A study on data hiding scheme for digital audio in amplitude-modulation domain

By Ngo Nhut Minh

A thesis submitted to
School of Information Science,
Japan Advanced Institute of Science and Technology,
in partial fulfillment of the requirements
for the degree of
Master of Information Science
Graduate Program in Information Science

Written under the direction of
Associate Professor Masashi Unoki

September, 2012

A study on data hiding scheme for digital audio in amplitude-modulation domain

By Ngo Nhut Minh (1010222)

A thesis submitted to
School of Information Science,
Japan Advanced Institute of Science and Technology,
in partial fulfillment of the requirements
for the degree of
Master of Information Science
Graduate Program in Information Science

Written under the direction of
Associate Professor Masashi Unoki

and approved by
Associate Professor Masashi Unoki
Professor Masato Akagi
Professor Jianwu Dang

August, 2012 (Submitted)

Copyright © 2012 by Ngo Nhut Minh

Acknowledgements

I would like to express my deep gratitude to all those who gave me the possibility to complete my thesis.

Especially, I would like to give my special thanks to my supervisor, Associate Professor Masashi Unoki whose instruction, encouragement, and support have helped me in all the time of study and research. I sincerely appreciate the way my supervisor has helped me in my Master thesis and my daily life in JAIST. His guidance allowed me to learn many skills as well as knowledge throughout my research.

I am grateful to my vice supervisor, Professor Masato Akagi who gave me many helpful comments. I am also grateful to Professor Yôiti Suzuki who suggested me doing this research and discussed many important points in my research. I would also like to pay my regards to all friends in my laboratory for a great help.

In addition, I express my thanks to FIVE-JAIST program that gave me an opportunity to study in a comfortable environment. I have got a great support from this program. I am greatly thankful to the Professors who gave me interesting lectures when I was in the first year of Master course.

Last but not least, I would like to give many thanks to my family and my close friends who always encourage and support me.

Contents

List of Figures

Chapter 1

Introduction

1.1 Overview and Motivation

Amplitude modulation (AM) is the most typical and basic technique used in telecommunications to strengthen the power of signals so that they can survive being transmitted from broadcaster to listeners via AM radio link. AM has been applied to many fields in daily life such as AM radio services since AM receivers is very simple and low-cost. There are many broadcasting services using AM radio systems, e.g., broadcasting news, information programs (NHK World Radio Japan), and educational programs (BBC Learning English).

As transmitters and receivers for AM are simple, AM radio broadcasting is a robust communication technology and thus useful in case of disasters. Another advantage of AM radio system is that AM radio receivers worked just with batteries while other devices such as TVs do not. Therefore, AM radio is really a useful device in case of disasters because the power suppliers are easily stopped working when there are earthquake and tsunami. Moreover, some emergency alert systems have been working on broadcasting warnings via AM radio services. Audible emergency messages are used to attract the attention of people engaged in various everyday activities, e.g., office workers and car drivers. Therefore, these messages need to be perceived accurately and efficiently by everyone during emergencies. Regular AM radio programs are interrupted and replaced by information about disasters during emergencies. The warning messages are broadcast to all listeners even in areas that are not affected by disasters. The warning systems would be improved if only the listeners in the associated areas could hear the announcements and alert messages can be announced while regular radio programs are still working.

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 information in the same channel [1]. There have been many approaches to watermarking for digital-audio content, which can precisely and robustly embed and detect data. If the methods of watermarking can be applied to AM radio systems, the emergency alert system can embed warning messages as additional data into AM signals. The embedded data can then be used as alert signals to attract the attention of people in emergencies, in the same AM channel. The embedded data can also be used as code for an area that is being affected by a disaster. Based on the code, the warning system could identify the area that needs to be informed while listeners in unaffected areas would not be informed.

In addition, some of advanced AM radio systems such AM signalling system (AMSS) [2, 3, 4] embed a small amount of information to existing AM broadcasting signal. The embedded information could be service information that allows receivers to identify the AM station. It helps the users to select the station by name as well as to switch to other versions of the same radio service, e.g., digital, AM or FM channels if they are available. However, since the capacity of embedded information in this system is not high, it is limited in adding very short message into AM signal. This problem could be overcome by applying audio data hiding to AM signal in order to transmit much more additional data in the same channel.

1.2 Purpose

This thesis proposes a novel data-hiding scheme for digital-audio in the AM domain to transmit additional information in the AM signals being transmitted to receivers. The watermarking method based on cochlear delay (CD) proposed by Unoki et al. [5, 6] is used in this scheme. The CD-based method uses a non-blind detection process to extract embedded data in which original and watermarked signals must be available on the receiver side. This is inconvenient in practical applications because we have to use two broadcasting systems. Therefore, we consider modulating both the original and watermarked signals with the same carrier. The proposed data hiding scheme must be compatible with existing AM radio receivers so that we can widely apply it in practice. The particular users in the proposed scheme could receive the embedded message in addition to the audio signal while the users that use a standard AM receiver can extract only the audio signal. The proposed scheme could be used to broadcast audio signals and embedded data to receiver over a AM radio link. The proposed scheme could be applied to construct an advanced emergency alert system or used to transmit service information of AM radio station in addition to audio signals.

Another purpose of this thesis is to improve embedding capacity of the method of audio watermarking based on cochlear delay. A watermarking scheme with high capacity is needed in the proposed data hiding scheme because it offers a possibility to embed additional data with large size. The proposed scheme could be used to transmit additional media content such as images and sounds. Improvement of CD-based watermarking method is based on concept of sub-band coding. We take the advantage of embedding data into sub-band of audio signal to increase embedding capacity.

1.3 Organization of thesis

The rest of this thesis are divided into six chapters as below

Chapter 2 begins by a survey of methods of audio watermarking that have been proposed in recent years. The method of audio watermarking based on CD is also described in this chapter. In addition, we introduce amplitude modulation techniques that are used in communication systems and we also present models of modulation and demodulation in AM.

Chapter 3 describes the approach for the proposed data hiding scheme in AM domain and the challenges that we must overcome to construct a data hiding system for digital audio in AM domain.

Chapter 4 presents the implementation of the proposed scheme. We describe the general scheme in which the are two main phases, one on the transmitter side and one on the receiver side. We also present novel processes of modulation and demodulation called double-modulation and double-demodulation that we propose to build up the data hiding scheme.

Chapter 5 is the evaluation of the proposed scheme. We evaluate the quality of demodulated signals that users extract from AM signal generated by the proposed scheme. We also investigate the effectiveness of CD-based watermarking in this scheme.

Chapter 6 improves the CD-based method of watermarking with regard to embedding capacity. We present a method based on concept of sub-band coding to increase embedding capacity with CD-based watermarking.

The last chapter, chapter 7, summarizes and concludes this thesis. Further research directions of this thesis are also mentioned in this chapter.

Chapter 2

Background

2.1 Digital-audio watermarking

Recently, researchers have shown an increased interest in digital-audio watermarking [7, 8, 1] to propose a method for protecting copyright information of audio content [9] as well as transmitting inaudible data in the host audio signal. Many approaches have been proposed basically based on the following requirements [5]:

(a) **inaudibility** (watermarks should be imperceptible to audiences),

(b) **confidentiality** (secure and accurate concealment of embedded data),

(c) **robustness** (watermarking methods should be robust against various attacks, e.g., resampling, compression), and

(d) **blind detection** (detecting watermark without original signals).

The first requirement seems to be the most important since the change in the audio signal should not be recognized by listeners. If the sound quality of the original signal is affected, the original content may lose its commercial value. The second requirement ensures that listeners do not know if the audio signal contains watermarks. The third requirement ensures that the watermarking methods are invulnerable to deliberate attempts to forge, remove, or invalidate watermarks [9].

In general, there are two types of audio watermarking with regard to detection schemes, blind detection and non-blind detection watermarking. As its name would suggest, a blind watermarking scheme can extract watermarks without use of the original signal. Otherwise, a non-blind watermarking scheme needs to use the original signal to extract watermarks in detection process. The second type of watermarking may not be so useful in practice because it requires double storage space and double transmission for watermark detection. However, there are trade-offs between detection schemes and other properties of the watermarking scheme such as sound quality and accuracy of detected watermarks. Using original signals in detection process could offer a possibility of increasing sound quality and accuracy of watermarking scheme. Another application of non-blind watermarking scheme is to verify the owner of digital content in a copyright dispute [9] or inversion attack. It should be noted that these requirements are dependent on each class of applications. Different applications demand different types of watermarking schemes with different requirements.

The most basic and simple technique of audio watermarking have been based on least significant bit (LSB) [1]. Watermarks are embedded into original audio signals by manipulating the sample data points of digital-audio in relation to binary data of watermarks in this approach. Although watermarks could be embedded by this approach at a high data rate, it is not robust against various manipulations (e.g., downsampling, upsampling, bit compression, MP3 compression). Other methods of digital-audio watermarking have been based on the characteristics of human auditory system (HAS) to embed inaudible watermarks in to audio signals. These kinds of watermarking could be categorized into two groups: one processed in amplitude domain and one processed in phase domain. Typical methods in amplitude domain have basically been based on manipulating signals in the amplitude spectrum. There are, for instance, the spread spectrum approach (e.g. the direct spread spectrum (DSS) [7] and the secure spread spectrum [10]). These methods have been used to directly embed watermarks into the amplitude of digital-audio signals and detect the embedded watermarks from the watermarked signal without using the original signal. Spread spectrum methods such as DSS are relatively more robust than others since watermarks are spread throughout whole frequencies. However, this method does not satisfy the four requirements, especially inaudibility requirement.

A variety of watermarking schemes based on manipulating phase spectra or group delay of signals have been reported in recent years. There are, for example, echo-hiding approach proposed by Gruhl et al. in [11] and a method based on periodical phase modulation (PPM) [12, 13]. Echo-hiding approaches have been used to directly embed data into audio signal as time shifts. The advantages of these approaches are embedding watermarks into original signals with less distortion and lower computational complexity. Although they could satisfy the requirement of inaudibility, they have drawback in confidentiality (anyone can easily detect echo information). The PPM approach was based on aural capabilities in which PPM is relatively inaudible to human. However, as phase modulation randomly disrupts the phase spectra of high frequency components, the modulated components (embedded data) may be detected by human in watermarked pulse-like sounds. This is because human can perceive raid phase variations related to long and rapid group delays in sound [14].

The typical watermarking methods used in LSD, DSS, ECHO, and PPM approaches could partially satisfy the four requirements. However, it is very difficult to achieve an audio watermarking scheme that can satisfy all requirements, especially inaudibility and robustness simultaneously.

Based on human-auditory perception, method of digital-audio watermarking based on cochlear delay (CD) characteristic proposed by Unoki et al. could be considered as the state-of-the-art method in audio watermarking field [5, 15, 6, 16]. Experimental results revealed that it can used to embed inaudible watermarks into sound signal and detect watermarks from watermarked signals successfully. The results obtained from computer simulation also showed that this method satisfies most of the requirements: inaudibility, confidentiality, robustness, and high capacity. The next section presents a detailed explanation about this method.

2.2 Method of digital-audio watermarking based on Cochlear delay

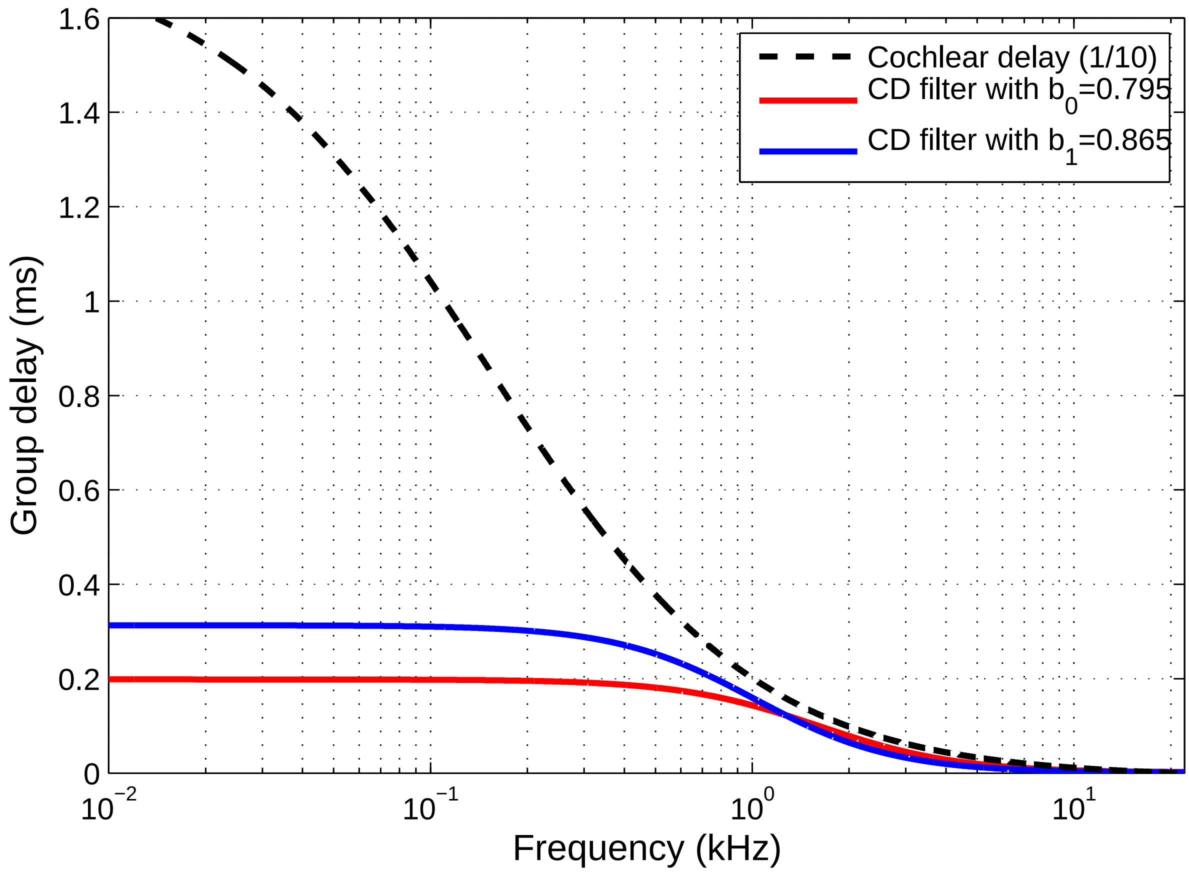


Figure 2.1: Cochlear delay and Group delay characteristics of IIR All-pass filter.

This method of audio watermarking was based on the properties of human cochlear. The cochlear is a fluid-filled cavity which receives vibrations caused by sound signals. It and other parts of the auditory system help us perceive sound contents. Researchers have shown that different frequency components of sound signals excite different positions in the cochlear (see [5] and references therein). The low frequency components require more time to reach the corresponding place where is near to the apex of the cochlear while the high frequency components excite the place where is near the base of the cochlear (the apex is further than the base from the outer ear). The difference in travelling time throughout the cochlear between low frequency components and high frequency components is referred to as “Cochlear delay”. Related studies on human perception with regard to Cochlear delay have been investigated by Aiba et al. [17, 18]. Their results suggested that the human auditory system cannot distinguish sound with enhanced delay and non-processing sound. The results have motivated the author to propose an inaudible watermarking scheme based on Cochlear delay characteristic.

The method of audio watermarking based on CD embeds data $s(k)$ into sound $x(n)$ by controlling the two different group delays of CD (phase information) of the original signal in relation to bit data being embedded ("1" and "0"). The key idea behind this method is that enhancing group delays related to CD does not affect human perception of the target sound. Figure 2.2 has a block diagram of the (a) embedding and (b) detection process for the CD-based method.

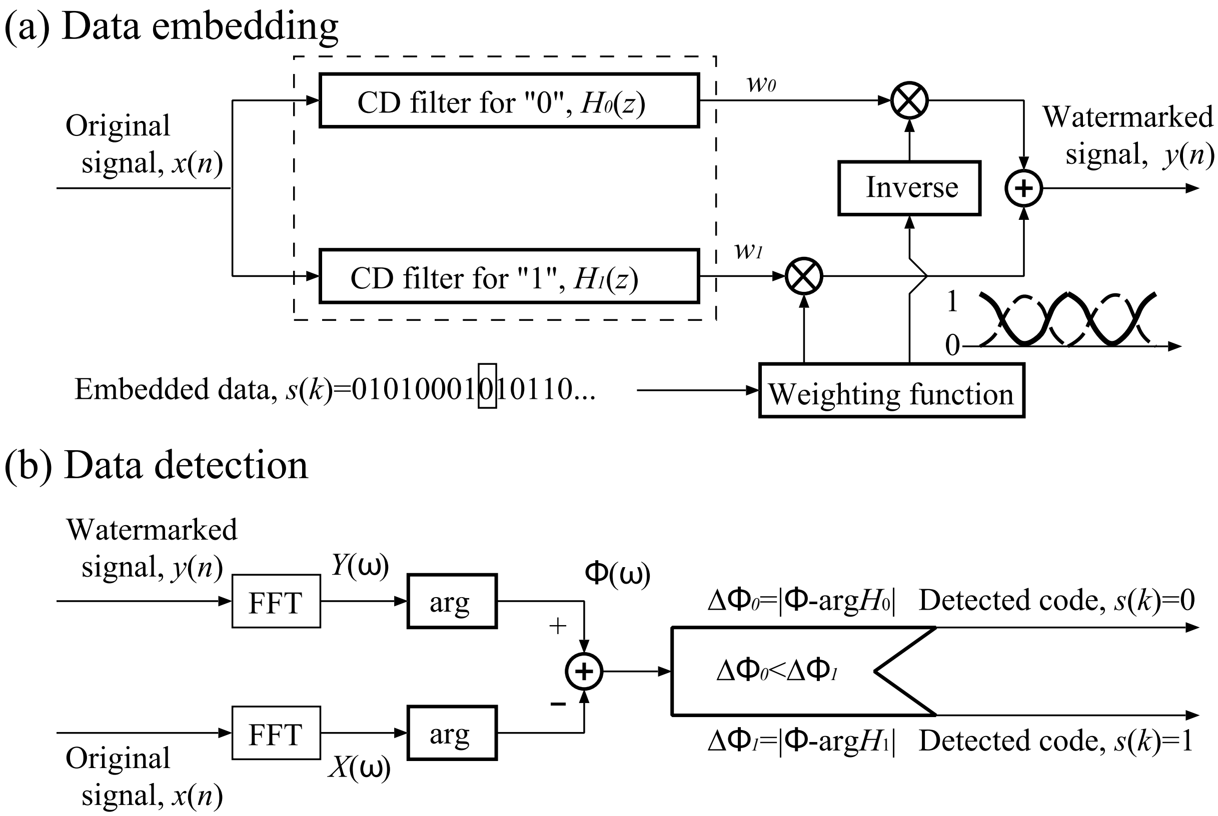


Figure 2.2: Method of Digital-audio Watermarking based on Cochlear delay characteristic: (a) data-embedding and (b) data-detection process.

An IIR all-pass filter is usually used to control delays in which amplitude spectra are passed equally without any loss. This method uses two IIR all-pass filters $H_0(z)$ and $H_1(z)$ to enhance the group delay of the original signal to embed data $s(k)$ into original signal $x(n)$. The phase characteristics of these two filters are modeled as Cochlear delay as shown in Fig.2.1. Figure 2.2 shows the block diagram of (a) watermarking process and (b) detection process of CD-based method. The outputs of phase enhancing processes are two signals, $w_0(n)$ and $w_1(n)$. The two intermediate signals, $w_0(n)$ and $w_1(n)$, are not affected in terms of human perception. The intermediate signals, $w_0(n)$ and $w_1(n)$, are then decomposed into overlapping segments (rate of overlap is 0.5). Finally, these signal segments are merged together in relation to watermark $s(k)$ (e.g., "010010101100110") as in Eq. (2.1), which results in watermarked signal $y(n)$. A weighting ramped cosine function was used to avoid discontinuity with this method between the marked segments in the watermarked signal. This processing is usually used as an overlap-add (OLA) method in short-term Fourier transformations.

$$y(n) = \begin{cases} w_0(n), \text{ if } s(k) = 0 \\ w_1(n), \text{ if } s(k) = 1, \end{cases} \tag{2.1}$$

where $(k-1)\Delta W \leq n < k\Delta W$. n is the sample index, k is the frame index, and $\Delta W = f_s/N_{bit}$ is the frame length. f_s is the sampling frequency of the original signal and N_{bit} is bit rate per second (bps) of embedded data.

The detection process of this method is a non-blind detection scheme. The original and watermarked signals in the detection process are first decomposed into overlapping segments using the same window function used in the data embedding process as shown in Fig. 2.2(b). The phase differences, $\phi(\omega)$s, are calculated between the segments of the original signal and those of the watermarked signal as in Eq. (2.2) . The phase difference of each segment, $\phi(\omega)$, is then compared to the group delays of $H_0(z)$ and $H_1(z)$ to detect bits of "0" or "1" as in Eqs. (2.3), (2.4), and (2.5). If the phase difference, $\phi(\omega)$, is closer to the group delay of $H_0(z)$ than that of $H_1(z)$, then bit "0" is detected. Otherwise, bit "1" is detected.

$$\phi(\omega_m) = \arg(\mathrm{FFT}[y(n)]) - \arg(\mathrm{FFT}[x(n)]), \tag{2.2}$$

$$\Delta\Phi_0 = \sum_m \left|\phi(\omega_m) - \arg(H_0(e^{j\omega_m}))\right|, \tag{2.3}$$

$$\Delta\Phi_1 = \sum_m \left|\phi(\omega_m) - \arg(H_1(e^{j\omega_m}))\right|, \tag{2.4}$$

$$\hat{s}(k) = \begin{cases} 0, \Delta\Phi_0 < \Delta\Phi_1 \\ 1, \text{otherwise} \end{cases} \tag{2.5}$$

Figure 2.3 has result of objective evaluations of CD-base method in comparison to the other methods. The result showed that PEAQs were over the evaluation threshold (¿-1 dB). It also revealed that LSDs increased as bit rate and that they were under the evaluation threshold. In addition, bit detection rate were greater than the evaluation criterion (75%) as bit rate increased from 4 to 1024 bps. The result suggested that the CD-based method could be used as non-blind inaudible watermarking, which can satisfy all evaluation thresholds.

2.3 Amplitude modulation

2.3.1 Definition

In telecommunication systems, modulation is usually used to transmit information (e.g., sound, image, and text message). It is used to gain certain advantages such as far-distance

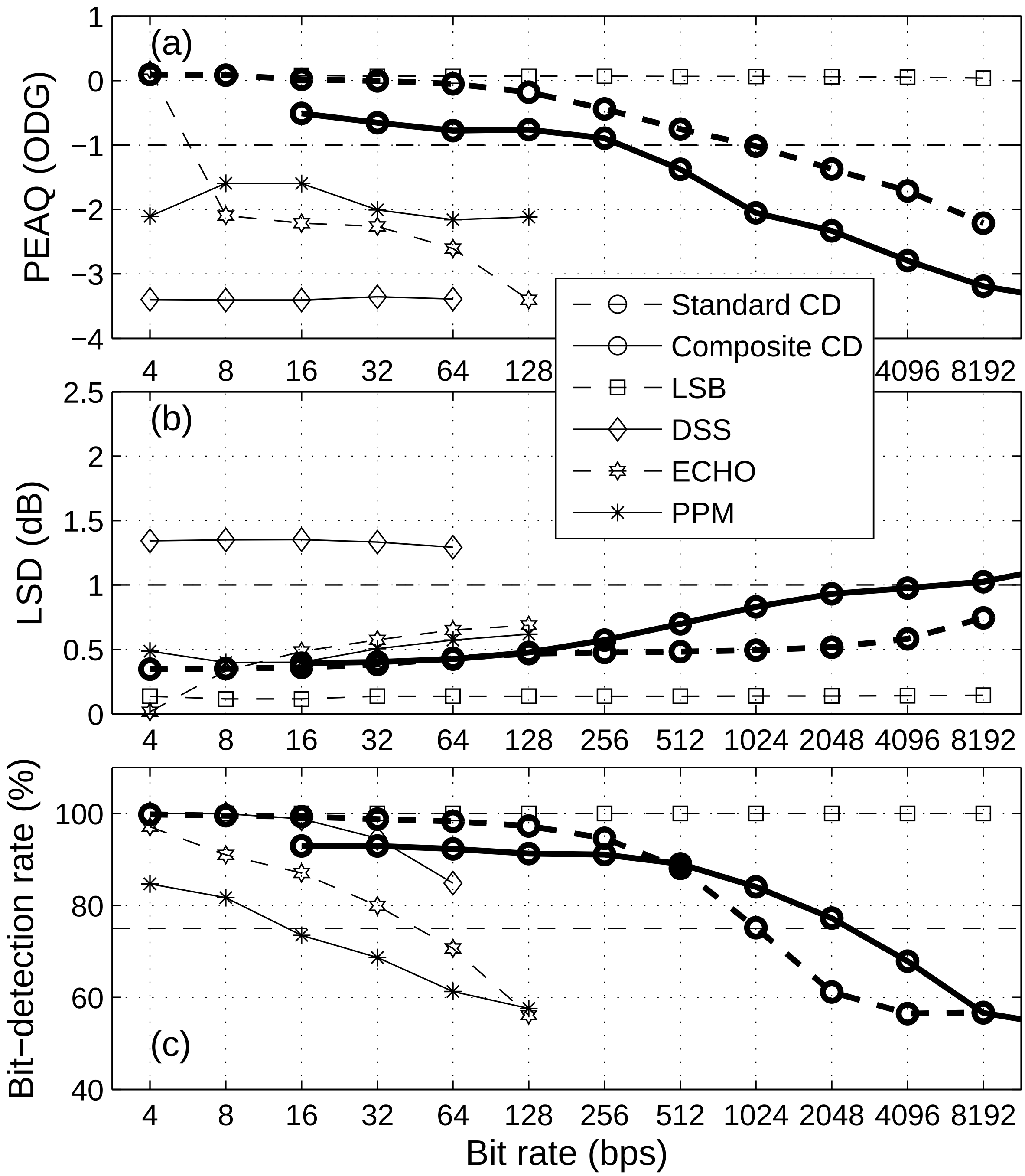
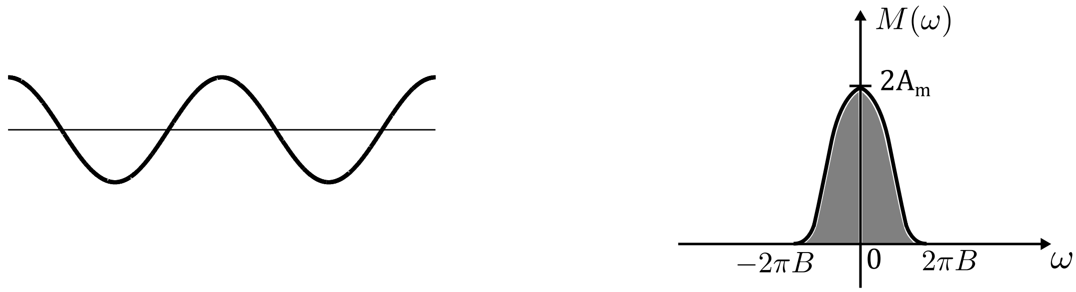
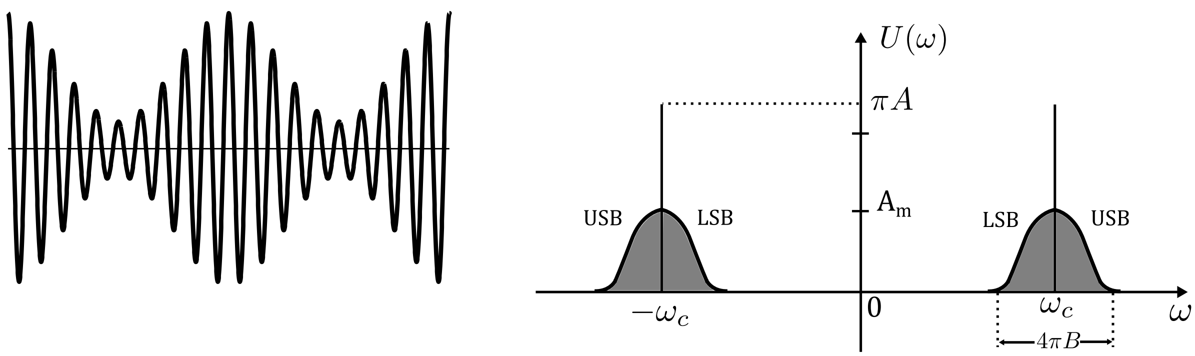


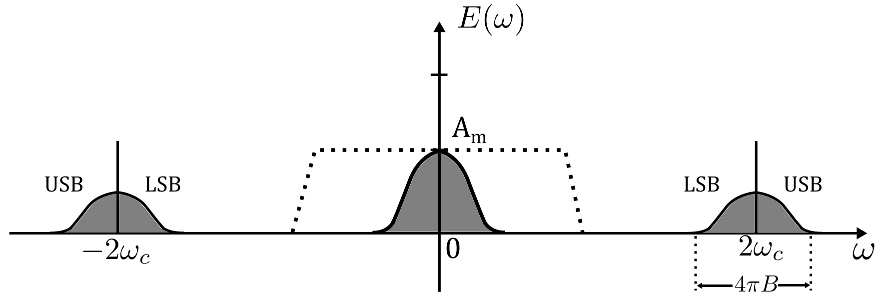
Figure 2.3: Objective evaluations of CD-based method compared with other methods.



(a) Message signal $m(t)$



(b) Modulated signal $u(t)$



(c) Spectrum of $e(t) = u(t)c(t)$

Figure 2.4: DSB-WC Modulation and Demodulation: (a) Message signal, (b) Modulated signal, and (c) Spectrum of product of Modulated and Carrier signals.

communication and transmitting over a radio wave. Contrast to communication using modulations is baseband communication. The term **baseband** is used to designate the band of frequencies of the information signal. In baseband communication, information signals (called baseband signals) are directly transmitted from transmitters to receivers over media. Because the baseband signals have sizeable power at low frequencies, they cannot be transmitted over a radio wave. The size of antennas that radiate the signal in waveform is directly proportional to the wavelength of the signal. Long-haul communication over a radio wave requires modulation to enable efficient power radiation using antennas of reasonable dimensions.

Amplitude modulation (AM) is the most typical and basic technique used in telecommunication for transmitting information via radio carrier waves. A carrier signal is a sinusoid signal of high frequency which is usually represented as follow

$$c(t) = \cos(\omega_c t), \tag{2.6}$$

in which the frequency in Hz is given by

$$f_c = \frac{\omega_c}{2\pi}. \tag{2.7}$$

The AM of the message signal $m(t)$ (Fig. 2.4(a)) which produces the modulated signal $u(t)$ (Fig. 2.4(b)) is defined as [19]

$$u(t) = [A + m(t)]\cos(\omega_c t) \tag{2.8}$$

In this formulation, the amplitude of the carrier is varied in proportion to the message signal $m(t)$.

The AM in Eq. (2.8) is the basic scheme of AM. It is called double-sideband with carrier (DSB-WC). There are other types of modulation in communication systems such as double-sideband suppressed carrier modulation (DSB-SC), single sideband modulation (SSB), etc.. In this thesis, we present only the basic type of AM. For further study, there are many references that provide comprehensive knowledge on modulation schemes [19, 20, 21].

2.3.2 Properties

Amplitude modulation simply shifts the spectrum of $m(t)$ to the carrier frequency. That means if

$$m(t) \Longleftrightarrow M(\omega), \tag{2.9}$$

then

$$u(t) \Longleftrightarrow \pi A[\delta(\omega + \omega_c) + \delta(\omega - \omega_c)] + \frac{1}{2}[M(\omega + \omega_c) + M(\omega - \omega_c)]. \tag{2.10}$$

We observe that if the bandwidth of $m(t)$ is B Hz, then the bandwidth of the modulated signal $u(t)$ is $2B$ Hz as shown in Fig. 2.4(b).

We also note that the spectrum of the modulated signal centered at ω_c is composed of two parts: a portion called lower sideband (LSB) lying below ω_c and a portion called upper sideband (USB) lying above ω_c. Similarly, the spectrum centered at $-\omega_c$ has LSB and USB.

2.3.3 AM modulator

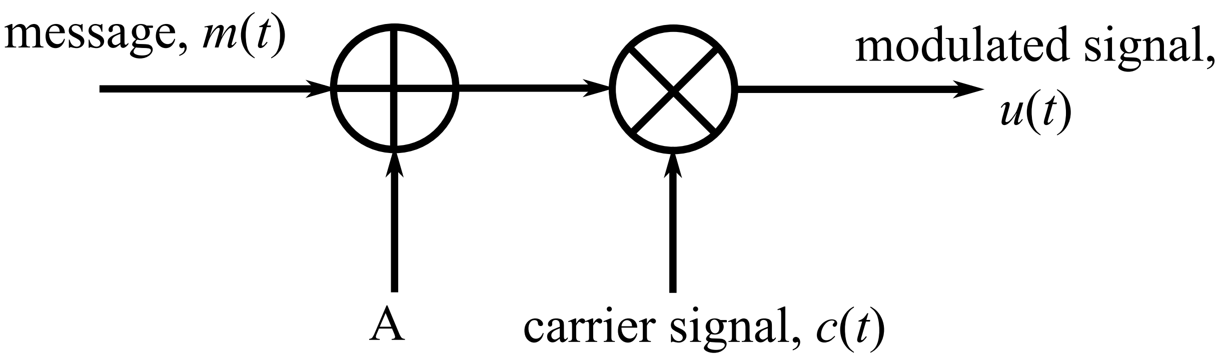


Figure 2.5: AM Modulator for DSB-WC.

Mathematically, the modulated signal is produced by using Eq. (2.8). The flow chart of an AM modulator is shown in Fig. 2.5. In practical communication systems, there are several kinds of AM modulators that use electric circuits. However, the output of such kinds of modulators is basically as in Eq. (2.8). In this thesis, we do not present the practical modulators which are designs for electric circuits. For comprehensive study, readers can refer to references [19, 20, 21].

2.3.4 AM demodulator

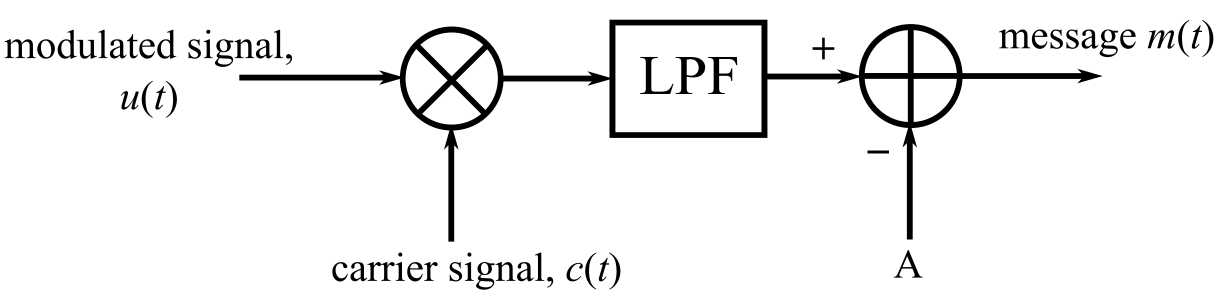


Figure 2.6: Synchronous Demodulator (Product Detector) for DSB-WC.

This section presents mathematical models for demodulating modulated signals.

In principle, there are two types of demodulations, synchronous demodulation and asynchronous demodulation. They are dependent on the use of the carrier in demodulation. The receiver must generate a carrier in phase and frequency synchronism for synchronous demodulation while there is no need to use a carrier in asynchronous demodulation.

Synchronous demodulation

This kind of demodulation may be referred to as a coherent or product detector. Figure 2.6 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.

$$\begin{aligned} e(t) &= u(t)c(t) && (2.11)\\ &= [A + m(t)]\cos^2(\omega_c t) && (2.12)\\ &= \frac{1}{2}[A + m(t)] + \frac{1}{2}[A + m(t)]\cos(2\omega_c t) && (2.13)\end{aligned}$$

Let us denote $m'(t) = A + m(t)$. Then,

$$e(t) = \frac{1}{2}m'(t) + \frac{1}{2}m'(t)\cos(2\omega_c t) \tag{2.14}$$

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

$$E(\omega) = \frac{1}{2}M'(\omega) + \frac{1}{4}[M'(\omega + 2\omega_c) + M'(\omega - 2\omega_c)] \tag{2.15}$$

The spectrum of $E(\omega)$ is shown in Fig. 2.4(c). This shows that the spectrum of the signal $e(t)$ consists of three components. The first component is the message spectrum. The two later components, which are the modulated signal of $m(t)$ with the carrier frequency $2\omega_c$, are centered at $\pm 2\omega_c$.

The signal $e(t)$ is then filtered by a lowpass filter with the cut-off frequency of f_c which results in the output $\frac{1}{2}m'(t)$. We can fully get $m'(t)$ by multiplying the output by two. We can also remove the inconvenient fraction 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

$$m(t) = m'(t) - A \tag{2.16}$$

Asynchronous demodulation

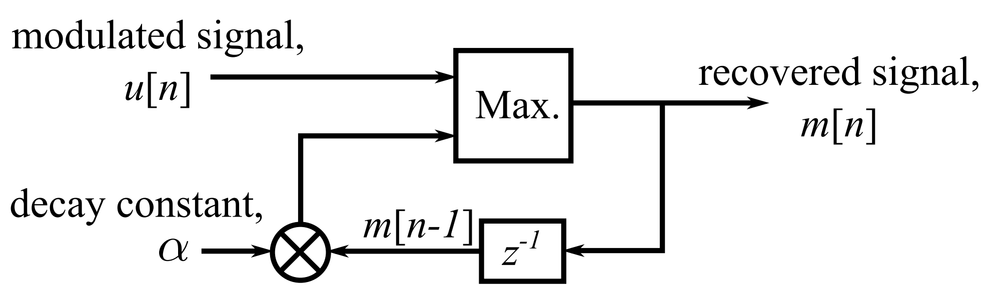


Figure 2.7: Asynchronous Demodulator (Envelope Detector) for DSB-WC.

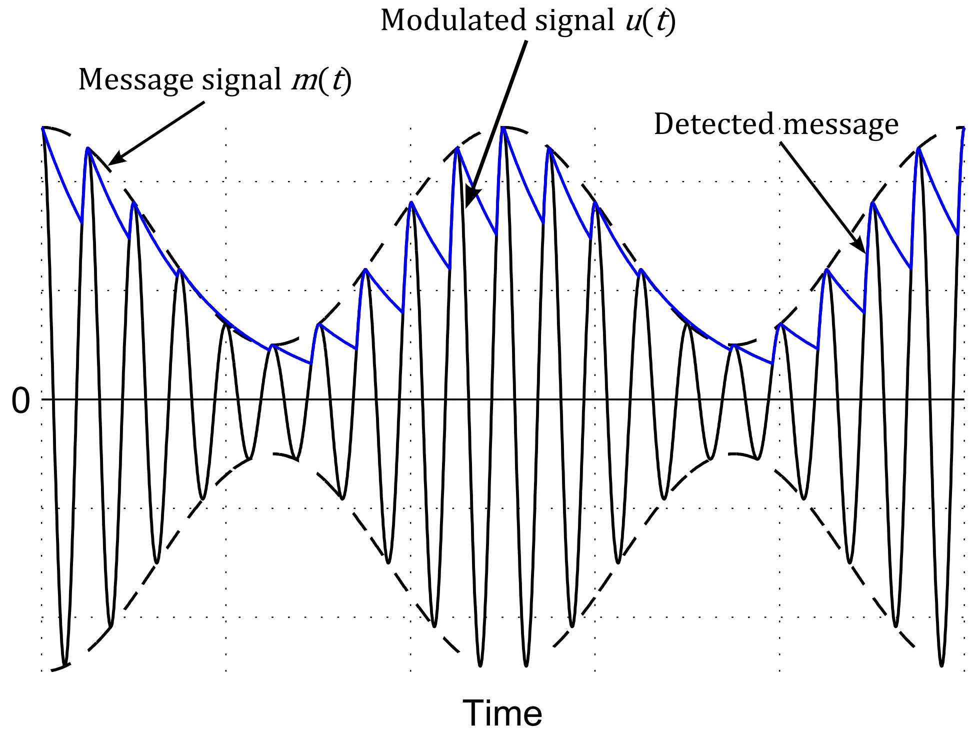


Figure 2.8: An example of Envelope Detector.

An asynchronous demodulator may be referred to as an envelope detector. For envelope detector, the modulated signal $u(t)$ must satisfy the requirement that

$$A + m(t) \geq 0, \forall t \tag{2.17}$$

An electric circuit is specially designed for detecting the envelope from an modulated signal. An enveloped detector can be simulated by the block diagram shown in Fig. 2.7. When the input is greater than the previous output, the output tracks the input. When the input falls below the output, the output decays exponentially with a decay constant $\alpha = \exp(-\frac{T}{RC})$, where $T = 1/fs$, RC is a constant, and the initial value $m[0] = 0$. The modulated signal and the envelope detector output are as shown in Fig. 2.3.4. The specification of envelope detector in realistic systems could be referred in [19] Section 4.3 p. 168.

Chapter 3

Proposed scheme

3.1 Approach

In communication systems such as AM radio broadcasting systems, the audio signal is modulated with a carrier before it is transmitted to receivers. This is because the base-band signals have sizable power at low frequencies while modulated signals have strong power enough to be able to be transmitted over radio link [19]. At the receivers, the modulated signals are demodulated to extract the audio signals. Based on this principle of communication systems, it suggests two approaches to embedding data into AM radio signal. The first approach is to directly modify modulated signals to embed data. This approach is very complex because we must process much more sample data point. Another approach is to embed data into audio signals and then modulate the audio signals that were embedded with data. This approach could be easily implemented by employing an available audio watermarking system. This thesis focuses on this approach to construct data hiding scheme for digital-audio in AM domain.

Figure 3.1 shows the approach to embedding data into AM radio signal. Audio signals are embedded with watermarks by using an available method of audio watermarking before it is modulated for transmitting. There have been many methods of audio watermarking in recent years. The method of audio watermarking should satisfy the requirements of

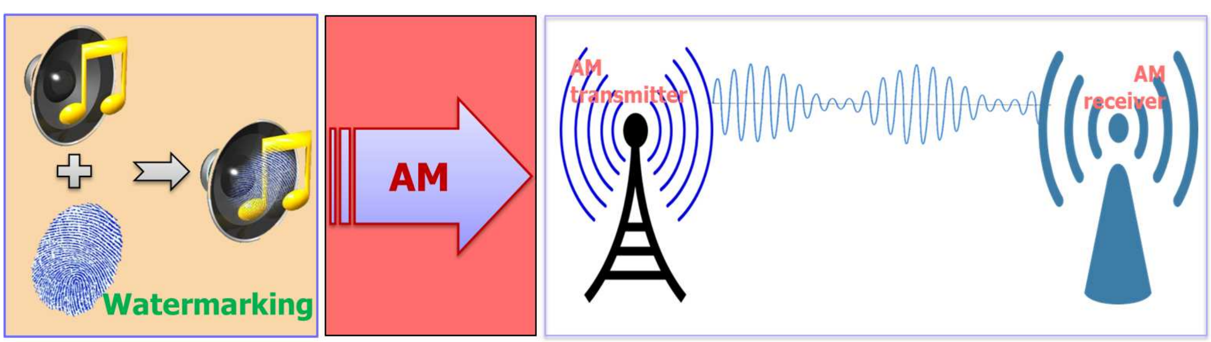


Figure 3.1: Approach to Embedding data into AM radio signal.

inaudibility, robustness, and high capacity to be able to embed data for audio signals in this scheme. After the audio signal is embedded with a watermark, the watermarked signal is modulated by a modulation process to prepare for transmitting. If the non-blind audio watermarking is used to embed watermark, the original signal must also be modulated and transmitted to receivers so that the embedded watermark could be detected. The modulated signal is then sent to receivers through an AM radio wave.

When the receivers in this scheme receive the modulated signal, they can demodulate the received signal and then extract the embedded watermark from the demodulated signal by using a demodulation process and a watermark detector. The extracted audio signal and the detected watermark are then used for desired purposes.

As presented in Section 2.2, the method of audio watermarking based CD characteristic has been the state-of-the-art method which satisfies most of requirements of an audio watermarking system: inaudibility, confidentiality, and robustness. We employed CD-based method to embed watermarks into audio signals in our proposed data hiding scheme. CD-based method, however, has a drawback of non-blind detection that requires the original signal in detection process. We overcame this problem by modulating both the original and watermarked signals in the same carrier. We proposed the novel double-modulation and double-demodulation to modulate both the original and watermarked signals with the same carrier and then demodulate the modulated signal to extract them. The advantage of double-modulation is that we only need one transmitter for transmitting both the original and watermarked signals to receivers. The drawback of blind detection in CD method could be cancelled out by the proposed double-modulation/double-demodulation. The proposed scheme combined the advantages of CD method and the advantage of novel double-modulation/double-demodulation technique to construct a perfect data hiding scheme for AM radio broadcasting system.

3.2 Challenging Issues

There are some issues that should be taken into account in with the proposed approach. The sound quality of audio signals that are carried by AM signal could be affected due to double-modulation process. The embedded data could also be affected by the double-modulation/double-demodulation process. The proposed scheme should be compatible with the vast majority of standard AM receivers as a requirement for widely applying it in practice. The embedding capacity of the watermarking system should increase so that we can transmit large-size media such as images, other audio in addition to text message.

In summary, we aim to construct a data hiding scheme with the following requirements

- Modulating both the original and watermarked signals with the same carrier,
- Sound quality of extracted signals kept at high level and precisely detecting watermarks,
- Compatible with standard AM receivers, and
- High capacity of embedded data.

Chapter 4

Implementation

4.1 General scheme

The proposed data-hiding scheme for AM radio signals is shown in Fig. 4.1. There are two main phases in this scheme: (a) data-embedding and a double-modulation process on the transmitter side and (b) a double-demodulation and data-detection process on the receiver side.

On the transmitter side (Fig. 4.1(a)), the CD-based method of watermarking is used to embed data, $s(k)$, into audio signal $x(n)$. The output of this step is the watermarked signal, $y(t)$. The original and watermarked signals, $x(n)$ and $y(n)$, are double-modulated as LSB and USB, respectively, with the carrier, $c(n)$. This is based on AM modulation with the method of double sidebands with the carrier (DSB-WC) but the difference between DSB-WC and the proposed scheme is that both sidebands are different in this scheme while LSB and USB in DSB-WC are the same. The double-modulated signal, $u(n)$, which conveys both the original and watermarked signals, $x(n)$ and $y(n)$, is used for broadcasting to receivers.

At the receiver (Fig. 4.1(b)), the double-modulated signal, $u(n)$, is double-demodulated to extract the conveyed signals, $\hat{x}(n)$ and $\hat{y}(n)$. The extracted signals, $\hat{x}(n)$ and $\hat{y}(n)$, are then used to detect embedded data $\hat{s}(k)$ by using CD-based watermarking.

The two following sections present the implementation of the double-modulation and double-demodulation processes used in this scheme. The implementation of method of audio watermarking based on CD (embedding and detection processes) are presented in Section 2.2. The standard modulation and demodulation are referred to as the AM techniques currently used in AM communication systems. They are presented in Section 2.3.

4.2 Double-modulation process

The flow for the double-modulation process is shown in Fig. 4.2. The standard modulation is defined as AM technique currently used in AM radio system. The original and

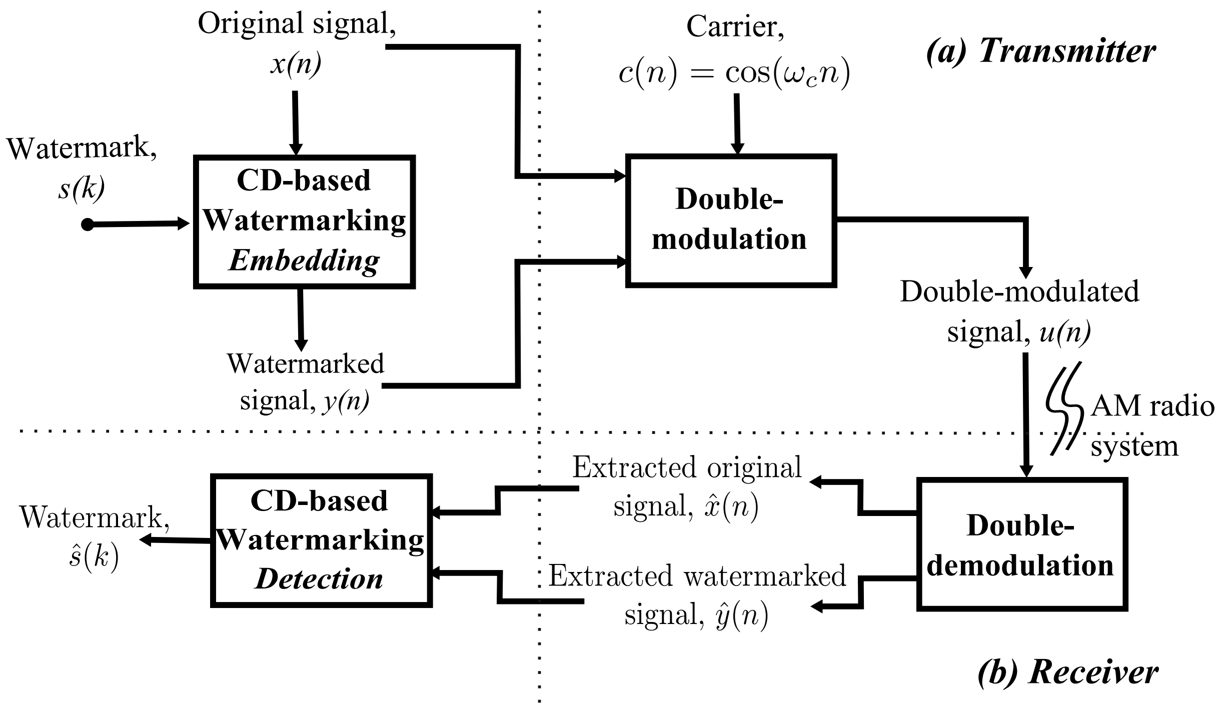


Figure 4.1: General scheme for Digital-audio Data hiding in AM domain.

watermarked signals, $x(n)$ and $y(n)$, are double-modulated as lower and upper sidebands with the carrier in four steps:

Step 1: Original signal $x(n)$ and watermarked signal $y(n)$ are modulated with the standard AM technique. The standard-modulated signals $u_1(n)$ and $u_2(n)$ are obtained as the outputs of standard AM modulation processes.

Step 2: The standard-modulated signals, $u_1(n)$ and $u_2(n)$, are then transformed with a fast Fourier transform (FFT). The outputs correspond to frequency spectra $U_1(\omega)$ and $U_2(\omega)$ of $u_1(n)$ and $u_2(n)$. Each spectrum contains two components, i.e., lower and upper sidebands.

Step 3: The low sideband component in $U_1(\omega)$ and the upper sideband component in $U_2(\omega)$ are merged into $U(\omega)$.

Step 4: Finally, $U(\omega)$ is transformed into $u(n)$ by using an inverse FFT. The double-modulated signal, $u(n)$, contains both original and watermarked signals $x(n)$ and $y(n)$.

4.3 Double-demodulation process

The double-demodulation process extracts signals $x(n)$ and $y(n)$ from the double-modulated signal, $u(n)$. They are original signal $x(n)$ in the lower sideband and watermarked signal $y(n)$ in the upper sideband. The block diagram of the double-demodulation process is

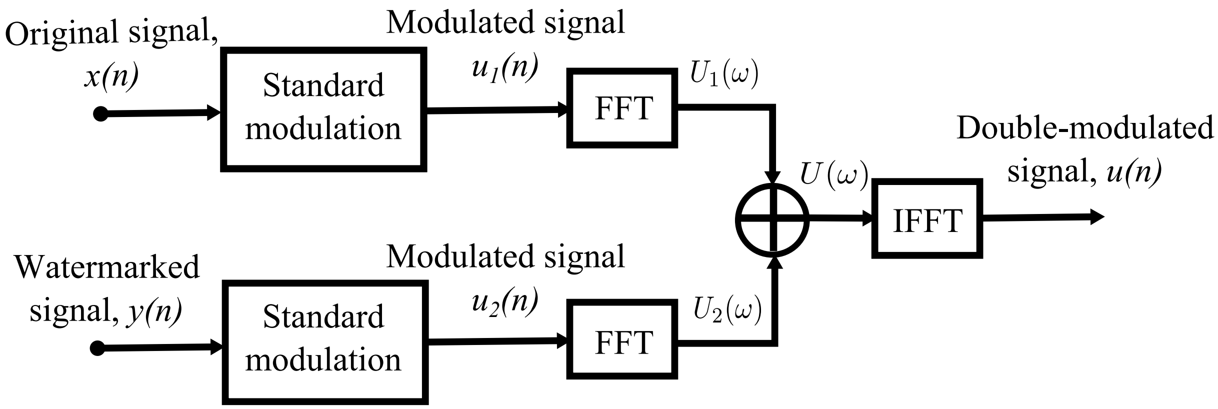


Figure 4.2: Flow chart of Double-modulation process.

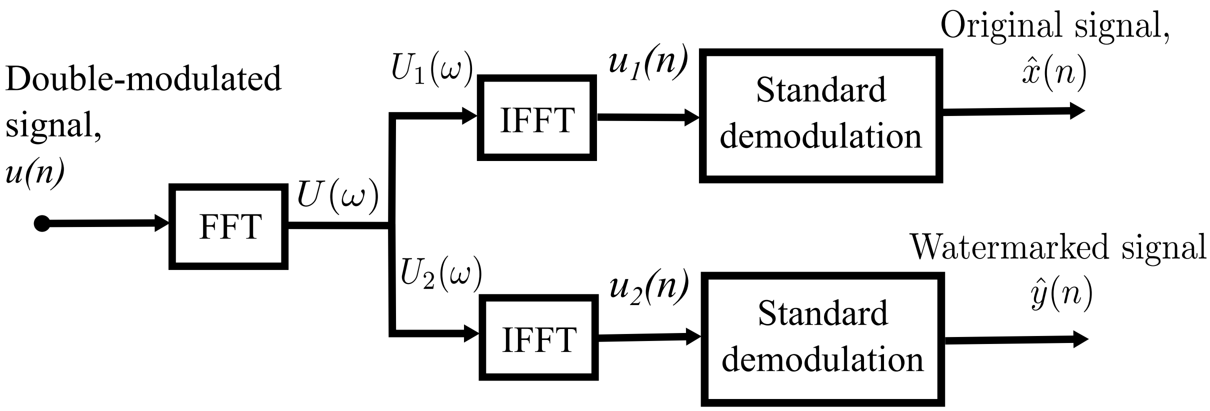


Figure 4.3: Flow chart of Double-demodulation process.

shown in Fig. 4.3. The standard demodulation is defined as AM demodulation technique currently used in AM radio system. The double-demodulation process involves four steps.

Step 1: The double-modulated signal, $u(n)$, is transformed into frequency spectrum $U(\omega)$ by FFT. $U(\omega)$ has the spectrum of the modulated signal, which contains two components, i.e., lower and upper sidebands.

Step 2: $U(\omega)$ is decomposed into $U_1(\omega)$ and $U_2(\omega)$; the first contains the lower sideband and the second contains the upper sideband. The upper sideband of $U_1(\omega)$ and the lower sideband of $U_2(\omega)$ are equal to zero.

Step 3: $U_1(\omega)$ is then transformed into $u_1(n)$ and $U_2(\omega)$ into $u_2(n)$ by the inverse FFT.

Step 4: Finally, the original signal, $\hat{x}(n)$, is extracted by demodulating signal $u_1(n)$ with standard synchronous demodulation technique. Since $u_1(n)$ only contains one sideband, the standard-demodulated signal of $u_1(n)$ is multiplied by two to fully recover the original signal. Similarly, the extracted watermarked signal, $\hat{y}(n)$, is obtained by multiplying the standard-demodulated signal of $u_2(n)$ by two.

Chapter 5

Evaluation

5.1 Database and Criteria of Evaluation

We conducted computer simulations to evaluate the proposed scheme for AM broadcasting systems. We investigated the sound quality of extracted original and watermarked signals in double-modulation/demodulation with the proposed scheme. We also evaluated sound quality with regard to the inaudibility and accuracy of watermark detection with the signals extracted from the double-modulated signals in the proposed scheme. In addition, we investigated the sound quality of the signals that were extracted from the double-modulated signals by using standard demodulators in order to test the low-level compatibility of the proposed scheme with current AM radio broadcasting system.

We used all 102 tracks of the RWC music database [22] as the original sound signals. These music tracks had a sampling frequency of 44.1-kHz, were 16 bits, and had two channels. The carrier frequency was 250 kHz. The sampling frequency of the AM system was 1000 kHz. The same watermarks "feb" were embedded into the original sound signal by using the CD-based method. The data rate was from 4 to 1024 bps.

We used objective evaluations, i.e., the signal-to-noise ratio (SNR), log spectrum distortion (LSD) [23], and perceptual evaluation of audio quality (PEAQ) [24] to measure sound quality of the target signals. SNR is used to compare the level of a clean signal to the level of noise. The higher SNR signal has better sound quality. SNR is defined in dB by

$$\mathrm{SNR} = 10\log_{10}\left\{\frac{\sum_{n=1}^{N}[x(n)]^2}{\sum_{n=1}^{N}[y(n)-x(n)]^2}\right\} \qquad \text{(dB)}, \tag{5.1}$$

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

LSD is used to measure the distance or distortion between two spectra. A lower value for LSD indicates a better result and vice versa. LSD is defined by

$$\mathrm{LSD} = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left[10\log_{10}\frac{|Y(\omega,k)|^2}{|X(\omega,k)|^2}\right]^2} \qquad \text{(dB)}, \tag{5.2}$$

where $X(\omega, k)$ and $Y(\omega, k)$ are short-time Fourier transform of the clean and observed signals, respectively, in which the overlap rate of 0.6 was chosen for this evaluation. k is the frame index and K is the number of frames.

PEAQ was used to measure degradation in audio quality according to the objective difference grade (ODG) which ranges from from -4 to 0. ODG indicates sound quality of target signals as shown in Table 5.1

Table 5.1: ODG and Sound quality Degradation

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

The evaluation thresholds for the SNR, LSD, and PEAQ criteria which correspond to 20 dB, 1 dB, and -1 ODG, respectively were chosen to evaluate the sound quality of the signals in this experiment.

The accuracy of watermark detection was measured by bit detection rate which is defined as the ratio between the number of correct bits and the number of total bits of the detected watermark. An evaluation threshold of bit detection rate below 75% was chosen in this experiment.

5.2 Sound quality of Double-demodulated signals

The results of sound quality of the signals extracted from the double-modulated signals are plotted in Figs. 5.1 and 5.2 as a function of bit-rate. When the bit-rate was 4 bps, which is a critical condition for the CD-based method of watermarking [5], the SNR, LSD, and PEAQ of the extracted original and extracted watermarked signals corresponded to approximately 55 dB, 0.15 dB, and -0.08 ODG (objective difference grade).

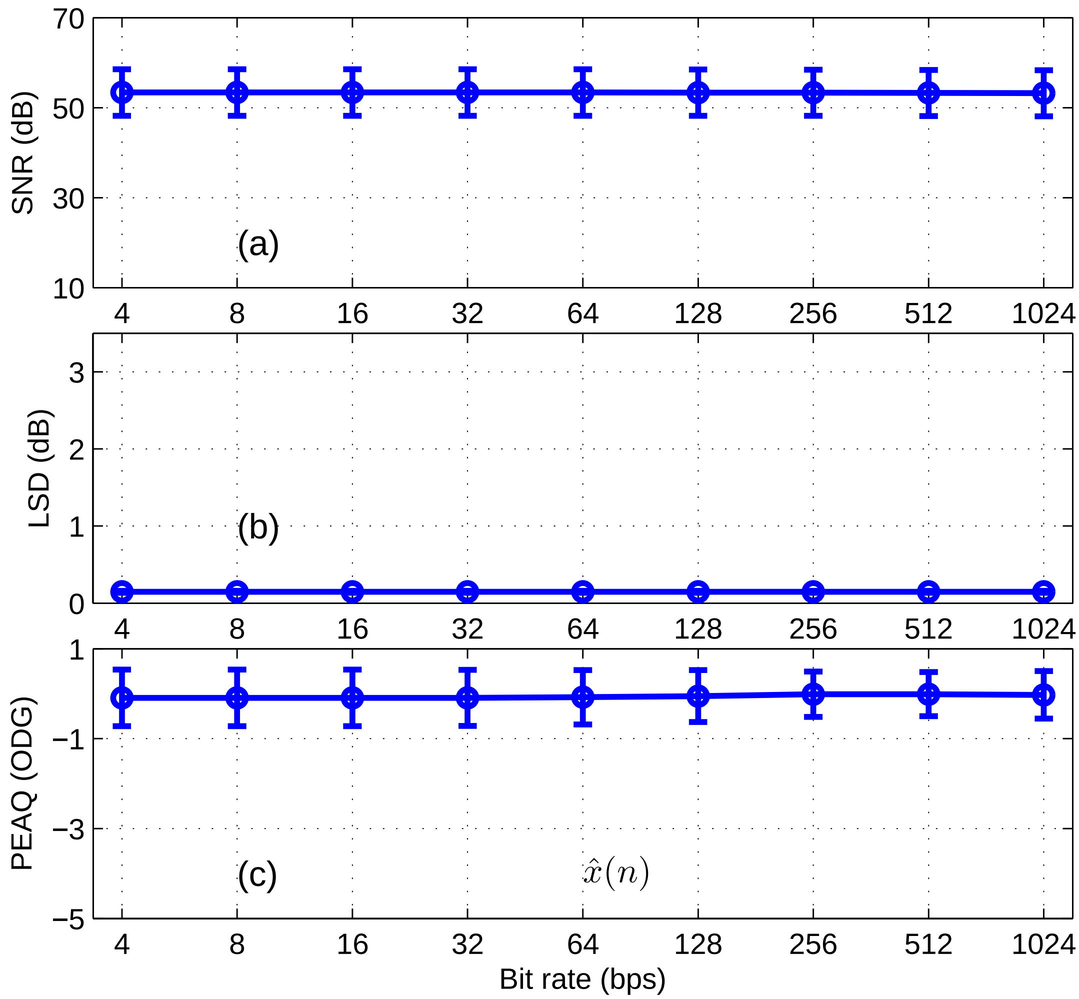


Figure 5.1: Results of Objective evaluations of Extracted Original signal as a function of Bit rate: (a) SNR, (b) LSD, and (c) PEAQ

The SNR and LSD had significantly high values in practice. The PEAQ was about -0.08 ODG, which is imperceptible indicating that the difference in the signals was not able to be perceived. These results indicated that the extracted signals were not distorted by the double-modulation and double-demodulation process. Since the bit-rate only affected the watermarked signal in relation to the original signal, when the bit-rate increased from 4 to 1024 bps the SNR, LSD, and PEAQ did not decrease.

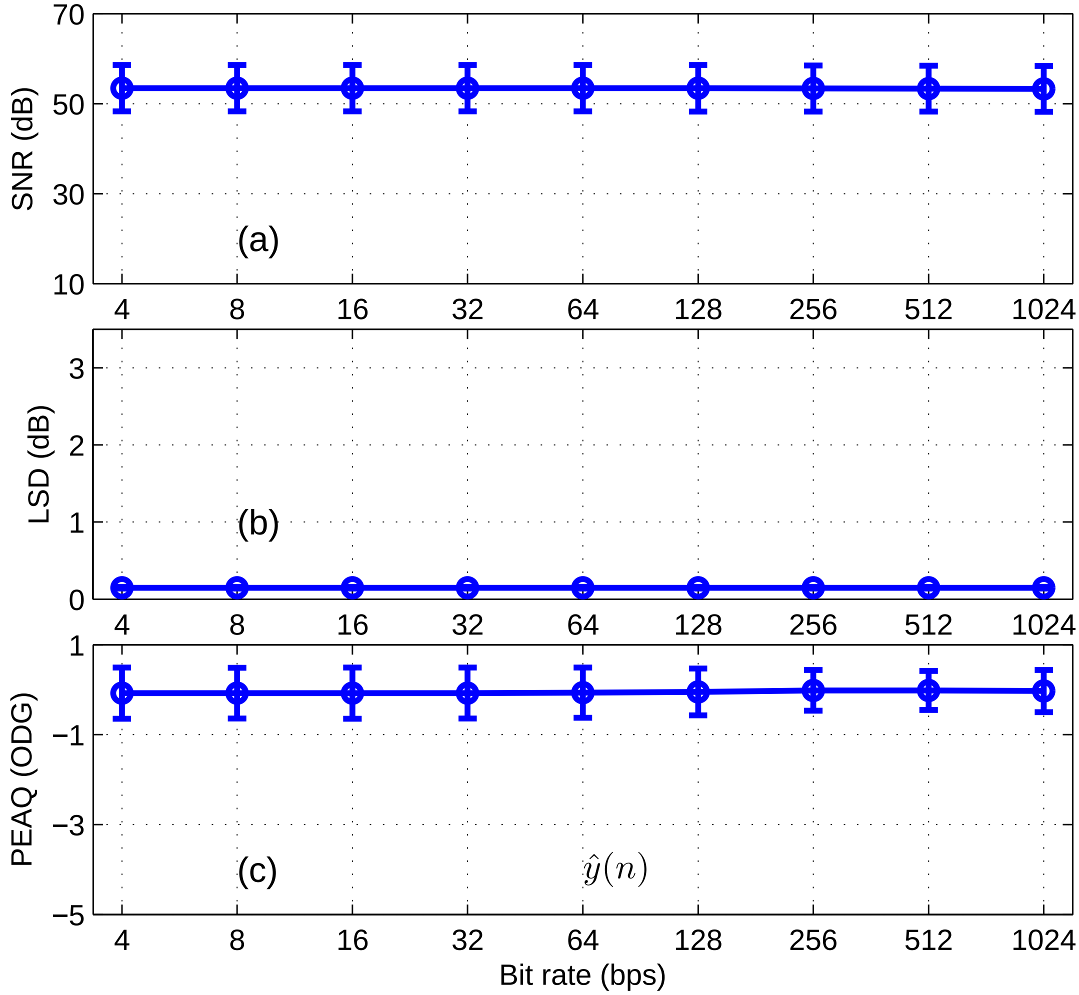


Figure 5.2: Results of Objective evaluations of Extracted Watermarked signal as a function of Bit rate: (a) SNR, (b) LSD, and (c) PEAQ

5.3 Inaudibility and Accuracy

The sound quality with regard to the inaudibility of the watermarked signals that were extracted from the double-modulated signals and accuracy of watermark detection with the signals that were extracted from the double-modulated signals are shown in Fig. 5.3. The bit-detection rate was greater than 99.5% when the bit-rate increased from 4 to 256 bps. It decreased dramatically when the bit-rate was 512 and 1024 bps. The PEAQs of extracted watermarked signal with extracted original signal were greater than −1 ODG and the LSDs were less than 0.5 dB.

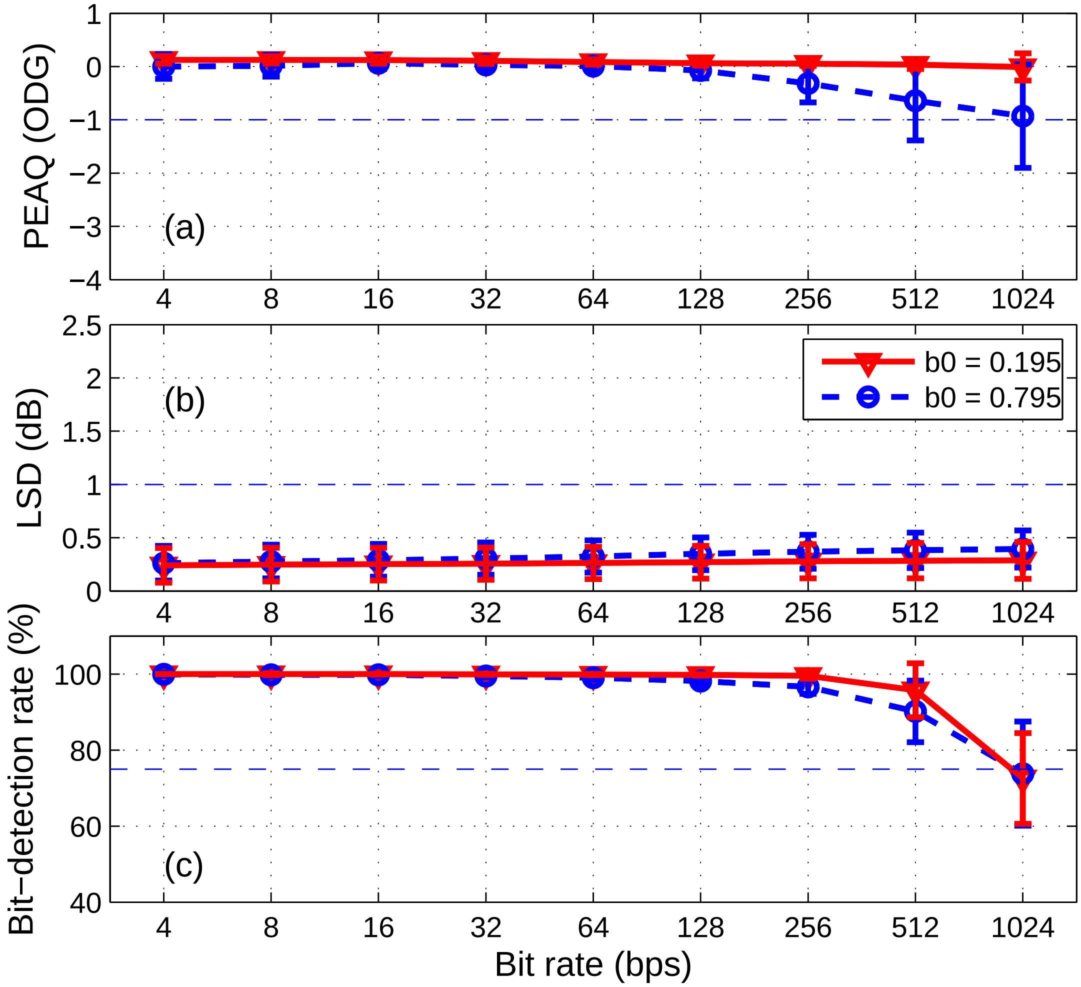


Figure 5.3: Result of Objective evaluations (PEAQ, LSD, Bit-detection rate) as functions of Bit rate.

These results demonstrated that the bit-detection rate and inaudibility of watermarked signal with this scheme were the same as that with the watermarking method. It indicates that the CD-based method of watermarking could be applied to the AM domain without distortion and that our proposed scheme could embed data into the AM audio signal and could precisely and robustly detect embedded data.

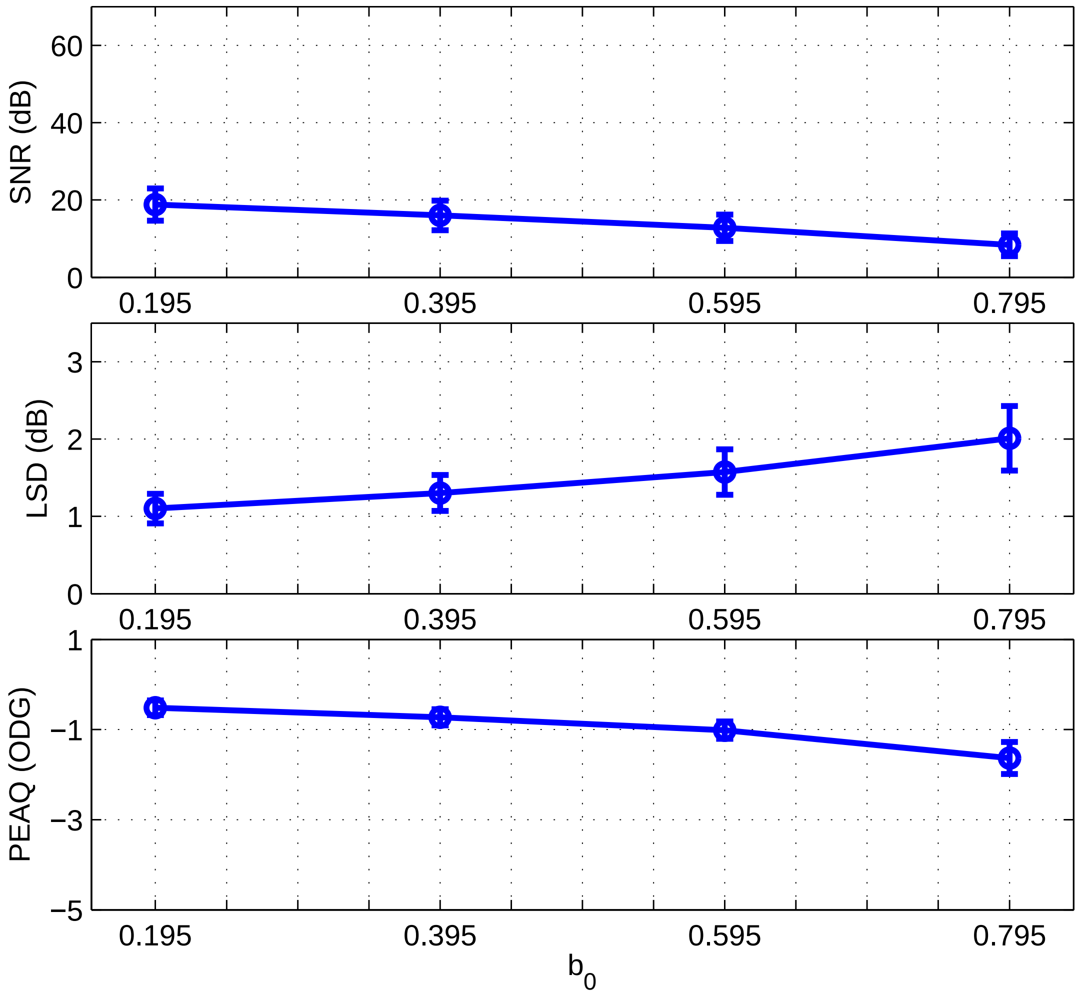


Figure 5.4: Sound quality of Signals extracted with Standard Demodulation technique.

5.4 Evaluation of Low-level compatibility

There are a vast majority of AM radio receivers that recover audio signals from AM radio signals through standard AM techniques (envelope detection and coherent detection). The data-hiding scheme should produce a modulated signal that can be demodulated by standard AM radio devices. The difference between two sidebands of the double-modulated signal is phase shift according to CD characteristics. Therefore, if the difference in phase of the original and watermarked signals could be properly reduced, the double-modulated signal could be demodulated with the standard technique with less distortion. Of course, the quality of the double-demodulated signal should not be smeared.

We utilized b_0 and b_1 with the CD-based method of watermarking to find the most suitable values for low-layer compatibility with our proposed scheme. Figure 5.4 presents the results of objective tests of standard-demodulated signals according to b_0 where $b_1 = b_0{+}0.07$ [5]. The results demonstrated that the sound quality of the standard-demodulated signals decreases as b_0 increases. Figure 5.3 shows that the PEAQ and LSD of extracted

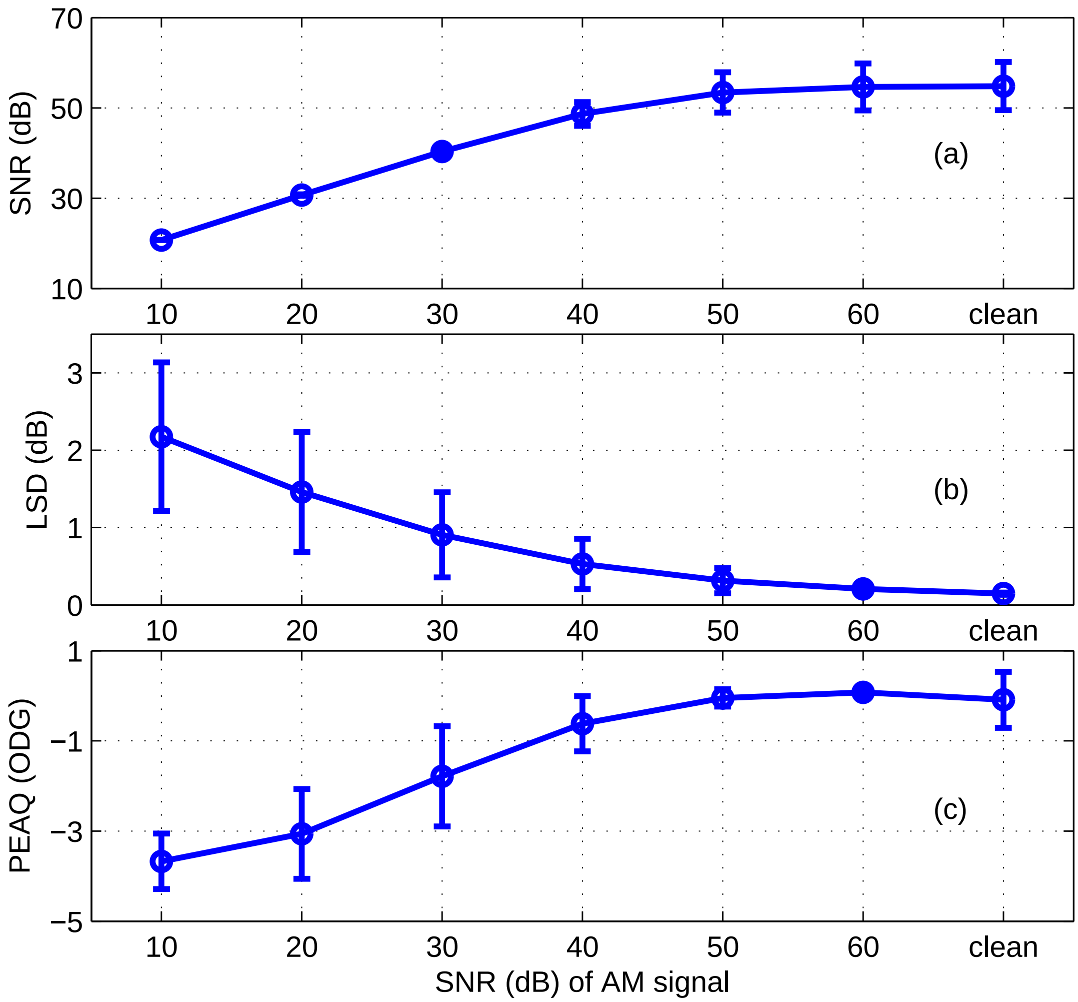


Figure 5.5: Results of Objective evaluations (SNR, LSD, and PEAQ) of Extracted Original signals against External noise.

watermarked signal with extracted original signal and bit-detection of watermarking in case $b_0 = 0.195$ are better than those in case $b_0 = 0.795$. The sound quality of double-demodulated signals remains unchanged under these conditions. Thus, smaller values for b_0 and b_1 should be chosen. It seems that $b_0 = 0.195$ and $b_1 = 0.265$ are reasonable for the proposed scheme; however, we need to consider these further because of the trade-off between sound quality and bit detection.

5.5 Sound quality of Extracted signals with regard to External noise

The AM double-modulated signal was transmitted through the air which may have been affected by external noise. We considered distortion of the extracted signals when the double-modulated signal was subjected to white noise. Figures 5.5 and 5.6 plots the

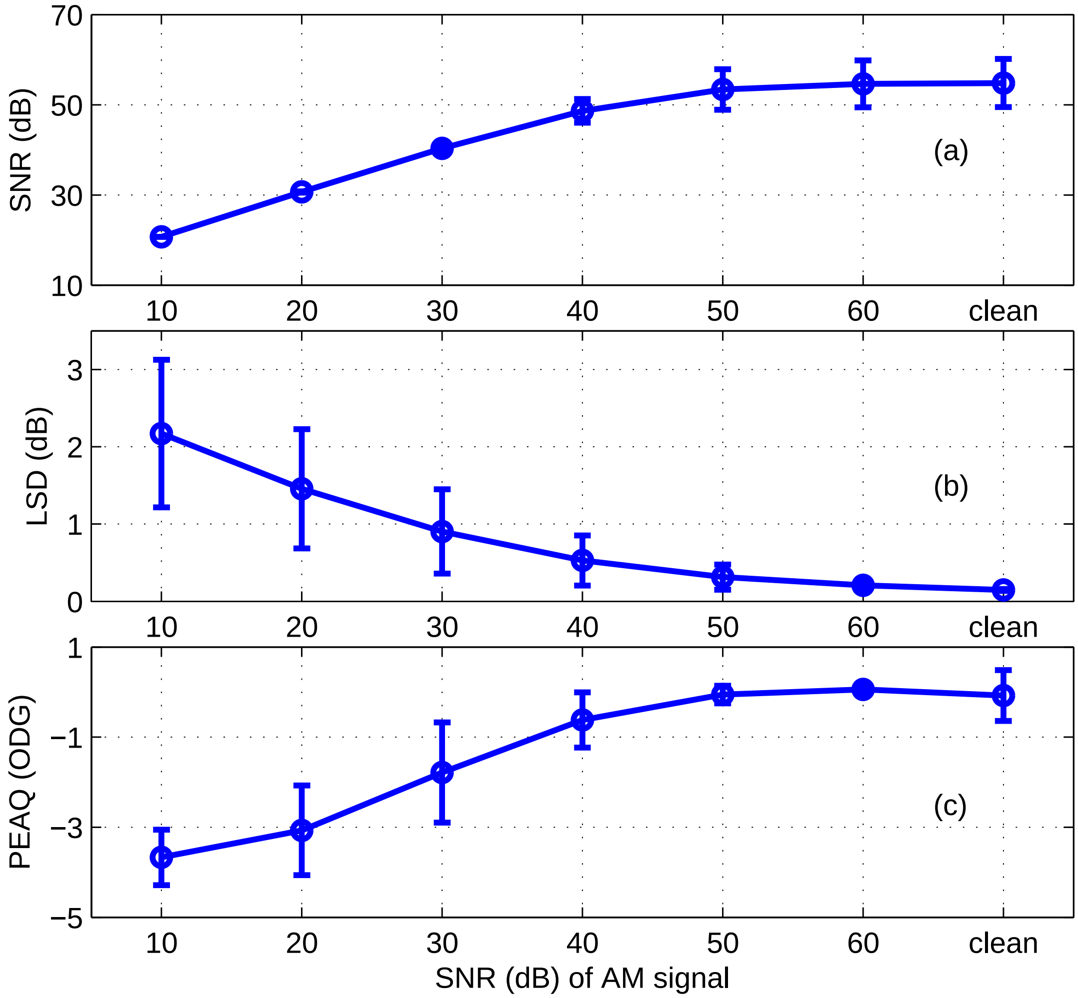


Figure 5.6: Results of Objective evaluations of Extracted Watermarked signals against External noise.

results for the objective tests of extracted original and extracted watermarked signals. The horizontal axis indicates the SNR of the double-modulated signal. The extracted signals were most distorted when the noise level was high (SNR $<$ 30 dB). However, when the noise level decreased (SNR $\geq$ 30 dB), the SNRs, LSDs, and PEAQs of $\hat{x}(n)$ and $\hat{y}(n)$ were significantly improved (40 dB, 0.9 dB, and -1.8 ODG). The bit-detection rate for high-level noise was less than 98.2% and for low-level noise (SNR $\geq$ 30 dB) it was greater than 99.2%. These results indicated that the proposed scheme can robustly extract signals from a double-modulated signal that is affected by low-level noise.

Chapter 6

An Improvement on Embedding capacity

We employed the method of digital-audio watermarking based on cochlear delay (CD) characteristics to embed watermarks into the original signals in the proposed data hiding scheme. The CD-based method satisfied the requirements of inaudibility, robustness, and confidentiality. It has the embedding capacity about 1 kbps. However, if it could have much higher capacity of watermarks, the proposed data hiding scheme could transmit much more data in AM signals. This section present an approach to watermarking in sub-bands to improve embedding capacity of the CD-based method.

6.1 Model concept

The CD-based method embeds data into the audio signal by enhancing its phase delay. The original signal is changed slightly by phase shifting according to CD. We may take advantage of this to apply the CD-based method to sub-band signals. We investigated the feasibility of applying the CD-based method of watermarking in sub-bands to increase embedding capacity. We took into consideration whether the watermarks could be distorted by using decomposition and synthesis processes with a filterbank. We also took into account the inaudibility of watermarked signals in the improved method. This approach offers a possibility of increasing embedding capacity with the proposed method as multiples of the number of sub-bands.

The improved method was based on the concept of sub-band coding. The signals are decomposed into a number of sub-band signals and then each sub-band signal is processed independently. The proposed method of watermarking based CD embeds data by enhancing the phases of the original signal according to CD. A bit data (e.g. '1') is embedded as the difference between the phases of the original signal and those of the watermarked signal in a frame. The phases of all frequency components are enhanced to represent a embedded bit '0' or '1'. The improved method enhances the phases of a number of consecutive frequency components (a sub-band) to embed a bit data. Thus, a number of bits could be embedded by enhancing the phases of sub-band frequency

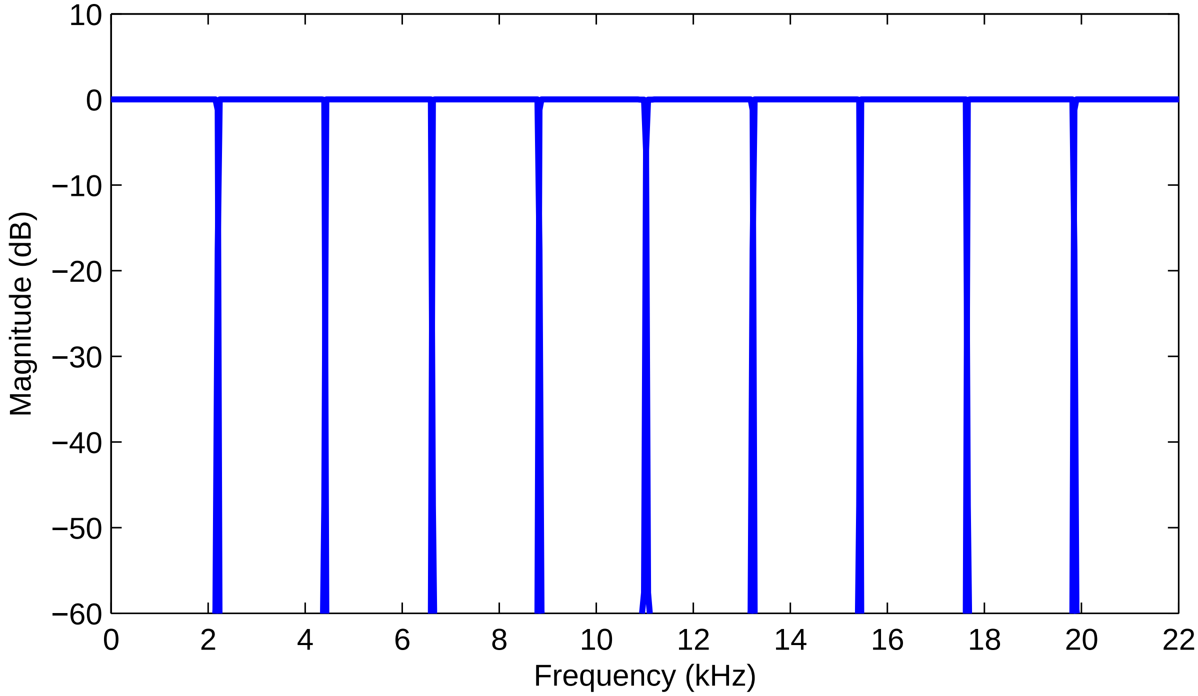


Figure 6.1: Frequency characteristics of CB-BPFB (10 channels).

components in a frame. Figures 6.4 and 6.5 have block diagrams for the improved method. A filterbank is used to decompose the original signal into sub-band signals. The watermark is embedded into sub-bands by using the CD-based method. The watermarked signal is a combination of watermarked sub-band signals. The embedding capacity of the method could be increased as multiples of the number of sub-bands.

6.2 Implementation

6.2.1 Filterbank

Filterbank is an array of band-pass filters that decomposes the signal into sub-band signals. We used two kinds of constant-bandwidth (CB) filterbanks in this study. The sub-band signals were resynthesized into a signal by using inverse processing in the filterbank after watermarking. The first filterbank was a conventional band-pass filterbank (BPFB) in which each filter was a finite impulse response (FIR) band-pass filter. A band-pass filter is usually designed by convoluting the filter kernels (impulse response function) of a low-pass filter and a high-pass filter. An FIR band-pass filter has the transfer function:

$$H(z) = d_0 + d_1 z^{-1} + d_2 z^{-2} + ... + d_N z^{-N}, \tag{6.1}$$

where N is the order of the filter. The filterbanks were designed with the window method [25]. If $w(n)$ denotes a window, where $0 \leq n \leq N$, and the impulse response of the ideal

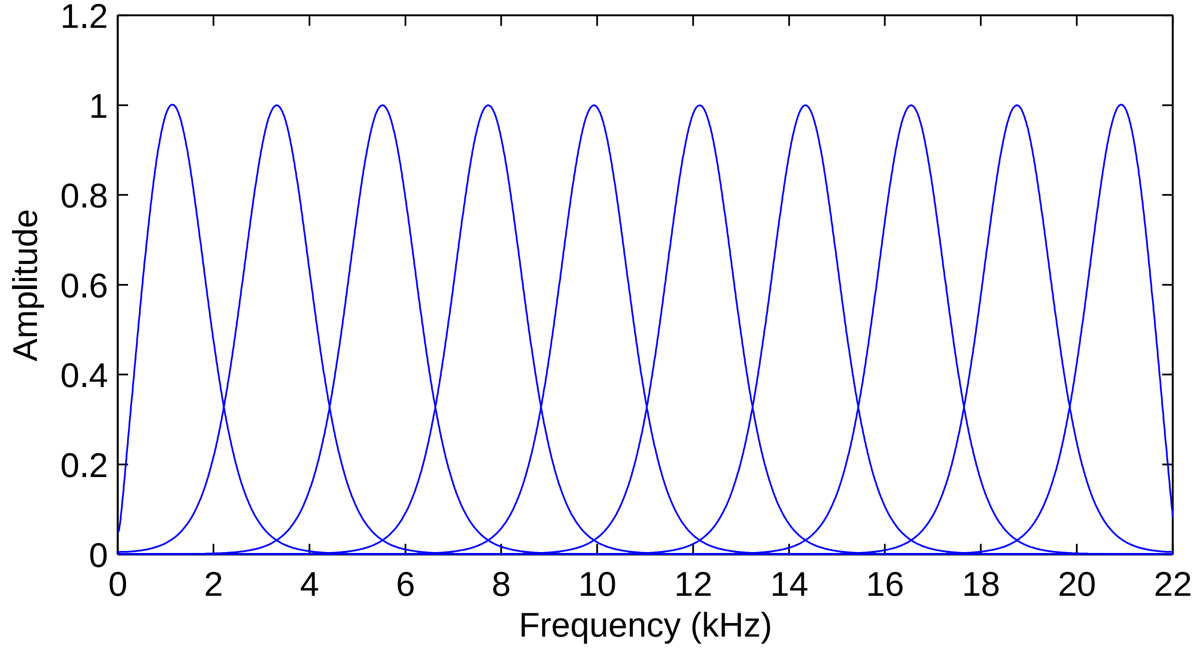


Figure 6.2: Frequency characteristics of CB-GTFB (10 channels).

filter is $h(n)$, then the filter coefficients are given by

$$d(n) = w(n)h(n), \qquad 0 \leq n \leq N. \tag{6.2}$$

The window function is Hamming window. The filtering process was implemented by fast Fourier transform. The frequency characteristics of CB-BPFB are shown in Fig. 6.1.

The second kind of filterbank was a Gammatone filterbank (GTFB). Its impulse response is given by:

$$gt(t) = At^{M-1} \exp(-2\pi b_f \mathrm{ERB}(f_0)t) \cos(2\pi f_0 t), t \geq 0, \tag{6.3}$$

where A, b_f, and M are parameters of the Gammatone filter, and $At^{M-1} \exp(-2\pi b_f \mathrm{ERB}(f_0)t)$ is the amplitude term represented by the Gamma distribution, f_0 is the center frequency, and $ERB(f_b)$ in this paper was set to CB. The frequency characteristics of GTFB are shown in Fig. 6.2.

Figure 6.3 shows the sound quality of re-synthesized signals by CB-BPFB and CB-GTFB with 10, 20, 50, 100 and 200 channels. The sound quality of re-synthesized signals when using CB-BPFB decreased as the number of channels increased. In contrast, the sound quality of CB-GTFB increased as the number of channels increased.

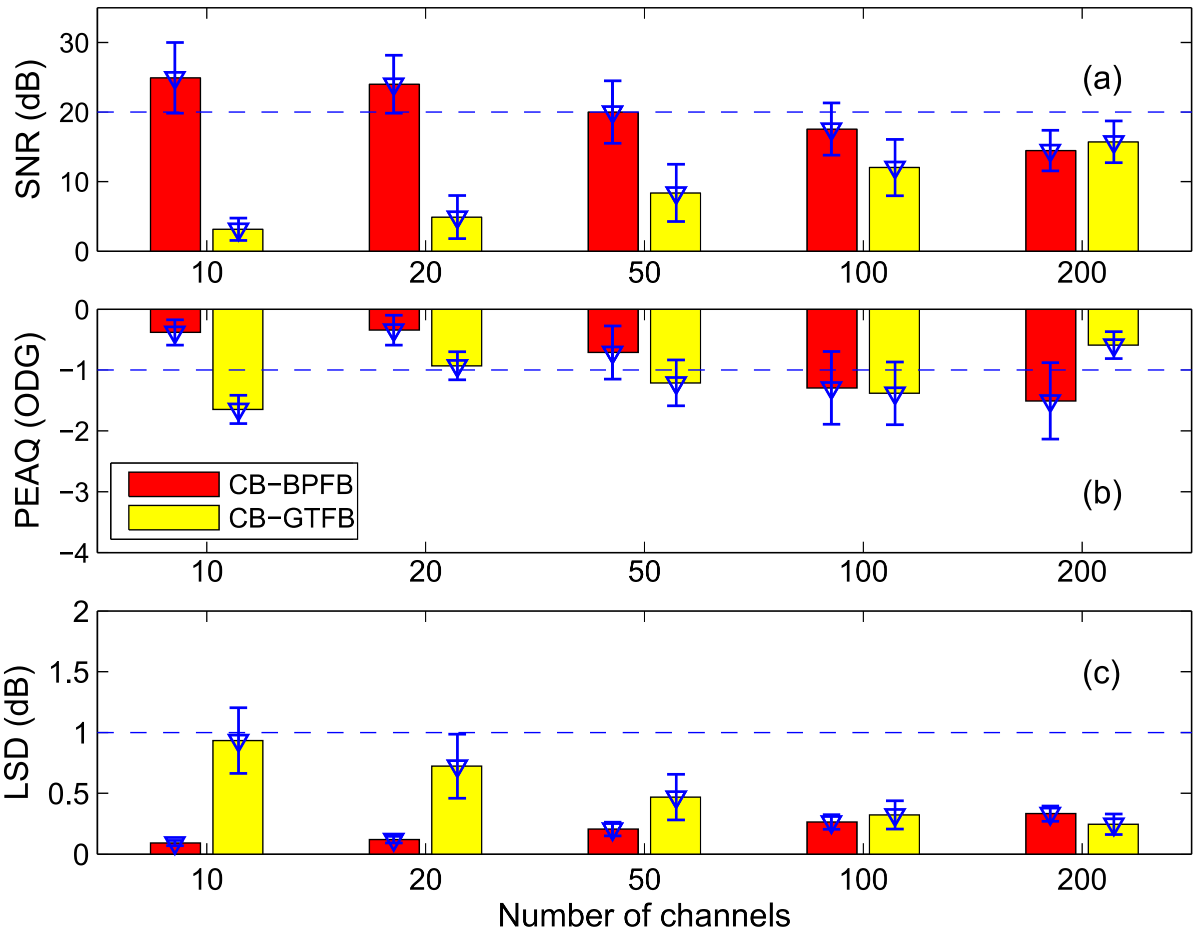


Figure 6.3: Sound quality of Re-synthesis signal. (a) SNR, (b) PEAQ and (c) LSD

6.2.2 Embedding in Sub-bands

There is a block diagram of the data embedding process in sub-bands in Fig. 6.4. The watermark is embedded into the original signal in four steps:

Step 1: Original signal $x(n)$ is decomposed into m sub-band signals $x_i(n)(i = 1, 2, ..., m)$ using the filterbank. The bandwidths of the sub-bands are the same, i.e., CB.

Step 2: The watermark in the binary string, $s(k)$, is alternately and equally split into m binary strings $s_i(k)(i = 1, 2, ..., m)$ to be embedded in m sub-band signals.

Step 3: Each sub-band signal $x_i(n)$ is embedded with watermarks $s_i(k)$ by using the CD-based method. The outputs obtained are the watermarked sub-band signals $y_i(n), (i = 1, 2, ..., m)$.

Step 4: The watermarked signal is created by combining watermarked sub-band signals $y_i(n)(i = 1, 2, ..., m)$ using inverse processing in the filterbank used in Step 1.

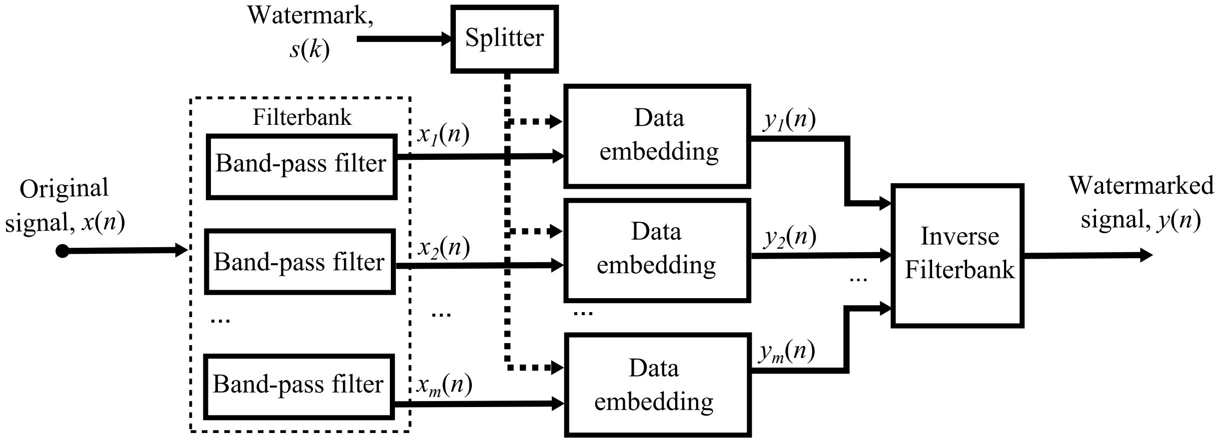


Figure 6.4: Block diagram for Embedding data in sub-bands.

6.2.3 Watermark Detection

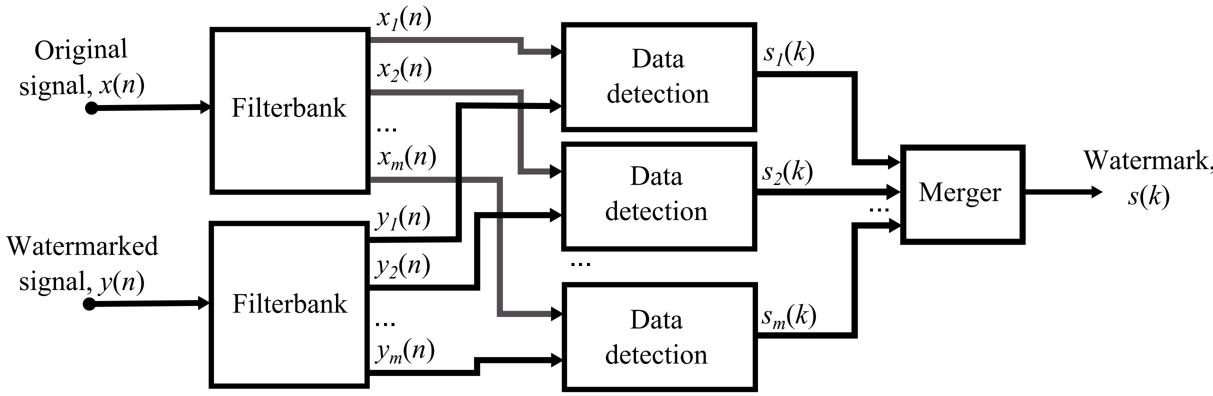


Figure 6.5: Block diagram for Data detection in sub-bands.

Figure 6.5 has a block diagram of the detection process in sub-bands. Watermarks are detected in each sub-band and are then merged to obtain the whole watermark. The details on this process are as follows:

Step 1: Both the original and the watermarked signals, $x(n)$ and $y(n)$, are first decomposed into sub-band pairs $x_i(n)$ and $y_i(n)(i = 1, 2, ..., m)$ by using the same filterbank in the embedding process.

Step 2: Each sub-band pair, $x_i(n)$ and $y_i(n)$, are processed by the detection process of the CD-method to detect watermarks $s_i(k), (i = 1, 2, ..., m)$.

Step 3: Watermark $s(k)$ is obtained by merging $s_i(k)(i = 1, 2, ..., m)$ according to the way the watermark is split in the embedding process.

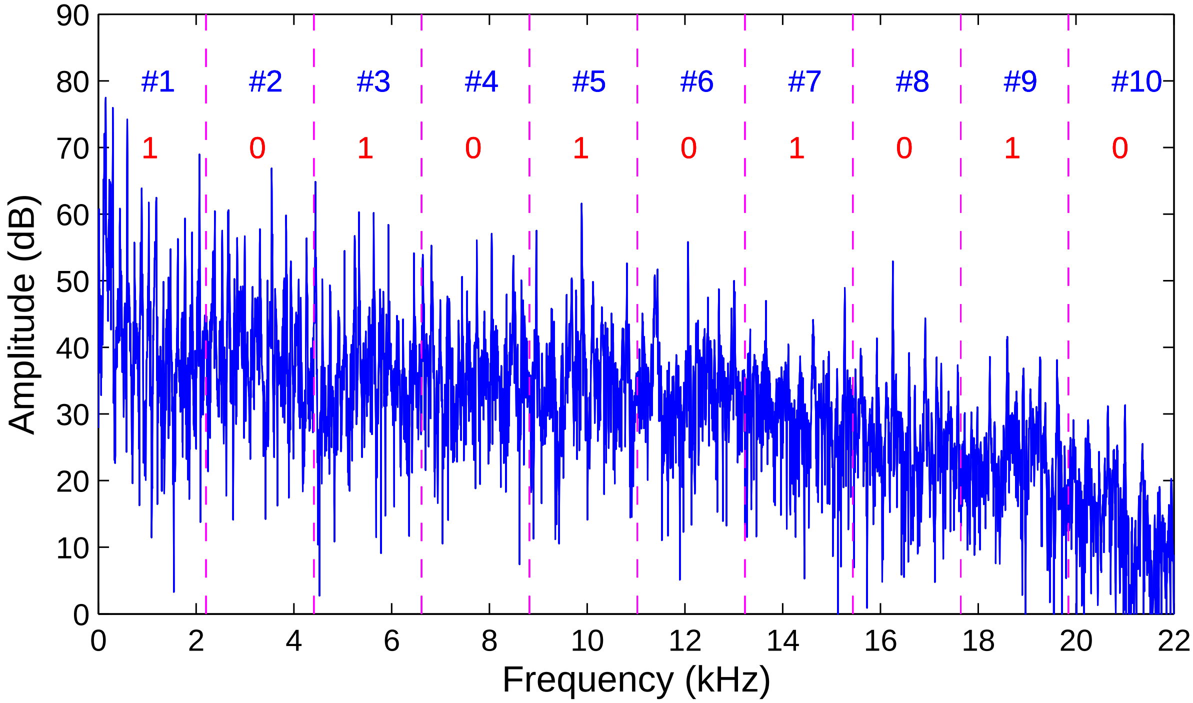


Figure 6.6: Frequency Response of Original signal.

6.2.4 Example

This section presents an example of embedding and detecting the watermark by the improved method. The number of channels is 10. The first frame of the original signal is demonstrated in this example. The watermark is the binary string of ten bits ‘1010101010’. The original signal has a sampling frequency of 44.1 kHz.

Figure 6.6 shows the frequency response of the first frame of the original signal. Firstly, the original signal was decomposed into ten sub-bands. The sub-bands were numbered from 1 to 10 as shown in Fig. 6.6. The frequency responses of the sub-bands (1-6) were shown in Fig. 6.7. The watermark ‘1010101010’ was alternately and equally distributed into ten sub-bands as shown by red-colored character in Fig. 6.6. The watermark was then embedded into each sub-band by using the proposed method of watermarking based on CD. Finally, the watermarked signal was created by synthesizing the watermarked sub-bands. The watermarked signal then contained ten bits of watermark ‘1010101010’.

The phase difference $\Phi(\omega)$ in each sub-band between the original signal and the watermarked signal was compared with the phases $\arg H_0(\omega)$ and $\arg H_1(\omega)$ of two filters $H_0(\omega)$ and $H_1(\omega)$ to detect embedded bit ‘0’ or ‘1’. Figure 6.8 shows the phase difference $\Phi(\omega)$ and the phases $\arg H_0(\omega)$ and $\arg H_1(\omega)$ of two filters $H_0(\omega)$ and $H_1(\omega)$. The binary characters in red color were the detected bits. The first sub-band shows that $\Phi(\omega)$ matched with $\arg H_1(\omega)$. Therefore, the bit ‘1’ was detected in the first sub-band. The second sub-band shows that $\Phi(\omega)$ matched with $\arg H_0(\omega)$ and then the bit ‘0’ was detected. Similarly, the remaining eight bits ‘10101010’ were detected by matching $\Phi(\omega)$

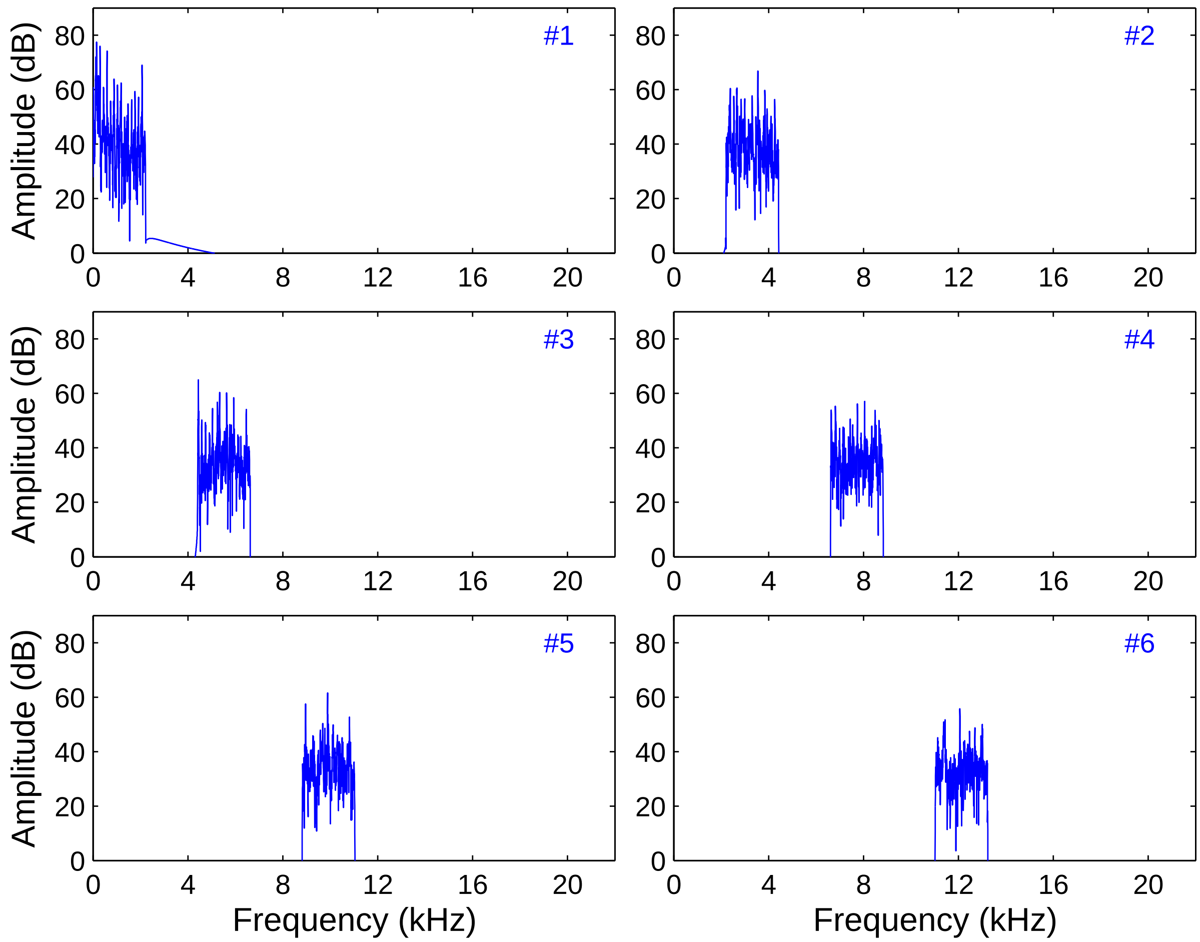


Figure 6.7: Frequency Response of a few Sub-bands of Original signal (1-6).

with $\arg H_0(\omega)$ or $\arg H_1(\omega)$.

6.3 Evaluation

We carried out simulations to objectively evaluate the sound quality of watermarked signals and bit-detection rate in the improved method. We also compared the improved method with the proposed method. We used all 102 music tracks in the real world computing (RWC) music database [22] as the original signals, which were the same as those in the proposed method. These music tracks had a sampling frequency of 44.1 kHz, were 16 bits, and had two channels. The same watermarks 'AIS-lab. Japan Advanced Institute of Science and Technology' were embedded into the original audio signal by using the improved methods. The sub-band bit rate defined as the bit rate of embedded data in sub-bands ranged from 4 to 1024 bps and there were 10, 20, 50, 100 and 200 sub-bands.

We used log spectral distortion (LSD) [23] and perceptual evaluation of audio quality (PEAQ) [24] to evaluate the sound quality of the watermarked signals. LSD was used

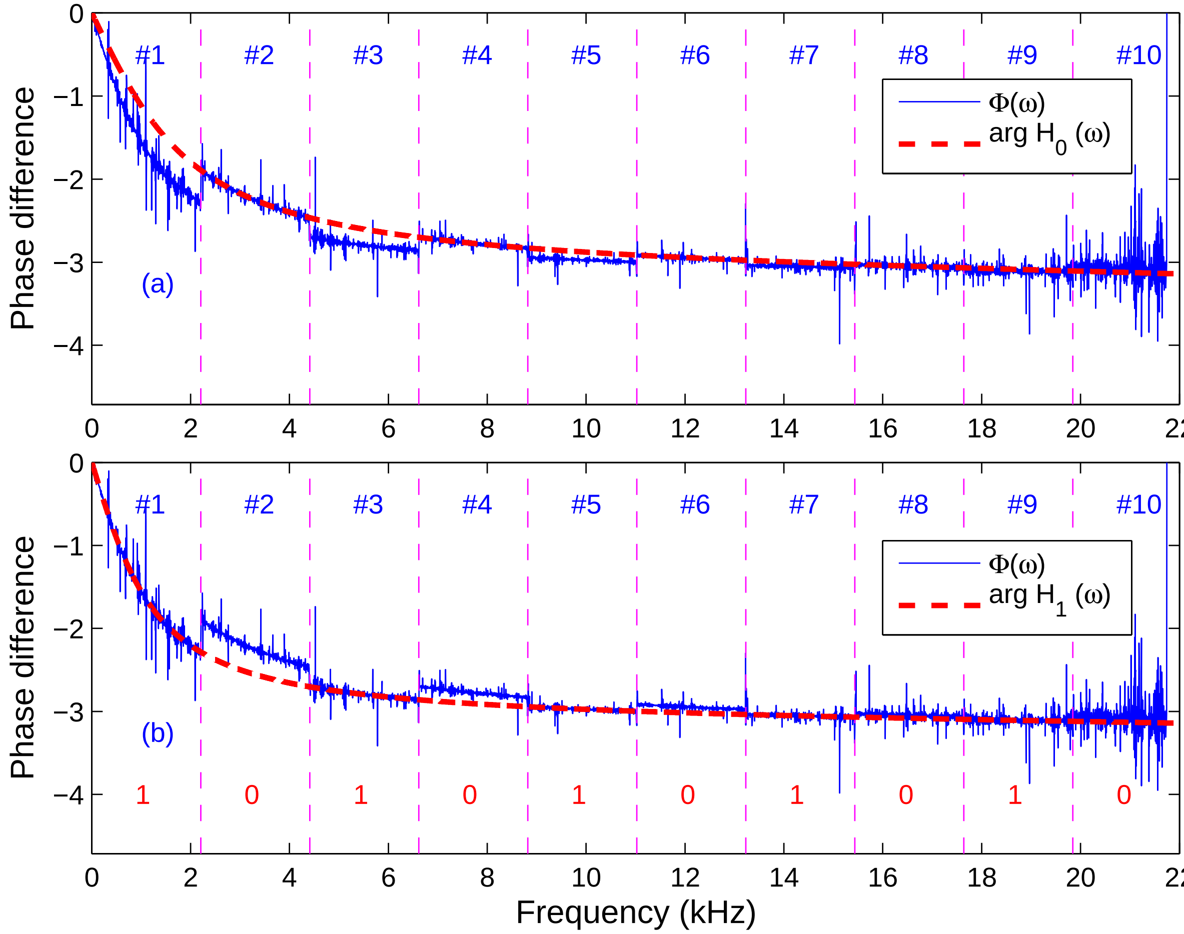


Figure 6.8: Comparison with Phase difference between Original signal and Watermarked signal with Phase characteristics of two filters $H_0(\omega)$ and $H_1(\omega)$.

to measure the distance or distortion between two spectra in which a lower value for LSD indicates a better result and vice versa. PEAQ was used to measure degradation in audio quality according to the objective difference grade (ODG) ranging from -4 (the worst quality) to 0 (the best quality). The evaluation thresholds for the LSD and PEAQ criteria corresponded to 1 dB and -1 ODG, which were the same as those in Unoki and Hamada [5] and Unoki et al. [6]. The subband bit-detection rate was referred to as the bit-detection rate of the CD method in sub-bands, which were independently calculated for each $s_i(k)$. The total bit-detection rates were calculated for the whole watermark $s(k)$ after the watermarks in sub-bands $s_i(k), (i = 1, 2, ..., m)$ were merged.

The average LSD and PEAQ of the watermarked signals were approximately 0.60 dB and -0.35 ODG under conditions where the sub-band bit rate was 4 bps and there were 20 channels. The averages of sub-band bit-detection rates are almost 100% and the standard deviations are around 0.02%. These results suggested that the watermarks that were embedded into sub-bands by using the CD-based method could survive through synthesis

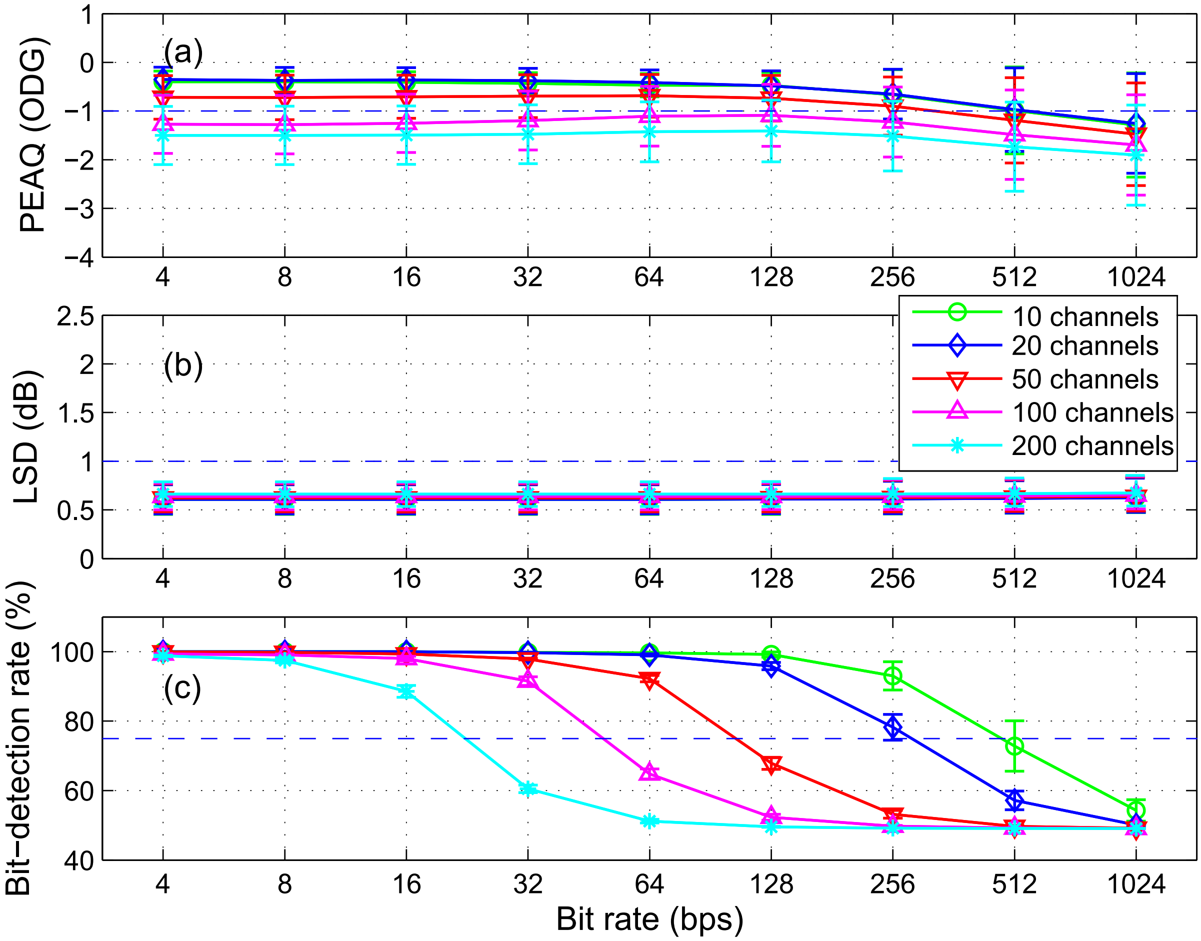


Figure 6.9: Objective evaluations of Improved method with CB-BPFB: (a) PEAQ, (b) LSD, and (c) PEAQ.

and decomposition processes. The results also ensured that the sound quality of the watermarked signals was not distorted by the improved method.

Figure 6.9 has the results for our objective evaluations of the improved method using CB-BPFB with five conditions of the number of channels. The PEAQs of the improved method with 10, 20 and 50 channels were mostly higher than the evaluation threshold when the bit rates were less than 256 bps. The PEAQs with 100 and 200 channels were all worse than the evaluation threshold. The LSDs, on the other hand, were all better than the evaluation threshold. The sub-band bit-detection rates with the improved method were all greater than 92% under the conditions of the number of channels 10, 20 and 50 and bit rates less than 64 bps. These results suggested that the improved method with 50 channels and at a bit rate of 64 bps could be used to embed watermarks into digital-audio. The improved method could satisfy the confidentiality requirement under these conditions. The total bit rate could be increased to 3.2 kbps.

The results of objective evaluations of the improved method using CB-GTFB are shown

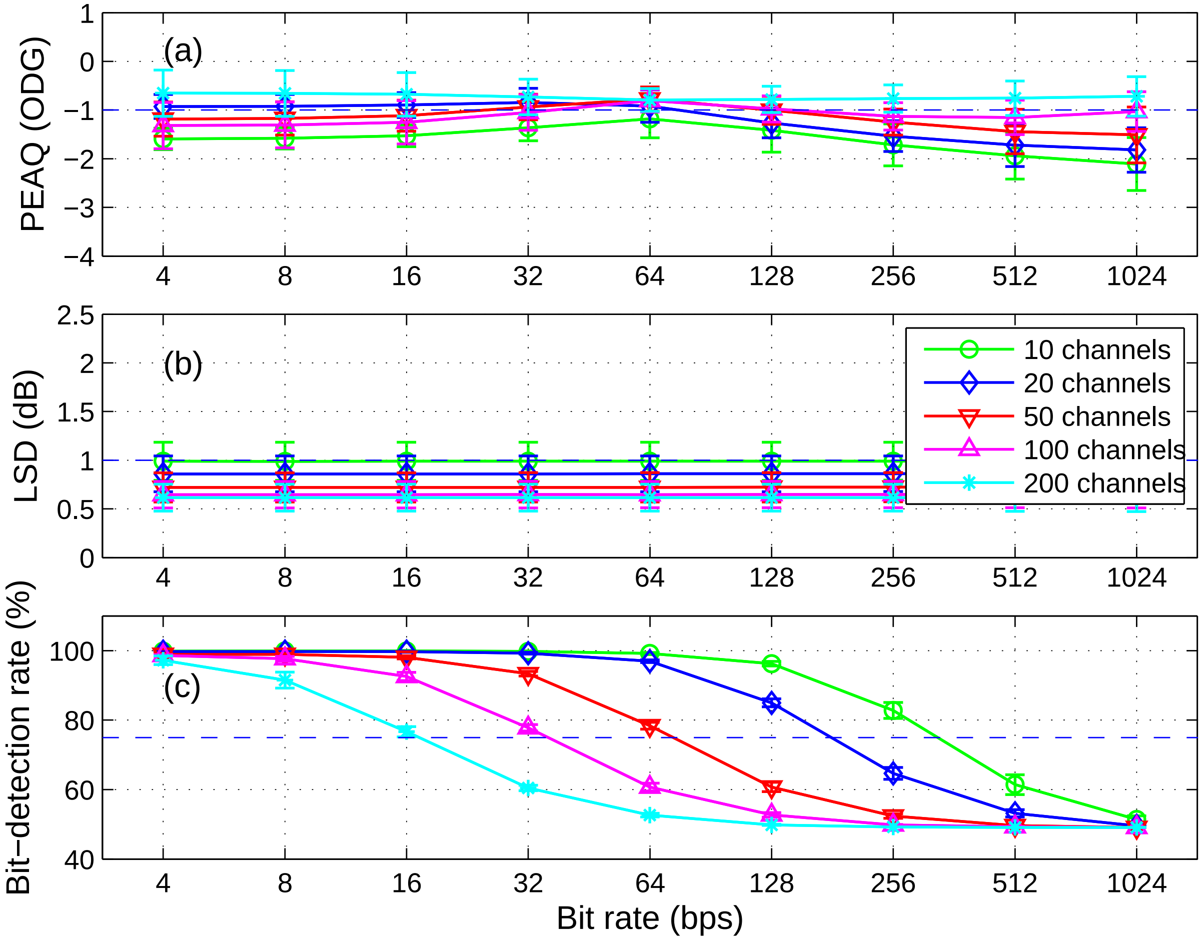


Figure 6.10: Objective evaluations of Improved method with CB-GTFB: (a) PEAQ, (b) LSD, and (c) PEAQ.

in Fig. 6.10. The PEAQs with 200 channels were all higher than −1 ODG, while the PEAQs with 10, 20, 50, and 100 were slightly lower than the evaluation threshold. The results of LSDs were almost better than the evaluation criteria. The sub-band bit detection rates with 10 and 20 were better that those with other numbers of channels. The bit-detection rate of the improved method with CB-GTFB at the bit rate of 128 bps and the under condition of 20 channels was 85%. It revealed that the improved method could be used to embed data into digital-audio at the total bit rate of 2.56 kbps.

We compared the improved method (CB-BPFC) with the proposed method by carrying out objective evaluations (PEAQ, LSD and a bit-detection test). Figure 6.11 have the results for our objective evaluations of the proposed and improved methods with 50 channels. The sound quality of the proposed method is better than the improved method when the bit rates were less than 512 bps. The total bit detection rates of the improved method, on the other hand, were higher than the proposed method under all conditions of bit rate. These results showed that the embedding capacity of the improved method

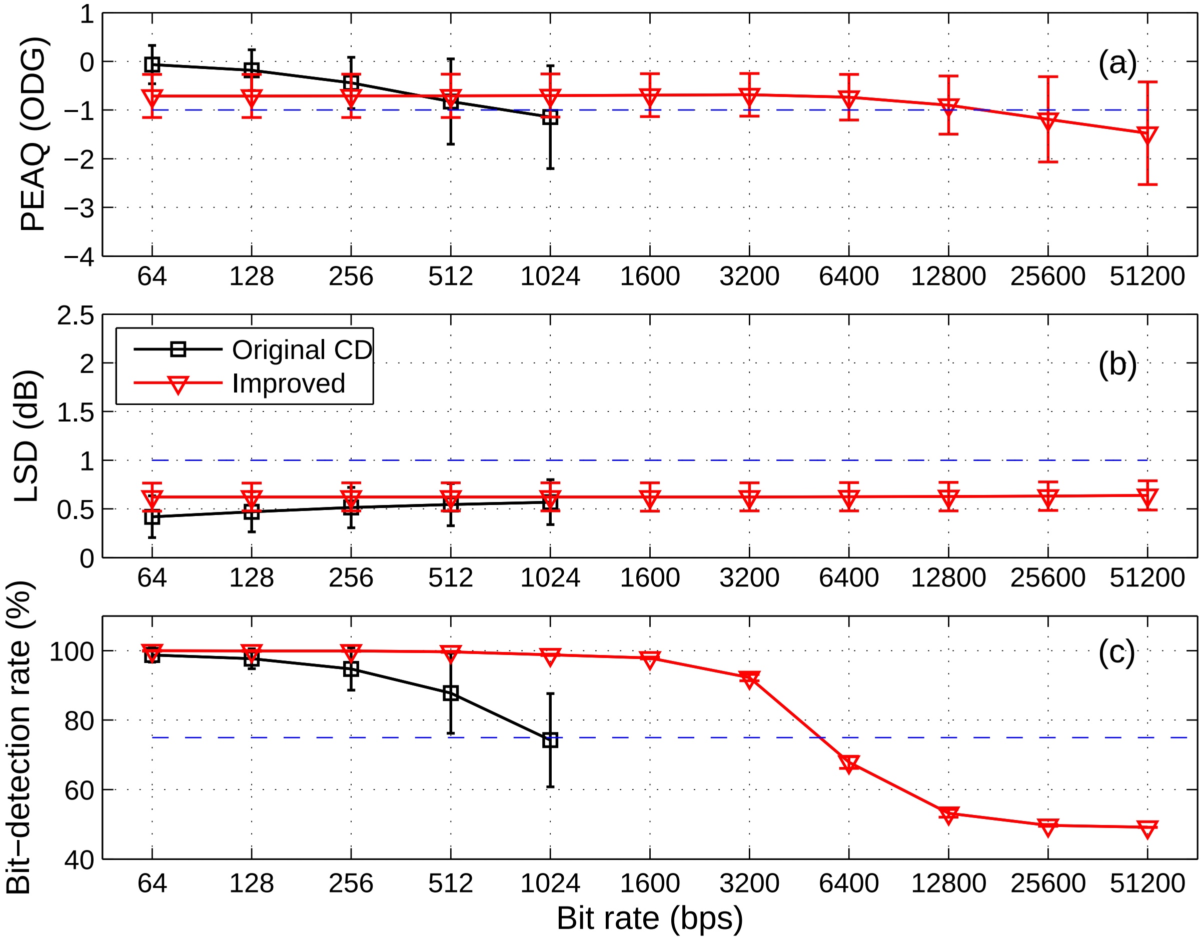


Figure 6.11: Comparative evaluations of Improved method with Proposed CD-method.

could be much higher increased.

6.4 Related works

There were also other approaches to increasing embedding capacity based on designing CD filters with parallel and cascade architectures [26, 27, 28]. These approaches proposed an optimal architecture of CD filters to increase embedding capacity and to improve sound quality of watermarked signals as well. The authors controlled phase delays of original sound signals by several CD filters (2^L in [26], or L in [27], where L is the number of bits per frame) to embed more bits while the original CD method controlled only two phase delays to embed a bit per frame. The approach using the parallel architecture of CD filter could increase the embedding capacity to 384 bps while the embedding capacity of the approach using the cascade architecture amounted to over 384 bps.

This thesis also proposed an approach to improving embedding capacity of CD method. This approach was based on the concept of sub-band coding. The original signal was decomposed into the sub-band signals and then these sub-band signals were embedded with watermarks. The differences between this approach and above approaches are that this approach controlled only two CD filters and it simultaneously controlled multiple frequency bands to embed more bits. We designed efficient filterbanks to decompose the audio signals into sub-band signals and to synthesize the watermarked sub-band signals after the processes of watermarking to ensure that sound quality of the watermarked signals could be kept at high level. The embedding capacity of our proposed approach could be increased to 3.2 kbps and sound quality were higher than evaluation criteria.

Chapter 7

Conclusion

7.1 Summary

Motivated from demands of an efficient emergence alert system as well as hiding data into AM radio signals, this thesis proposed a data hiding scheme for AM radio broadcasting system. The proposed scheme can basically be applied to transmit additional data besides audio content in AM radio signals. Playing an important role in this scheme, the method of audio watermarking based on CD was used to embed inaudible data into audio content before the audio is further processed to be far transmitted over a radio link. Since, the employed method of watermarking, CD-based method, is a non-blind watermarking scheme, which requires a double transmission bandwidth, we overcame this issue by proposing novel modulation and demodulation techniques namely double-modulation and double-demodulation, respectively. The double-modulation embeds both the original and watermarked signals as LSB and USB, respectively, with the same carrier signal. After that, receivers can use the double-demodulation to extract these signals and then retrieve the embedded data after that. Although the proposed scheme generates AM radio signals having different sidebands in their frequency spectra, standard receivers are still able to extract audio content from the AM signals using standard DSB-WC demodulation techniques.

We conducted simulations on the computer to evaluate the effectiveness of the proposed scheme. We firstly confirmed the ability of the processes of double-modulation and double-demodulation by measuring sound quality of the audio signals that were extracted from the AM signals. The results of SNR, LSD, and PEAQ showed that these signals could be properly extracted. The average values of SNRs, LSDs, PEAQs were about 54 dB, 0.15 dB, and -0.08 ODG, respectively, which are all better than the evaluation thresholds. Secondly, we evaluated the inaudibility and the accuracy of CD-method that was used to detect embedded data from the extracted original and watermarked signals. The inaudibility was investigated by calculating sound quality of the extracted watermarked signals in reference to the extracted original signals. The results of LSDs and PEAQs reconfirmed that CD-method could be applied to the proposed scheme to embed inaudible data. The results of bit-detection rate revealed that the accuracy of the watermark detection process could be kept as high as that in the CD-based watermarking system for digital-audio. Thirdly, we evaluated sound quality of the extracted signals from the AM signals using the standard DSB-WC demodulator to check the compatibility of the proposed data hiding scheme with the standard system. It was revealed that the standard receiver could acquire audio content at a highly reasonable level of distortion. Finally, we paid attention on the distortion of the extracted signals which was caused by external white noise. We measured the SNRs, LSDs, and PEAQs of the extracted signals under several conditions of white noise in which the SNRs of AM signals ranged from 10 to 60 dB. These scores showed that sound quality of the extracted signals were degraded as the levels of the external noise were relatively high (SNR $<$ 30 dB).

The other task of this thesis is to improve embedding capacity of CD-based method of audio watermarking. It is based on the concept of sub-band coding to embed the data into the sub-band signals instead of the undecomposed signal. The objective of this task is to obtain a method of audio watermarking having high embedding capacity so that it can be used to transmit much more information or to convey the media data of large size such as photos and speeches. We also carried out the experiments to objectively evaluate the improved method and to compare it with the original CD-based method. Sound quality (LSD, PEAQ) and bit-detection rate of the improved method were assessed under different conditions of the number of sub-bands. The sound quality of the watermarked signals in the improved method was mostly lower than those in the original method. However, they were greater than the evaluation criteria and the embedding capacity could be improved to the bit rate of 3.2 kbps.

7.2 Contribution

In summary, it can be said that the proposed data hiding scheme well satisfy the four requirements mentioned in Section 3.2 and it makes two important contributions as described below.

This thesis aims to propose a data hiding scheme for AM radio broadcasting system with high capacity of embedded data. A novel data hiding scheme was proposed and implemented successfully. The proposed scheme has low-level compatibility with a vast majority of AM radio receivers that use the standard demodulation technique to extract audio signals. The proposed scheme has possibility of acting as a data-hiding-enabled AM-radio-broadcasting-system. It can be used as an emergency alert system in cases of disasters because the advantage that AM radio is simple and popularly used nowadays. Emergency messages could be hidden into AM radio signals to announce the disaster to people. The alert message could be announced to attract the attention of listeners while the regular radio program is broadcast simultaneously. Moreover, the proposed scheme can broadcast additional information of radio service such as radio station information, advertisement, etc. These information could be used to improve AM radio service, for example, offering the function of switching over to a digital or FM version of the same service [3]. It moves towards reaching the advanced radio service that “The radio knows what it is listening to” [2].

With regard to embedding capacity, this thesis proposed an approach to improve CD-based method of audio watermarking. The improved method was based on the concept of sub-band coding in order to embed watermark into sub-band signals. The embedding capacity could be increased by multiple of the number of sub-bands. The experimental results confirmed that the embedding capacity was much increased in comparison with the original method. The improved method could be used to embed not only text message but also the media content of large size into audio signals as well as AM radio signals.

7.3 Future work

In the proposed scheme, double-modulation was implemented in frequency domain to combine the spectra of the original and watermarked signals so that the modulated signals can carry both two signals as LSB and USB. Similarly, double-demodulation process was also implemented in frequency domain to separate the spectra to extract two signals from LSB and USB. Double-modulation and double-demodulation algorithms must wait for a full frame of the signal coming to process. The frame length of the signal needs to be long enough so that the signals are correctly processed. Therefore, there is a period of time delay in retrieving the audio content and embedded data. To deal with this problem, the proposed double-modulation and double-demodulation could be implemented in time domain. Essentially, the required frame length of the signal is much shorter when the signal is processed in time domain. As a result, the output of the double-demodulator could be in quick response to the incoming AM signal.

Regarding to the improved method of watermarking based on concept of sub-band coding, only the properties of inaudibility and embedding capacity have been investigated by carrying out the experiments. Robustness of the improved method needs to be considered to confirm the improvement of this method in comparison with the original method. It offers a room for further investigation and study in this research.

Bibliography

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

[2] A. Murphy and R. Poole. AM signalling system: AMSS - does your radio know what it is listening to? Technical report, BBC Research & Development, UK, 2006.

[3] L. Cornell. The AM signalling system (AMSS). Technical report, Principal Systems Architect, BBC New Media & Technology, U.K., 2007.

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

[5] M. Unoki and D. Hamada. Method of digital-audio watermarking based on cochlear delay characteristics. *Int. J. Inn. Com. Inf., and Cont.*, 6(3(B)):1325–1346, 2010.

[6] M. Unoki, K. Imabeppu, D. Hamada, A. Haniu, and R. Miyauchi. Embedding limitations with digital-audio watermarking method based on cochlear delay characteristics. *J. Information Hiding and Multimedia Signal Processing*, 2(1):1–23, 2011.

[7] L. Boney, A.H. Tewfik, and K.N. Hamdy. Digital watermarks for audio signals. In *Multimedia Computing and Systems, 1996., Proceedings of the Third IEEE International Conference on*, pages 473 –480, jun 1996.

[8] M. Arnold. Audio watermarking: features, applications and algorithms. In *Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on*, volume 2, pages 1013 –1016 vol.2, 2000.

[9] S. Craver, N. Memon, B.-L. Yeo, and M.M. Yeung. Resolving rightful ownerships with invisible watermarking techniques: limitations, attacks, and implications. *Selected Areas in Communications, IEEE Journal on*, 16(4):573 –586, may 1998.

[10] I.J. Cox, J. Kilian, F.T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. *Image Processing, IEEE Transactions on*, 6(12):1673 –1687, dec 1997.

[11] A. Lu D. Gruhl and W. Bender. Echo hiding. In *Proceedings of Information Hiding 1st Workshop*, pages 295–315, 1996.

[12] R. Nishimura and Y. Suzuki. Audio watermark based on periodical phase shift. *J. Acoust. Soc. Jpn.*, 60(5):269–272, 2004.

[13] A. Takahashi, R. Nishimura, and Y. Suzuki. Multiple watermarks for stereo audio signals using phase-modulation techniques. *IEEE Trans. Signal Processing*, 53(2):806–815, 2005.

[14] B. C. J. More. *An introduction to the psychology of hearing.* Academic Press, 2003.

[15] M. Unoki and D. Hamada. Audio watermarking method based on the cochlear delay characteristics. In *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2008 Sixth International Conference on*, pages 616–619, aug. 2008.

[16] K. Kondo. *Multimedia Information Hiding Technologies and Methodologies for Controlling Data.* IGI Global, 2012.

[17] E. Aiba and M. Tsuzaki. Perceptual judgement in synchronization of two complex tones: Relation to the cochlear delays. *Acoust. Sci. & Tech.*, 28(5):357–359, 2007.

[18] Eriko Aiba, Minoru Tsuzaki, Satomi Tanaka, and Masashi Unoki. Judgment of perceptual synchrony between two pulses and verification of its relation to cochlear delay by an auditory model,. *Japanese Psychological Research*, 50(4):204–213, 2008.

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

[20] L. Couch. *Digital and Analog Communication Systems.* Upper Saddle Riverm NJ: Prentice Hall, 2007.

[21] J. Proakis. *Fundamentals of Communication Systems.* Upper Saddle Riverm NJ: Prentice Hall, 2005.

[22] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka. Music genre database and musical instrument sound database. In *Proc. ISMIR 2003*, pages 229–230, 2003.

[23] Jr. Gray, A. and J. Markel. Distance measures for speech processing. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24(5):380–391, 1976.

[24] P. Kabal. An examination and interpretation of ITU-RBS.1387: Perceptual evaluation of audio quality. TSP Lab technical report, Dept, Elect. Comp. Eng.

[25] *Programs for Digital Signal Processing.* IEEE Press, New York, 1979. Algorithm 5.2.

[26] Kuniaki Imabeppu, Daiki Hamada, and Masashi Unoki. Embedding limitations with audio-watermarking method based on cochleardelay characteristics. In *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2009 Sixth International Conference on*, pages 82–85, 2009.

[27] Masashi Unoki, Toshizo Kosugi, Atsushi Haniu, and Ryota Miyauchi. Design of IIR all-pass filter based on cochlear delay to reduce embedding limitations. In *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on*, pages 526–529, 2010.

[28] Toshizo Kosugi, Atsushi Haniu, Ryota Miyauchi, Masashi Unoki, and Masato Akagi. Study on suitable-architecture of IIR all-pass filter for digital-audio watermarking technique based on cochlear-delay characteristic. In *2011 International Workshop on Nonlinear Circuits, Communication and Signal Processing*, pages 135–138, 2011.

Publication

1. Nhut Minh Ngo, Shengbei Wang, and Masashi Unoki, "Method of Digital-audio Watermarking Based on Cochlear Delay in Sub-bands," Proc. ITC-CSCC2012, July 2012.

2. Nhut Minh Ngo, Masashi Unoki, Ryota Miyauchi, and Yiti Suzuki, "Data-hiding scheme for digital-audio in amplitude modulation domain," Proc. IIHMSP2012, pp. 114-117, July 2012 [Best Paper Award].

3. Nhut Minh Ngo, Shengbei Wang, and Masashi Unoki, "Method of digital-audio watermarking based on cochlear delay in sub-bands," IEICE Tech. Rep. EA2012-57, pp. 19-24, August 2012.

4. Nhut Minh Ngo, Masashi Unoki, Ryota Miyauchi, and Yiti Suzuki, "Study on data hiding scheme for digital-audio in amplitude modulation domain," Proc. 2012 ASJ Autumn meeting, 2-Q-b7, Nagano, Sept. 2012.

5. Shengbei Wang, Nhut Minh Ngo, and Masashi Unoki, "Digital-audio watermarking based on cochlear delay in sub-bands," Proc. 2012 ASJ Autumn meeting, Nagano, Sept. 2012.