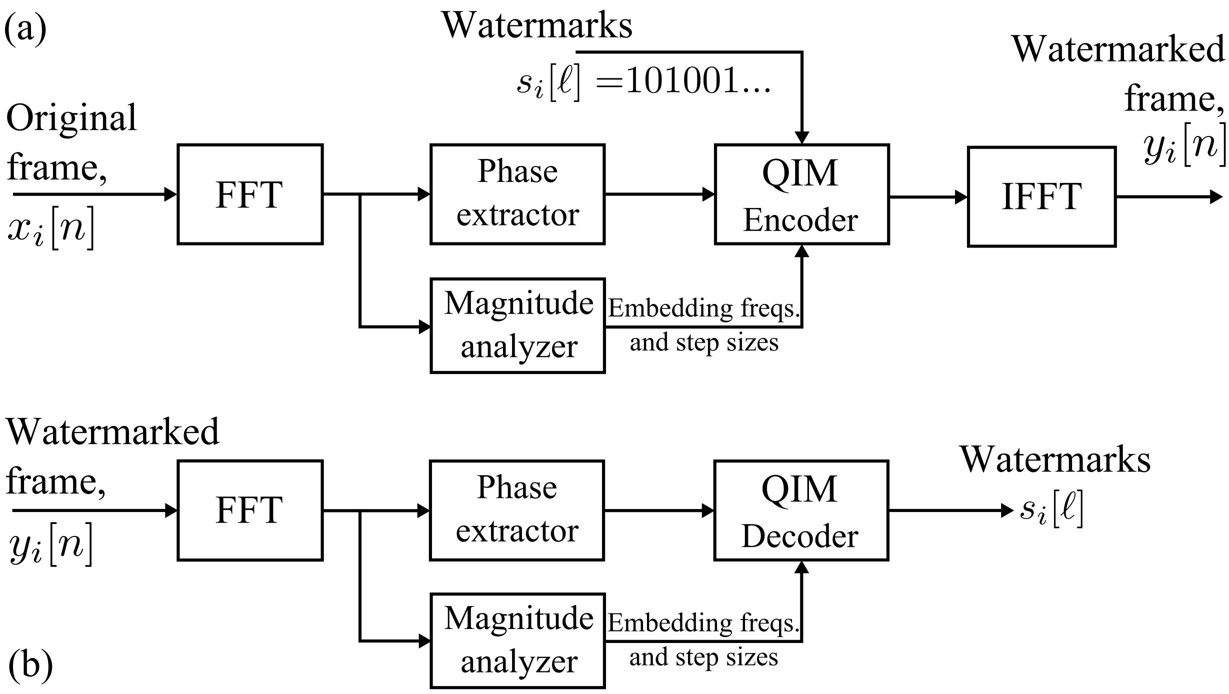


Figure 4.2: Proposed scheme of audio watermarking: (a) embedding process and (b) detection process

are not used for embedding at all. (v) To further reduce detection error, ECC is employed to correct a number of errors.

In this chapter, we investigate whether (i)—(iii) can help increase robustness and inaudibility simultaneously and whether (iv) and (v) can further lower bit error rate by using ECC to encode watermarks before embedding process.

4.2.3 Watermark embedding

The embedding process starts with frame segmentation of the original signal, $x(n)$ into frames $x_i(n)$ with a fixed frame size. Watermark bits $s_i(\ell)$ are embedded into audio frame $x_i(n)$. Figure 4.2(a) depicts a block diagram of the four steps that embed the watermark into an audio frame as follows. Figure 4.3 shows an illustrative diagram of this process.

Step 1. Original frame $x_i(n)$ is transformed into the Fourier spectrum $X_i(\omega)$ by fast Fourier transform (FFT). Magnitude spectrum $|X_i(\omega)|$ and phase spectrum $\angle X_i(\omega)$ are calculated.

Step 2. We select the embedding components as the meaningful frequency components in the range of $[0.001, F]$ kHz, i.e., $f \in [0.001, F]$kHz, $|X_i(f)| > \epsilon$. The watermark bits are embedded into only these selected components to increase reliability.

For each embedding component, the amount of phase modification that is quantified by a QIM step size is determined based on its magnitude. Firstly, the magnitudes are normalized to 1 and linearly divided into L ranges in which each range has a corresponding

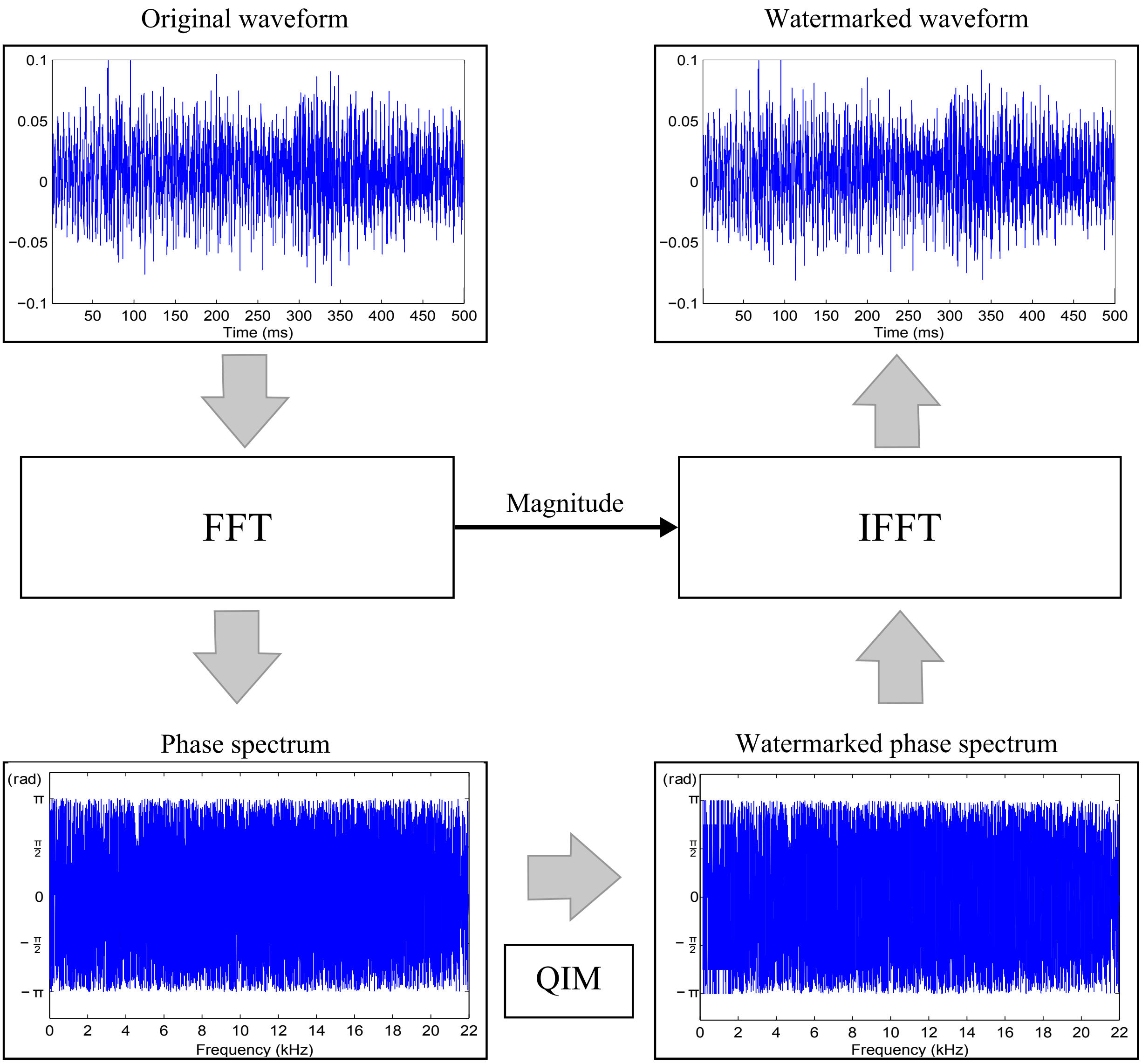


Figure 4.3: Process of watermark embedding: conversion of a frame of time-domain samples into the Fourier domain, QIM watermark addition, and conversion back to the time-domain.

QIM step size. The higher range has a smaller QIM step size.

Suppose that we have a set of L QIM step sizes, $\{\Delta_1, \Delta_2, ..., \Delta_L\}$, where $\Delta_1 > \Delta_2 > ... > \Delta_L$. The QIM step size for an embedding component f, namely Δ, is determined as follows.

$$|\hat{X}_i(f)| = \frac{|X_i(f)|}{M}$$
$$\Delta = \Delta_u, \tag{4.4}$$

where $u = \lceil L|\hat{X}_i(f)|\rceil$ (if $u > L$, $u = L$) and M is the parameter that represents for the maximum amplitude of most frequency components.

Step 3. The bits $s_i(\ell)$ are encoded into the phase of the selected components by (4.1) and a quantized phase spectrum $\angle Y_i(\omega)$ is obtained. Although each bit can be embedded in only one component, it is embedded in several components to increase robustness. The bit rate is adjusted by changing the number of components for each bit.

Step 4. The magnitude spectrum, $|X_i(\omega)|$ and the quantized phase spectrum, $\angle Y_i(\omega)$, are combined into Fourier spectrum $Y_i(\omega)$ which is then transformed into time domain signal $y_i(n)$ by inverse Fourier transform (IFFT).

Finally, all the processed frames are combined together to yield a watermarked signal $y(n)$.

4.2.4 Watermark detection

The detection process also starts with frame segmentation of the watermarked signal, $y(n)$ into frames $y_i(n)$ with the same frame size as in the embedding process. Figure 4.2(b) shows a block diagram of the process that detects watermark bits from a watermarked frame involving three steps as follows.

Step 1. Watermarked frame $y_i(n)$ is firstly transformed into $Y_i(\omega)$ by FFT. Phase spectrum $\angle Y_i(\omega)$ is calculated.

Step 2. The embedding frequency components and corresponding QIM step sizes are determined as in Step 2 in the embedding process.

Step 3. The embedding components are decoded by (4.3) to extract all the bits. The output bits, $s_i(\ell)$, are determined by majority decision, e.g., if the number of '0', N_0, are greater the number of '1', N_1, the output is '0'.

These steps are repeated until we reach the final frame.

4.2.5 Frame synchronization

The detection process works with an assumption that the frame positions are available. In practice, the frame positions might be unavailable, so we have to identify the starting point

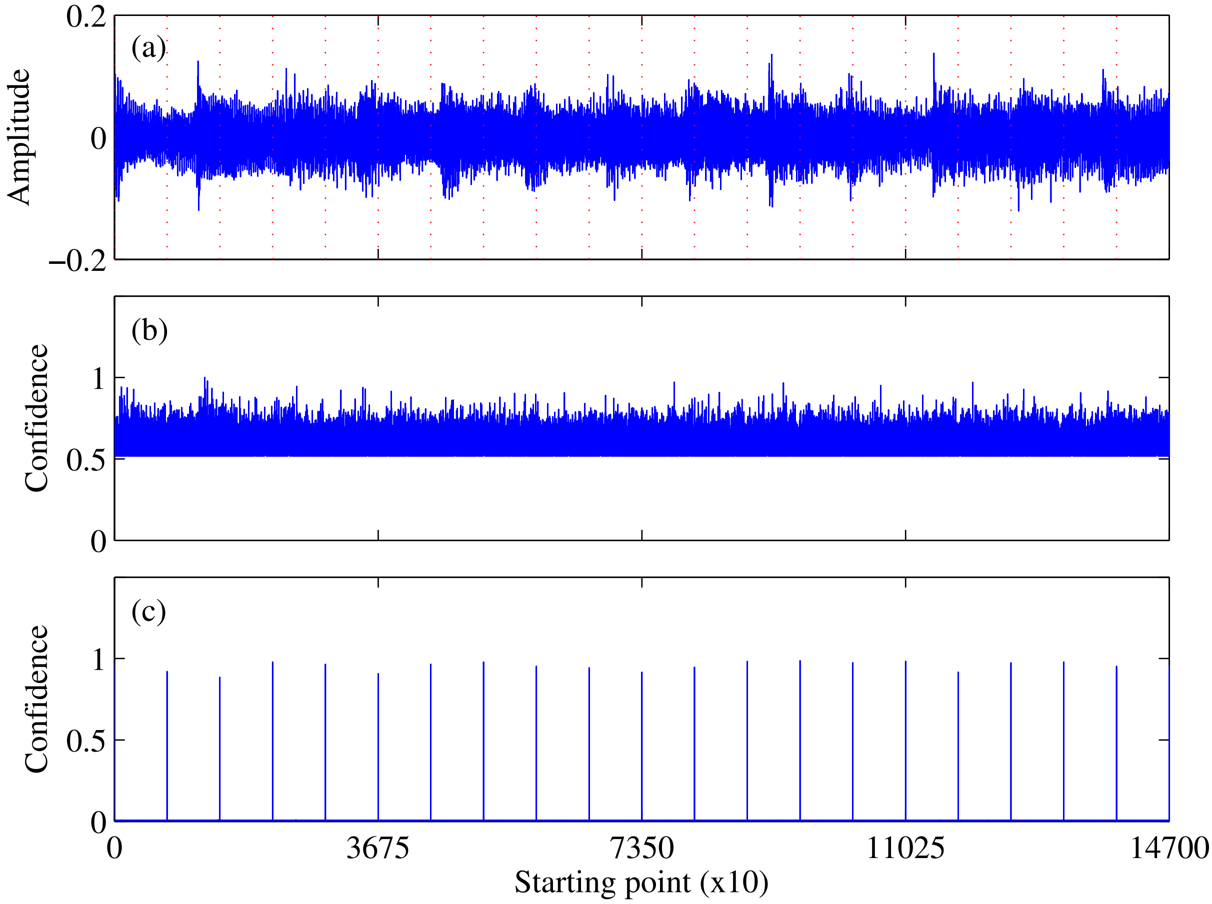


Figure 4.4: Frame synchronization in the case frame length = 7350 points: (a) original signal and (b) watermarked signal

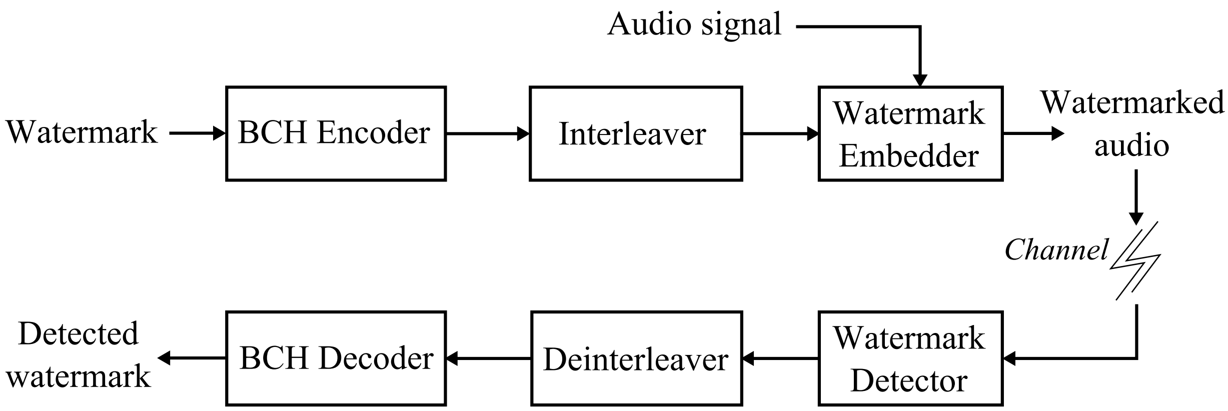


Figure 4.5: Proposed framework of audio watermarking in incorporation with ECC

before detecting watermarks. It is noteworthy that a bit is detected from a watermarked frame by majority decision. N_0 and N_1 are compared to decide the output bit, $s_i(\ell)$. We can see that N_0 much greater than N_1 implies that the probability $P(s_i(\ell) = \text{'0'})$ is much higher. In other words, the confidence that $s_i(\ell)$ is correctly detected is higher. In general, we define the detection confidence of a bit by: $\delta_i(\ell) = \max(N_0/N_1, N_1/N_0)$.

We can search for a correct frame position over a frame length of signal. It is obvious that if we select a correct frame i, $\sum_\ell \delta_i(\ell)$ is maximized. Figure 4.4 depicts an illustration of frame position detection. We calculate detection confidences over 32 frames in two cases: (a) without embedded watermark and (b) with embedded watermark. The detection confidences are normalized to 1. There is no obvious peak in Case (a) while very high peaks occur at the correct frame-starting points in Case (b). The search procedure is performed at the beginning of the detection process. Once the starting point is determined, all the frame positions can be synchronized.

4.3 Incorporation of ECC to the System

Figure 4.5 shows a diagram of the proposed framework of audio watermarking with ECC. The watermark is firstly encoded by a BCH encoder after which certain codewords are obtained. We choose BCH codes because they are binary error-correcting codes with excellent properties, such as simple implementation, not requiring statistic information of the watermark [113]. In order to improve the performance of ECC against burst errors, the BCH codewords are interleaved to distribute the errors into different codewords. The interleaved codewords are then embedded into audio signals. At the receiver side, the watermark detector is firstly used to extract the interleaved codes. Then the extracted codewords are deinterleaved and deinterleaved codewords are finally decoded by BCH decoder. Watermark embedding and watermark detection processes are presented in the previous sections. The next two subsections give descriptions of BCH codes and the

Table 4.1: Typical BCH codes with the length of less than 2^{10}

n_0	k_0	t_0	n_0	k_0	t_0	n_0	k_0	t_0	n_0	k_0	t_0
7	4	1	63	57	1	255	215	5	1023	1013	1
15	11	1	63	45	3	255	187	9	1023	913	11
15	7	2	63	36	5	255	155	13	1023	818	21
15	5	3	63	24	7	255	115	22	1023	728	30
31	26	1	127	120	1	511	502	1	1023	523	55
31	21	2	127	92	5	511	412	11	1023	453	63
31	16	3	127	71	9	511	331	21	1023	348	87
31	11	5	127	50	13	511	250	31	1023	308	93
31	6	7	255	247	1	511	211	41	1023	278	102

interleaving technique.

4.3.1 BCH codes

BCH codes [114] form a class of parameterized error-correcting codes which have been applied to many applications, such as satellite communications, DVD players, and two-dimensional bar codes. The principal advantages of BCH codes is that they are binary codes with excellent minimum distance properties, and can be decoded via an elegant algebraic method which allows very simple electronic hardware to perform the task. BCH codes are also highly flexible, allowing control over block length and acceptable error thresholds.

A BCH code is represented by (n_0, k_0, t_0) in which n_0 is the code length, k_0 is the data length, and t_0 is the number of correctable bits [113]. If the number of errors occurring during transmission is less than or equal to t_0, all the errors can be corrected. Otherwise, BCH code fails to correct the errors. Several typical BCH codes with the length of less than 2^{10} are shown in Tab. 4.1.

BCH codes add additional parity bits to the data bits, hence when applied to audio watermarking, BCH codes reduce the bit rate of watermark. The higher value of t_0 creates a stronger BCH code, i.e., it can correct more errors. However, k_0 becomes smaller which reduces the actual bit rate of watermark. A suitable BCH code can help improve reliability. In practice, the strategy for section of BCH code's parameters is firstly fixing the criteria for sound quality and bit error rate, then selecting the parameters (n_0, k_0, t_0) so that those criteria could be met and the bit rate is maximized.

Table 4.2: Configuration on the QIM step sizes with regard to the normalized magnitude

	QIM step sizes (Δ)				
Normalized magnitude	(0, 0.2]	(0.2, 0.4]	(0.4, 0.6]	(0.6, 0.8]	(0.8, 1]
Set 1	$\frac{\pi}{2}$	$\frac{\pi}{4}$	$\frac{\pi}{6}$	$\frac{\pi}{8}$	$\frac{\pi}{10}$
Set 2	$\frac{\pi}{3}$	$\frac{\pi}{5}$	$\frac{\pi}{7}$	$\frac{\pi}{9}$	$\frac{\pi}{11}$
Set 3	$\frac{\pi}{4}$	$\frac{\pi}{6}$	$\frac{\pi}{8}$	$\frac{\pi}{10}$	$\frac{\pi}{12}$

4.3.2 Interleaving

In audio watermarking systems, errors typically occur in bursts rather than independently. Interleaving multiple codewords can be used to improve performance of error correcting codes. If the number of errors due to a burst are greater than the error-correcting code's capacity, the error-correcting code cannot recover the original codeword successfully. Interleaving shuffles the source symbols across several codewords in order to create a more uniform distribution of errors, reducing the effect of burst errors.

4.4 Evaluations

4.4.1 Database and condition

Experiments were carried out to evaluate inaudibility and robustness of the proposed method with 102 RWC music tracks [110] that have a sampling frequency of 44.1 kHz and 16-bit quantization. The frame size is set to 500 ms. The FFT size is equal to the frame size and the rectangle window was used. The watermarks were randomly generated. The parameters, F, ϵ, L, and M, were determined by experimental analysis and set to 1.6 kHz, 10^{-4}, 5, and 0.005 respectively. The QIM step sizes are chosen as integer divisions of π to reduce wrapping errors. We investigated the proposed method with three sets of five QIM step sizes as shown in Tab. 4.2.

Inaudibility was tested by PEAQ [33] which rates sound quality by the ODG from -4 (very annoying) to 0 (imperceptible). Detection accuracy was measured by BDR, the ratio between the numbers of correct bits and total bits. Robustness was investigated with the following processing: MP3 128 kbps, MP4 96 kbps, adding white Gaussian noise 36 dB, requantization 8 bits, resampling 22 kHz and 16 kHz, and bandpass filtering with passband [0.1, 6] kHz and stopband attenuation -12 dB/octave.

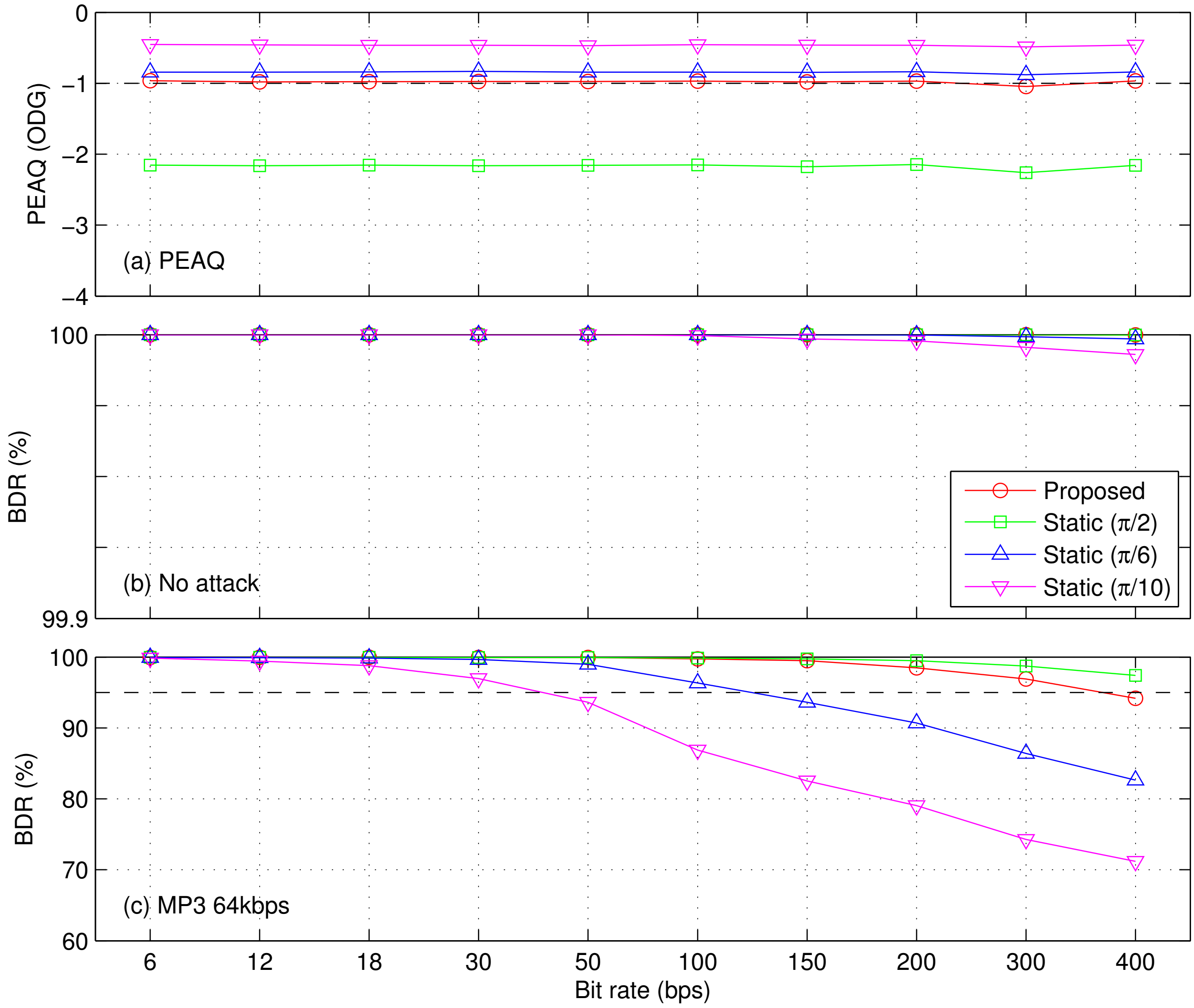


Figure 4.6: Sound quality (PEAQ) and BDR with respect to bit rate in comparison with the static scheme: (a) PEAQ, (b) BDR without attack, and (c) BDR against MP3 attack

4.4.2 Effectiveness of the proposed method

We carried out experiments to test the PEAQ and BDR of the proposed method (DPC) in comparison with an audio data hiding method based on a static phase coding (SPC) scheme, in which the QIM step size is fixed for all the frequency components. The first QIM step set (Set 1) was used for the proposed method. We investigated the SPC scheme with the QIM step size of $\pi/2$, $\pi/6$, and $\pi/10$.

Figure 4.6 shows the results of the PEAQ and the BDR. For the SPC scheme, the PEAQ increases from -2.17 ODG to -0.46 ODG while the BDR decrease in both the cases of no attack and MP3 attack as the QIM step size decreases from $\pi/2$ to $\pi/10$. Especially, the BDR drastically decreases when the bit rate increases. These results suggest that the SPC scheme meets a trade-off among inaudibility, robustness, and capacity.

In contrast, the proposed method has the PEAQ approximate to -1 ODG and a high

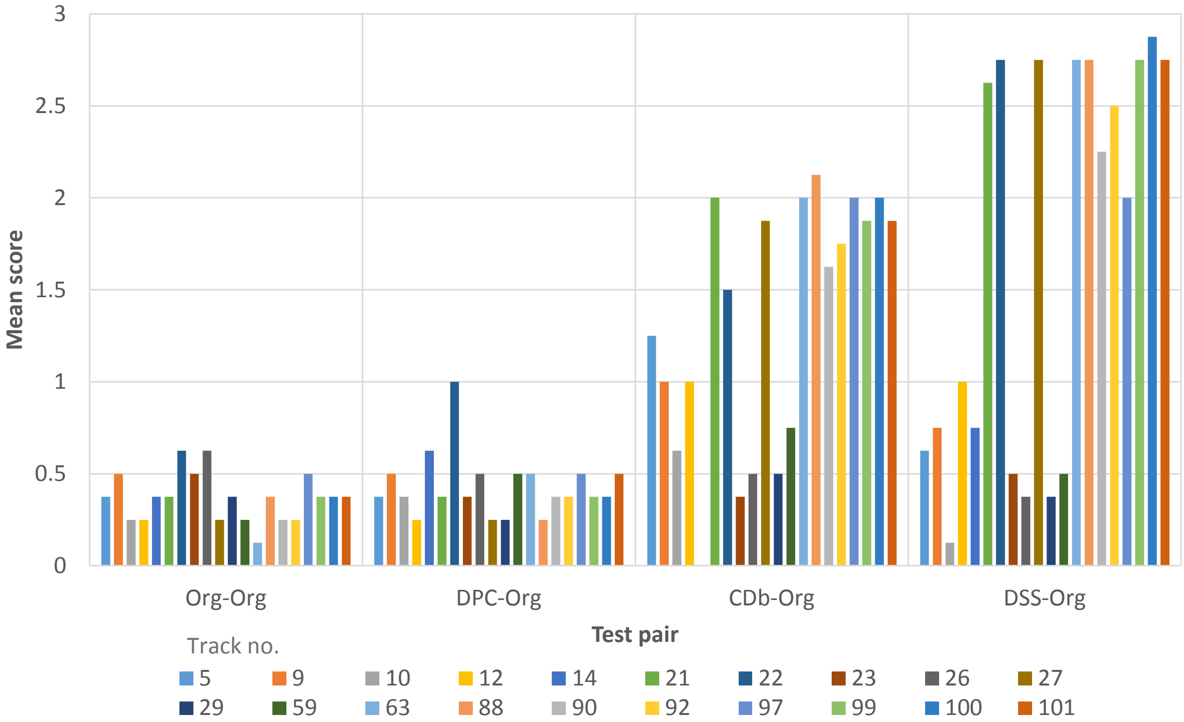


Figure 4.7: Results of subjective listening test.

BDR for both the cases of no attack and MP3 attack, even at high bit rate (up to 300 bps). The results reveal that the proposed method has a better trade-off among inaudibility, robustness, and capacity than the SPC scheme. Accordingly, the proposed DPC scheme is effective for audio data hiding in obtaining a reasonable trade-off among the properties.

4.4.3 Subjective listening test

The subjective listening test was carried out with the settings as the same as those in Sect. 3.3.3. For each track, there are four pairs in a random order, in which the first one in the pair is the original track (Org) and the other one is the original track or an watermarked track created by the proposed method (DPC), the CDb method [112], or the DSS method [17]. Each subject evaluated the similarity of 80 test pairs in total, which is 20 tracks multiplied by 4 pairs (Org-Org, DPC-Org, CDb-Org, and DSS-Org).

Figure 4.7 shows the mean scores of four test pairs for each track. The mean scores of the pairs DPC-Org are approximate to those of the pairs Org-Org, revealing that the distortion caused by the watermark embedding of the proposed method is hardly perceptible. In contrast, the distortion caused by the CDb method and the DSS method is slightly perceptible and perceptible for most of the tracks.

4.4.4 Robustness of the proposed method

We carried out experiments to evaluate the robustness of the proposed method. Figure 4.8 shows the results of PEAQ and BDR of the proposed method with three sets of QIM step size. The bit rate was varied from 6 to 400 bps. All the PEAQs are greater than -1 ODG (not annoying) and the sound quality of watermarked signals remains unchanged as the bit rate increases. The sound quality becomes better when the QIM step size decreases from the values in Set 1 to those in Set 3.

In the cases of no attack and resampling, the BDRs are greater than 99.9% for all the bit rates and do not change much among three sets of QIM step size. In the cases of MP3, MP4, adding white noise, and requantization, the BDRs are greater than 99% for the bit rates less than or equal to 200 bps with Set 1 and slightly decrease with Set 2 and Set 3. The bandpass filtering seems to be the strongest attack which makes the BDRs around 90% for the bit rates less than or equal to 200 bps with Set 1 and decrease much more with Set 2 and Set 3.

These results suggest that the proposed method is effective with regard to inaudibility and robustness. The proposed method provides good sound quality in the watermarked signals and high robustness against most types of processing.

4.4.5 Effectiveness of ECC

We evaluated the effectiveness of incorporation of ECC into the watermarking system in the case of MP3 64 kbps. We chose to investigate MP3 attack because it is popularly used in practice and is the strongest attack except for bandpass filtering. If ECC can correct errors after MP3 compression, it can also correct errors from the other attacks. The five codes with the length of 1023 and different values of k_0 have been used.

Figure 4.9 shows the bit-error rate (BER) after BCH decoding, with respect to embedding bit rate with three sets of QIM step size. The results show that the system can extract watermarks without any detection error at a bit rate of 102, 51, and 28 bps with Set 1, Set 2, and Set 3, respectively. Compared with the case that ECC is not used to encode watermarks, the incorporation of ECC is remarkably effective in correcting all the errors at relatively high bit rates.

4.5 Summary

We proposed an audio watermarking method based on dynamic phase coding and ECC. Watermarks are embedded into the phase of somewhat low frequency components. The QIM step size for each component is adjusted according to the magnitude to balance inaudibility and robustness. BCH coding is applied in encoding watermarks before embedding process to increase reliability. We carried the subjective listening test and the objective evaluation to evaluate the inaudibility and robustness of the proposed method. The first experiment has been carried to confirm the effectiveness of the DPC scheme. The second experiment is to investigate the robustness of the proposed method under various kinds of attacks. The last experiment has been conducted to evaluate the effectiveness of the incorporation of ECC into the data hiding system.

The experimental results suggest that the proposed DPC scheme is more effective than the SPC scheme, resulting in a better trade-off among the properties. The results also reveal that the watermarked signals produced by the proposed method have high sound quality and the embedded watermarks are robust against various attacks. The incorporation of ECC is effective for audio watermarking to carry more reliable watermark in practice.

The proposed method is capable of embedding watermarks into audio signals at a bit rate of 102 bps with the accuracy of watermark detection of 100 %. It has been revealed that the proposed method could satisfy the requirements of *inaudibility*, *blindness*, *robustness*, *high capacity*, and *high reliability*.

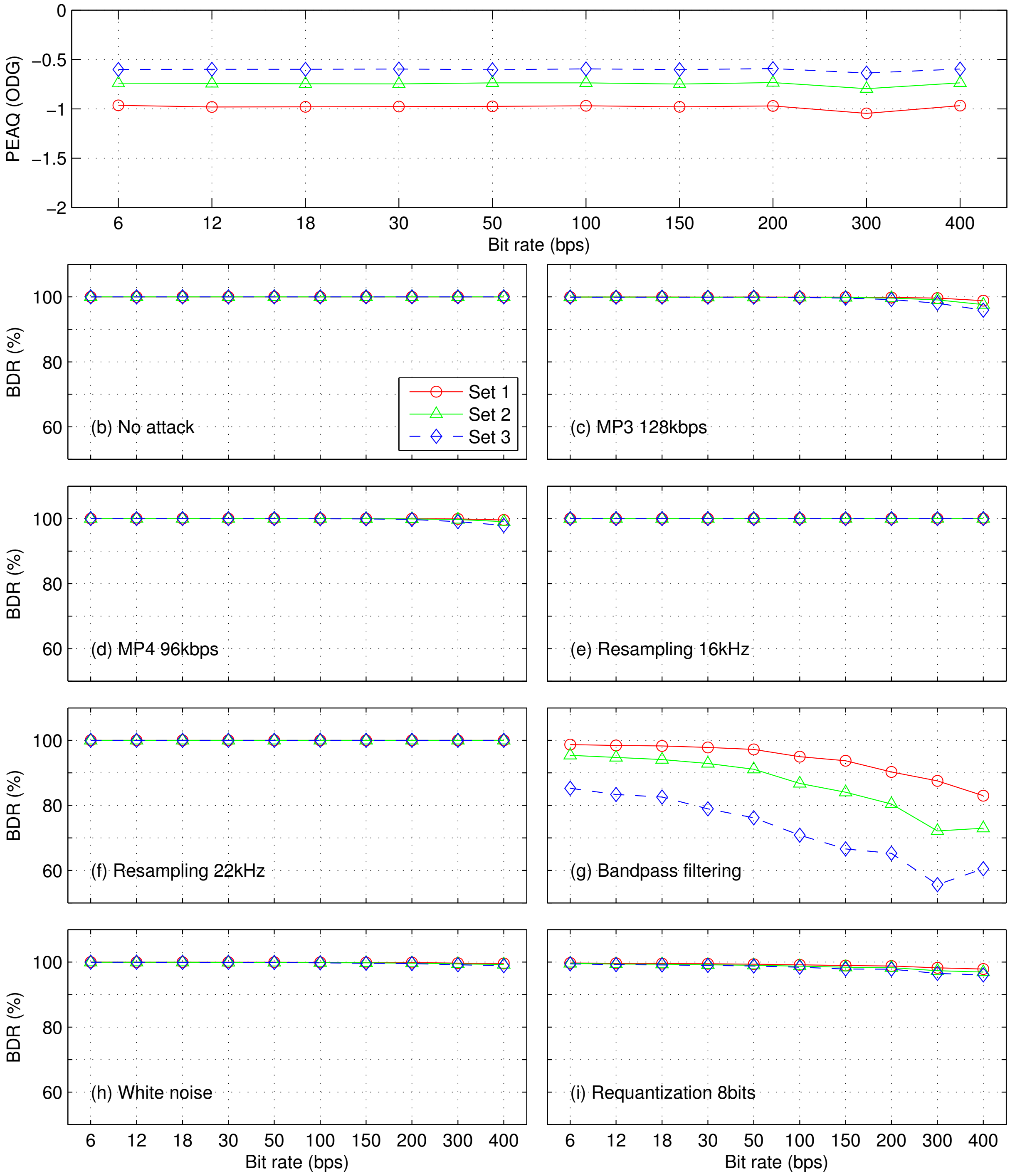


Figure 4.8: Sound quality (PEAQ) and BDR with respect to bit rate in comparison with the static scheme: (a) PEAQ, (b) BDR without attack, and (c)–(i) BDR against attacks

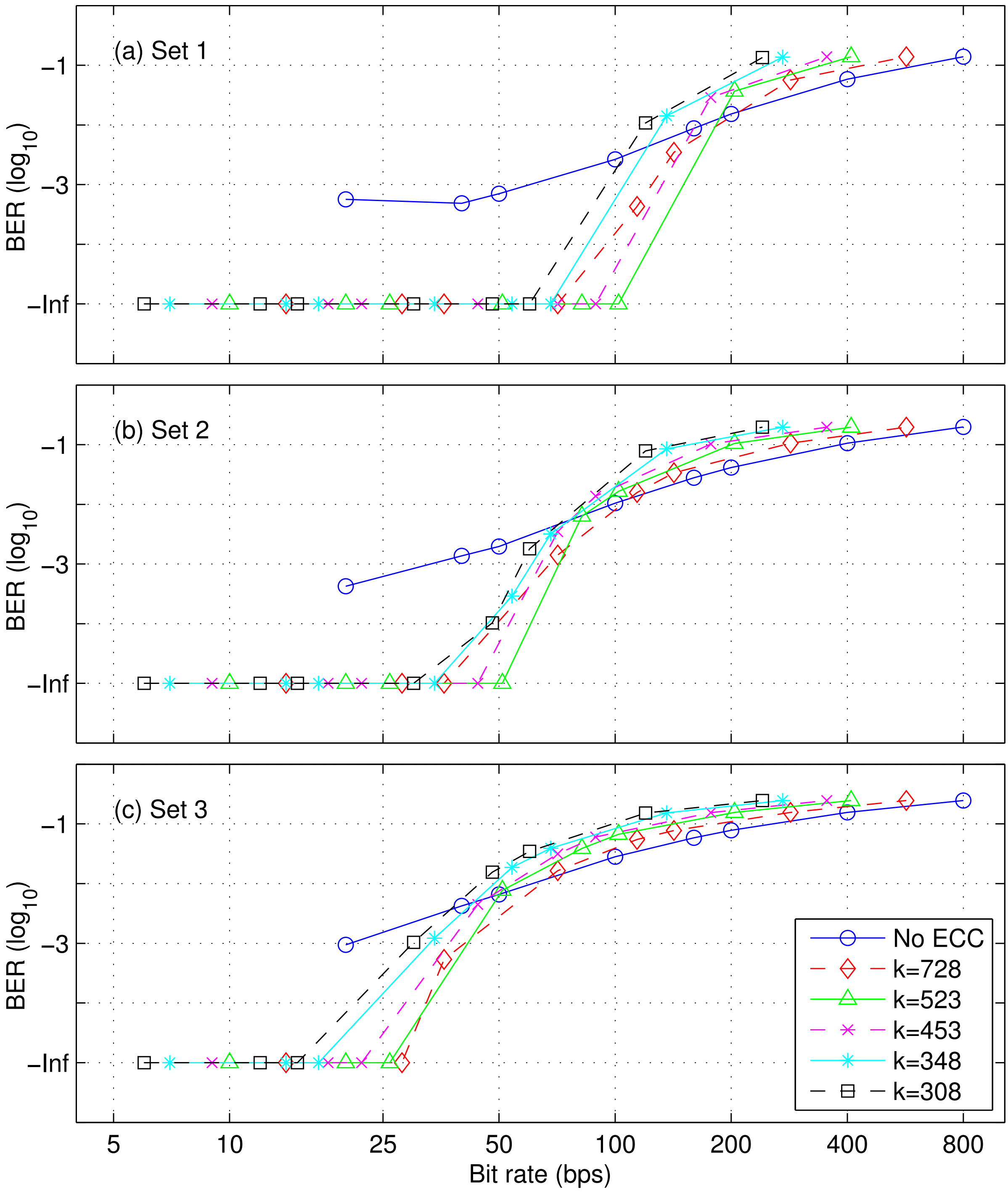


Figure 4.9: BER after BCH decoding in the case of MP3 attack with respect to embedding bit rate

Chapter 5

Applications of Audio Data Hiding

5.1 Application to Copy Prevention

Human culture history has proven that music plays an essential role in our daily lives. Listening to music and creating it have always accompanied with the development of society. Music brings people together, is a key to creativity, and is a universal language. Scientific research has been conducted to improve our experiences on listening to music. Compact discs have been invented to store music. MP3 compression makes music digitally stored in smaller size. Mechanism of human perception on music has also been studied to explore the myths behind that make music fuels our mind. Music is an invaluable asset of human, thus protecting music is always an important task.

As an effort to combat copyright infringement, copy protection prevents publication of copyrighted digital data to file sharing services. Unlike conventional copy prevention that restricts duplication of digital audio on users' computer that has been proven impossible [8], this technique is made feasible by enforcing file sharing services to comply with the copyright law. Users cannot upload digital audio to fire sharing services without rightful ownership. Figure 5.1 sketches a schematic frame of the application of copy prevention. In this application, unique identity or digital fingerprint of copyrighted audio is embedded as watermark into the audio signal. The digital fingerprint should be unchanged with imperceptible modification in the audio signal. When the audio signal is uploaded to a file sharing website, the embedded watermark is extracted and used to verify if it is legal to be uploaded. Media sites like YouTube can analyze files and compare their fingerprints with a database of copyrighted material and stop users from uploading these files [12].

The watermark should be very robust and secure to withstand against attempts to distort it. This application offers a possibility to limit widespread sharing of copyrighted digital audio. However it cannot stop peer-to-peer file sharing, which transfers digital audio directly between users.

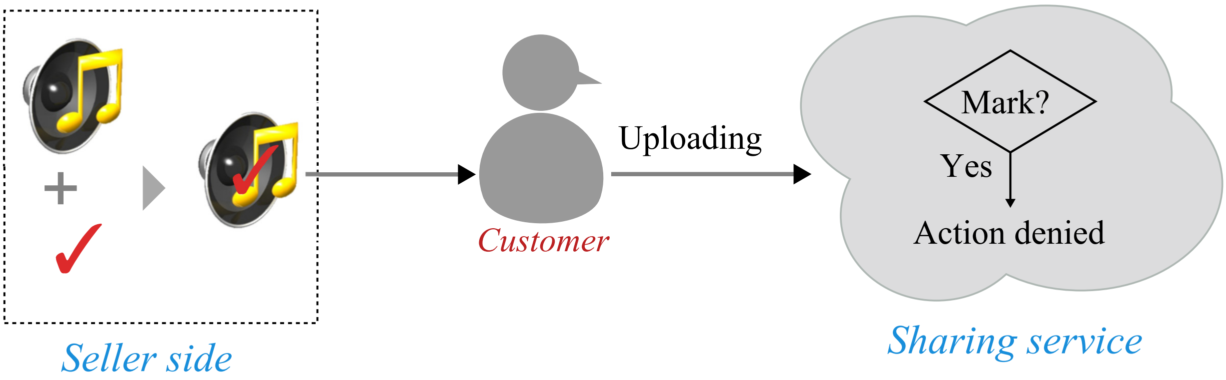


Figure 5.1: Schematic framework of the application of copy prevention

5.2 Application to Annotated Audio

In this application, watermark is used as a channel carrying complementary information for digital audio content. Content information about an audio signal is interesting to users, especially in music applications, helping enrich users' experience on listening to music. Another usage scenario is to enable broadcasters to enhance entertainment experiences for their consumers. The watermark is expected to have high capacity in order to convey data as much as possible. Inaudibility is also desired to maintain the high sound quality. Figure 5.2 shows a schematic framework of this application.

Watermark can be employed as a channel for carrying any kind of information [44–47]. For example, it could be the content information which describes for a music track or the information that adds value for radio or TV programs.

Content information is defined as information about an audio signal that is interesting for users, especially in music applications. Depending on specific application, several levels of content information could be defined. Some examples of interest we can mention are: content information describing a song (timbre, rhythmic, or harmonic description), meta-data describing a music track (title, composer, created date, performer, performance date, lyrics, album cover art, copyright notice).

Broadcasters can embed text to their radio or TV programs to enhance entertainment experiences for consumers [15, 48, 49]. The text is then displayed on a small screen of the radio receiver or TV. Here are some examples that we can think about: station identification, current time, advertisement, breaking news, weather forecast, stock prices.

These applications requires to transfer an as-much-as-possible amount of information, thus embedding capacity is highly necessary. Inaudibility should also be satisfied to maintain the high sound quality of audio signals.

Figure 5.2: Schematic framework of the application of annotated audio

5.3 Application to AM Radio Broadcast

5.3.1 Introduction

Radio is a simple and typical technique for the wireless communication of signals through free space. The baseband signal, which represents text, images, or audio, is conveyed by a carrier signal. The amplitude, frequency, or phase of carrier signals is modified in relation to baseband signals by a process called modulation. The two most typical kinds of analog modulation technique are AM and frequency modulation (FM), and we have corresponding radio systems, i.e., AM and FM radios. AM radio has some advantages over FM radio, such as its use of narrow channel bandwidth and wider coverage area, while FM radio provides a better sound quality of the conveyed sound at the expense of a using larger channel bandwidth [115]. AM and FM radios have been used popularly in our daily life to receive radio services broadcasting news, entertainment, and educational programs. When radio users listen to radio programs, all the information is usually provided by audio signals only. However, if additional information such as programs, weather forecasting, news, advertisements, etc. was also provided digitally, as in recently introduced smart TV broadcasting, such information could improve the utility of radio services as well.

As radio transmitters and receivers for radio are simple in construction, radio is a robust communication technology and thus useful in the event of natural disasters (e.g., earthquakes or typhoons). Radio receivers can operate for many hours on just a few batteries while other devices such as televisions and PCs cannot. Radio is a particularly useful device during and after disasters in which power suppliers can easily be cut off. Emergency Alert Systems (EAS) [116] and Earthquake Early Warning systems (EEW) [117, 118] have been working on broadcasting warnings via radio services. Audible emergency warnings are used to attract the attention of people engaged in everyday activities (e.g., office workers and car drivers). These emergency warnings allow workers to take protective actions such as slowing and stopping trains or taking steps to protect important infrastructure, and provide people a few seconds to take cover [119, 120]. The warning messages need to be accurately and quickly understood by everyone during emergencies, though, so complementing audible messages with additional digital information expressed in characters is needed to more effectively broadcast warnings during and even after disasters. The application of multimedia to conventional radio will therefore be useful particularly in the event of emergencies where the ordinary Internet does not work well.

In a nutshell, supplementing the sound (speech) information conveyed by radio waves with additional digital information is likely to be very useful in emergency situations. Data hiding for radio signals is therefore important for public safety. Some advanced radio systems such as Radio Data System (RDS) [121] and AM Signalling System (AMSS) [122–124] have been proposed to embed a certain amount of digital information within broadcast signals. RDS embeds digital information into a subcarrier of the broadcast signal in FM radio broadcasts. The data rate of the RDS is 1187.5 bps [121]. RDS allows some functions to be implemented in FM radio services, such as Program Identification, Program Service name, and Alternative Frequency list. AMSS provides broadly similar functionality to that offered by RDS, and uses low bit-rate phase modulation of the AM carrier to add a small amount (about 47 bps [123]) of digital information to the broadcast signal. The envelope detector used in conventional AM receivers, however, does not respond to AMSS. While AMSS has the limitations of a low data rate and incompatibility with conventional radio receivers, RDS can be effectively used to embed digital information with a high data rate into radio signals in FM radio broadcasts. This drawback motivated us to search for a new method of information hiding for AM signals.

Data-hiding techniques such as audio watermarking have been proposed in recent years to protect the copyright information of public digital-audio content and to transmit digital information in the same channel [55]. Many approaches have been taken towards watermarking for digital audio content, which can embed and then precisely detect data. These approaches have been implemented in computer and the results suggest that digital information can be embedded into digital audio with little or no effect on the listener's perception of the target audio. It has also been reported that the bit rate of embedded data is reasonably high, in the hundreds of bits per second (bps) range [55, 59]. We can take these advantages of watermarking techniques and apply them to an AM radio system to create a data hiding scheme for AM radio signals.

This section presents a data-hiding scheme for AM radio signals. Digital information is embedded into audio signals instead of modulated signals as is done in RDS and AMSS systems. The carrier signal is then modulated by the audio signals and transmitted to receivers. Although methods of embedding data and detecting data have been taken care of by watermarking techniques, some issues might arise due to the modulation and demodulation processes in an AM radio system. For instance, the inaudibility of watermarked signals and the accuracy of embedded data detection could be reduced. Additionally, the vast majority of users listen to AM radio using conventional receivers. The hiding system should have downward compatibility, i.e. it should yield AM signals that can be detected by conventional receivers in order to be widely applied in practice. Suitably-equipped receivers will be able to extract embedded data in addition to audio signals while conventional receivers will be able to extract only an audio signal. The proposed scheme could be used to broadcast audio signals and embedded data to receivers over an AM radio link, and applied to construct a high utility AM radio service able to multi-modally distribute essential information during emergencies.

The rest of this section is organized as follows. Section 5.3.2 provides a basic explanation of the amplitude modulation technique and the concept underlying the data-hiding scheme. In Sect. 5.3.3, we explain the proposed data-hiding scheme. Section 5.3.4 gives the results from an evaluation of the proposed scheme. Discussion and conclusion are presented in Sect. 5.3.5.

5.3.2 Background

Amplitude modulation in radio broadcast

Modulation is a typical technique used to transmit information (e.g., text, image, and sound) in telecommunication. It is used to gain certain advantages such as far-distance communication and transmission of signals over radio waves. The baseband signal, which carries the information, cannot be directly transmitted over a radio wave because it has sizable power at low frequencies. The size of antennas that radiate the waveform signal is directly proportional to the signal wavelength. Long-haul communication over a radio wave requires modulation to enable efficient power radiation using antennas of reasonable dimensions [125].

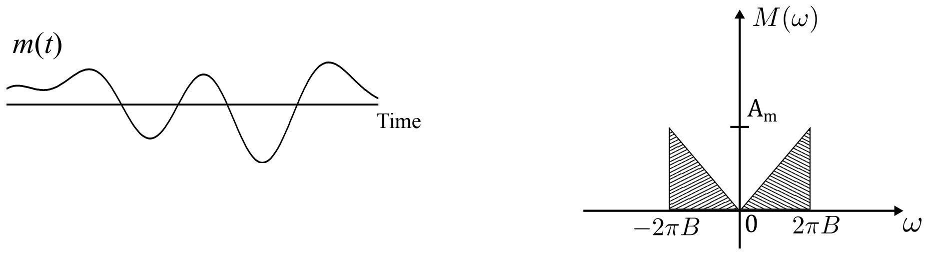
AM varies the amplitude of a carrier signal $c(t)$ which is usually a sinusoidal signal of high frequency in proportion to the message signal $m(t)$ according to the formula,

$$u(t) = [A + m(t)]c(t) = [A + m(t)]\cos(\omega_c t), \tag{5.1}$$

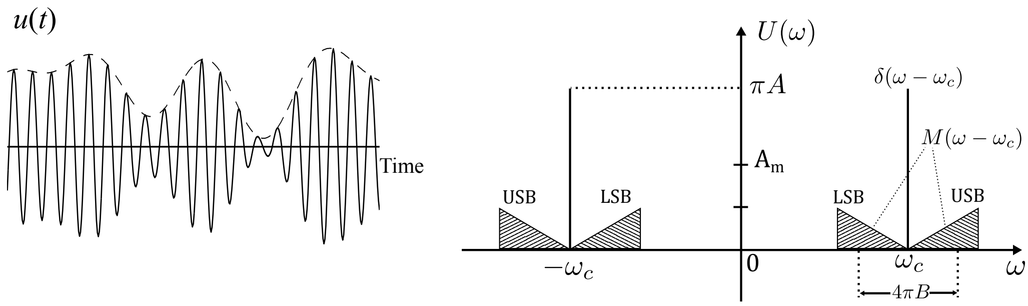
where ω_c is angular frequency (rad/s) of the carrier signal and $u(t)$ is referred to as an AM signal. The modulation depth is defined as $\frac{\max(|m(t)|)}{A}$.

Equation (5.1) represents the basic AM scheme, namely a double-sideband with carrier (DSB-WC). There are other types of AM such as double-sideband suppressed carrier (DSB-SC) and single-sideband modulation (SSB) which can be referred to in [125].

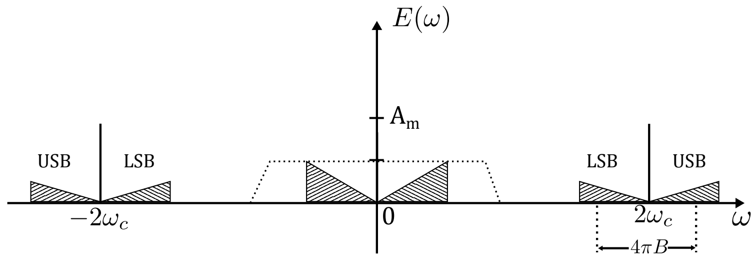
The waveform and spectrum of the message signal and the modulated signal are de-



(a) Message signal $m(t)$



(b) Modulated signal $u(t)$



(c) Demodulated signal

Figure 5.3: Waveform and spectrum of signals in DSB-WC method: (a) message signal, (b) modulated signal, and (c) demodulated signal.

picted in Figs. 5.3(a) and 5.3(b), respectively. AM simply shifts the spectrum of $m(t)$ to the carrier frequency. Suppose $M(\omega)$ is the spectrum of the message signal $m(t)$; the spectrum of the AM signal is then expressed by

$$U(\omega) = \pi A\delta(\omega + \omega_c) + \pi A\delta(\omega - \omega_c) + \frac{1}{2}M(\omega + \omega_c) + \frac{1}{2}M(\omega - \omega_c). \tag{5.2}$$

We observe that if the bandwidth of $m(t)$ is B Hz, then the bandwidth of the modulated signal $u(t)$ is $2B$ Hz. The spectrum of the modulated signal centered at ω_c is composed of two parts: a portion called the lower sideband (LSB) lying below ω_c and a portion called the upper sideband (USB) lying above ω_c. Similarly, the spectrum centered at $-\omega_c$ also has an LSB and USB.

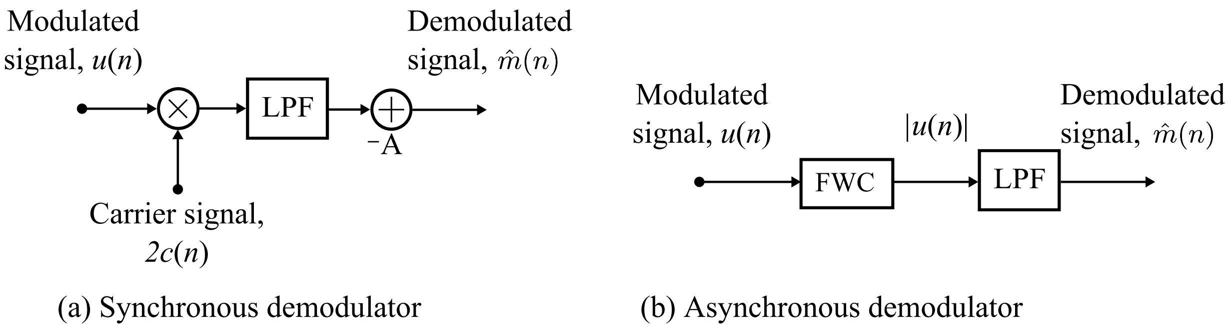


Figure 5.4: Block diagram of demodulation methods: (a). synchronous demodulator and (b) asynchronous demodulator for DSB-WC.

Demodulation of the AM signal

There are two types of demodulation: synchronous demodulation and asynchronous demodulation. They differ in the use of a carrier signal in demodulation processes. The receiver must generate a carrier signal synchronized in phase and frequency for synchronous demodulation, but the carrier signal is not needed in asynchronous demodulation.

Synchronous Demodulation This kind of demodulation can be referred to as a coherent or product detector. Figure 5.4(a) shows a flow chart of a synchronous demodulator. At the receiver, we multiply the incoming modulated signal by a local carrier of frequency and phase in synchronism with the carrier used at the transmitter.

$$\begin{aligned} e(t) = u(t)c(t) &= [A + m(t)]\cos^2(\omega_c t) \\ &= \frac{1}{2}[A + m(t)] + \frac{1}{2}[A + m(t)]\cos(2\omega_c t) \end{aligned} \tag{5.3}$$

Let us denote $m_a(t) = A + m(t)$. Then,

$$e(t) = \frac{1}{2}m_a(t) + \frac{1}{2}m_a(t)\cos(2\omega_c t) \tag{5.4}$$

The Fourier transform of the signal $e(t)$ is

$$E(\omega) = \frac{1}{2}M_a(\omega) + \frac{1}{4}[M_a(\omega + 2\omega_c) + M_a(\omega - 2\omega_c)] \tag{5.5}$$

The spectrum $E(\omega)$ consists of three components as shown in Fig. 5.3(c). The first component is the message spectrum. The two other components, which are the modulated signal of $m(t)$ with carrier frequency $2\omega_c$, are centered at $\pm 2\omega_c$.

The signal $e(t)$ is then filtered by a lowpass filter (LPF) with a cut-off frequency of f_c to yield $\frac{1}{2}m_a(t)$. We can fully get $m_a(t)$ by multiplying the output by two. We can also get rid of the inconvenient fraction $\frac{1}{2}$ from the output by using the carrier $2\cos(\omega_c)$ instead of

$\cos(\omega_c)$. Finally, the message signal $m(t)$ can be recovered by $\hat{m}(t) = m_a(t) - A$.

Asynchronous Demodulation. An asynchronous demodulator can be referred to as an envelope detector. For an envelope detector, the modulated signal $u(t)$ must satisfy the requirement that $A + m(t) \geq 0, \forall t$. A block diagram of the asynchronous demodulator is shown in Fig. 5.4(b). The incoming modulated signal, $u(t)$, is passed through a full wave rectifier (FWR) which acts as an absolute function. The FWR output which is the absolute value of $u(t)$, $|u(t)|$, is then filtered by a low-pass filter resulting in the demodulated signal, $\hat{m}(t)$.

A synchronous demodulator can decode over-modulated signals and the modulated signals produced by DSB-SC and SSB . A signal demodulated with a synchronous demodulator should have signal to noise ratio higher than that of the same signal demodulated with an asynchronous demodulator. However, the frequency of the local oscillator must be exactly the same as the frequency of the carrier at the transmitter, or else the output message will fade in and out in the case of AM, or be frequency shifted in the case of SSB . Once the frequency is matched, the phase of the carrier must be obtained, or else the demodulated message will be attenuated. On the other hand, an asynchronous demodulator does not need to generate the carrier signal at the receiver, thus it is simple to implement in practical applications.

5.3.3 Data Hiding Scheme for AM Radio Broadcast

Concept of data hiding scheme

Telecommunication systems, such as AM radio broadcasting systems, modulate the carrier signal with the audio signal for transmission of the audio signal to receivers. Modulation is advantageous because the baseband signals have sizable power at low frequencies while modulated signals have sufficiently strong power to be transmitted over radio links [125]. At the receivers, the modulated signals are demodulated to extract the audio signals. Based on this principle of modulation, two approaches to embedding data into AM radio signals seem feasible. The first approach is to directly modify modulated signals to embed data. With this approach, however, it is difficult to ensure downward compatibility. Another approach is to embed data into audio signals and then modulate the carrier signal with the embedded audio signals. This approach could be implemented by employing a proposed audio watermarking scheme. Our work focuses on this approach to construct data hiding scheme for digital-audio in AM domain.

Figure 5.5 shows the approach to embedding data into AM radio signals. Audio signals are embedded with data by using an available method of audio watermarking before they are modulated for transmitting. Many methods of audio watermarking have been devel-

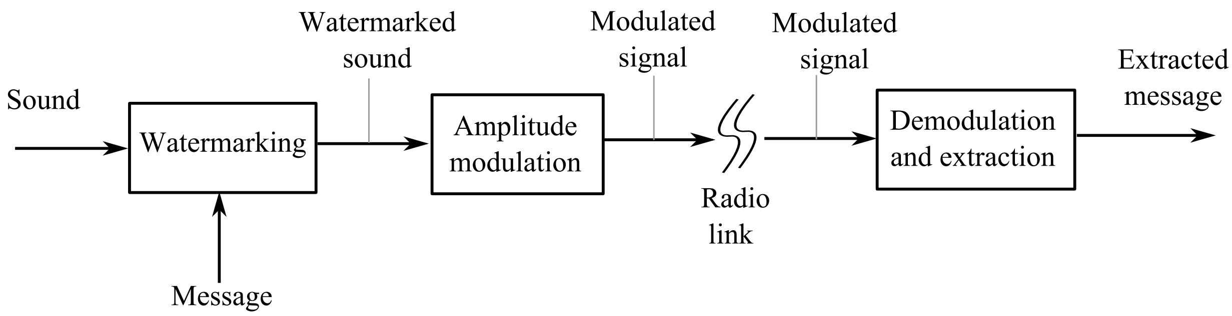


Figure 5.5: Approach to a data-hiding scheme for AM radio signals.

oped in recent years. The method of audio watermarking used to embed data into audio signals in this scheme should satisfy the requirements of inaudibility, robustness, and high capacity. After the audio signal is embedded with data, the embedded signal is modulated by a modulation process. The modulated signal is then sent to receivers through an AM radio link. When the receivers in this scheme receive the modulated signal, they can demodulate the received signal and extract the embedded message from the demodulated signal by using a demodulation process and a data extractor. The extracted audio signal and the detected data are then used for the desired purposes.

We employed the proposed method of audio watermarking based on DPC because it enable a high embedding capacity, provides high quality of watermarked audio, and is robust against signal modification. To efficiently apply the proposed method to AM radio broadcast system, the algorithm was designed to yield AM signals that can be reacted to by conventional radio devices.

In the same approach, there is a data hiding method for AM radio signals that employs a non-blind watermarking method (namely cochlear delay (CD)-based watermarking) [126]. Although the watermarking method is non-blind, which needs the original signal available in the detection proposed, the method in [126] overcame this drawback by introducing a double-modulation scheme, which modulates both the original signal and the watermarked signal in the same carrier. Using a non-blind method can benefit a better performance on sound quality of watermarked signals, detection accuracy, or bit rate. On the other hand, the proposed scheme in this dissertation has an advantage from the blind detection that it has less computational complexity.

Implementation

The proposed data-hiding scheme for AM radio signals is shown in Fig. 5.6. There are two main phases in this scheme: (a) data-embedding and a modulation process on the transmitter side and (b) a demodulation and data-detection process on the receiver side. On the transmitter side (Fig. 5.6(a)), the method of watermarking based on DPC is used to embed data, $s(k)$, into audio signal $x(n)$. The output of this step is the watermarked

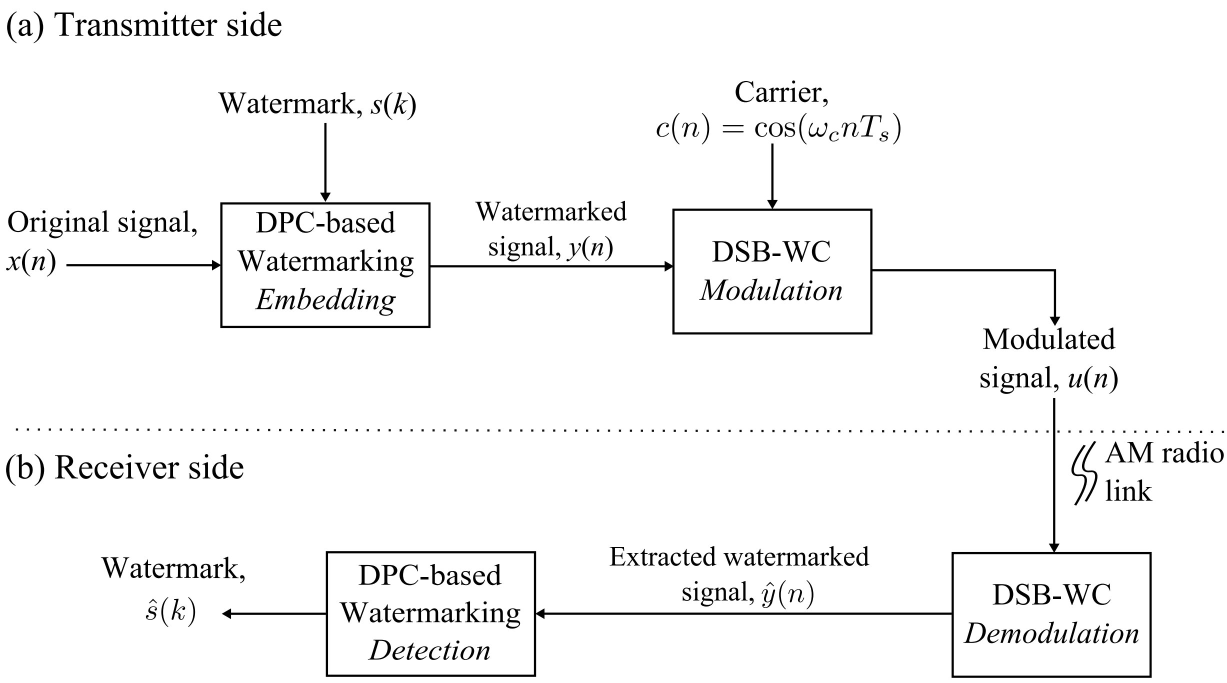


Figure 5.6: General scheme for digital audio data hiding in the AM domain.

signal, $y(n)$. The watermarked signal $y(n)$ is modulated with the carrier, $c(n)$ by the modulation process of the DSB-WC scheme. The modulated signal, $u(n)$, which conveys the watermarked signal $y(n)$ is used for broadcasting to receivers.

For the particular receivers that are suitably-equipped with the demodulator and the watermark detector as shown in Fig. 5.6(b), they can extract the conveyed signals from the modulated signals and the embedded data therein. The modulated signal, $u(n)$, is first demodulated to extract the conveyed signals $\hat{y}(n)$ by the demodulation process of the DSB-WC scheme. The extracted signal $\hat{y}(n)$ is then used to detect the embedded data $\hat{s}(k)$ by using the detection process of DPC-based watermarking. For conventional receivers that use an envelope detector to extract signals, they are able to extract the conveyed signals as the watermarked signal from the modulated signal.

5.3.4 Evaluation

Database and condition

We conducted computer simulations to evaluate the proposed data hiding scheme for AM radio signals. We evaluated the sound quality of the watermarked signals that were extracted from demodulation process. We investigated the sound quality of extracted watermarked signals and accuracy of watermark detection to confirm feasibility of applying a watermarking method based on DPC to AM radio signals. The proposed method is compared with [126] to study the advantages and disadvantages between the two methods.

In addition, we investigated the sound quality of demodulated signals that were output from of a standard demodulator to confirm the low-level compatibility of the proposed scheme with standard AM radio receivers.

We used all 102 tracks of the RWC music database [110] as the original signals. These music tracks had a sampling frequency of 44.1-kHz, were 16-bit quantized, and had two channels. The carrier frequency was 531 kHz. The sampling frequency was 5000 kHz. The same watermark "JAIST-AIS" was embedded into the original signal. The data rates were from 4 to 1024 bps.

We used objective evaluations: the SER, LSD [32], and PEAQ [33] to measure the sound quality of the target signals. SER was used to compare the level of a clean signal to the level of error, which is defined as the same as SNR. A higher SER signal indicated better sound quality. LSD was used to measure the distance or distortion between two spectra. A lower LSD value indicates a better result. PEAQ is used to measure quality degradation in audio according to the ODG which ranges from −4 to 0. ODG indicates the sound quality of target signals as shown in Table 2.1. Evaluation thresholds for the SER, LSD, and PEAQ corresponding to 20 decibel (dB), 1 dB, and −1 ODG, respectively, were chosen to evaluate the sound quality of the signals in these simulations.

The accuracy of watermark detection was measured by the bit detection rate, which is defined as the ratio between the number of correct bits and the total number of bits of the detected watermark. The evaluation threshold for the bit detection rate was 90%.

Effectiveness of the proposed scheme

The sound quality results for the signals extracted from the modulated signals are plotted in Fig. 5.7 as a function of bit rate. When the bit rate was 4 bps, the SER, LSD, and PEAQ of the extracted signals by [126] and the proposed scheme corresponded to approximately 53.36 dB, 0.15 dB, and −0.09 ODG.

The SER and LSD had significantly high values in practice. The PEAQ was about −0.09 ODG, which is imperceptible, i.e., no different in the signals could be perceived. These results indicate that the extracted signals were not distorted by the modulation process. When the bit-rate increased from 4 to 1024 bps, the SER, LSD, and PEAQ remained relatively unchanged. This confirms that the quality of watermarked signals conveyed in modulated signals is independent on the bit rate.

The sound quality the watermarked signals extracted from the modulated signals and the accuracy of watermark detection with the extracted signals are shown in Fig. 5.8. With [126], the bit-detection rate was greater than 95 % when the bit-rate increased from 4 to 256 bps. It decreased dramatically when the bit-rates were 512 and 1024 bps. In contrast, the bit detection rates with the proposed scheme are all greater than 99 %. While the PEAQs of the extracted watermarked signal were greater than −1 ODG and the LSDs were

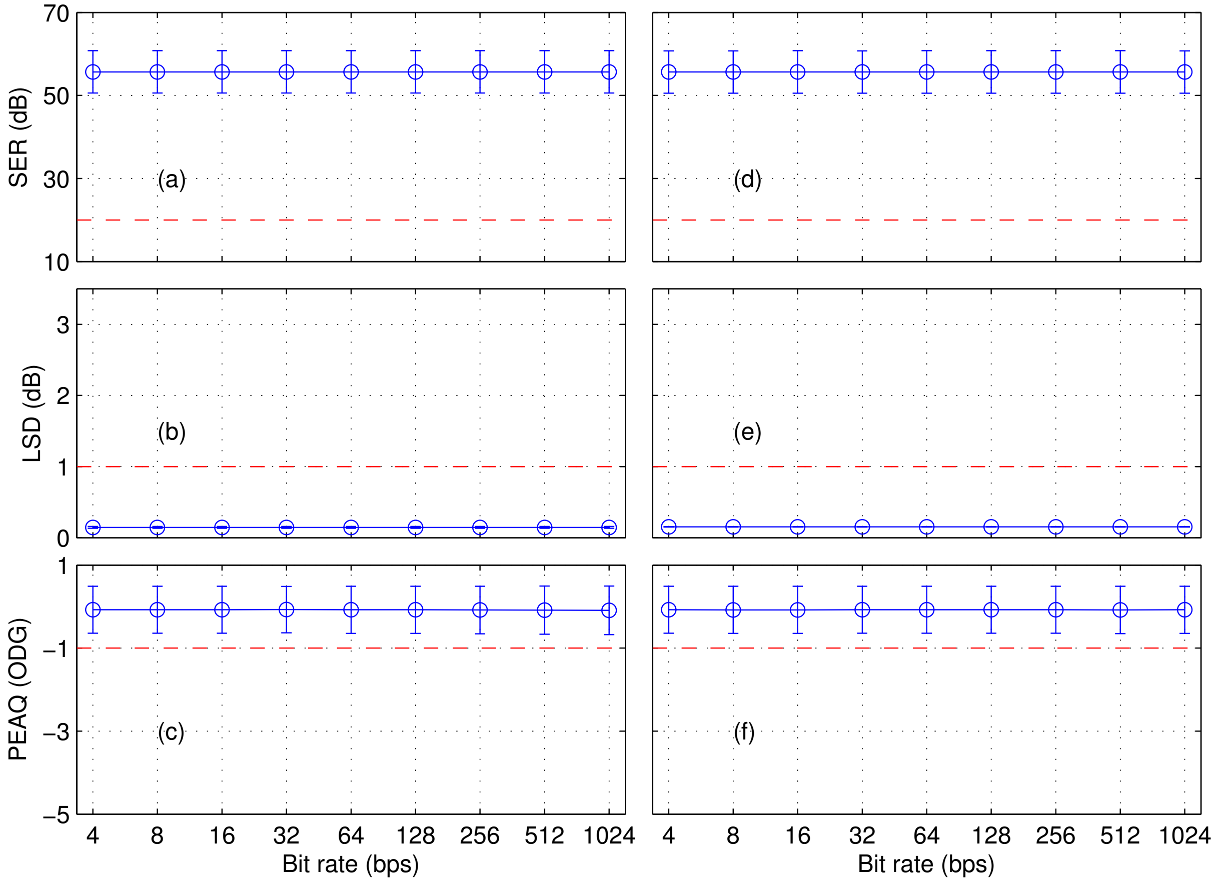


Figure 5.7: Results from objective evaluations of extracted signals as a function of bit rate: (a), (b), and (c) provide results for extracted watermarked signals by [126] and (d), (e), and (f) provide results for extracted watermarked signals by the proposed scheme.

less than 0.5 dB with [126], the proposed scheme has the lower PEAQ but still approximate to the evaluation criterion and the similar LSD.

These results demonstrate that the bit-detection rate and inaudibility of watermarked signals with the proposed scheme were the same as with the watermarking method. This indicates that the DPC-based method of watermarking could be applied to the AM domain. On the other hand, the proposed scheme has higher bit detection rate and bit rate but lower sound quality in comparison with [126].

Performance of hiding system under noise

The modulated signal was transmitted through the air and may have been affected by external white noise. We examined the distortion of extracted signals when the modulated signal was subjected to white noise. Figure 5.9 plots the results from the objective tests of the extracted watermarked signals. The horizontal axis shows the SER of the modulated signal which indicates the level of noise. With [126], the extracted signals were most distorted when the noise level was high (SER $<$ 30 dB). However, when the noise level

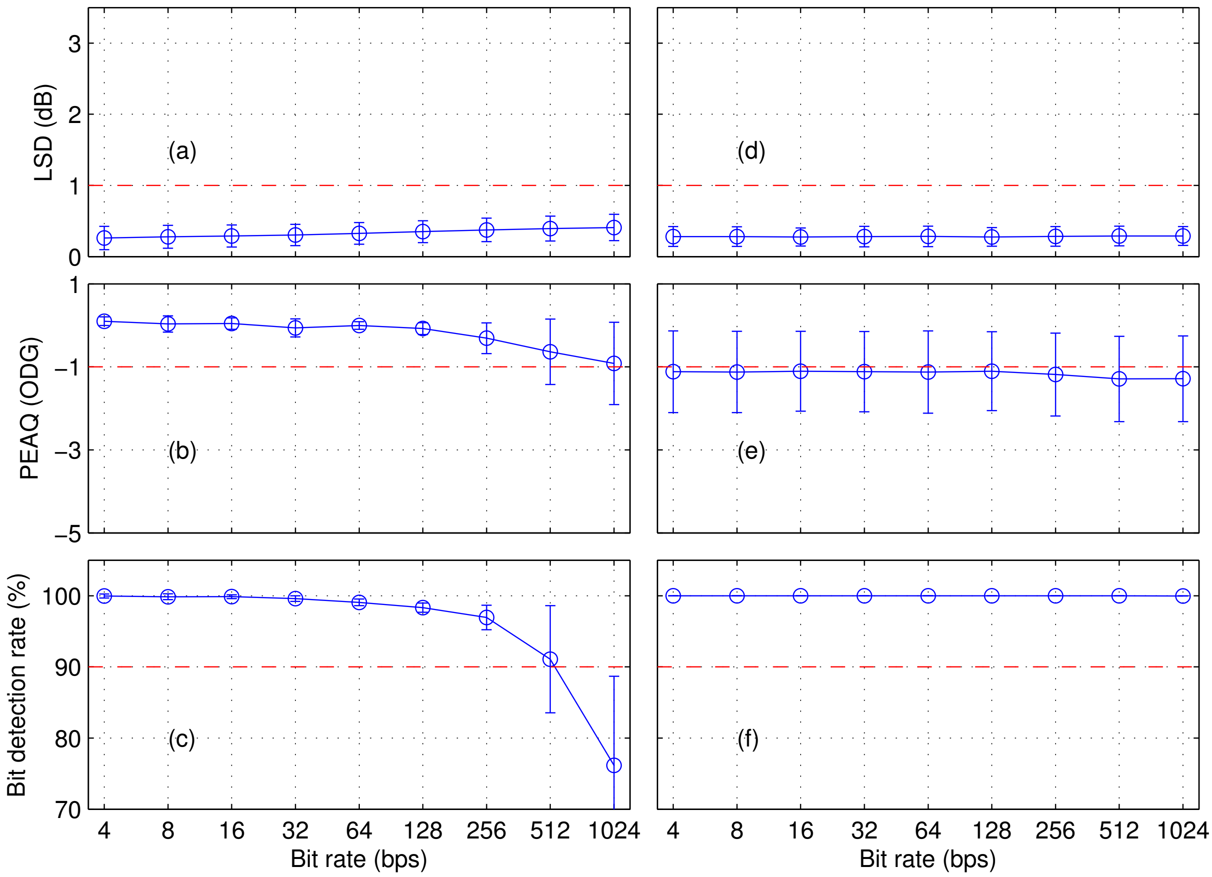


Figure 5.8: Results from objective evaluations of the sound quality of watermarked signals and bit detection rate as functions of the bit rate: (a), (b), and (c) provide results of [126] and (d), (e), and (f) provide results of the proposed scheme.

decreased (SER ≥ 30 dB), the SERs, LSDs, and PEAQs were significantly better (≥ 40 dB, ≤ 0.9 dB, and ≥ -1.8 ODG). The bit-detection rates for high-level noise were less than 98.2% and for low-level noise (SER ≥ 30 dB) they were greater than 99.2%.

With the proposed scheme, the extracted signals were most distorted when the noise level less than < 20 dB while when the noise level decreased (SER ≥ 20 dB), the SERs, LSDs, and PEAQs were significantly better. The bit-detection rates were all greater than 99%. These results indicated that the proposed scheme can correctly extract signals from a modulated signal that is affected by noise.

Low level compatibility

A vast majority of AM radio receivers extract audio signals from AM radio signals by using standard AM techniques (the envelope detector and the product detector). The data-hiding scheme should produce a modulated signal that can be demodulated by standard AM radio devices. Table 5.1 shows the sound quality of signals extracted by using an envelope demodulator in comparison with the method in [126]. The bit rate was set to

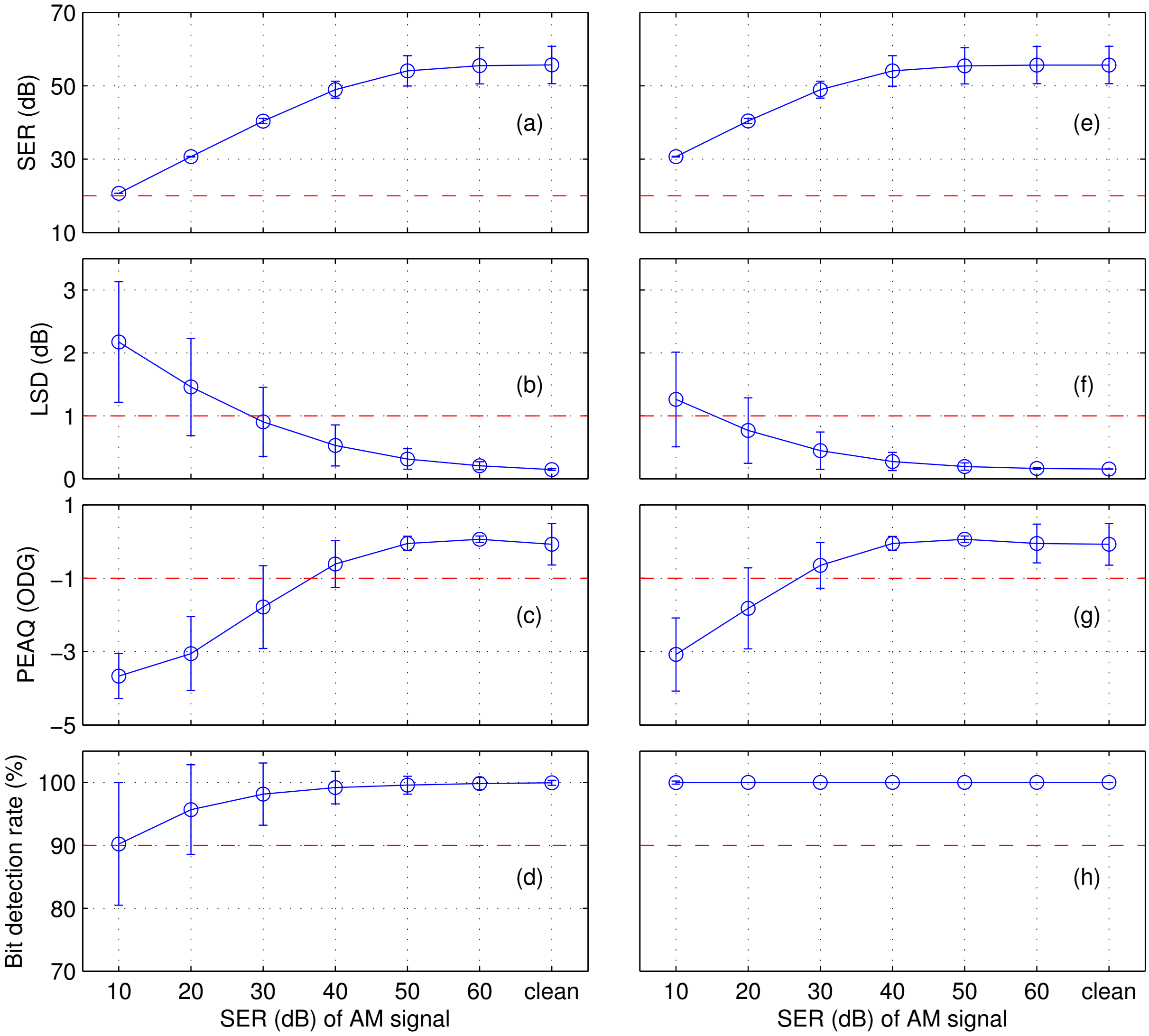


Figure 5.9: Results from objective evaluations of extracted signals against external noise: (a), (b), and (c) provide results for extracted watermarked signals by [126] and (d), (e), and (f) provide results for extracted watermarked signals by the proposed scheme.

4 bps in this experiment. The results show that the proposed method has the SNR almost the same with [126]. The proposed method has a better LSD but a worse PEAQ.

5.3.5 Summary

To enable a more efficient emergency alert system as well as high utility AM radio service, this dissertation proposed a data-hiding scheme for AM radio broadcasting systems. The proposed scheme can be used to transmit additional digital information along side audio content in AM radio signals. Playing an important role in this scheme, the digital audio watermarking method based on DPC was used to embed an inaudible message into audio content before the audio was further processed for long-distance transmission over a radio

Table 5.1: Sound quality of signals extracted using standard demodulators

	SER (dB)	LSD (dB)	PEAQ (ODG)
[126]	19.0	1.09	−0.55
Proposed	19.2	0.29	−1.11

link. Standard receivers are able to extract audio content from the AM signals.

We conducted computer simulations to evaluate the effectiveness of the proposed scheme. We first confirmed the effect of the modulation process by measuring the sound quality of audio signals that were extracted from the AM signals. The SER, LSD, and PEAQ results showed that these signals could be properly extracted. These results confirmed that we can precisely extract both the watermarked signals from the modulated signal. Second, we evaluated the inaudibility of watermarked signals and the accuracy of the watermark detection process. The LSD and PEAQ results confirmed that the DPC-method could be used with the proposed scheme to embed inaudible message. The bit detection rate results revealed that the watermark detection accuracy could be kept as high as that in the DPC-based watermarking system for digital-audio. Third, we evaluated the sound quality of the signal extracted from the AM signal using standard demodulators to check the compatibility of the proposed data-hiding scheme. We found that the standard receivers could acquire audio content at a reasonable level of distortion. Finally, we looked at the distortion of the extracted signals caused by external white noise. The scores showed that the sound quality of the extracted signals was degraded when the the external noise level was relatively high (SER < 20 dB).

The proposed method has higher detection accuracy and higher bit rate but lower sound quality in comparison with the method in [126] which employs a non-blind watermarking scheme. Compared to conventional techniques, such as AMSS, for embedding additional digital information into AM radio signals, the proposed method offers better embedding capacity and compatibility. The embedding capacity of the proposed scheme is about 1024 bps while that of AMSS is 47 bps. Conventional radio devices can demodulate modulated signals to extract audio signals in the proposed scheme, but will not respond to the modulated signals in AMSS.

The proposed scheme can be used to develop a hidden-message AM radio system. Such a system can be used as an emergency alert system in the event of natural disasters. Moreover, it can broadcast additional digital information as part of radio services such as programs, weather forecasting, news, advertisements, etc., thus providing high utility AM radio service.

Chapter 6

Conclusions

6.1 Summary

Audio data hiding has been born as not only a potential solution for securing digital audio value but also a promising technology for enriching users' experience on audio content. Due to the nature of audio signals, researchers in this field have had to confront numerous challenging issues related sensitivity of the HAS and variety of harmful attacks in order to create a new method. Traditional methods have attempted to incorporate a mathematic model into embedding algorithm to minimize the distortion and maintain the robustness but this approach has faced with the trade-off between these two properties. This dissertation aims to explore relation between characteristics of sound and mechanism of the HAS to tackle the major challenges. The phase characteristics and the variability of audio signals have been exploited to propose an efficient method of audio data hiding with reasonable trade-off in the properties.

The fact that the HAS is incredibly sensitive with small change in the sound and audio signals usually suffer a lot of signal-processing operations has caused audio data hiding difficulty in obtaining reasonable trade-off between inaudibility and robustness. Phase has been intensively explored for embedding watermark thanks to its advantage in resistance to common processing. Although the HAS appears sensitive to relative phase of audio signals, the modification of phase only induces perceptible distortion when the phase relation between frequency components is dramatically changed. Inaudible phase coding can be achieved by controlling the modification of phase to be sufficiently small and slowly varying. This inspired us to investigate a dynamic phase-manipulation scheme for audio data hiding with the two objectives: obtaining reasonable trade-off in properties of audio data hiding and wide applicability to practical problems.

The general idea is adaptively controlling the amount of phase modification so that the watermark does not result in inaudible distortion. Since the phase modification of a frequency component causes distortion in sound in a manner that is directly proportional to the magnitude of that component, the amount of phase modification is thus adjusted according to the magnitude. Furthermore, most audio signals in realistic environments are variant with time, so the modification is adaptively adjusted to local characteristics to attain stable performance. More precisely, depending on each frame, frequency components with large magnitude will have smaller amount of phase modification whereas smaller-magnitude frequency components will have somewhat higher amount of phase modification. This mechanism ensures the watermark to be hardly perceptible by the HAS and robust against attacks simultaneously. As a result, the proposed method could obtain better trade-off between inaudibility and robustness. The idea has been realized by two approaches, giving two novel methods: APM and DPC.

The method based APM has been implemented using the second-order IIR APF and based on the following observations: (i) each frame has a different energy distribution in frequency domain and (ii) the APF phase response which quantifies for the amount of phase modification has a peak at the pole-zero frequency. The frequency components around the pole-zero frequency are more affected than the outside components. In other words, if the pole-zero frequency is set in higher-energy frequency-region, the phase modulation causes higher distortion in the frame. In contrast, if the pole-zero frequency lies on lower-energy frequency-region, the frame suffers less distortion. Therefore, the pole-zero frequency of the APF is adaptively set to the low-energy frequency-region to ensure the best quality for the watermarked signal. The inaudibility can be traded for the robustness by setting the pole-zero frequency to higher-energy frequency region. The simulation have been carried out with various kinds of audio signals to evaluate the proposed method with regard to inaudibility and robustness. The results reveal that the watermark causes distortion between not annoying and imperceptible according to ODG scores. In the case without attack, the proposed method achieves a bit rate of 100 bps with 99 % accuracy. In the case with attack, the embedding bit rate is 6 bps with 94.3 % accuracy. Although the results are quite promising and satisfy several requirements, this approach still has limit in trade-off between robustness and high embedding capacity. Thus, the proposed APM scheme has successfully verified the effectiveness of the adaptive phase-manipulation rather than has wide applicability.

The method based on DPC has been implemented using the QIM technique and based on the following principle: (i) most of the audio information is concentrated on moderately low frequencies and (ii) the amount of phase modification causes the distortion directly proportional to the magnitude. Accordingly, to assure the inaudibility and the robustness, the moderately low frequency-region is chosen for embedding and for each embedding component the QIM step that quantifies for the amount of phase modification is determined based on the magnitude. Moreover, to increase the reliability of watermark detection, the technique of ECC is incorporated to encode the watermark before embedding. The

Table 6.1: Comparison between the two methods with regard to the four properties.

	APM	DPC
Inaudibility (PEAQ)	−0.69	−0.97
Blind detection	✓	✓
Robustness	✓	✓
Capacity (bps)	6	102

experimental results demonstrate that the watermark cause not annoying distortion in the watermarked signals. The watermark is robust against various attacks, especially MP3 compression 64 kbps, while achieving a bit rate of 200 bps with 98.5 % accuracy. The incorporation of ECC is remarkably effective in correcting all the errors, resulting in an embedding bit rate of 102 bps with 100 % accuracy. The results verify that the proposed DPC scheme is capable of achieving an excellent trade-off between inaudibility and robustness and that the proposed method satisfies all the requirements: inaudibility, robustness, blind-detection ability, high embedding capacity, and reliability.

The proposed DPC method was applied to the applications of copy prevention of digital-audio, annotated digital-audio, and information carrier over AM radio broadcast. The application of copy prevention was proposed to limit widespread sharing of copyrighted digital audio over public file sharing services. The digital fingerprint of the copyrighted audio is embedded as watermark into the audio signal. Whenever the watermarked signal is uploaded to a file sharing service, the embedded watermark can help verify if the uploading is illegal. Once the misconduct is determined, the service stops the uploading immediately. The application of annotated audio can help enhance users' experience on listening to music or radio broadcast. The watermark is used as an information channel that carries complementary information in order to enrich the playback or the radio program. The application of information carrier over AM radio broadcast was proposed to construct a high utility AM radio service. This application supplements the sound (speech) information conveyed by radio waves with additional digital information which is likely to be very useful in emergency situations where the ordinary Internet is not available, as well as improve the convenience of radio services.

Table 6.1 shows a comparison on properties of the two proposed method. The APM method has an advantage of high sound quality but a limitation in capacity while the DPC method has high capacity but lower sound quality. The proposed method can be widely applied to practical applications, demonstrating that the two objectives in this dissertation have been successfully achieved.

6.2 Contributions

Being inspired from the interesting characteristics of phase and the mechanism on phase perception of the HAS, this study investigated the feasibility of the dynamic phase-manipulation scheme for audio data hiding. The overall contribution of this dissertation is an efficient and widely applicable method of audio data hiding for solving social concerns on security of digital audio content. This is the combination of the novel sophisticated methodology for research on audio data hiding, the exploration of suitable acoustic features for watermarking, and the exploitation of psychoacoustic knowledge. Throughout the study, several findings have been obtained and the major contributions can be summarized as follows.

(1) The first contribution is a proposal of sophisticated methodology for research on audio data hiding. As audio data hiding is a challenging task, most methods in the literature are stuck with the trade-off in the requirements and have limitation in applicability. The proposed methodology describes the process of creating a new method of audio data hiding systematically. In order to realize a method of audio data hiding with efficiency and applicability, the methodology proposes four criteria for the process, i.e., acoustic features, embedding technique, parameters, and adjustable algorithm. These criteria ensure that a new method could obtain reasonable trade-off and have necessary properties for satisfying practical applications.

(2) By applying the proposed methodology, two contributions on the field of audio data hiding were derived. The second contribution of this study is the realization of the method of audio data hiding based on APM scheme. This method introduces the new concept of APM to deal with the discontinuity issue in the approach of audio data hiding based on phase modulation. The proposed scheme not only minimizes the distortion caused by phase-modification but also eliminates the discontinuity at frame boundaries to achieve the high sound quality. Although the proposed method has limitation in embedding capacity, it outperforms the other methods in the same approach.

(3) The third contribution in this dissertation is the implementation of the method of audio data hiding based on DPC. The proposed method exploits the relation between the magnitude and the distortion caused by phase modification of a frequency component and adopts the technique of QIM to give out a new concept of dynamic quantization of phase for audio data hiding. The proposed method controls the amount of phase modification sufficiently and adaptively according to the magnitude to achieve the high sound quality while keeping the watermark resistant to the attacks. The effectiveness has been verified with the good inaudibility, the high robustness, and the competitive reliability, i.e., 100 % accuracy, at 102 bps. The obtained results show that the proposed method is efficiently practical, becomes the best in the same approach, and is comparable with the state-of-

the-art in the literature.

(4) The last contribution of this study is the two applications of the proposed method of audio data hiding based on DPC to digital audio protection and audio entertainment and an application of a general audio data hiding method to AM radio broadcasting systems. It verifies again the effectiveness and the applicability of the proposed method and demonstrates that audio data hiding is a promising technology for solving with social concerns.

6.3 Future Research

In this dissertation, we have proposed an efficient and applicable method of audio data hiding based on dynamic phase-manipulation. The proposed method utilizes the knowledge from psychoacoustic field and the sophisticated methodology to achieve better trade-off in the properties of audio data hiding. It has been demonstrated that the proposed method is effective and capable of solving practical problems. However, beyond the scope of this study is a number of avenues for future research.

(1) For the proposed method based on APM, the phase shift of the phase response can be more flexibly controlled by using higher order of IIR APF to build a better adaptive phase modulator. It offers a way to adjust the amount of phase modification by not only setting the pole-zero frequency to the frequency region of interest but also changing the phase-shift level of the APF's phase response. More flexibly adaptive phase-modulator could better balance the inaudibility and the robustness of embedded watermark. In addition, the sound quality of watermarked signals could be enhanced after extracting the embedded watermark by reversibility. An inverse filter of the APF can be designed to restore the watermarked signal to the original by performing inverse filtering [112].

(2) For the proposed method based on DPC, better sound quality could be obtained by keeping phase relation between unresolved frequency components. As Moore et al. [30] pointed out that the HAS is very sensitive to the relative phase of unresolved components, the timbre of the watermarked signal could be maintained as the original by preserving the phase relation. Besides, the current adaptation scheme that adjusts the amount of phase modification according to the magnitude in a linear scale could be replaced by a new scheme that adapts to a logarithmic scale of the magnitude as the the HAS's dynamic range of pressure sensitivity.

(3) As the proposed method has many required properties, it could be applied to other applications such as usage control, fingerprinting, and covert communication. These applications would help provide more solutions for existing social problems.

Appendix A

Experimental Details of Attacks

In the experiments on the robustness of the proposed methods, the following configurations for the attacks were used.

AWGN. White Gaussian noise is added into watermarked signals with an SNR of 36 dB.

Re-sampling. Watermarked signals are downsampled to 22.05 kHz and 16 kHz and then upsampled to 44.1 kHz.

MP3. Watermarked signals are changed to MP3 formats (128 kbps or 64 kbps) and then changed back to waveform format.

MP4. Watermarked signals are changed to MP4 formats (96 kbps) and then changed back to waveform format.

Re-quantization. Watermarked signals are requantized by 8 bits.

Bandpass filtering. Watermarked signals are filtered by a bandpass filter with passband [0.1, 6] kHz and stopband attenuation −12 dB/octave.

Each attack has a different effect on sound quality of the target signal. Figures A.1–A.3 respectively show the SNR, the LSD, and the PEAQ of the attacked signals. All the tracks of the RWC music databased were used in this experiment.

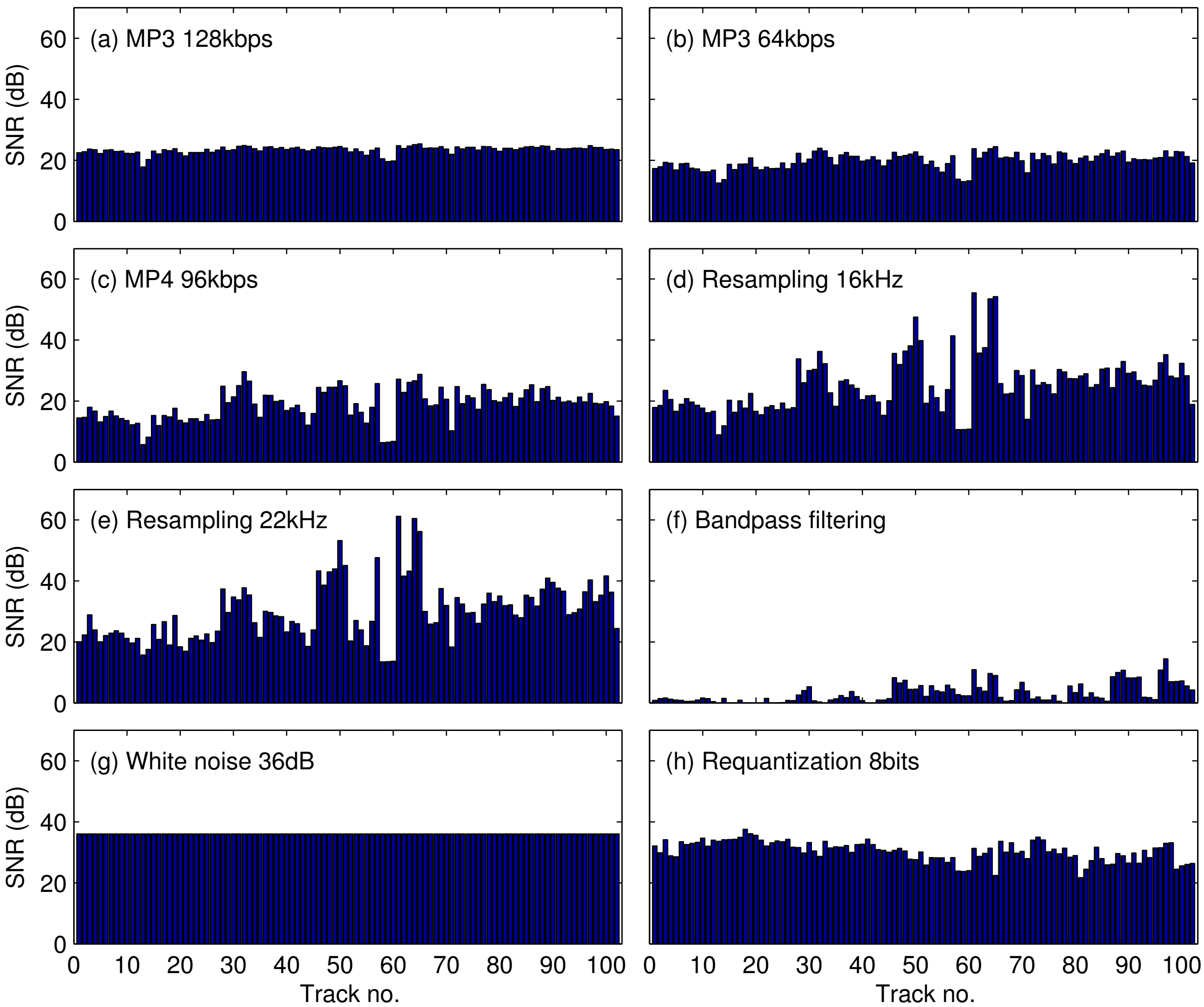


Figure A.1: SNR of the signals under different types of attacks

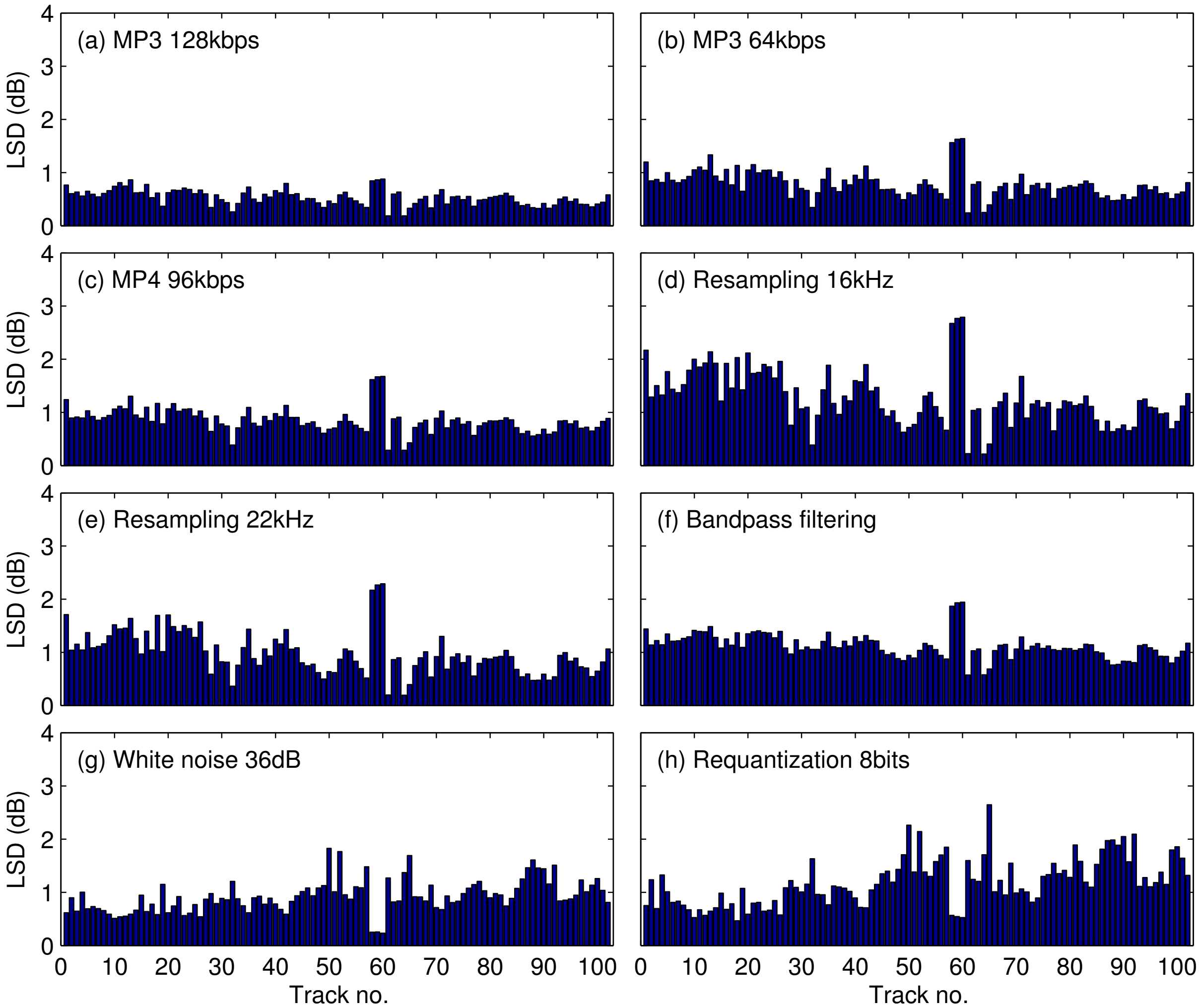


Figure A.2: LSD of the signals under different types of attacks

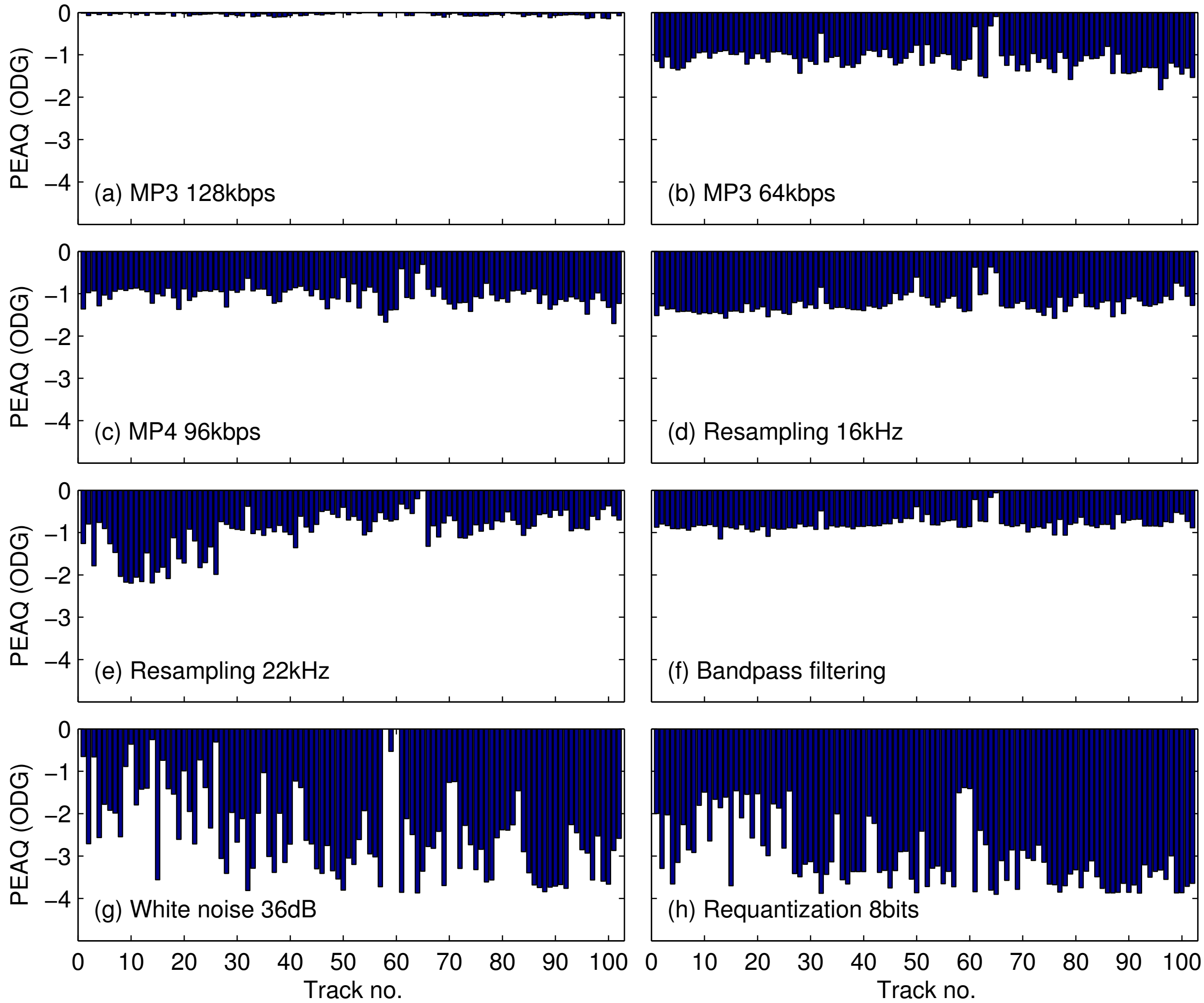


Figure A.3: PEAQ of the signals under different types of attacks

Bibliography

[1] B. Kahle, "Internet Archive: Universal Access to all Knowledge." https://archive.org/details/SDForumBK, 2004. [Online; accessed 05-May-2015].

[2] Wikipedia, "Digital revolution — Wikipedia, The Free Encyclopedia." http://en.wikipedia.org/wiki/Digital_Revolution, 2015. [Online; accessed 05-May-2015].

[3] "The digital revolution ahead for the audio industry." Business Week, New York, March 1981, p. 40D.

[4] H. R. Varian, "Copying and copyright," *Journal of Economic Perspectives*, vol. 19, no. 2, pp. 121–138, 2005.

[5] P. Samuelson, "Good news and bad news on the intellectual property front," *Communications of The ACM*, vol. 42, no. 3, pp. 19–24, 1999.

[6] E. Pfanner, "Music industry counts the cost of piracy." http://nyti.ms/1Dn1UMe, 2010. The New York Times, [Online; access 06-May-2015].

[7] A. Zentner, "File Sharing and International Sales of Copyrighted Music: An Empirical Analysis with a Panel of Countries," *Berkeley Electronics, Topics in Economic Analysis Policy*, vol. 5, no. 1, 2005.

[8] B. Schneier, "The futility of digital copy prevention." https://www.schneier.com/crypto-gram/archives/2001/0515.html#3, 2001. Crypto-Gram Newsletter, [Online; accessed 06-May-2015].

[9] N. Cvejic and T. Seppänen, *Digital Audio Watermarking Techniques and Technologies*. IGI Publishing Hershey, USA in press, 2007.

[10] C. I. Podilchuk and E. J. Delp, "Digital watermarking: Algorithms and applications," *IEEE Signal Processing Magazine*, vol. 18, no. 4, pp. 33–46, 2001.

[11] A. Nikolaidis, S. Tsekeridou, A. Tefas, and V. Solachidis, "A survey on watermarking application scenarios and related attacks," in *Proc. of International Conference on Image Processing*, pp. 991–994, 2001.

[12] W. Fenlon, "How digital fingerprinting works." http://computer.howstuffworks.com/digital-fingerprinting.htm, 2011. HowStuffWorks.com, [Online; 11-May-2015].

[13] J. Fridrich, M. Goljan, and A. C. Baldoza, "New fragile authentication watermark for images," in *Proc. of Internation Conference on Image Processing*, 2000.

[14] D. Kirovski and H. S. Malvar, "Robust covert communication over a public audio channel using spread spectrum," in *Proc. of International Workshop on Information Hiding, LNCS 2137*, pp. 354–368, 2001.

[15] D. W. Alliance, "Digital watermarking plays exciting role in enhancing entertainment experiences for consumers through television, magazines and mobile devices." http://www.digitalwatermarkingalliance.org/pr_10JUL2012.asp, 2012. [Online; access 07-May-2015].

[16] D. Gruhl, A. Lu, and W. Bender, "Echo hiding," in *Proc. of Information Hiding, LNCS 1174*, pp. 295–315, 1996.

[17] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1673–1687, 1997.

[18] M. Arnold, "Audio watermarking: features, applications and algorithms," in *Proc. of International Conference on Multimedia and Expo*, vol. 2, pp. 1013–1016, 2000.

[19] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1423–1443, 2001.

[20] A. Takahashi, R. Nishimura, and Y. Suzuki, "Multiple watermarks for stereo audio signals using phase-modulation techniques," *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 806–815, 2005.

[21] K. Sonoda and O. Takizawa, "Audio watermarking using qim on wavelet packet coefficients," in *Proc. of International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 72–75, 2009.

[22] M. Narimannejad and S. M. Ahadi, "Watermarking of speech signal through phase quantization of sinusoidal model," in *Proc. of Iranian Conf. on Elec. Engineering*, pp. 1–4, 2011.

[23] B. Y. Lei, I. Y. Soon, and Z. Li, "Blind and robust audio watermarking scheme based on svddct," *Elsevier Journal of Signal Processing*, vol. 91, no. 8, pp. 1973–1984, 2011.

[24] I. Natgunanathan, Y. Xiang, Y. Rong, W. Zhou, and S. Guo, "Robust Patchwork-Based Embedding and Decoding Scheme for Digital Audio Watermarking," *IEEE Transactions on Audio, Speech, and Language Processing*, no. 8, pp. 2232–2239, 2012.

[25] I. Cox and J.-P. Linnartz, "Some general methods for tampering with watermarks," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 4, pp. 587–593, 1998.

[26] J. D. Gordy, "Performance evaluation of digital watermarking algorithms." Master Thesis, The University of Calgary, 2000.

[27] S. Craver, N. Memon, B.-L. Yeo, and M. Yeung, "Resolving rightful ownerships with invisible watermarking techniques: limitations, attacks, and implications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 4, pp. 573–586, 1998.

[28] S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*. California Tech. Pub, 1997.

[29] K. Ozawa, Y. Suzuki, and T. Sone, "Monaural phase effect on timbre of two signals," *Journal of Acoustic Society American*, vol. 93, no. 2, pp. 1007–1011, 1993.

[30] B. C. J. Moore, *An Introduction to the Psychology of Hearing, 6th ed*. Brill Academic Pub, 2013.

[31] X. Dong, M. F. Bocko, and Z. Ignjatovic, "Data hiding via phase manipulation of audio signals," in *Proc. of International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 377–380, 2004.

[32] A. J. Gray and J. Markel, "Distance measures for speech processing," *Proc. of International Conference on Acoustics Speech and Signal Processing*, vol. 24, no. 5, pp. 380–391, 1976.

[33] P. Kabal, "An examination and interpretation of ITU-RBS.1387: Perceptual evaluation of audio quality," tech. rep., Department of Electrical Computer Engineering, 2002.

[34] P. Y. Liew and M. A. Armand, "Inaudible watermarking via phase manipulation of random frequencies," *Journal of Multimedia Tools and Applications*, vol. 35, no. 3, pp. 357–377, 2007.

[35] A. Nishimura, M. Unoki, K. Kondo, and A. Ogihara, "Objective evaluation of sound quality for attacks on robust audio watermarking," in *Proc. of Meeting on Acoustics*, pp. 1–9, 2013.

[36] L. de C. T. Gomes, P. Cano, E. Gmez, M. Bonnet, and E. Batlle, "Audio watermarking and fingerprinting: For which applications," *Journal of New Music Research*, vol. 32, no. 1, pp. 65–81, 2003.

[37] M. Unoki and R. Miyauchi, "Detection of Tampering in Speech Signals with Inaudible Watermarking Technique," in *Proc. of International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 118–121, 2012.

[38] C. P. Wu and C. C. J. Kuo, "Fragile speech watermarking based on exponential scale quantization for tamper detection," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 3305–3308, 2002.

[39] X. Huang, I. Echizen, and A. Nishimura, "A new approach of reversible acoustic steganography for tampering detection," in *Proc. of International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 538–542, 2010.

[40] R. Tachibana, "Sonic Watermarking," *EURASIP Journal on Advances in Signal Processing*, vol. 13, pp. 1955–1964, 2004.

[41] Y. Nakashima and R. Tachibana and N. Babaguchi, "Watermarked Movie Soundtrack Finds the Position of the Camcorder in a Theater," *IEEE Transactions on Multimedia*, vol. 11, no. 3, pp. 443–454, 2009.

[42] X.-W. Li, B.-L. Guo, X.-X. Wu, and L.-D. Li, "On collusion attack for digital fingerprinting," *Journal of Information Hiding and Multimedia Signal Processing, Ubiquitous International*, vol. 2, no. 4, pp. 366–376, 2011.

[43] D. W. Alliance, "Broadcast monitoring." `http://www.digitalwatermarkingalliance.org/app_broadcast.asp`, 2015. [Online; access 12-May-2015].

[44] S. Wu, J. Huang, D. Huang, and Y. Q. Shi, "Efficiently self-synchronized audio watermarking for assured audio data transmission," *IEEE Transactions on Broadcasting*, vol. 51, no. 1, pp. 69–76, 2005.

[45] C. B. N. Moreau and P. Dymarski, "Controlling the inaudibility and maximizing the robustness in an audio annotation watermarking system," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 5, pp. 1172–1182, 2006.

[46] N. Lazic and P. Aarabi, "Communication over an acoustic channel using data hiding techniques," *IEEE Transactions on Multimedia*, vol. 8, no. 5, pp. 918–924, 2006.

[47] H. S. Yun, K. Cho, and N. S. Kim, "Acoustic data transmission based on modulated complex lapped transform," *IEEE Signal Processing Letters*, vol. 17, no. 1, pp. 67–70, 2010.

[48] M. Lytynoja, N. Cvejic, A. Keskinarkaus, E. Lhetkangas, and T. Seppnen, "Mobile commerce from watermarked broadcast audio," in *Proc. of International Conference on Consumer Electronics, Digest of Technical Papers*, pp. 5–6, 2006.

[49] M. Arnold, X.-M. Chen, P. Baum, U. Gries, and G. Doërr, "A phase-based audio watermarking system robust to acoustic path propagation," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 3, pp. 411–425, 2014.

[50] K. Hofbauer and H. Hering, "Noise robust speech watermarking with bit synchronisation for the aeronautical radio," in *Proc. of International Workshop on Information Hiding, LNC S4567*, pp. 252–266, 2007.

[51] D. V. Roosbroek, "EATMP Communications Strategy - Volume 2 Technical Description (6th ed)," 2006. Eurocontrol.

[52] E. Tavakoli, I. T. B. V. V. M. B. Shamsollahi, and R. Sameni, "Audio watermarking for covert communication through telephone system," in *Proc. of International Symposium on Signal Processing and Information Technology*, pp. 955–959, 2006.

[53] P. Karthigaikumar, K. Baskaran, and K. J. Kirubavathy, "Hardware implementation of audio watermarking—covert communication," in *Proceedings of IEEE International Conference on Computer and Automation Engineering*, pp. 588–592, 2010.

[54] P. Karthigaikumar, K. J. Kirubavathy, and K. Baskaran, "FPGA based audio watermarking—Covert communication," *Elsevier Journal of Microelectronics*, vol. 42, no. 5, pp. 778–784, 2011.

[55] N. Cvejic and T. Seppänen, *Digital audio watermarking techniques and technologies.* Idea Group Inc., 2007.

[56] N. Cvejic and T. Seppänen, "Increasing Robustness of LSB Audio Steganography by Reduced Distortion LSB Coding," *Journal of Universal Computer Science*, vol. 11, no. 1, pp. 56–65, 2005.

[57] W. Cao and S. L. Y. Yan, "Bit Replacement Audio Watermarking Using Stereo Signals," in *Proc. of International Conference on New Trends in Information and Service Science*, pp. 603–606, 2009.

[58] F. S. Wei and D. Qi, "Robustness of One Bit Per Sample audio watermarking," in *Proc. of International Symposium on Circuits and Systems*, pp. 932–935, 2009.

[59] L. Boney, A. H. Tewfik, and K. N. Hamdy, "Digital watermarks for audio signals," in *International Conference on Multimedia Computing and Systems*, pp. 473–480, 1996.

[60] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *Journal of IBM System*, vol. 35, no. 3–4, pp. 313–336, 1996.

[61] H. S. Malvar and D. A. F. Florêncio, "Improved Spread Spectrum: A New Modulation Technique for Robust Watermarking," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 898–905, 2003.

[62] Z. Ying, X. Zhengguang, and H. Benxiong, "Channel Capacity Analysis of the Generalized Spread Spectrum Watermarking in Audio Signals," *IEEE Signal Processing Letters*, vol. 22, no. 5, pp. 519–523, 2015.

[63] D. Kirovski and H. Malvar, "Robust spread-spectrum audio watermarking," in *Proc. of International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1345–1348, 2001.

[64] D. Kirovski and H. S. Malvar, "Spread-Spectrum Watermarking of Audio Signals," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 1020–1033, 2003.

[65] H. Xing and M. S. Scordilis, "Improved spread spectrum digital audio watermarking based on a modified perceptual entropy psychoacoustic model," in *Proc. of SoutheastCon*, pp. 283–286, 2005.

[66] H. Malik, A. Khokhar, and A. Rashid, "Robust audio watermarking using frequency selective spread spectrum theory," in *Proc. of International Conference on Acoustics, Speech, and Signal Processing*, pp. V–385–358, 2004.

[67] H. Malik, R. Ansari, and A. Khokhar, "Robust audio watermarking using frequency-selective spread spectrum," *IET Information Security*, vol. 2, no. 4, pp. 129–150, 2008.

[68] H. Oh, J. W. Seok, J. W. Hong, and D. H. Youn, "New echo embedding technique for robust and imperceptible audio watermarking," in *Proc. of International Conference on Acoustic, Speech, and Signal Processing*, pp. 1341–1344, 2001.

[69] H.-O. Oh, H. W. Kim, J. W. Seok, J. W. Hong, and D. H. Youn, "Transparent and robust audio watermarking with a new echo embedding technique," in *Proc. of International Conference on Multimedia and Expo*, pp. 317–320, 2001.

[70] H. J. Kim and Y. H. Choi, "A novel echo-hiding scheme with backward and forward kernels," *IEEE Transactions on Circuits and Systems Video Technology*, vol. 13, no. 8, pp. 885–889, 2003.

[71] B.-S. Ko, R. Nishimura, and Y. Suzuki, "Time-spread echo method for digital audio watermarking," *IEEE Transactions on Multimedia*, vol. 7, no. 2, pp. 212–221, 2005.

[72] O. T.-C. Chen and W.-C. Wu, "Highly Robust, Secure, and Perceptual-Quality Echo Hiding Scheme," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 3, pp. 629–638, 2008.

[73] Y. Xiang, D. Peng, I. Natgunanathan, and W. Zhou, "Effective pseudonoise sequence and decoding function for imperceptibility and robustness enhancement in time-spread echo-based audio watermarking," *IEEE Transactions on Multimedia*, vol. 13, no. 1, pp. 2–13, 2011.

[74] Y. Xiang, I. Natgunanathan, D. Peng, W. Zhou, and S. Yu, "A dual-channel time-spread echo method for audio watermarking," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 383–392, 2012.

[75] G. Hua, J. Goh, and V. L. L. Thing, "Time-Spread Echo-Based Audio Watermarking With Optimized Imperceptibility and Robustness," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 2, pp. 227–239, 2015.

[76] S.-T. Chen, G.-D. Wu, and H.-N. Huang, "Wavelet-domain audio watermarking scheme using optimisation-based quantisation," *IET Signal Processing,*, vol. 4, no. 6, pp. 720–727, 2010.

[77] J. Huang, Y. Wang, and Y. Shi, "A blind audio watermarking algorithm with self--synchronization," in *Proc. of International Symposium on Circuits and Systems*, pp. 627–630, 2002.

[78] X.-Y. Wang and H. Zhao, "A Novel Synchronization Invariant Audio Watermarking Scheme Based on DWT and DCT," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4835–4840, 2006.

[79] X. Wang, W. Qi, and P. Niu, "A New Adaptive Digital Audio Watermarking Based on Support Vector Regression," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 8, pp. 2270–2277, 2007.

[80] H.-T. Hu, L.-Y. Hsu, and H.-H. Chou, "Perceptual-based DWPT-DCT framework for selective blind audio watermarking," *Elsevier Journal of Signal Processing*, vol. 105, pp. 316–327, 2014.

[81] H.-T. Hu and L.-Y. Hsu, "Robust, transparent and high-capacity audio watermarking in DCT domain," *Elsevier Journal of Signal Processing*, vol. 109, pp. 226–235, 2015.

[82] H. L. F. Helmholtz, *On the Sensations of Tone, 2nd ed.* Dover Publications, 1954.

[83] D. A. Nelson and R. C. Bilger, "Pure-tone octave masking in normal-hearing listeners," *J. of Speech and Hearing Research*, vol. 17, no. 2, 1974.

[84] M. Arnold, P. G. Baum, and W. Voebing, "A phase modulation audio watermarking technique," in *Proc. of Information Hiding Workshop, LNCS 5806*, pp. 102–116, 2009.

[85] X.-M. Chen, G. Doerr, M. Arnold, and P. G. Baum, "Efficient coherent phase quantization for audio watermarking," in *Proc. of International Conference Acoustics Speech and Signal Processing*, pp. 1844–1847, 2011.

[86] M. Unoki and D. Hamada, "Method of digital-audio watermarking based on cochlear delay characteristics," *International Journal of Innovative Computing, Information and Control*, vol. 6, no. 3(B), pp. 1325–1346, 2010.

[87] R. Nishimura and Y. Suzuki, "Audio watermark based on periodical phase shift," *Journal of The Acoustic Society Japan*, vol. 60, no. 5, pp. 269–272, 2004.

[88] A. Takahashi, R. Nishimura, and Y. Suzuki, "Multiple watermarks for stereo audio signals using phase-modulation techniques," *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 806–815, 2005.

[89] N. Khademi, M. A. Akhaee, S. M. Ahadi, M. Moradi, and A. Kashi, "Audio watermarking based on quantization index modulation in the frequency domain," in *Proc. of International Conference on Signal Processing and Communication*, pp. 1127–1130, 2007.

[90] L. Gang, A. N. Akansu, and M. Ramkumar, "MP3 resistant oblivious steganography," in *Proc. of International Conference on Acoustic, Speech, and Signal Processing*, pp. 1365–1368, 2001.

[91] D. Megas, J. Serra-Ruiz, and M. Fallahpour, "Efficient self-synchronised blind audio watermarking system based on time domain and FFT amplitude modification," *Elsevier Journal of Signal Processing*, vol. 90, no. 12, pp. 3078–3092, 2010.

[92] S.-K. Lee and Y.-S. Ho, "Digital audio watermarking in the cepstrum domain," *IEEE Transactions on Consumer Electronics*, vol. 46, no. 3, pp. 744–750, 2000.

[93] H.-T. Hu and W.-H. Chen, "A dual cepstrum-based watermarking scheme with self--synchronization," *Elsevier Journal of Signal Processing*, vol. 92, no. 4, pp. 1109–1116, 2012.

[94] B. Lei, I. Y. Soon, F. Zhou, Z. Li, and H. Lei, "A robust audio watermarking scheme based on lifting wavelet transform and singular value decomposition," *Signal Processing*, vol. 92, no. 9, pp. 1985–2001, 2012.

[95] B. Lei, I. Y. Soon, and E.-L. Tan, "Robust svd-based audio watermarking scheme with differential evolution optimization," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 11, pp. 2368–2378, 2013.

[96] V. B. K, I. Sengupta, and A. Das, "An adaptive audio watermarking based on the singular value decomposition in the wavelet domain," *Elsevier Journal of Digital Signal Processing*, vol. 20, no. 6, pp. 1547–1558, 2010.

[97] M. D. Swanson, B. Zhu, A. H. Tewfik, and L. Boney, "Robust audio watermarking using perceptual masking," *Elsevier Journal of Signal Processing*, vol. 66, no. 3, pp. 337–355, 1998.

[98] P. Bassia, I. Pitas, and N. Nikolaidis, "Robust Audio Watermarking in the Time Domain," *IEEE Transactions on on Multimedia*, vol. 3, no. 2, pp. 232–241, 2001.

[99] W.-N. Lie and L.-C. Chang, "Robust and High-Quality Time-Domain Audio Watermarking Based on Low-Frequency Amplitude Modification," *IEEE Transactions on on Multimedia*, vol. 8, no. 1, pp. 46–59, 2006.

[100] S. Xiang and J. Huang, "Histogram-Based Audio Watermarking Against Time-Scale Modification and Cropping Attacks," *IEEE Transactions on Multimedia*, vol. 9, no. 7, pp. 1357–1372, 2007.

[101] C. Wang, X. Ma, X. Cong, and F. Yin, "An Audio Watermarking Scheme with Neural Network," in *Proc. of International Symposium on Neural Networks, LNCS 3497*, pp. 795–800, 2005.

[102] Y. Bai, S. Bai, C. You, and B. Liu, "A blind audio watermarking algorithm based on fft cofficients quantization," in *Proc. of International Conference on Artificial Intelligence and Education*, 2010.

[103] Q. Zhang, Y. Huang, J. Deng, and L. Zheng, "A dither modulation with distortion compensation audio watermarking algorithm based on dwt," *Journal of Computational Information Systems*, vol. 7, no. 1, pp. 106–113, 2011.

[104] E. Aiba, M. Tsuzaki, S. Tanaka, and M. Unoki, "Judgment of perceptual synchrony between two pulses and verification of its relation to cochlear delay by an auditory model," *Japanese Psychological Research*, vol. 50, no. 4, pp. 204–213, 2008.

[105] J. A. Yasi, "An algorithm for extracting the relative phase of harmonics from a periodic digital signal." `http://physics.illinois.edu/undergrad/reu/2004/Yasi.pdf`, 2004.

[106] W. A. Yost, A. N. Popper, and R. R. Fay, *Auditory perception of sound sources.* Springer, 2008.

[107] K. Brandenburg, "MP3 and AAC Explained," in *Proc. of International Conference on High-Quality Audio Coding*, 1999.

[108] L. R. Rabiner, R. W. Schafer, and C. M. Rader, "The chirp z-transform algorithm," *IEEE Trans. Audio and Electroacoustics*, vol. 17, no. 2, pp. 86–92, 1969.

[109] P. Bassia, I. Pitas, and N. Nikolaidis, "Robust audio watermarking in the time domain," *IEEE Transactions on on Multimedia*, vol. 3, no. 2, pp. 232–241, 2001.

[110] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "Music genre database and musical instrument sound database," in *Proc. of International Conference on Music Information Retrieval*, pp. 229–230, 2003.

[111] M. Unoki and D. Hamada, "Embedding limitations with digital-audio watermarking method based on cochlear delay characteristics," *International Journal of Innovative Computing, Information and Control*, vol. 2, no. 1, pp. 1–23, 2011.

[112] M. Unoki and R. Miyauchi, "Robust, blindly-detectable, and semi-reversible technique of audio watermarking based on cochlear delay characteristics," *IEICE Transactions on Information and System*, vol. E98-D, no. 1, pp. 38–48, 2015.

[113] S. Lin and D. J. Costello, *Error Control Coding, 2nd ed.* Prentice Hall, 2004.

[114] R. C. Bosea and D. K. Ray-Chaudhuria, "On a class of error correcting binary group codes," *Information and Control*, vol. 3, no. 1, pp. 68–79, 1960.

[115] B. Peterson, L. Smith, J. Gruszynski, and W. Patstone, "Spectrum analysis amplitude and frequency modulation," *Hewlett Packard Test and Measurement Digital Application Note Library*, 1996.

[116] L. K. Moore, "The Emergency Alert System (EAS) and All-Hazard Warnings," *Congressional Research Service*, 2010.

[117] P. Gasparini, G. Manfredi, and J. Zschau, *Earthquake Early Warning Systems*. Springer, 2007.

[118] R. M. Allen, P. Gasparini, O. Kamigaichi, and M. Böse, "The status of earthquake early warning around the world: An introductory overview," *Seismological Research Letters*, vol. 80, no. 5, pp. 682–693, 2009.

[119] United States Geological Survey, "Earthquake early warning system." http://earthquake.usgs.gov/research/earlywarning, 2013. [Online; accessed 11-May-2015].

[120] Japan Meteorological Agency, "Earthquake early warnings." http://www.jma.go.jp/jma/en/Activities/eew.html, 2013. [Online; 12-May-2015].

[121] D. Kopitz and B. Marks, *RDS: The Radio Data System*. Artech House, 1999.

[122] A. Murphy and R. Poole, "AM Signalling System: AMSS - Does your radio know what it is listening to?," tech. rep., BBC Research and Development, U.K., 2006.

[123] L. Cornell, "The AM signalling system (AMSS)," tech. rep., Principal Systems Architect, BBC New Media and Technology, U.K., 2007.

[124] "Technical specification: ETSI TS 102 386, AM signalling system (AMSS)." Digital Radio Mondiale (DRM).

[125] B. P. Lathi, *Modern Digital and Analog Communication Systems*. Oxford University Press, Inc., 1998.

[126] N. M. Ngo, M. Unoki, R. Miyauchi, and Y. Suzuki, "Data hiding scheme for amplitude modulation radio broadcasting systems," *Journal of Information Hiding and Multimedia Signal Processing, Ubiquitous International*, vol. 5, no. 3, pp. 324–341, 2014.

Publications

Journal

[1] Nhut M. Ngo and Masashi Unoki, "Method of Audio Watermarking Based on Adaptive Phase Modulation," IEICE Transactions on Information and Systems, 2015 (conditionally accepted)

[2] Nhut M. Ngo, Masashi Unoki, Ryota Miyauchi, and Yôiti Suzuki, "Data Hiding Scheme for Amplitude Modulation Radio Broadcasting Systems," Journal of Information Hiding and Multimedia Signal Processing, Ubiquitous International, vol. 5, no. 3, pp. 324–341, 2014

Lecture Note

[3] Nhut M. Ngo and Masashi Unoki, "Watermarking for Digital Audio Based on Adaptive Phase Modulation," in Proc. of International Workshop on Digital-forensics and Watermarking, LNCS 9023, pp. 105–119, 2015

International Conference

[4] Nhut M. Ngo and Brian M. Kurkoski and Masashi Unoki, "Robust and Reliable Audio Watermarking Based on Dynamic Phase Coding and Error Control Coding," Proc. of 23rd European Signal Processing Conference (EUSIPCO2015), pp. 2316–2320, Nice, France, 2015

[5] Nhut M. Ngo and Masashi Unoki, "Robust and Reliable Audio Watermarking Based on Phase Coding," in Proc. of 40th International Conference on Audio, Speech, and Signal Processing (ICASSP2015), pp. 345–349, Brisbane, Australia, 2015

[6] Masashi Unoki, Jessada Karnjana, Shengbei Wang, Nhut M. Ngo, and Ryota Miyauchi, "Comparative Evaluations of Inaudible and Robust Watermarking for

Digital Audio Signals," in Proc. of The 21st International Congress on Sound and Vibration, 2014

[7] Nhut M. Ngo, Masashi Unoki, Ryota Miyauchi, and Yôiti Suzuki, "Data-hiding Scheme for Digital-audio in Amplitude Modulation Domain," in Proc. of International Conference on Information Hiding and Multimedia Signal Processing (IIHMSP2015), Piraeus, Greece, pp. 114–117, 2012

[8] Nhut M. Ngo, Shengbei Wang, and Masashi Unoki, "Method of digital-audio watermarking based on cochlear delay in sub-bands," in Proc. of International Technical Conference on Circuits/Systems, Computers and Communications, Sapporo, Japan, 2012

Domestic Conference

[9] Nhut M. Ngo, Brian M. Kurkoski, and Masashi Unoki, "Robust Audio Watermarking Based on Dynamic Phase Coding and Error Control Coding", IEICE Technical Report IT2015-16,EMM2015-16, vol. 115, no.37, pp. 81–86, May 2015

[10] Nhut M. Ngo and Masashi Unoki, "Study on Watermarking for Digital Audio Based on Adaptive Phase Modulation," , IEICE Technical Report EMM2014-102, pp. 149–154, March 2015

[11] Nhut M. Ngo and Masashi Unoki, "On Inaudibility of Cochlear-Delay-based Watermarking Method with Blind Detection," IEICE Technical Report EMM2013-114 pp. 87–92, March 2014

[12] Jessada Karnjana, Masashi Unoki, Shengbei Wang, Nhut M. Ngo, and Ryota Miyauchi, "Comparative Evaluations of Inaudible and Robust Watermarking Methods for Digital-Audio Signals," IEICE Technical Report EMM2013-110, pp. 63-68, March 2014

[13] Nhut M. Ngo and Masashi Unoki, "Second-Order Allpass Filter for Method of Watermarking Based on Cochlear Delay," Proc. 2013 ASJ Autumn meeting, 3-Q-2, pp. 791–794, Toyohashi, Sept. 2013

[14] Nhut M. Ngo, Masashi Unoki, Ryota Miyauchi, and Yôiti Suzuki, "Data hiding scheme for digital-audio in AM radio broadcasting systems," IEICE Technical Report EA2012-93 EMM2012-74, pp. 63–68, Nov. 2012

[15] Nhut M. Ngo, Masashi Unoki, Ryota Miyauchi, and Yôiti Suzuki, "Study on data hiding scheme for digital-audio in amplitude modulation domain," Proc. 2012 ASJ Autumn meeting, Nagano, Sept. 2012

[16] Shengbei Wang, Nhut M. Ngo, Masashi Unoki, "Digital-audio Watermarking Based on Cochlear Delay in Sub-bands," Proc. 2012 ASJ Autumn meeting, Nagano, Sep. 2012

[17] Nhut M. Ngo, Shengbei Wang, and Masashi Unoki, "Method of digital-audio watermarking based on cochlear delay in sub-bands," IEICE Technical Report EA2012-57, pp. 19–24, Aug. 2012