공학박사학위논문

Audio Data Hiding for Acoustic Data Transmission in Reverberant Aerial Space

반향 환경에 강인한 음향 데이터 전송을 위한 오디오 정보 은닉 기법 연구

2014 년 2 월

서울대학교 대학원

전기·컴퓨터 공학부

조 기 호

공학박사학위논문

Audio Data Hiding for Acoustic Data Transmission in Reverberant Aerial Space

반향 환경에 강인한 음향 데이터 전송을 위한 오디오 정보 은닉 기법 연구

2014 년 2 월

서울대학교 대학원

전기·컴퓨터 공학부

조 기 호

Audio Data Hiding for Acoustic Data Transmission in Reverberant Aerial Space

반향 환경에 강인한 음향 데이터 전송을 위한 오디오 정보 은닉 기법 연구

지도교수 김 남 수

이 논문을 공학박사 학위논문으로 제출함

2013 년 11 월

서울대학교 대학원

전기·컴퓨터 공학부

조 기 호

조기호의 공학박사 학위논문을 인준함

2013 년 12 월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

Abstract

In this dissertation, audio data hiding methods suitable for acoustic data transmission are studied. Acoustic data transmission implies a technique which communicates data in short-range aerial space between a loudspeaker and a microphone. Audio data hiding method implies a technique that embeds message signals into audio such as music or speech. The audio signal with embedded message is played back by the loudspeaker at a transmitter and the signal is recorded by the microphone at a receiver without any additional communication devices. The data hiding methods for acoustic data transmission require a high level of robustness and data rate than those for other applications.

For one of the conventional methods, the acoustic orthogonal frequency division multiplexing (AOFDM) technique was developed as a reliable communication with reasonable bit rate. The conventional methods including AOFDM, however, are considered deficient in transmission performance or audio quality. To overcome this limitation, the modulated complex lapped transform (MCLT) is introduced in the second chapter of the dissertation. The system using MCLT does not produce blocking artifacts which may degrade the quality of the resulting data-embedded audio signal. Moreover, the interference among adjacent coefficients due to the overlap property is analyzed to take advantage of it for data embedding and extraction.

In the third chapter of the dissertation, a novel audio data hiding method for the acoustic data transmission using MCLT is proposed. In the proposed system, audio signal is transformed by the MCLT and the phases of the coefficients are modified to embed message based on the fact that human auditory perception is more sensitive to the variation in magnitude spectra. In the proposed method, the perceived quality

of the data-embedded audio signal can be kept almost similar to that of the original audio while transmitting data at several hundreds of bits per second (bps). The experimental results have shown that the audio quality and transmission performance of proposed system are better than those of the AOFDM based system. Moreover, several techniques have been found to further improve the performance of the proposed acoustic data transmission system which are listed as follows: incorporating a masking threshold (MM), clustering based decoding (CLS), and a spectral magnitude adjustment (SMA).

In the fourth chapter of the dissertation, an audio data hiding technique more suitable for acoustic data transmission in reverberant environments is proposed. In this approach, sophisticated techniques widely deployed in wireless communication is incorporated which can be summarized as follows: First, a proper range of MCLT length to cope with reverberant environments is analyzed based on the wireless communication theory. Second, a channel estimation technique based on the Wiener estimator to compensate the effect of channel is applied in conjunction with a suitable data packet structure. From the experimental result, the MCLT length longer than the reverberation time is found to be robust against the reverberant environments at the cost of the quality of the data-embedded audio. The experimental results have also shown that the proposed method is robust against various forms of attacks such as signal processing, overwriting, and malicious removal methods.

However, it would be the most severe problem to find a proper window length which satisfies both the inaudible distortion and robust data transmission in the reverberant environments. For the phase modification of the audio signal, it would be highly likely to incur a significant quality degradation if the length of time-frequency transform is very long due to the pre-echo phenomena. In the fifth chapter, therefore, segmental SNR adjustment (SSA) technique is proposed to further modify the spectral components for attenuating the pre-echo. In the proposed SSA technique, segme-

natal SNR is calculated from short-length MCLT analysis and its minimum value is limited to a desired value. The experimental results have shown that the SSA algorithm with a long MCLT length can attenuate the pre-echo effectively such that it can transmit data more reliably while preserving good audio quality. In addition, a good trade-off between the audio quality and transmission performance can be achieved by adjusting only a single parameter in the SSA algorithm.

If the number of microphones is more than one, the diversity technique which takes advantage of transmitting duplicates through statistically independent channel could be useful to enhance the transmission reliability. In the sixth chapter, the acoustic data transmission technique is extended to take advantage of the multi-microphone scheme based on combining. In the combining-based multichannel method, the synchronization and channel estimation are respectively performed at each received signal and then the received signals are linearly combined so that the SNR is increased. The most noticeable property for combining-based technique is to provide compatibility with the acoustic data transmission system using a single microphone. From the series of the experiments, the proposed multichannel method have been found to be useful to enhance the transmission performance despite of the statistical dependency between the channels.

Keywords: acoustic data transmission, audio data hiding, modulated complex lapped transform (MCLT), reverberation, segmental SNR adjustment, multichannel.
Student Number: 2009-30212

Contents

Abstract **i**

List of Figures **ix**

List of Tables **xv**

Chapter 1 Introduction **1**

1.1 Audio Data Hiding and Acoustic Data Transmission 1

1.2 Previous Methods 4

1.2.1 Audio Watermarking Based Methods 4

1.2.2 Wireless Communication Based Methods 6

1.3 Performance Evaluation 9

1.3.1 Audio Quality 9

1.3.2 Data Transmission Performance 10

1.4 Outline of the Dissertation 10

Chapter 2 Modulated Complex Lapped Transform **13**

2.1 Introduction 13

2.2 MCLT 14

2.3 Fast Computation Algorithm 18

2.4 Derivation of Interference Terms in MCLT 19

2.5 Summary 24

Chapter 3 Acoustic Data Transmission Based on MCLT 25

3.1 Introduction 25

3.2 Data Embedding 27

3.2.1 Message Frame 27

3.2.2 Synchronization Frame 29

3.2.3 Data Packet Structure 32

3.3 Data Extraction 32

3.4 Techniques for Performance Enhancement 33

3.4.1 Magnitude Modification Based on Frequency Masking . . . 33

3.4.2 Clustering-based Decoding 35

3.4.3 Spectral Magnitude Adjustment Algorithm 37

3.5 Experimental Results 39

3.5.1 Comparison with Acoustic OFDM 39

3.5.2 Performance Improvements by Magnitude Modification and Clustering based Decoding 47

3.5.3 Performance Improvements by Spectral Magnitude Adjustment 50

3.6 Summary 52

Chapter 4 Robust Acoustic Data Transmission against Reverberant Environments 55

4.1 Introduction 55

4.2 Data Embedding 56

4.2.1 Data Embedding 57

4.2.2 MCLT Length 58

4.2.3 Data Packet Structure 60

4.3 Data Extraction 61

4.3.1 Synchronization 61

4.3.2 Channel Estimation and Compensation 62

4.3.3 Data Decoding 65
4.4 Experimental Results 66
4.4.1 Robustness to Reverberation 69
4.4.2 Audio Quality 71
4.4.3 Robustness to Doppler Effect 71
4.4.4 Robustness to Attacks 71
4.5 Summary 75

Chapter 5 Segmental SNR Adjustment for Audio Quality Enhancement 77
5.1 Introduction 77
5.2 Segmental SNR Adjustment Algorithm 79
5.3 Experimental Results 83
5.3.1 System Configurations 83
5.3.2 Audio Quality Test 84
5.3.3 Robustness to Attacks 86
5.3.4 Transmission Performance of Recorded Signals in Indoor Environment 87
5.3.5 Error correction using convolutional coding 89
5.4 Summary 91

Chapter 6 Multichannel Acoustic Data Transmission 93
6.1 Introduction 93
6.2 Multichannel Techniques for Robust Data Transmission 94
6.2.1 Diversity Techniques for Multichannel System 94
6.2.2 Combining-based Multichannel Acoustic Data Transmission 98
6.3 Experimental Results 100
6.3.1 Room Environments 101
6.3.2 Transmission Performance of Simulated Environments . . . 102

6.3.3 Transmission Performance of Recorded Signals in Reverberant Environment 105
6.4 Summary 106

Chapter 7 Conclusions **109**

Bibliography **113**

국문초록 **121**

List of Figures

Figure 1.1 Brief concept diagram of acoustic data transmission. 2

Figure 1.2 Four categories of data hiding [2]. Acoustic data transmission can be categorized as an over embedded communication. 2

Figure 1.3 Embedding process of acoustic OFDM. 8

Figure 1.4 Frame structure of acoustic OFDM. 8

Figure 2.1 Interference terms $(\mathbf{A}_m)_{kl}$ in the reconstructed MCLT frame with respect to l. 16

Figure 2.2 Time-frequency representation of MCLT coefficients. Modification of the MCLT coefficients corresponding to the shaded blocks can contribute to interfering the MCLT coefficient at the hatched block. 17

Figure 3.1 Block diagram of proposed acoustic data transmission system. 27

Figure 3.2 Data embedding at message frames with guard coefficients. 29

Figure 3.3 Procedure obtaining the MCLT coefficients of the reconstructed frame for data-embedded audio signal. 30

Figure 3.4 Data embedding at synchronization frames whose equivalent window is drawn as a triangle. Synchronization sequence is embedded the coefficients corresponding to transparent windows and white blocks. 31

Figure 3.5 Structure of a data packet consisting of synchronization and message frame block. 32
Figure 3.6 Procedure of computing the correlation by frequency domain filtering [23]. 33
Figure 3.7 Example of MCLT magnitude spectrum (solid line) and masking threshold (dotted line). 34
Figure 3.8 Example of constellation diagram of received message coefficients. 36
Figure 3.9 Grouping example of MCLT coefficients in the SMA algorithm. 39
Figure 3.10 MUSHRA test scores with 95% confidence intervals for five configurations; hidden reference (no data embedded); proposed method; AOFDM; and two low pass filtered anchors. 42
Figure 3.11 Example of error between original and processed audio with acoustic OFDM signal. The error bursts at near 600 and 1624 at which the positions represent the boundary of an OFDM frame. 44
Figure 3.12 Example of spectrum for acoustic OFDM signal: (a) the exact position and (b) near GI position. 45
Figure 3.13 MUSHRA test scores with 95% confidence intervals for five configurations; hidden reference (no data embedded); magnitude modified; unaltered magnitude; and two low pass filtered anchors. 48
Figure 3.14 Comparison of the difference of the MUSHRA test scores between magnitude modified and magnitude unaltered audio clips. 48

Figure 3.15 MUSHRA test scores with 95% confidence intervals for five configurations: hidden reference, data-embedded audio signal without and with SMA, and two anchors. 51

Figure 4.1 Embedding procedure of proposed audio data hiding method for acoustic data transmission system. 57

Figure 4.2 Structure and frame indices representation of a data packet. A data packet consists of several blocks containing a pilot (or synchronization) frame and message frames. Data embedding is performed only on the data frames (white blocks). 61

Figure 4.3 Data extraction procedure of proposed audio data hiding method for acoustic data transmission system. 62

Figure 4.4 Room environment with location of the loudspeaker and microphone. Height of the room is 3 m. At the below of each installation, Cartesian coordinate is written. 68

Figure 4.5 BER of the data transmission systems in simulated room with various MCLT length. 70

Figure 5.1 Consonant intelligibility as a function of window duration for the magnitude-only and phase-only a-Consonant-a stimuli [40]. 78

Figure 5.2 Embedding procedure of proposed audio data hiding method with segmental SNR adjustment algorithm. 79

Figure 5.3 Block diagram of segmental SNR adjustment (SSA) algorithm. 79

Figure 5.4 Example of segmental SNR and target SNR in MCLT domain. The segmental SNR is upper-limited to the target segmental SNR. 81

Figure 5.5 Example of pre-echo where the length of MCLT frame is 0.37 second: (a) host signal, (b) data-embedded signal, (c) data-embedded signal with segmental SNR adjustment algorithm. 82

Figure 5.6 MUSHRA test scores with 95% confidence intervals according to test music clips. M1-M16 on horizontal axis represents the name of each music clip. 85

Figure 5.7 BER of the data transmission systems in a real room with respect to the distance between loudspeaker and microphone. 88

Figure 5.8 BER with 1/3 convolutional coding of the data transmission systems with respect to the distance between loudspeaker and microphone. 89

Figure 5.9 BER with 1/3 convolutional coding as a function of BER (solid line). Dotted line denotes the identity line. 90

Figure 6.1 Examples of channel configurations. 95

Figure 6.2 Block Diagram of space-frequency block coding (SFBC) for multichannel acoustic data transmission. 97

Figure 6.3 Block Diagram of combining-based method for multichannel acoustic data transmission. 99

Figure 6.4 Room environment with location of two loudspeakers and microphones. Height of the room is 3 m. At the below of each installation, Cartesian coordinate is written. 101

Figure 6.5 BER of the data transmission systems in a simulated room with equal gain combining method: (a) single loudspeaker and (b) stereo loudspeakers. 104

Figure 6.6 BER of the data transmission systems in a real room with equal gain combining method: (a) single loudspeaker and (b) stereo loudspeakers. 108

List of Tables

Table 3.1 System parameters of audio data hiding based on MCLT . . 40

Table 3.2 System parameters of acoustic OFDM 40

Table 3.3 Audio clips list 41

Table 3.4 Object difference score for proposed method and AOFDM with various overlap length 43

Table 3.5 BER of acoustic OFDM and the proposed acoustic data transmission system 46

Table 3.6 BER of the proposed acoustic data transmission systems with magnitude modification and clustering-based decoding . . . 49

Table 3.7 System parameters for spectral magnitude adjustment algorithm 50

Table 3.8 BER of the proposed acoustic data transmission systems with spectral magnitude adjustment algorithm 52

Table 4.1 Audio clips list 67

Table 4.2 Parameters of system configurations 68

Table 4.3 Estimated maximum delay of simulated channel with various distance 70

Table 4.4 Objective difference score with various MCLT length 71

Table 4.5 BER versus the length of MCLT with different Doppler frequency 72

Table 4.6 BER of attacked test audio clips with respect to the length of MCLT 75

Table 5.1 Parameters of system configurations 83
Table 5.2 Objective difference score of test audio clips 86
Table 5.3 BER of attacked test audio clips 87

Table 6.1 Mapping with spcae-frequency block codes for two loudspeakers 96
Table 6.2 RERR of the data transmission systems in a simulated room with different combining methods 103
Table 6.3 Average RERR of the data transmission systems in a simulated room with combining method versus distance between two microphones of a receiver 103
Table 6.4 RERR of the data transmission systems in a real room with different combining methods 105
Table 6.5 Average RERR of the data transmission systems in a simulated room with combining method versus distance between two microphones of a receiver 106

Chapter 1

Introduction

1.1 Audio Data Hiding and Acoustic Data Transmission

Audio data hiding (or information hiding) has been widely applied in many areas such as audio watermarking for copyright protection, steganography, covert communication, and broadcast monitoring [1], [2]. Apart from these traditional applications, the audio data hiding techniques can be also deployed as a fundamental framework for acoustic data transmission of which a brief implementation is illustrated in Fig. 1.1. Acoustic data transmission implies a method that sends a message signal through aerial space by playing it back using a loudspeaker at a transmitter and receives the signal by recording it using a microphone at a receiver without any additional communication devices. For instance, this method makes it possible to receive a data stream while listening to some music sound, which had been modulated to embed the intended data. What is important in this scenario is that the modulation should not modify the original music sound severely such that it can be perceived differently. In addition, increasing use of mobile devices such as smart phones demands the use

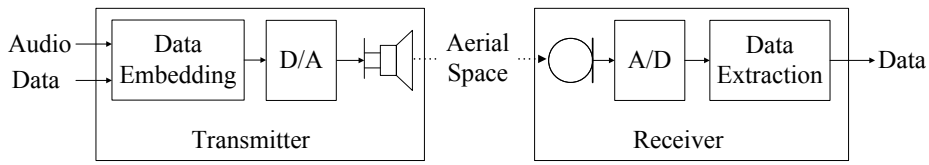


Figure 1.1 Brief concept diagram of acoustic data transmission.

	Coverwork Dependent Message	**Coverwork Independent Message**
Existence Hidden	Steganographic (covert) Watermarking	Steganography (Covert Communication)
Existence Known	Non-steganographic (overt) Watermarking	Overt Embedded Communications

Figure 1.2 Four categories of data hiding [2]. Acoustic data transmission can be categorized as an over embedded communication.

of sound for data transmission [10], [15]. This technique can be applied to various applications, e.g., querying an audio track on a radio, automatic check-in a store, localizing a pirate camcorder in a cinema [3], and providing additional information during a TV show or an advertisement.

The data hiding techniques can be subdivided into four categories based on the dependency of the messages on the coverworks, host (or original) audio in this work, (watermarking vs. non-watermarking) and awareness of the existence of the messages (non-steganographic vs. steganographic) as shown in Fig. 1.2 [2]. For the case of the audio watermarking, for instance, the messages are dependent on the contents of host signal and their existence may be known to user or not. For another example, the

messages for covert communication are independent from the host signal and the existence of the messages is unknown to the ordinary users. Acoustic data transmission can be categorized as an overt embedded communication; the messages are unrelated to the host audio signal (non-watermarking). Moreover, the users are assumed to be aware of the existence of the messages (non-steganographic) because they should be able to extract the messages from the audio signal if they want.

There are three important requirements for a practical application of the acoustic data transmission technique: inaudibility, robustness, and data rate [1], [2]. Inaudibility implies that the data-embedded audio signal should be perceptually indistinguishable from the original audio signal. At least, the listener should not be annoyed when listening to the data-embedded audio. Robustness means the ability to detect the messages in the distorted audio signal at the receiver. To employ the audio data hiding method for acoustic data transmission, the embedded data should be extracted from the received signal distorted by not only the typical signal processing procedures but also the noisy and reverberant acoustic environments. Data rate refers to the amount of data delivered in a unit time. For practical applications, a reasonable range of data rate sufficient to send short literal messages like web URL or contents identification code is needed. Because these requirements have a contradictory relation, it is hard for the data hiding system to improve all requirements. For example, if an acoustic data transmission system improves data rate, imperceptibility and robustness may be degraded. The main difference of the acoustic data transmission from the typical audio data hiding lies on the fact that the former requires a higher level of robustness and data rate than the latter. Security, which is one of the most important topics for steganographic audio watermarking or covert communication [9], [30], however, is not an essential issue of the acoustic data transmission because the existence of the messages should be assumed to be known.

1.2 Previous Methods

1.2.1 Audio Watermarking Based Methods

There have been a number of acoustic data transmission schemes motivated by audio watermarking such as echo hiding [5], spread spectrum [3], [6], [7], and adding sparse multi-carrier signals [8]. Generally, the audio watermarking-based methods cannot support sufficient data rates to transmit some useful messages while they can provide good quality of data-embedded audio.

Echo hiding

The echo hiding is one of the well-known audio watermarking methods and it has been applied to the acoustic data transmission in conjunction with low-density parity-check (LDPC) coding [5]. It is motivated by the well-known fact that the human auditory system is insensitive to a small amount of echoes [4]. The data-embedded audio signal $w(n)$ is produced by convolving the original audio signal $s(n)$ with the echo kernels $k(n)$, which is given by

$$w(n) = s(n) * k(n). \tag{1.1}$$

The echo kernel $k(n)$ can be defined by

$$k(n) = \delta(n) + \alpha\delta(n - \Delta_i), \tag{1.2}$$

where α and Δ_i represent the amplitude of echo and the delay samples having different value according to the input data i.

For the acoustic data transmission purpose, the LDPC coding is applied in order to correct the error arising from the effects in the aerial space at the cost of data rate. From the experimental results, the detection error rate of this technique showed about 0.1 at a distance of 4 m when the measured sound pressure level was 80 dB at the position of the receiver. The relatively higher error rate stems from the existence

of the variety echoes such as damping oscillation in the loudspeakers or the room impulse response. Furthermore, the data rate is as low as 8 bits per second (bps), which can be considered inappropriate for sending useful messages.

Spread spectrum

The spread spectrum watermarking, the most popular audio watermarking technique, has also been applied to the acoustic data transmission [6], [7]. Here, a pseudo-noise (PN) sequence is shaped according to a psychoacoustic model and added to the host audio signal generally in the frequency domain to embed the data. The frequency component vector of watermarked audio $\vec{\mathbf{Y}}$ is related with that of the original audio $\vec{\mathbf{X}}$ as follows:

$$\vec{\mathbf{Y}} = \vec{\mathbf{X}} + b\vec{\mathbf{p}}, \tag{1.3}$$

where $b \in \{-1, 1\}$ is a binary data to be embedded and $\vec{\mathbf{p}}$ is a pseudo-random sequence. The usage of the frequency masking model make the watermarked audio more indistinguishable from the original audio while maintaining the power of the embedded information.

In the acoustic data transmission method based on spread spectrum in discrete Fourier transform domain [6], the transmission performance from the recorded audio signal was evaluated at a distance of 1 m. The bit error rate of this technique was 0.0949 at a data rate of 213 bits per minutes and 0.0187 at a data rate of 64 bits per minutes. The method based on the spread spectrum watermarking in discrete cosine transform (DCT) domain with the Reed-Solomon code [7] can transmit the data at a rate of 76.6 bps and its detection rate showed around 0.1 at a distance of 2 m. Although they might show relatively better performance than the echo hiding method [5] in terms of the error rate and data throughput, the performance of the spread spectrum based methods is considered unsatisfactory for data transmission purposes.

Adding sparse multi-carrier signals

A technique to add sparse multi-carrier signals to the host audio signal [8] was also proposed, which is similar to the spread spectrum approach in its specific implementation. Usually the watermark signal in frequency domain is zero except one frequency index in a critical band. Psychoacoustics model is also employed in this method and it decides the optimal gain of the watermark signal for each critical band to improve the audio quality, and as a result the watermarked signal shows a good perceptual quality for non-speech signals. In [8], the cumulative distribution function of the room impulse response is utilized to design an optimal filter bank to cope with the reverberant environments. The most important parameter for designing the filter bank is the symbol interval and the authors choose 40 msec which represents a good compromise for a large variety of audio signals, including speech and most music genres. However, the transmission performance is considered doubtful because all of the BER results are higher than 0.1 even in the simulated room environment. The throughput is also not satisfactory, moreover, which is 2.9 bps at most.

1.2.2 Wireless Communication Based Methods

On the other hand, wireless communication-based methods have been also developed for acoustic data transmission. In short, wireless communication based methods can guarantee a higher data rate with more reliability than the watermarking based methods at the cost of degrading the audio quality.

Generic Purpose

One of the early methods produces digital communication signals and then allocates them at certain frequency bands and temporal positions in order to imitate music [10]. While this method provides high data rates, the modulated sound usually becomes annoying even though the parameters associated with ASK and FSK

are selected so as to obtain a relatively pleasant sound quality. Furthermore, it can transfer data only with the pre-designed sound, which restricts applications of these techniques within narrow areas. In [11], the optimal symbol length of the acoustic differential binary phase shift keying (DBPSK) signal in air was experimentally examined, and the recommended data rates versus signal-to-noise ratio (SNR) were given. The experimental results in [11] have found that the symbol duration should be longer 0.3 msec due to the limited temporal resolution of the loudspeaker.

Acoustic OFDM

As a reliable method providing a reasonable bit rate, the acoustic orthogonal frequency division multiplexing (AOFDM) technique was developed as a reliable communication with reasonable bit rate [14]-[15]. The AOFDM inherits several advantages of the multi-carrier modulation scheme in wireless communication. Each sub-carrier in the multi-carrier modulation scheme is modulated by a conventional digital modulation to carry data. While the wireless communication based on the conventional single carrier modulation schemes is usually suffered by frequency selective fading, the individual sub-carrier of the multi-carrier modulation scheme is only affected by the flat fading, which can be equalized with simple one-tap equalizer.

The embedding procedure of AOFDM is shown in Fig. 1.3. A band stop filter cuts off a certain frequency region of the host (original) audio signal. Each sub-carrier of the OFDM signal is modulated using differential binary phase shift keying (DBPSK) so that each sub-carrier carries a binary data. In order to minimize the perceptual quality degradation, the power of each sub-carrier is adjusted to be same with the corresponding frequency spectrum of the host audio signal. The additional cyclic prefix as the guard interval are then inserted in front of each OFDM frame as shown in Fig. 1.4 in order to reduce the interferences by adjacent frames (inter-symbol interference) [33]. Moreover, as can be seen in this figure, original signal can be optionally

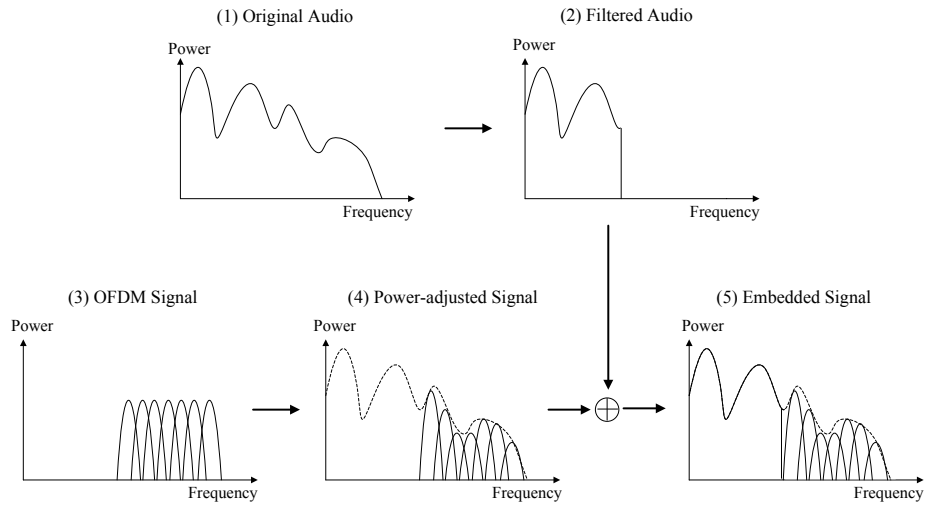


Figure 1.3 Embedding process of acoustic OFDM.

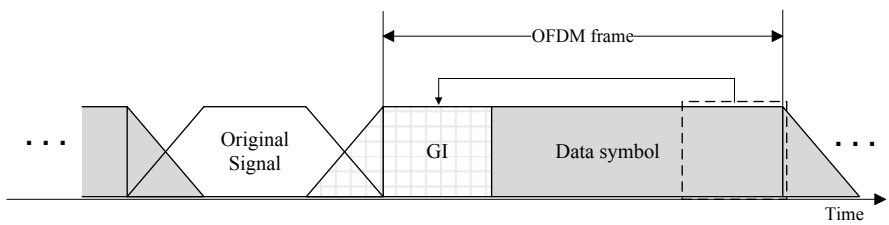


Figure 1.4 Frame structure of acoustic OFDM.

concatenated for the better audio quality and windowing is applied at the boundary of adjacent frames for the smooth signal connection. Finally, OFDM signal which has the same spectral power as the host audio signal is added at the cut-off frequency region. The synchronization signal which based on the spread spectrum watermarking with frequency masking model is additionally embedded at the low frequency region.

From the experimental results, the quality of some data-embedded audio such as rock music was reported comparable to that of the MP3 encoded audio signal. For the rock music signal, moreover, the bit error rate is almost zero at a distance up to

2 m at approximately 240 bps with the convolutional coding of 1/3 coding rate. As a result, the AOFDM succeeds in achieving robust transmission with sufficient data rate by adopting techniques widely used in wireless communication. Even though the transmission performance of AOFDM has been reported much better than that of the previous techniques, however, the quality degradation of the data-embedded audio is inevitable especially for the low-powered signal such as speech.

1.3 Performance Evaluation

In order to evaluate the performance of the audio data hiding systems proposed in this dissertation, various kinds of experiments concerned with audio quality and data transmission performance are conducted.

1.3.1 Audio Quality

In this dissertation, both of the objective quality score based on the computational algorithm and the subjective one obtained by human perception are used to evaluate and compare the audio quality of the data-embedded audio signals.

Objective Method

In this dissertation, the objective difference grade (ODG) was calculated by using the perceptual evaluation of audio quality (PEAQ) method [47]. The ODG score ranges from -4 to 0, where each digit score indicates that the perceived audio quality is very annoying, annoying, slightly annoying, perceptible but not annoying, or imperceptible.

Subjective Method

For subjective audio quality evaluation, MUSHRA test [48] was conducted. In the MUSHRA test, each listener compared the sixteen reference signal (host audio

signal) with eight differently processed test audio clips for each reference signal: hidden reference (no modification), data-embedded audio signals processed through different data hiding systems or configurations, and two anchor signals obtained by low pass filtering (3.5 kHz and 7.0 kHz). In the MUSHRA test, each listener compares the test sounds, the hidden reference and anchor signals with the reference and gives a score between 0 and 100 depending on the perceived quality.

1.3.2 Data Transmission Performance

In this dissertation, data transmission performance of each data hiding system was evaluated in a variety of acoustic environmental conditions and signal processing attacks. The data transmission performance was measured in terms of the bit error rate (BER) which refers to the ratio of the number of bits successfully detected to the total number of bits in each audio clip.

1.4 Outline of the Dissertation

In this dissertation, an audio data hiding technique based on MCLT to possibly overcome the limitations of the conventional approaches to acoustic data transmission is presented. Moreover, the extended system incorporating various methods based on the audio signal processing and wireless communication are proposed to make the acoustic data transmission system more suitable for practical applications. The rest of this dissertation is organized as follows. In Chapter 2, MCLT is introduced and the interference among adjacent coefficients is analyzed. In Chaper 3, the basic form of the data embedding and extraction method based on MCLT and some techniques for robust data transmission is described. In Chapter 4, the revised structure of a data packet to apply the channel estimation and data embedding procedure for reverberant environment are proposed. To be robust against the reverberant channel, a proper range of MCLT length are described and verified by experimental results. In Chap-

ter 5, the segmental SNR adjustment (SSA) technique is presented with the trade-off parameter controlling audio quality and data transmission performance. The multi-channel techniques for acoustic data transmission are shown in Chapter 6. Finally, Chapter 7 concludes this dissertation.

Chapter 2

Modulated Complex Lapped Transform

2.1 Introduction

The modulated complex lapped transform (MCLT) is a complex extension of the modulated lapped transform (MLT) [24], which is also known as the modified discrete cosine transform (MDCT) [25]. The MCLT technique applies a cosine-modulated filter bank that maps overlapping blocks into complex-valued blocks of transform coefficients. In most applications, the MCLT can be used in place of a DFT filter bank. The lapped transforms including MCLT have the perfect reconstruction property by overlapping with adjacent frames even after the modification for the transform coefficients. Due to this property, they have been used in most audio coding systems, such as Dolby AC-3 or MPEG-2 Advanced Audio Coding. Since each MCLT frame overlaps by half with the adjacent ones, the MCLT-based approach reduces the blocking artifacts. For this reason, the perceived quality of the data-embedded audio signal can be kept almost similar to that of the original audio.

2.2 MCLT

In this section, specific notations and formulations for MCLT are presented. MCLT generates M complex-valued coefficients from the $2M$-length frame of real-valued input signal $x(n)$. The i-th input frame which is shifted by M samples is denoted by a vector $\vec{\mathbf{x}}_i = [x(iM), x(iM+1), ..., x(iM+2M-1)]^T$ with T denoting the transpose of a vector or matrix, and the MCLT coefficient vector $\vec{\mathbf{X}}_i = [X_i(0), X_i(1), ..., X_i(M-1)]^T$ corresponding to the i-th input frame is given by [28]

$$\vec{\mathbf{X}}_i = \vec{\mathbf{X}}_{c,i} - j\vec{\mathbf{X}}_{s,i} \tag{2.1}$$

$$= (\mathbf{C} - j\mathbf{S})\mathbf{W}\vec{\mathbf{x}}_i \tag{2.2}$$

where $\vec{\mathbf{X}}_{c,i}$ and $\vec{\mathbf{X}}_{s,i}$ represent the real (cosine) and imaginary (sine) parts of $\vec{\mathbf{X}}_i$, respectively. In (2.2), $\mathbf{C}$ and $\mathbf{S}$ denote the $M \times 2M$ cosine and sine modulation matrices whose (k, n)-th elements are defined by

$$(\mathbf{C})_{kn} = \sqrt{\frac{2}{M}} \cos\Big[\Big(n + \frac{M+1}{2}\Big)\Big(k + \frac{1}{2}\Big)\frac{\pi}{M}\Big] \tag{2.3}$$

$$(\mathbf{S})_{kn} = \sqrt{\frac{2}{M}} \sin\Big[\Big(n + \frac{M+1}{2}\Big)\Big(k + \frac{1}{2}\Big)\frac{\pi}{M}\Big], \tag{2.4}$$

respectively with $j = \sqrt{-1}$. The diagonal matrix $\mathbf{W}$ is a $2M \times 2M$ window matrix whose n-th main diagonal element is commonly designed as

$$(\mathbf{W})_{nn} = -\sin\Big[\Big(n + \frac{1}{2}\Big)\frac{\pi}{2M}\Big]. \tag{2.5}$$

The inverse MCLT of $\mathbf{X}_i$ is given by

$$\vec{\mathbf{y}}_i = \mathbf{W}(\beta_c \mathbf{C}^T \vec{\mathbf{X}}_{c,i} + \beta_s \mathbf{S}^T \vec{\mathbf{X}}_{s,i}) \tag{2.6}$$

where β_c and β_s are arbitrary values that satisfy $\beta_c + \beta_s = 1$. In this work, $\beta_c = \beta_s = \frac{1}{2}$ is chosen.

To obtain the reconstructed signal, the inverse MCLT frames are overlapped and added by M samples (half of the length of an MCLT frame) with its adjacent frames.

Let $\hat{\vec{\mathbf{y}}}_i$ be the i-th reconstructed frame. Then,

$$\hat{\vec{\mathbf{y}}}_i = \begin{bmatrix} \vec{\mathbf{y}}_{2,i-1} \\ \hdashline \vec{\mathbf{0}} \end{bmatrix} + \begin{bmatrix} \vec{\mathbf{y}}_{1,i} \\ \hdashline \vec{\mathbf{y}}_{2,i} \end{bmatrix} + \begin{bmatrix} \vec{\mathbf{0}} \\ \hdashline \vec{\mathbf{y}}_{1,i+1} \end{bmatrix} \tag{2.7}$$

where $\vec{\mathbf{y}}_i = \begin{bmatrix} \vec{\mathbf{y}}_{1,i}^T & \vdots & \vec{\mathbf{y}}_{2,i}^T \end{bmatrix}^T$ with $\vec{\mathbf{y}}_{1,i}$ and $\vec{\mathbf{y}}_{2,i}$ being the M-length subvectors of $\vec{\mathbf{y}}_i$ and $\vec{\mathbf{0}}$ denotes an M-length zero vector.

By (2.2), (2.6), and (2.7), it can be shown that $\hat{\vec{\mathbf{Y}}}_i = [\hat{Y}_i(0), \hat{Y}_i(1), ..., \hat{Y}_i(M-1)]^T$, the MCLT coefficient vector obtained from the reconstructed frame $\hat{\vec{\mathbf{y}}}_i$, is formulated as follows [32]:

$$\begin{aligned} \hat{\vec{\mathbf{Y}}}_i &= \hat{\vec{\mathbf{Y}}}_{c,i} - j\hat{\vec{\mathbf{Y}}}_{s,i} \\ \hat{\vec{\mathbf{Y}}}_{c,i} &= \frac{1}{2}\vec{\mathbf{X}}_{c,i} + \frac{1}{2}[\mathbf{A}_{-1}\vec{\mathbf{X}}_{s,i-1} + \mathbf{A}_0\vec{\mathbf{X}}_{s,i} + \mathbf{A}_1\vec{\mathbf{X}}_{s,i+1}] \\ \hat{\vec{\mathbf{Y}}}_{s,i} &= \frac{1}{2}\vec{\mathbf{X}}_{s,i} + \frac{1}{2}[\mathbf{A}_1^T\vec{\mathbf{X}}_{c,i-1} + \mathbf{A}_0^T\vec{\mathbf{X}}_{c,i} + \mathbf{A}_{-1}^T\vec{\mathbf{X}}_{c,i+1}], \end{aligned} \tag{2.8}$$

where

$$\begin{aligned} \mathbf{A}_{-1} &= \mathbf{C}_1\mathbf{W}_1\mathbf{W}_2\mathbf{S}_2^T \\ \mathbf{A}_0 &= \mathbf{C}\mathbf{W}\mathbf{W}\mathbf{S}^T \\ \mathbf{A}_1 &= \mathbf{C}_2\mathbf{W}_2\mathbf{W}_1\mathbf{S}_1^T \end{aligned} \tag{2.9}$$

with $\mathbf{C}_m, \mathbf{S}_m, \mathbf{W}_m$ being $M \times M$ submatrices given by

$$\begin{aligned} \mathbf{C} &= \begin{bmatrix} \mathbf{C}_1 & \vdots & \mathbf{C}_2 \end{bmatrix} \\ \mathbf{S} &= \begin{bmatrix} \mathbf{S}_1 & \vdots & \mathbf{S}_2 \end{bmatrix} \\ \mathbf{W} &= \begin{bmatrix} \mathbf{W}_1 & \vdots & \mathbf{O} \\ \hdashline \mathbf{O} & \vdots & \mathbf{W}_2 \end{bmatrix} \end{aligned} \tag{2.10}$$

in which $\mathbf{O}$ is an $M \times M$ zero matrix. As can be seen from (2.8), the terms $\frac{1}{2}[\mathbf{A}_{-1}\vec{\mathbf{X}}_{s,i-1} + \mathbf{A}_0\vec{\mathbf{X}}_{s,i} + \mathbf{A}_1\vec{\mathbf{X}}_{s,i+1}]$ and $\frac{1}{2}[\mathbf{A}_1^T\vec{\mathbf{X}}_{c,i-1} + \mathbf{A}_0^T\vec{\mathbf{X}}_{c,i} + \mathbf{A}_{-1}^T\vec{\mathbf{X}}_{c,i+1}]$ are added to $\vec{\mathbf{X}}_i$ when MCLT is applied to the reconstructed frame $\hat{\vec{\mathbf{y}}}_i$.

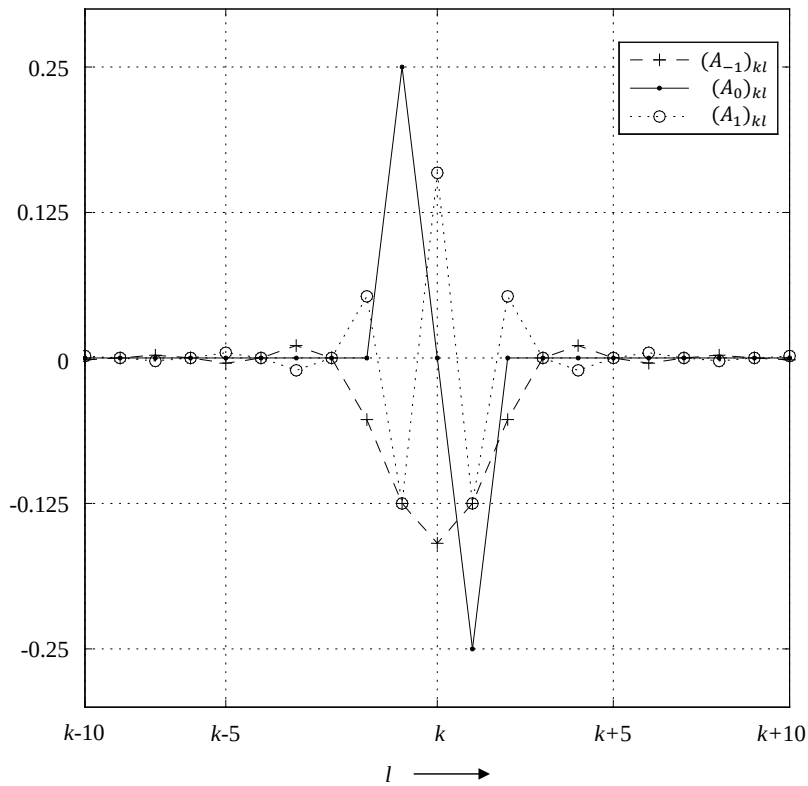


Figure 2.1 Interference terms $(\mathbf{A}_m)_{kl}$ in the reconstructed MCLT frame with respect to l.

When the frequency index k is not 0 or M, $\hat{Y}_i(k)$, which represents the k-th element of $\hat{\vec{\mathbf{Y}}}_i$, is given as follows (see Section 2.4):

$$\begin{aligned}\hat{Y}_i(k) = \frac{1}{2}X_i(k) + j\frac{1}{2}\Bigg[&(\mathbf{A}_{-1})_k\vec{\mathbf{X}}_{i-1} \\ &+ \frac{1}{2}X_i(k-1) - \frac{1}{2}X_i(k+1) + (\mathbf{A}_1)_k\vec{\mathbf{X}}_{i+1}\Bigg]\end{aligned} \tag{2.11}$$

where $(\mathbf{A}_m)_k$ represents the k-th row of $\mathbf{A}_m$. In (2.11), $(\mathbf{A}_m)_{kl}$ which represents the

(k, l)-th element of $\mathbf{A}_m$ is given by (see Section 2.4)

$$(\mathbf{A}_m)_{kl} = \begin{cases} (-m)\frac{(-1)^{l+d}}{\pi(2d-1)(2d+1)} & \text{if } |l-k| = 2d \\ \frac{(-1)^l}{4} & \text{else if } |l-k| = 1 \\ 0 & \text{otherwise,} \end{cases} \tag{2.12}$$

where d is a nonnegative integer. The examples of $(\mathbf{A}_m)_{kl}$ are plotted with respect to l in Fig. 2.1.

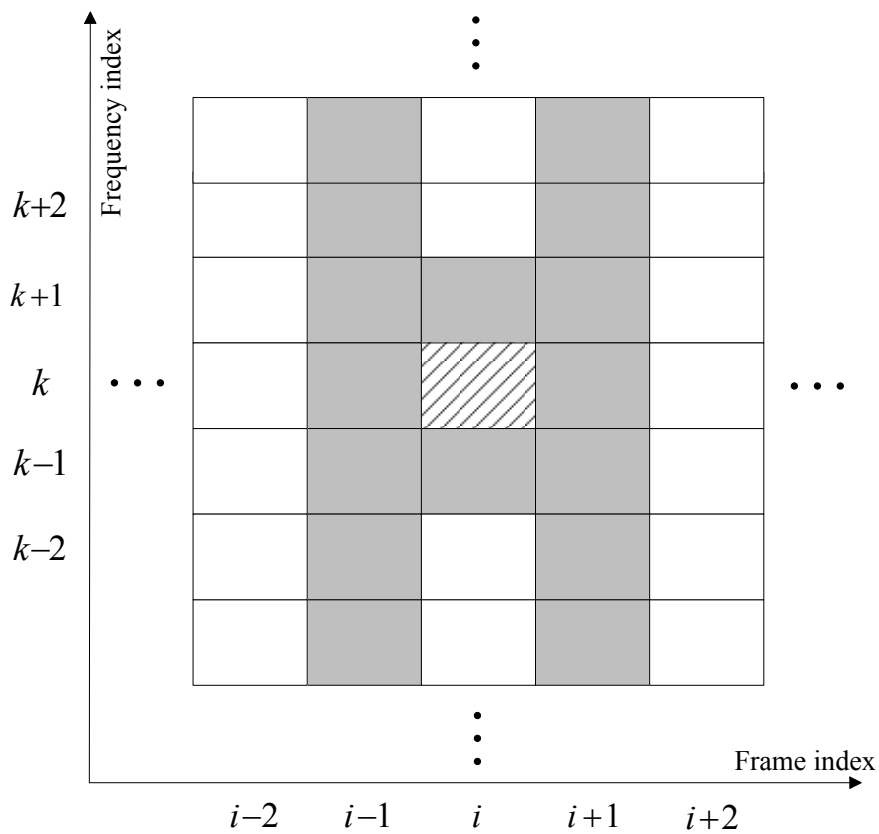


Figure 2.2 Time-frequency representation of MCLT coefficients. Modification of the MCLT coefficients corresponding to the shaded blocks can contribute to interfering the MCLT coefficient at the hatched block.

The coefficients which can contribute to interfering $\hat{Y}_i(k)$ are illustrated in Fig. 2.2. If the two adjacent frames and two adjacent coefficients $\vec{\mathbf{X}}_{i-1}$, $\vec{\mathbf{X}}_{i+1}$, $X_i(k-1)$, and $X_i(k+1)$ (the shaded blocks in Fig. 2.2) are not modified, $\hat{Y}_i(k)$ becomes the same with $X_i(k)$ due to the perfect reconstruction property of MCLT. Modification of these terms, however, gives rise to interferences resulting in a difference between $\hat{Y}_i(k)$ and $X_i(k)$. This interference is the common phenomenon observed in almost all kinds of lapped transforms such as discrete Fourier transform with tapered window.

2.3 Fast Computation Algorithm

For practical implementations, the fast computation algorithm of MCLT is indispensable to ensure that the application performs in almost real-time. Even though several fast MCLT algorithms have been proposed in various ways [26], [27], [28], the algorithms based on the fast fourier transform (FFT) with additional butterfly-like stage has known to be the most efficient [28]. After obtaining the FFT coefficients from the time-domain signal, each MCLT coefficient is computed from the two consecutive FFT coefficients. The k-th MCLT coefficient $X(k)$ is calculated in the following manner:

$$X(k) = jV(k) + V(k+1), \tag{2.13}$$

where

$$V(k) = c(k)U(k) \tag{2.14}$$

$$c(k) = W_8(2k+1)W_{4M}(k) \tag{2.15}$$

$$U(k) = \sqrt{\frac{1}{2M}} \sum_{n=0}^{2M-1} x(n)W_{2M}(kn) \tag{2.16}$$

$$W_M(k) = \exp\Big(-j\frac{2\pi k}{M}\Big) \tag{2.17}$$

where $U(k)$ denotes the $2M$-length FFT coefficients of the input signal $x(n)$. Since $x(n)$ is a real value, the number of the computation for $U(k)$ can be reduced by using M-length complex FFT [37]. The most important advantage of this algorithm is that the data shuffling, which might incur the bottleneck to access the external memory, is not required, whereas the other algorithms do the data shuffling [26], [27]. Moreover, it can reduce rounding errors in the fixed-point implementation since the magnitude of $c(k)$ is equivalent to one.

The fast computation algorithm of the inverse MCLT also can be conducted using the inverse FFT. Given the MCLT coefficients $X(k)$, the FFT coefficients $Y(k)$ for $1 \leq k \leq M-1$ are obtained as follows:

$$Y(k) = \frac{c^*(k)}{4}[X(k-1) - jX(k)], \tag{2.18}$$

where * represents the conjugate of the complex number. The real-value $Y(0)$ and $Y(M)$ are given by

$$Y(0) = \frac{1}{\sqrt{8}}[Re\{X(0)\} + Im\{X(0)\}] \tag{2.19}$$

$$Y(M) = -\frac{1}{\sqrt{8}}[Re\{X(M-1)\} + Im\{X(M-1)\}], \tag{2.20}$$

and the remaining values of $Y(k)$ can be obtained by using the conjugate symmetry property of real FFT such that

$$Y(2M-k) = Y^*(k). \tag{2.21}$$

Finally, the inverse FFT are computed to obtain the output signal $y(n)$. The total computational complexity takes $M(log_2 M + 1)$ multiplications and $M(3log_2 M + 3) - 2$ additions in this algorithm [28].

2.4 Derivation of Interference Terms in MCLT

The proof can be thought of separating $\hat{Y}_i(k)$ in (2.8) into three terms as follows:

$$\hat{Y}_i(k) = Y_i(k) + Y_{2,i-1}(k) + Y_{1,i+1}(k) \tag{2.22}$$

where the each term denotes the MCLT coefficient vector of $\vec{\mathbf{y}}_i$, $\left[\begin{array}{c:c} \vec{\mathbf{y}}_{2,i-1}^T & \vec{\mathbf{0}}^T \end{array}\right]^T$, and $\left[\begin{array}{c:c} \vec{\mathbf{0}}^T & \vec{\mathbf{y}}_{1,i+1}^T \end{array}\right]^T$ in (2.7), respectively. $Y_i(k)$ can be defined as follows:

$$Y_i(k) = \frac{1}{2}X_i(k) + \frac{1}{2}\left[(\mathbf{A}_0)_k\vec{\mathbf{X}}_{s,i} - j(\mathbf{A}_0^T)_k\vec{\mathbf{X}}_{c,i}\right]. \tag{2.23}$$

where $(\mathbf{A}_0)_k$ represents the k-th row of $\mathbf{A}_0$.

The simplest way to derive the elements of $\mathbf{A}_0$ is to utilize the procedure of the fast algorithm of MCLT calculating DFT coefficients and then converting them to the MCLT coefficients described in the previous section. Because the calcluations cascading DFT and IDFT results identity, only conversion calculations are considered in this derivation: MCLT-to-DFT and DFT-to-MCLT conversions. The MCLT-to-DFT conversion is given by

$$U_i(k) = \frac{c^*(k)}{4}[X_i(k-1) - jX_i(k)], \tag{2.24}$$

where $U_i(k)$ denotes the DFT coefficient calculated from MCLT coefficient $X_i(k)$, and $c(k) = W_8(2k+1)W_{4M}(k)$ where $W_M(k) = \exp(-j\frac{2\pi k}{M})$ [28]. The DFT-to-MCLT conversion is calculated by

$$Y_i(k) = c(k)U_i(k) + jc(k+1)U_i(k+1). \tag{2.25}$$

By substituting (2.24) to (2.25), we can see that

$$Y_i(k) = \frac{1}{2}X_i(k) + j\frac{1}{4}\left[X_i(k-1) - X_i(k+1)\right] \tag{2.26}$$

because $|c(k)|$ equals one, and this represents that $\mathbf{A}_0^T = -\mathbf{A}_0$ and its (k,l)-th element is given by

$$(\mathbf{A}_0)_{kl} = \begin{cases} \frac{1}{2} & \text{if } l = k-1 \\ -\frac{1}{2} & \text{else if } l = k+1. \end{cases} \tag{2.27}$$

The second and third terms of $\hat{Y}_i(k)$ in (2.22) denoted by $Y_{2,i-1}(k)$ and $Y_{1,i+1}(k)$ respectively are defined as follows:

$$Y_{2,i-1}(k) = \frac{1}{2}\left[(\mathbf{A}_{-1})_k\vec{\mathbf{X}}_{s,i-1} - j(\mathbf{A}_1^T)_k\vec{\mathbf{X}}_{c,i-1}\right] \tag{2.28}$$

and

$$Y_{1,i+1}(k) = \frac{1}{2}\left[(\mathbf{A}_1)_k \vec{\mathbf{X}}_{s,i+1} - j(\mathbf{A}_{-1}^T)_k \vec{\mathbf{X}}_{c,i+1}\right]. \tag{2.29}$$

To derive the elements of $\mathbf{A}_{-1}$ and $\mathbf{A}_1$, The (k,n)-th element of $\mathbf{S}_2$ is modified as follows:

$$\begin{aligned}(\mathbf{S}_2)_{kn} &= \sqrt{\frac{2}{M}} \sin\left[\left(n + M + \frac{M+1}{2}\right)\left(k + \frac{1}{2}\right)\frac{\pi}{M}\right] \\ &= \sqrt{\frac{2}{M}} \sin\left[f(2k+1,n)\frac{\pi}{4M} + \left(k + \frac{1}{2}\right)\pi\right] \\ &= \sqrt{\frac{2}{M}}(-1)^k\left[\frac{W_{8M}(f(2k+1,n))}{2} + \frac{W_{8M}(-f(2k+1,n))}{2}\right],\end{aligned} \tag{2.30}$$

where $f(k,n) = k(2n+M+1)$. The n-th diagonal element of $\mathbf{W}_2$ can be derived by

$$\begin{aligned}(\mathbf{W}_2)_{nn} &= \sin((n+\frac{1}{2})\frac{\pi}{2M} + \frac{\pi}{2}) \\ &= \cos((n+\frac{1}{2})\frac{\pi}{2M}) \\ &= \frac{W_{8M}(2n+1) + W_{8M}(-2n-1)}{2},\end{aligned} \tag{2.31}$$

and we can derive $(\mathbf{C}_1)_{ln}$ and $(\mathbf{W}_1)_{nn}$ in similar manner with $(\mathbf{S}_2)_{kn}$ and $(\mathbf{W}_2)_{nn}$ without loss of generality as follows:

$$(\mathbf{C}_1)_{ln} = \sqrt{\frac{2}{M}}\left[\frac{W_{8M}(f(2l+1,n))}{2} + \frac{W_{8M}(-f(2l+1,n))}{2}\right], \tag{2.32}$$

$$(\mathbf{W}_1)_{nn} = j\frac{W_{8M}(2n+1) - W_{8M}(-2n-1)}{2}. \tag{2.33}$$

From (2.9), the (l,k)-th element of $\mathbf{A}_{-1}$ is derived by multiplying these four

matrices as follows:

$$
\begin{aligned}
(\mathbf{A}_{-1})_{lk} &= \sum_{n=0}^{M-1} (\mathbf{C}_1\mathbf{W}_1)_{ln}(\mathbf{W}_2\mathbf{S}_2^T)_{nk} \\
&= j\frac{(-1)^k}{8M}\sum_{n=0}^{M-1}\Bigg[\Big(W_{8M}(f(2l+1,n)) + W_{8M}(-f(2l+1,n))\Big) \\
&\qquad\qquad \times \Big(W_{8M}(4n+2) - W_{8M}(-4n-2)\Big) \\
&\qquad\qquad \times \Big(W_{8M}(f(2k+1,n)) + W_{8M}(-f(2k+1,n))\Big)\Bigg],
\end{aligned}
\tag{2.34}
$$

which only two terms are nonzero by summation, those are

$$
\begin{aligned}
&W_{8M}(f(2k+1,n))W_{8M}(4n+2)W_{8M}(-f(2l+1,n)) \\
&= -jW_{4M}(f(k-l-1,n)) \\
&W_{8M}(f(2k+1,n))W_{8M}(-4n-2)W_{8M}(-f(2l+1,n)) \\
&= jW_{4M}(f(k-l-1,n))
\end{aligned}
\tag{2.35}
$$

and the other terms go to zero after summation for any value of k and l. Excluding zero terms, $(\mathbf{A}_{-1})_{lk}$ is calculated by

$$
\begin{aligned}
(\mathbf{A}_{-1})_{lk} = \frac{(-1)^k}{8M}\sum_{n=0}^{M-1}\Big[&W_{4M}(f(l-k-1,n)) + W_{4M}(f(l-k+1,n)) \\
&+ W_{4M}(f(k-l-1,n)) + W_{4M}(f(k-l+1,n))\Big],
\end{aligned}
\tag{2.36}
$$

and for each term, we can see that

$$
\begin{aligned}
&\sum_{n=0}^{M-1} W_{4M}(f(k',n)) \\
&= \sum_{n=0}^{M-1}\Big[\cos(k'(2n+M+1)\frac{\pi}{2M}) - j\sin(k'(2n+M+1)\frac{\pi}{2M})\Big].
\end{aligned}
\tag{2.37}
$$

In (2.37), each term can be decomposed by using the properties $\cos(A+B) = \cos A\cos B - \sin A\sin B$ and $\sin(A+B) = \sin A\cos B + \cos A\sin B$ and by setting $A = \pi k'/2M$ and $B = \pi k'n/M + \pi k'/2$. When $k' \ll M$, the sinusoidal function can be approximated as $\cos(\frac{\pi k'}{2M}) \simeq 1$ and $\sin(\frac{\pi k'}{2M}) \simeq \frac{\pi k'}{2M}$. Moreover, the summation of some sinusoidal function is calculated as follows:

$$\sum_{n=0}^{M-1} \cos(\frac{\pi k' n}{M}) = \begin{cases} M & \text{if } k' = 0 \\ 1 & \text{if } k' = 2d+1, d \in \mathbb{Z} \\ 0 & \text{if } k' = 2d, d \in \{\mathbb{Z} - 0\}, \end{cases} \tag{2.38}$$

and

$$\sum_{n=0}^{M-1} \sin(\frac{\pi k' n}{M}) = \begin{cases} \frac{2\pi}{k'M} & \text{if } k' = 2d+1, d \in \mathbb{Z} \\ 0 & \text{otherwise,} \end{cases} \tag{2.39}$$

thus yielding $\sum_{n=0}^{M-1} \sin(k'(2n+M+1)\frac{\pi}{2M}) = 0$ and

$$\begin{aligned} g(k') &= \sum_{n=0}^{M-1} \cos(k'(2n+M+1)\frac{\pi}{2M}) \\ &= \begin{cases} M & \text{if } k' = 0 \\ (-1)^{d+1}\frac{2M}{k'\pi} & \text{else if } k' = 2d+1, d \in \mathbb{Z} \\ 0 & \text{else if } k' = 2d, d \in \{\mathbb{Z} - 0\}, \end{cases} \end{aligned} \tag{2.40}$$

where $\mathbb{Z}$ represents the integer set.

Without loss of generality, we can simplify $(\mathbf{A}_{-1})_{lk}$ using terms $g(k')$ and also derive $(\mathbf{A}_1)_{lk}$ in the same manner as follows:

$$\begin{aligned} (\mathbf{A}_{-1})_{lk} &= \frac{(-1)^k}{4M}[g(|l-k|+1) + g(|l-k|-1)] \\ (\mathbf{A}_1)_{lk} &= -\frac{(-1)^l}{4M}[g(|l-k|+1) + g(|l-k|-1)], \end{aligned} \tag{2.41}$$

which yields that $\mathbf{A}_{-1} = -\mathbf{A}_1^T$ and (2.28) and (2.29) can be modified by

$$\begin{aligned} Y_{2,i-1}(k) &= j\frac{1}{2}(\mathbf{A}_{-1})_k \vec{\mathbf{X}}_{i-1} \\ Y_{1,i+1}(k) &= j\frac{1}{2}(\mathbf{A}_1)_k \vec{\mathbf{X}}_{i+1}, \end{aligned} \tag{2.42}$$

respectively. Therefore, (2.11) can be obtained by substituting each terms in (2.22) with (2.26) and (2.42), respectively. Finally, each element of matrix in (2.12) can be easily calculated by substituting $|l-k|$ in (2.41) to one, even integers including zero, and odd integers whose absolute value is greater than one.

2.5 Summary

In this chapter, MCLT and its fast computation algorithm is introduced with notations and the interference among adjacent coefficients is analyzed. Due to the properties of MCLT, the modification of MCLT coefficients does not introduce the blocking artifact, but it can contribute to interfering the adjacent coefficients. The audio data hiding method based on MCLT which will be introduced in next chapter should take advantage of these properties.

Chapter 3

Acoustic Data Transmission Based on MCLT

3.1 Introduction

It is doubtful to use the previous acoustic data transmission techniques introduced in Chapter 1.2 in a practical applications because they could not meet all of requirements for acoustic data transmission systems simultaneously. Generally, even though the audio watermarking-based techniques have robustness to signal processing distortions and malicious attacks, they cannot support sufficient bit rates to transmit some useful messages in a few seconds. Therefore, it is not a good idea to use current audio watermarking techniques as acoustic data transmission system. The system should accept a little quality degradation of data-embedded audio to increase the robustness and to transmit more data.

The experimental results in [11] can conclude that it is preferable to use longer symbol duration; techniques such as the acoustic orthogonal frequency division multiplexing (AOFDM) would be a good approach because it has much longer symbol

duration than the single-carrier modulation scheme. Even though the AOFDM technique can establish a reliable communication at a reasonable bit rate, some amount of audio distortion is inevitable since data embedding is achieved by modifying the original audio spectra. This quality degradation is caused by several components of the system such as the guard interval (GI) and the employed bandpass filter, and becomes more serious for low powered signals.

In this chapter, to overcome these limitations, a novel audio data hiding method for the acoustic data transmission using the modulated complex lapped transform (MCLT) is proposed. The MCLT is one of the most widely used lapped transform and it is well-known that the lapped transforms reduce the blocking artifacts induced by some modification of spectrum parameters which may degrade the quality of the resulting audio [31]. In the proposed system, audio signal is transformed by the MCLT and the phases of the coefficients are modified for data embedding. Using MCLT has an advantage that each MCLT frame overlaps half of the adjacent ones and the proposed system does not produce blocking artifacts which may degrade the quality of the resulting data-embedded audio signal [25]. Because of this overlapping property, the perceived quality of the data embedded audio signal can be kept almost similar to that of the original audio while transmitting data at several hundreds of bits per second (bps).

Moreover, several techniques to improve the performance of the proposed previous acoustic data transmission system. The newly proposed techniques are summarized as follows: First, the amplitudes of MCLT coefficients are modified to improve the detection performance for weak base audio signal. Second, the synchronization algorithm is calibrated based on a clustering method. Finally, a spectral magnitude adjustment (SMA) technique in embedding process to maintain the magnitude spectrum of the data-embedded audio signal to those of the host audio signal. The overall block diagram of the audio data hiding method which will be presented in this chapter

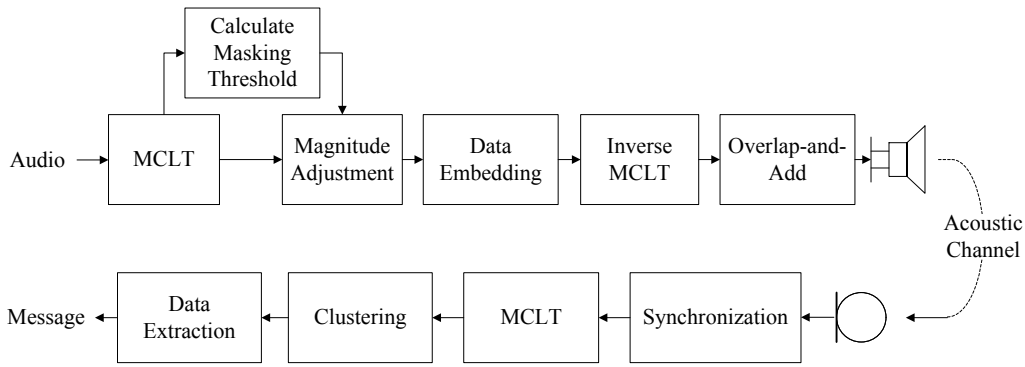


Figure 3.1 Block diagram of proposed acoustic data transmission system.

is shown in Fig. 3.1.

3.2 Data Embedding

In each data frame, the data corresponding to the synchronization sequence or message is embedded according to the procedure shown in Fig. 3.1. The embedding method at the message frames is different from that at the synchronization frames and both methods will be introduced in the following subsections.

3.2.1 Message Frame

In the proposed system, data is embedded by modifying the MCLT coefficients extracted from the audio signal. The data embedding strategy at the message frames is to modify the phases of the MCLT coefficients while the magnitudes are maintained unaltered based on the fact that human auditory perception is more sensitive to the variation in magnitude spectra [42].

Given a MCLT coefficient $\vec{\mathbf{X}}_i = [X_i(0), X_i(1), ..., X_i(M-1)]^T$, data embedding at the message frame is performed as follows:

$$X_i^D(k) = \begin{cases} |X_i(k)|d_i(k) & \text{if } k \in \mathbb{D} \\ X_i(k) & \text{otherwise,} \end{cases} \tag{3.1}$$

where $d_i(k) \in \{-1, 1\}$ depending on the input message bit and $\mathbb{D}$ is the set of the coefficient indices corresponding to the target frequency band. At receiver, the data is recovered by making a decision on whether the phase of the MCLT coefficient is closer to 0 or π. Although the phase of the received coefficient may be altered due to the interferences, it is still possible to decode the data successfully. From (2.11), we can see that the real part of the MCLT coefficient is not affected by the interferences if (3.1) is applied to all the consecutive frames. As a result, decoding message data at the receiver can be performed by examining the sign of the real part of MCLT coefficient on the target frequency lines.

Since the data is embedded in the limited frequency band and limited number of consecutive frames, the message coefficients which are close to the boundaries of the message-embedded region are interfered by unaltered MCLT coefficients which have complex values. To prevent decoding errors which due to the interference in vicinity of the boundary, the guard coefficients are embedded in the same way with (3.1) so that they do not affect to the sign of the adjacent data coefficients. An example of the location for the message and guard coefficients is illustrated in Fig. 3.2 where each block refers to an individual MCLT coefficient and white blocks denote unaltered MCLT coefficients. As can be seen in this figure, the number of the guard coefficients should be greater than two along the high and low frequency boundaries of the message-embedded region in order to guarantee the decoding errors due to the interference negligible. This comes from that the imaginary terms of adjacent frames only at from $k-2$-th to $k+2$-th frequencies can practically affect interfering the real terms at k-th frequency as can be seen Fig. 2.1.

To increase robustness against ambient noise, a single data bit is embedded in L MCLT coefficients. As L gets larger, the robustness of the system improves accordingly at the cost of reduced bit rate. The message $d_i(k)$ is spread by an L-length spreading sequence $s(m)$ whose elements are 1 or -1. In addition to the spread spec-

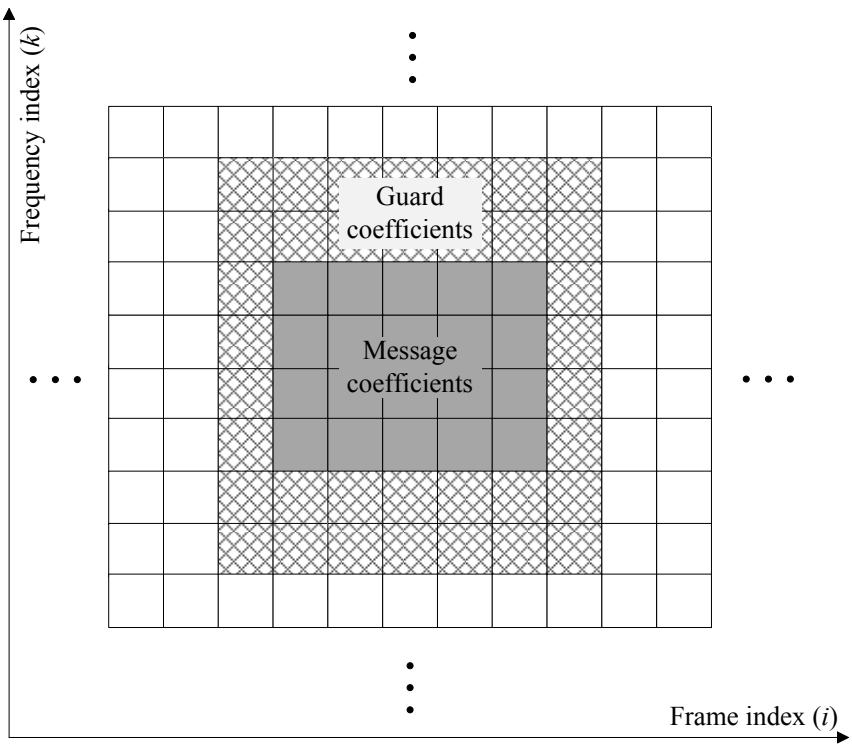


Figure 3.2 Data embedding at message frames with guard coefficients.

trum technique, some error recovery codes can be used in order to increase robustness. The modified MCLT coefficients are converted into a time-domain waveform by applying inverse MCLT and overlap-add with previous MCLT frame.

3.2.2 Synchronization Frame

In order to make a precise detection of the message, the receiver has to know the exact location of the analysis interval. In the proposed system, a synchronization sequence is used to identify the starting point of each analysis frame of the MCLT. A known synchronization bit is spread by a pseudo random sequence and embedded by modifying the phases of the MCLT coefficients to 0 or π.

To achieve a good correlation property, the phases of the reconstructed MCLT

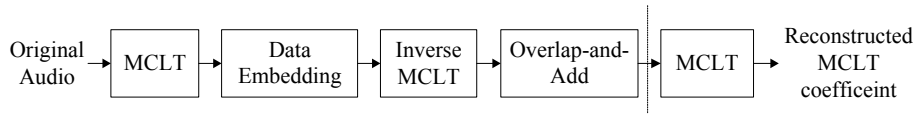


Figure 3.3 Procedure obtaining the MCLT coefficients of the reconstructed frame for data-embedded audio signal.

coefficients, which also can refer to the ideally received MCLT coefficients, should be 0 or π. However, if the synchronization sequence is embedded as given by (3.1), the original phases will be altered due to interferences described by (2.11). To prevent this, when embedding the synchronization sequence, (3.5) should be modified. Let $\hat{\tilde{\mathbf{Y}}}_i^D = [\hat{Y}_i^D(0), \hat{Y}_i^D(1), ..., \hat{Y}_i^D(M-1)]^T$ be the MCLT coefficients vector obtained at the receiver by following the procedure in Fig.3.3. Then, it can be derived that

$$\begin{aligned}\hat{Y}_i^D(k) = \frac{1}{2}X_i^D(k) + j\frac{1}{2}\Bigg[&(\mathbf{A}_{-1})_k\vec{\mathbf{X}}_{i-1} \\ &+ \frac{1}{2}X_i(k-1) - \frac{1}{2}X_i(k+1) + (\mathbf{A}_1)_k\vec{\mathbf{X}}_{i+1}\Bigg].\end{aligned} \tag{3.2}$$

To set the phase of ideally received MCLT coefficients to 0 or π, embedding of the synchronization sequence is modified as follows:

$$\begin{aligned}X_i^D(k) = 2|X_i(k)|p(k) - j\Bigg[&(\mathbf{A}_{-1})_k\vec{\mathbf{X}}_{i-1} \\ &+ \frac{1}{2}X_i(k-1) - \frac{1}{2}X_i(k+1) + (\mathbf{A}_1)_k\vec{\mathbf{X}}_{i+1}\Bigg], k \in \mathbb{S},\end{aligned} \tag{3.3}$$

where $p(k) \in \{-1, 1\}$ is the synchronization sequence known to both the transmitter and receiver, and $\mathbb{S}$ is the set of the coefficients for synchronization.

The synchronization sequence is embedded in every other frame and frequency line. From (3.3), we can see that, the two adjacent frames and two adjacent coefficients $\vec{\mathbf{X}}_{i-1}$, $\vec{\mathbf{X}}_{i+1}$, $X_i(k-1)$, and $X_i(k+1)$ contribute to interfering the MCLT coefficient $X_i^D(k)$. These coefficients are modified by embedding dummy data using (3.1) such that the magnitude of the subtracted interference terms in (3.3) is minimized. For this, the synchronization sequence is embedded in every other frame and

coefficient as shown in Fig. 3.4 where each block refers to a MCLT coefficient and shaded blocks denote the coefficients that are not modified. By applying (3.3) to (3.2), it can be shown that the interferences are cancelled and the ideally received MCLT coefficient becomes $|X_i(k)|p(k)$ as desired. In the proposed system, the synchronization sequence is embedded as in (3.3) while the content data is embedded following (3.1).

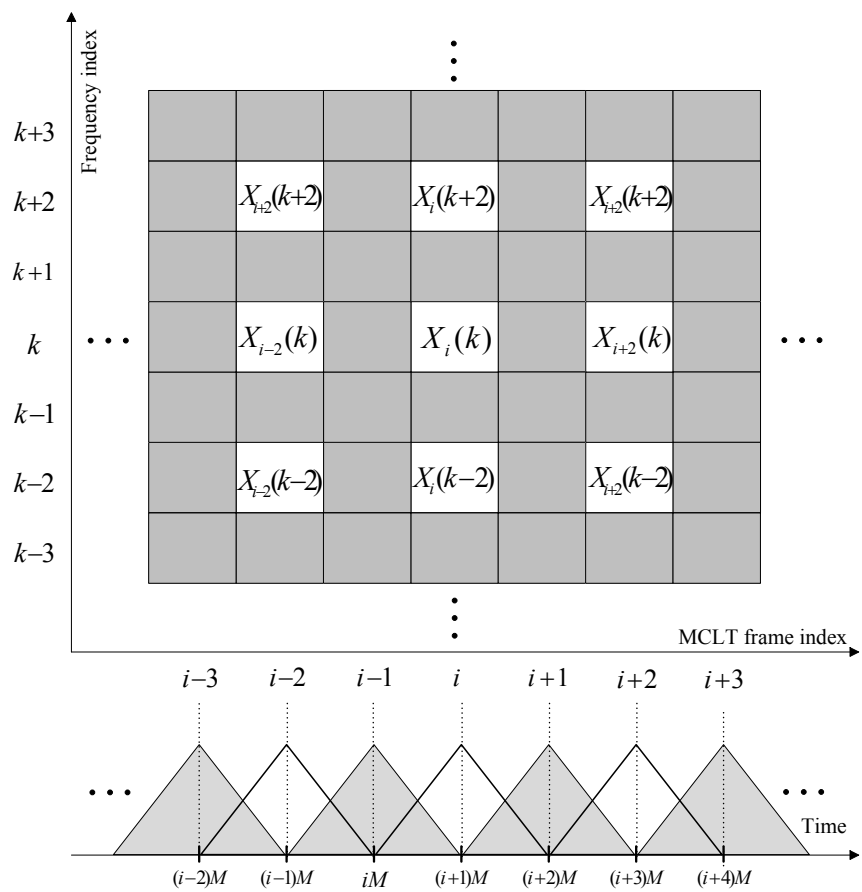


Figure 3.4 Data embedding at synchronization frames whose equivalent window is drawn as a triangle. Synchronization sequence is embedded the coefficients corresponding to transparent windows and white blocks.

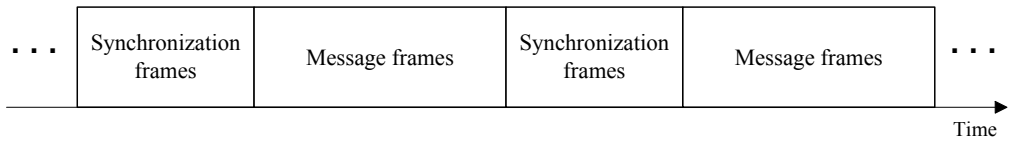


Figure 3.5 Structure of a data packet consisting of synchronization and message frame block.

3.2.3 Data Packet Structure

The structure of a data packet consisting of the synchronization and message frame block is shown in Fig. 3.5 where each block of synchronization and message consists of several successive frames. To detect the synchronization sequence at any point of the audio, the synchronization block is inserted in front of each message block.

3.3 Data Extraction

Before data extraction at the receiver, the received sound signal needs to be synchronized. The receiver computes the phase correlation between the known synchronization sequence $p(k)$ and the received MCLT coefficients extracted at all the possible analysis window locations and finds the time at which it achieves the maximum, and it becomes the starting time of the MCLT analysis frame. The estimated starting time $\hat{n}$ is given by

$$\hat{n} = \arg\max_{n} \sum_{k \in \mathbb{S}} \frac{\hat{Y}^R(k,n)p(k)}{|\hat{Y}^R(k,n)|} \tag{3.4}$$

where $\hat{Y}^R(k,n)$ is the k-th MCLT coefficient when the analysis window starts at time n. However, the computation of MCLT at every possible sample point leads to a heavy computational complexity, the correlation algorithm based on the frequency domain filtering shown in Fig. 3.6 is applied to reduce the computational burden [23].

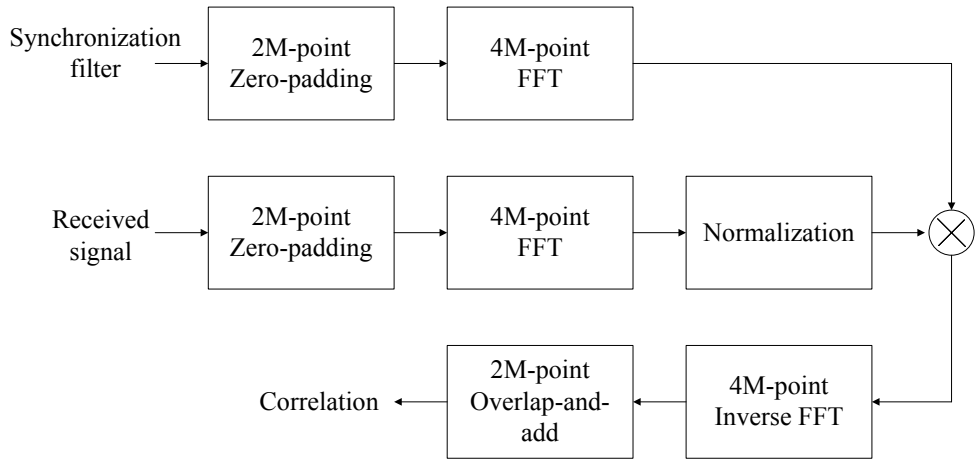


Figure 3.6 Procedure of computing the correlation by frequency domain filtering [23].

Once the received signal is synchronized, MCLT coefficients are extracted at each possible location of the analysis window and then despread. Finally, the received bit is decided according to the sign of the real part of the despread data coefficients.

3.4 Techniques for Performance Enhancement

3.4.1 Magnitude Modification Based on Frequency Masking

The performance of the acoustic data transmission system relies heavily on the spectral power of the base audio signal since a stronger signal component is likely to be less affected by the environment. The human auditory perception, however, is sensitive to the variation in magnitude spectra. For these reasons, we need to find a wise way to modify the magnitudes of the MCLT coefficients while maintaining a good audio quality.

One possible approach is to take advantage of the frequency masking effect [39]. Frequency masking effect is the phenomenon that a ceratin weak signal cannot be heard due to a strong signal in a nearby frequency region. The sound pressure level

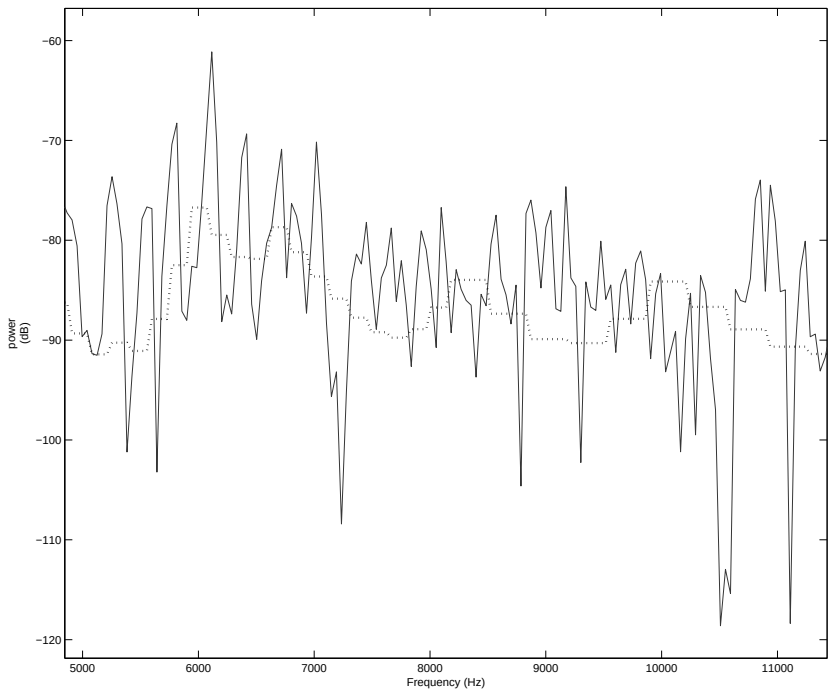


Figure 3.7 Example of MCLT magnitude spectrum (solid line) and masking threshold (dotted line).

below which the weak maskee is kept inaudible is called the masking threshold. In the proposed approach, the magnitudes of the MCLT coefficients are modified considering the masking threshold. By increasing the magnitude of the coefficients up to the level of masking threshold, we can improve the transmission performance while preventing audio quality degradation. When the data or synchronization sequences are embedded, the magnitude of the MCLT coefficients of the i-th input frame, $|X_i(k)|$ is replaced by $|X'_i(k)|$ in the following way:

$$|X'_i(k)| = \max\{|X_i(k)|, M(k)\} \tag{3.5}$$

where $M(k)$ represents the masking threshold [38] in the k-th frequency bin. An

example of the relationship between the magnitude of the MCLT coefficients and the masking threshold is illustrated in Fig. 3.7. In this figure, the magnitude spectrum and the masking threshold are drawn as solid line and dotted line, respectively.

3.4.2 Clustering-based Decoding

Exact timing recovery as known as synchronization at the receiver is difficult in acoustic data transmission due to a variety of acoustic interferences. Failure in fine synchronization will place the analysis window at a shifted position and cause many detection errors. Furthermore, since the digital-analog converter of the audio playback device and the analog-digital converter of the recording device cannot be perfectly synchronized, the audio signal at the receiver may experience a non-integer shift of sampling point. A time-shift of the analysis window results in a phase rotation of the received MCLT coefficients. The phase rotation when the analysis window is shifted from its original location by τ samples is given by

$$\phi(k) = 2\pi \frac{k + 0.5}{2M} \tau. \tag{3.6}$$

This phase rotation can change the sign of the real part of the received MCLT coefficient, which leads to decoding errors.

In order to make the decoding process robust to this phase rotation, a bit decoding algorithm that applies the k-means clustering technique [44] is proposed. Despread data coefficient of the i-th input frame $\hat{d}_i(k)$ is calculated from the received MCLT coefficients $\hat{Y}_i^R(k)$, which is given by

$$\hat{d}_i^R(k) = \frac{1}{L} \sum_{m=0}^{L-1} s(m) \frac{\hat{Y}_i^R(kL+m)}{|\hat{Y}_i^R(kL+m)|}, \quad kL + m \in \mathbb{D}, \tag{3.7}$$

which represents the mean value of the normalized MCLT coefficients. Each received data coefficient $\hat{d}_i^R(k)$ is then mapped to the nearest centroid of a two-codeword codebook. The two codewords respectively represent the data assigned to bits 0 and 1. The

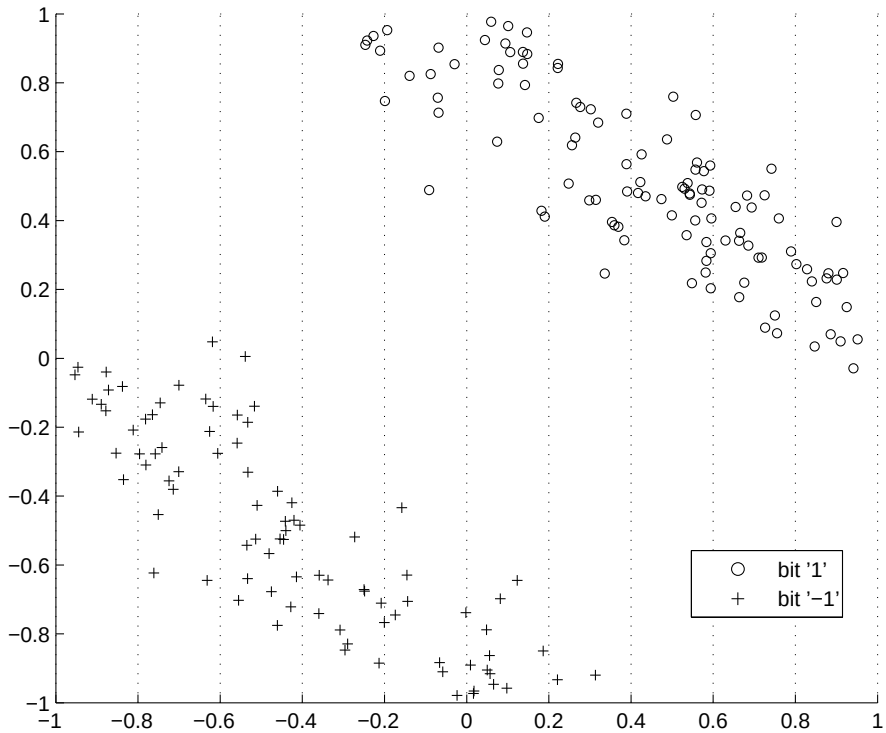


Figure 3.8 Example of constellation diagram of received message coefficients.

codebook is initialized by utilizing the normalized MCLT coefficients obtained in the synchronization frames. Let $\mu(l)$ denote the initial codeword for the synchronization bit l. Then,

$$\mu(l) = \frac{1}{|\mathbb{L}|} \sum_{k \in \mathbb{L}} \frac{\hat{Y}_i^R(k)}{|\hat{Y}_i^R(k)|}, \tag{3.8}$$

where $\mathbb{L}= \{k|p(k) = l, k \in \mathbb{S}\}$ and $|\mathbb{L}|$ is the number of elements in $\mathbb{L}$. The codebook is updated by recalculating the mean value of each separate cluster of the data coefficients extracted from a number of recent data frames.

An example of the constellation diagram of the despread message coefficients is shown in Fig. 3.8. The shape of each point '◯' or '+' represents the value on the

embedded data bit. In Fig. 3.8, you can see that the bit decoding method based on the sign of the real part of the despread coefficient makes a lot of errors. However, the clustering-based decoding method can decode the message bits without errors if it is possible to partition the constellation points off two groups as can be seen in this figure.

3.4.3 Spectral Magnitude Adjustment Algorithm

From a number of experiments, the data-embedded audio signal was observed to have different magnitude spectra from the original audio signal. This difference can be investigated by comparing $X_i(k)$ and $\hat{Y}_i^D(k)$ which represent the MCLT coefficients of the original audio signal and the ideally received data-embedded one following the procedure shown in Fig. 3.3, respectively.

To cope with the audio quality degradation while improving the performance, a spectral magnitude adjustment (SMA) algorithm is proposed in embedding process. To use the SMA algorithm, the structure of the message frames should be same with that of the synchronization frames. This algorithm modifies magnitudes of the MCLT coefficients to reduce the spectral difference between the original audio signals and the data-embedded ones. Since the magnitude of an MCLT coefficient can be altered by interferences from overlapping and adding, we should consider the effect of the interferences among adjacent MCLT coefficients when adjusting the magnitude. For this reason, the SMA approach should be an iterative algorithm, and it is summarized as follows:

1. Set the initial value of the scaling factor for the k-th MCLT coefficient in the i-th frame, $\alpha_{i,k}^{(0)} = 1$.

2. Apply the scaling factor to the original MCLT coefficient, $\tilde{X}_i^{(\nu)}(k) = \alpha_{i,k}^{(\nu)} X_i(k)$, where ν denotes the iteration number. This scaled coefficient substitutes for the original one.

3. Embed the data in the scaled MCLT coefficient $\tilde{X}_i^{(\nu)}(k)$ using (3.3).

4. Derive the ideally received MCLT coefficient $\hat{Y}_i^{(\nu)D}(k)$ from $\tilde{X}_i^{(\nu)}(k)$ using (3.2).

5. Given $\hat{Y}_i^{(\nu)D}(k)$ and the original MCLT coefficient $X_i(k)$, compute the magnitude ratio, $\gamma_{i,k}^{(\nu)} = |X_i(k)|/|\hat{Y}_i^{(\nu)D}(k)|$.

6. Update the scaling factor as $\alpha_{i,k}^{(\nu+1)} = \gamma_{i,k}^{(\nu)} \alpha_{i,k}^{(\nu)}$.

7. Repeat Steps 2 to 6 until the ratio $\gamma_{i,k}^{(\nu)}$ approaches close to 1.

After completing the above process, the final scaling factor $\alpha_{i,k}^{*}$ which makes $\hat{Y}_i(k)$ and $X_i(k)$ more similar is obtained. The scaling factor $\alpha_{i,k}^{*}$ is applied to each coefficient to obtain the final scaled MCLT coefficient $\tilde{X}_i(k) = \alpha_{i,k}^{*} X_i(k)$ and the data are embedded by using $\tilde{X}_i(k)$ instead of $X_i(k)$.

By clustering the coefficients into a number of groups according to spectral and temporal position and by applying a common scaling factor for each group, the computational efficiency of the SMA algorithm can be enhanced. In that case, the scaled MCLT coefficient at Step 2 of the SMA approach for m-th group is obtained by

$$\tilde{X}_i^{(\nu)}(k) = \alpha_m^{(\nu)} X_i(k), (i,k) \in \mathbb{G}_m \tag{3.9}$$

where $\mathbb{G}_m$ denotes the set of the frame and frequency indices pairs corresponding to the m-th group. In addition, the magnitude ratio at Step 5 of the SMA algorithm is calculated as follows:

$$\gamma_m^{(\nu)} = \frac{\sum_{(i,k)\in\mathbb{G}_m} |X_i(k)|}{\sum_{(i,k)\in\mathbb{G}_m} |\hat{Y}_i^{(\nu)D}(k)|}. \tag{3.10}$$

The SMA algorithm can be applied to both the synchronization and data frames. The example of clustering coefficients into four groups is illustrated in Fig. 3.9 where

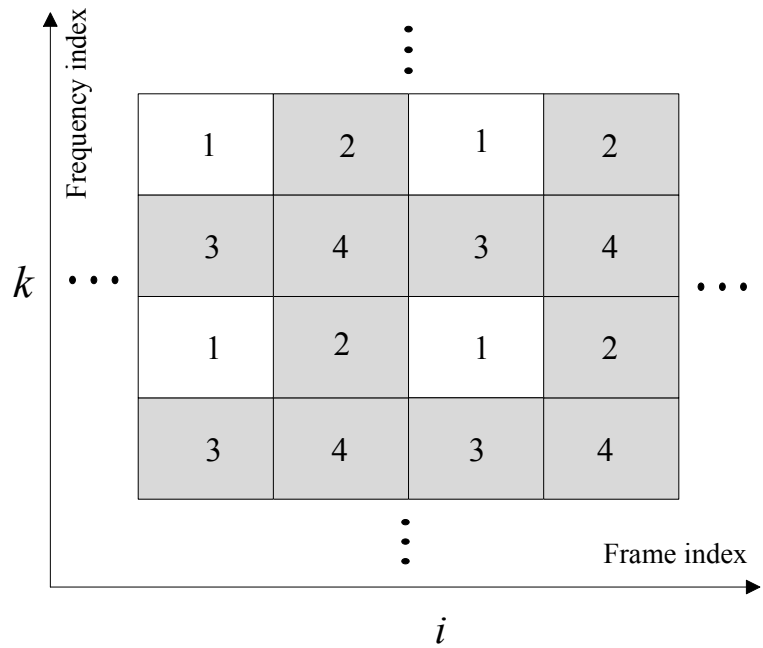


Figure 3.9 Grouping example of MCLT coefficients in the SMA algorithm.

each rectangular bin refers to an individual MCLT coefficient and the white bin denotes the data-embedded coefficient. The number in each rectangular bin in Fig. 3.9 represents the the group number m in which the coefficient is included.

3.5 Experimental Results

3.5.1 Comparison with Acoustic OFDM

To evaluate the performance of the proposed MCLT based system, subjective quality tests and transmission performance measurements were conducted. In the implementation of the proposed system, the data was embedded in the frequency range 6 - 8 kHz and transmitted at about 0.6 kbps. The AOFDM based data transmission system is also implemented for performance comparison [13]. The parameters of the AOFDM system were specified to have almost the same bit rate to that of the pro-

posed approach. The system configurations of the MCLT and AOFDM based techniques are listed in Tables 3.1 and 3.2. In this chapter, all of the experiments were performed for eight audio clips from rock, pop, jazz, and classical music genres each with length 40 seconds listed in Table 3.3.

Table 3.1 System parameters of audio data hiding based on MCLT

Sampling frequency	44.1 kHz
MCLT frame Size (M)	512 samples
Message frequency band ($\mathbb{D}$)	{149, 150, ..., 183, 184}
Synchronization frequency band ($\mathbb{S}$)	{149, 151, ... 183, 185}
Message block length	40 frames
Synchronization block length	12 frames
Message bits per frame	9 bits
Spreading length (L)	4
Average data rate (Previous)	596 bps

Table 3.2 System parameters of acoustic OFDM

Sampling frequency	44.1 kHz
Frequency band of data	6400 - 8000 Hz
OFDM frame length	2124 samples
Symbol length	1024 samples
Guard interval	600 samples
Overlap length	82 samples
Message block length	12 OFDM frames
Number of subcarriers	32
Data rate	589 bps

Table 3.3 Audio clips list

rock1	Wherever You Will Go (The Calling)
rock2	Faint (Linkin Park)
pop1	Uptown Girl (Westlife)
pop2	She Will Be Loved (Maroon 5)
jazz1	Here We Go Again (Ray Charles)
jazz2	So What (Miles Davis)
classical1	Eine Kleine Nachtmusik (Mozart)
classical2	Voice of Spring (Strauss II)

Subjective Quality Test

First, the perceived quality of the audio clips provided by the proposed system was compared with that by the AOFDM technique through the MUSHRA test [48]. In this test, nine listeners participated. The results are shown in Fig. 3.10 where the average score and 95 % confidence interval are displayed. From the results, it can be concluded that the quality of the data embedded audio clips provided by the proposed system was not much degraded from that of the original audio clips. In addition, it was shown that the proposed system outperformed the AOFDM based system irrespective of the music genres. Furthermore, particularly the improvement was remarkable for the classical music clips.

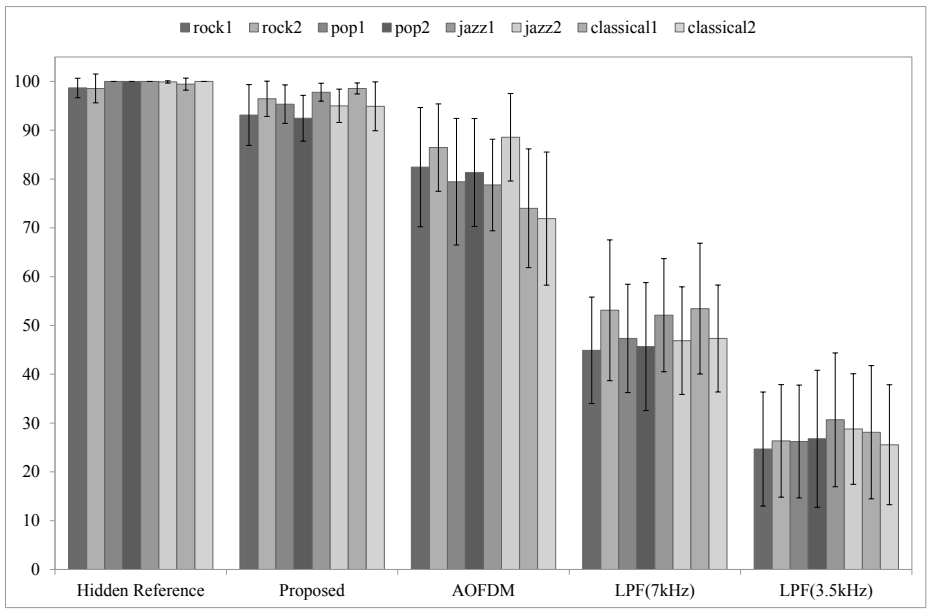


Figure 3.10 MUSHRA test scores with 95% confidence intervals for five configurations; hidden reference (no data embedded); proposed method; AOFDM; and two low pass filtered anchors.

To examine the audio quality for AOFDM according to the overlap length, PEAQ test was additionally performed to each audio clip and the ODG scores are shown in Table 3.4. From the results, we can see that the quality enhancement for AOFDM according to the overlap length is small. Moreover, the audio quality of AOFDM is significantly worse than the proposed method especially for jazz1, classical1, and classical2. The cause of the quality degradation of AOFDM is found that the data-embedded audio wave produces impulse noise at the first and the last part of each OFDM frame as can be seen in Fig. 3.11. The audio quality usually degrades seriously for low powered signals like speech and classical music [18]. This audio quality

Table 3.4 Object difference score for proposed method and AOFDM with various overlap length

test system	Proposed	AOFDM based	
overlap length	512	82	512
rock1	-0.4475	-0.5704	-0.5652
rock2	-0.3958	-0.5954	-0.5934
pop1	-0.4684	-1.0962	-0.7894
pop2	-0.3906	-0.8356	-0.6652
jazz1	-0.5351	-2.5797	-2.3404
jazz2	-0.4516	-0.6140	-0.6023
classical1	-0.3363	-2.1181	-2.0887
classical2	-0.3944	-3.0883	-3.0866
average	-0.4275	-1.4372	-1.3414

degradation is caused by several components of the system such as the guard interval (GI), the bandpass filter, and the overlap interval. Because the GI is just a copy of the last part of the frame, the frequency spectrum around the GI may not be similar to that of the original audio signal as can be seen in Fig. 3.12. In this figure, we can see the power at the data-embedded frequency region is extremely boosted. Moreover, the audio signal in the band-stop frequency band is not totally removed and this also makes the AOFDM system worse in terms of audio quality and transmission performance both.

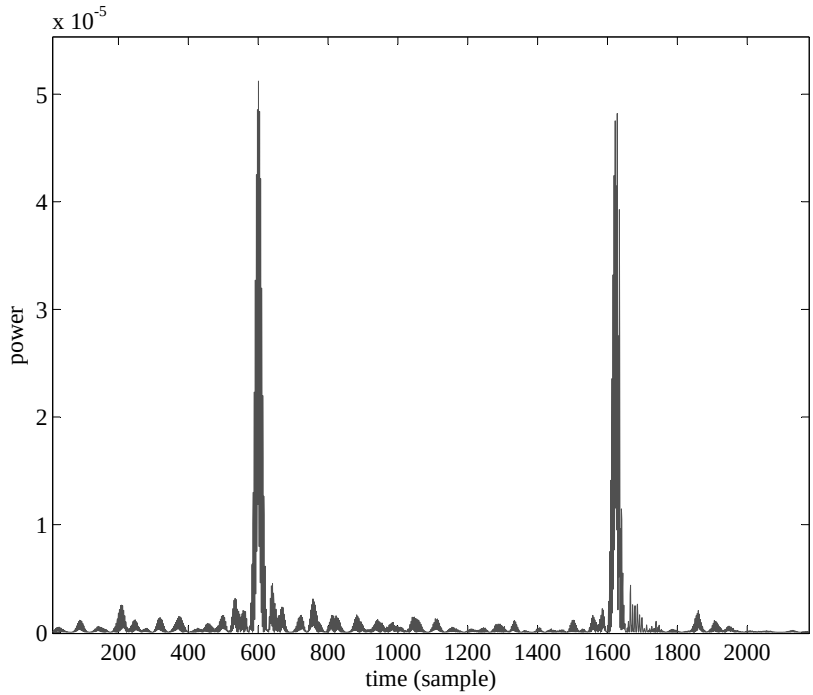
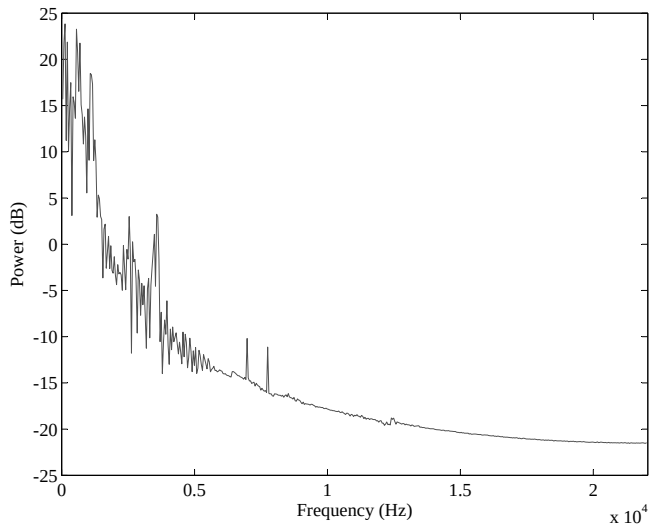
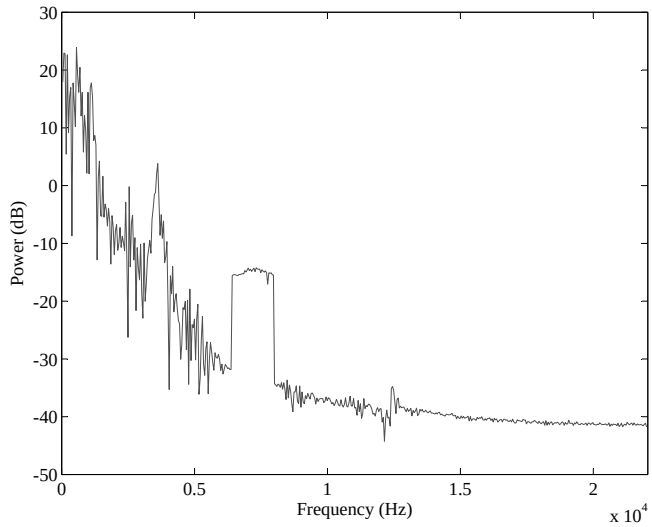


Figure 3.11 Example of error between original and processed audio with acoustic OFDM signal. The error bursts at near 600 and 1624 at which the positions represent the boundary of an OFDM frame.



(a)



(b)

Figure 3.12 Example of spectrum for acoustic OFDM signal: (a) the exact position and (b) near GI position.

Table 3.5 BER of acoustic OFDM and the proposed acoustic data transmission system

test system	Proposed			AOFDM based		
distance	1m	2m	3m	1m	2m	3m
rock1	0.0402	0.0734	0.1175	0.0310	0.1087	0.1951
rock2	0.0428	0.0787	0.1371	0.0426	0.1133	0.1904
pop1	0.0437	0.0721	0.1335	0.0431	0.1169	0.1931
pop2	0.0654	0.1032	0.1404	0.0434	0.1162	0.1935
jazz1	0.0834	0.1328	0.1778	0.0550	0.1306	0.2156
jazz2	0.0435	0.0713	0.1306	0.0375	0.1045	0.2096
classical1	0.0984	0.1462	0.1970	0.0626	0.1351	0.2070
classical2	0.0647	0.1311	0.2152	0.0642	0.1400	0.1997
average	0.0647	0.1011	0.1561	0.0474	0.1207	0.1997

Transmission Performance Test

Next, the bit error rate (BER) of the received data was measured at different distances between the loudspeaker and the microphone. In this experiment, any channel coding algorithm was applied. Table 3.5 shows the test results. Although the BER of the proposed system was a little higher comparing to that of the AOFDM based system at the distance of 1 m, the proposed system showed better transmission performance than the AOFDM based system as the transmission distance got longer. The proposed system compensates the overall delay of the received signal in the course of the synchronization procedure but does not provide a sophisticated method to avoid the multi-path interference which can decrease the transmission performance.

3.5.2 Performance Improvements by Magnitude Modification and Clustering based Decoding

The performance of the acoustic data transmission system incorporating magnitude modification and clustering-based decoding was compared to the system without any additional algorithms through a number of subjective quality and transmission performance tests. The implementation parameters of the system are specified in Table 3.1.

Subjective Quality Test

MUSHRA test [48] was performed to compare the perceived quality of the data-embedded audio clips generated from the magnitude modified MCLT coefficients with those obtained without magnitude modification. The results are shown in Fig.3.13 where the average scores of test audio clips are displayed in conjunction with 95 % confidence intervals. In Fig.3.14, the score difference between the magnitude modified audio clips and those obtained without magnitude modification is shown. From the results, it can be concluded that the magnitude modification approach based on frequency masking does not make a serious degradation in audio quality.

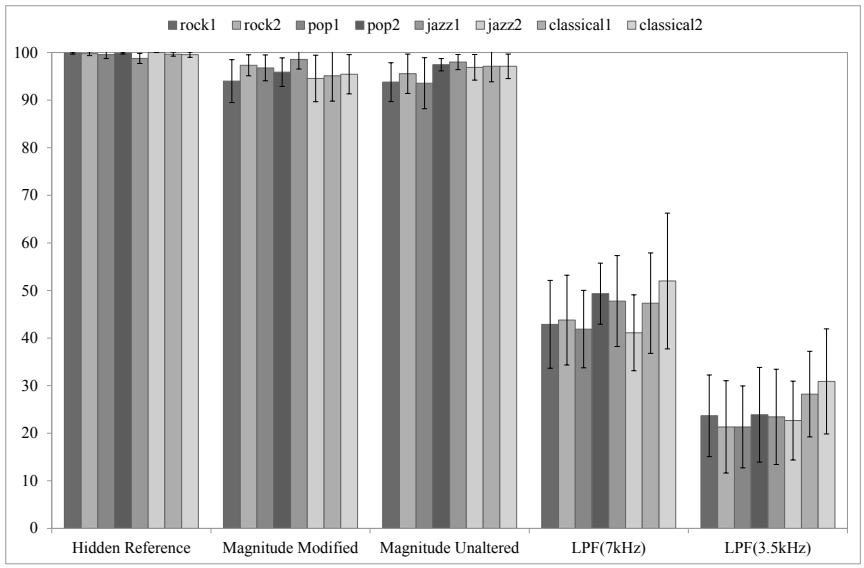


Figure 3.13 MUSHRA test scores with 95% confidence intervals for five configurations; hidden reference (no data embedded); magnitude modified; unaltered magnitude; and two low pass filtered anchors.

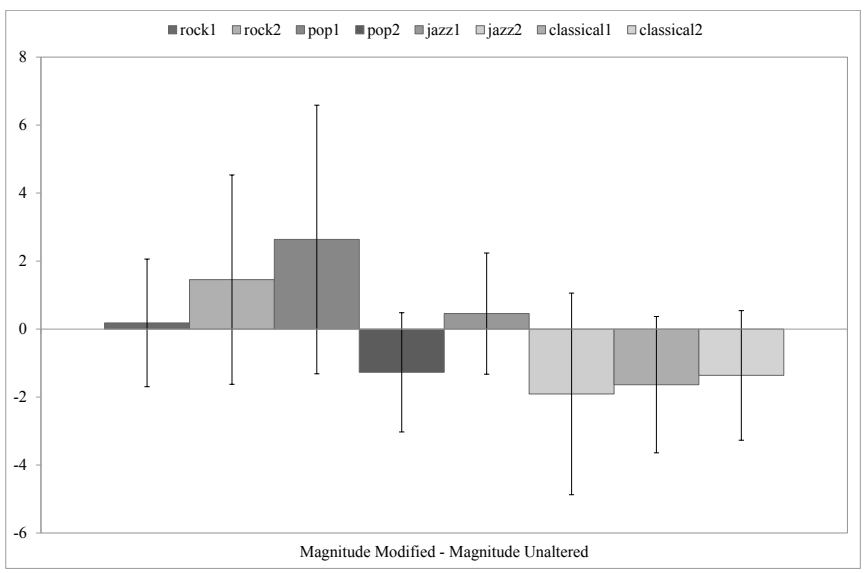


Figure 3.14 Comparison of the difference of the MUSHRA test scores between magnitude modified and magnitude unaltered audio clips.

Table 3.6 BER of the proposed acoustic data transmission systems with magnitude modification and clustering-based decoding

test system	Previous		Prev. + MM		Prev. + CLS	
distance	1m	3m	1m	3m	1m	3m
rock1	0.0211	0.0544	0.0236	0.0558	0.0167	0.0491
rock2	0.0097	0.0780	0.0116	0.0655	0.0080	0.0697
pop1	0.0082	0.0634	0.0083	0.0612	0.0068	0.0544
pop2	0.0212	0.0937	0.0161	0.0922	0.0170	0.0886
jazz1	0.1105	0.1859	0.0892	0.2003	0.0993	0.1757
jazz2	0.0334	0.1035	0.0269	0.0831	0.0235	0.0890
classical1	0.1071	0.2767	0.0958	0.2750	0.0955	0.2433
classical2	0.1290	0.1991	0.1327	0.1921	0.1181	0.2009
average	0.0550	0.1318	0.0505	0.1282	0.0481	0.1214

Transmission Performance Test

Bit error rate (BER) was measured to compare the transmission performance of the proposed system with that of the previous system. The audio clips played back from a loudspeaker were recorded by a microphone at various distances in a typical office room where there existed rather stationary noise generated from a number of fans. The measured average noise level in the room is 40dBSPL and the average audio signal level is set 65dBSPL at 1m from the loudspeaker. In this test, any channel coding algorithms are not applied in order to compare the transmission performance more fairly. The results are shown in Table 3.6 in which each method indicates the combination of the acoustic data transmission system with the magnitude modification (MM) and clustering decoding (CLS) approaches. In Table 3.6, one can see that

Table 3.7 System parameters for spectral magnitude adjustment algorithm

Sampling frequency	44.1 kHz
MCLT frame Size	512 samples
Data frequency band ($\mathbb{D}$)	{82, 84, ..., 198, 200}
Synchronization frequency band ($\mathbb{S}$)	{81, 83, ..., 138, 139}
Synchronization block length	12 frames
Message block length	80 frames
Message bits per frame	15 bits
Spreading length	4
Data rate	561 bps

the clustering decoding and magnitude modification techniques are efficient in reducing the BER. However, for the magnitude modification method did not get better for relatively calm audio clips (Jazz1 and Classics) at 3m from the loudspeaker because the masking threshold of calm audio clips are quite lower than that of loud audio clips.

3.5.3 Performance Improvements by Spectral Magnitude Adjustment

In this section, subjective quality, objective quality, and transmission performance tests to evaluate the performance of the proposed SMA algorithm. The system parameters are displayed in Table 3.7.

Subjective Quality Test

To examine the quality of data-embedded audio contents between the proposed SMA algorithm and the previous one, the MUSHRA test was conducted [48]. In this test, ten listeners participated. The result of the subjective audio quality test is shown in Fig. 3.15 with 95% confidential intervals of each score. As can be seen, the audio

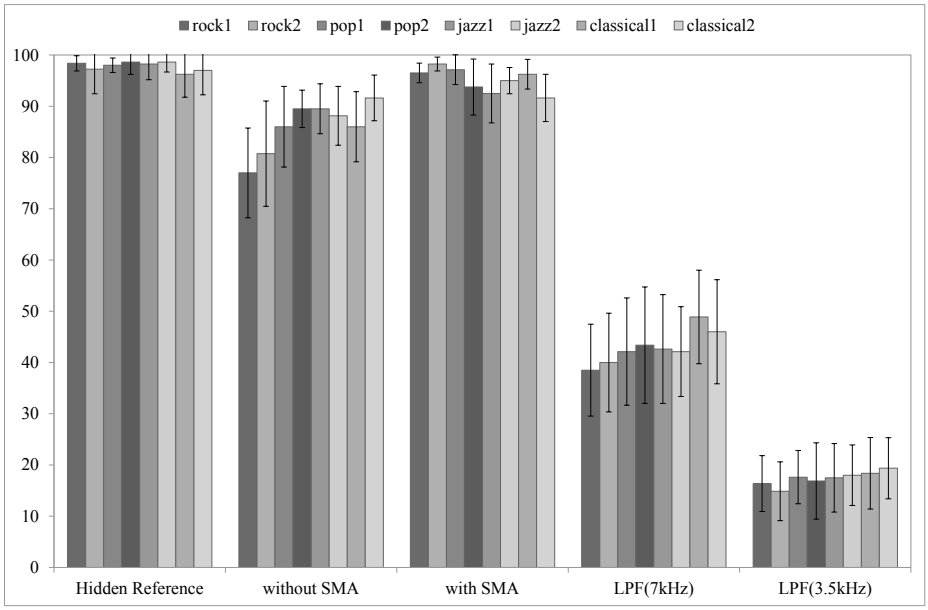


Figure 3.15 MUSHRA test scores with 95% confidence intervals for five configurations: hidden reference, data-embedded audio signal without and with SMA, and two anchors.

quality of the proposed method is mostly better than that of the previous one because the lowest confidential differential scores are positive except for “jazz1” clip.

Transmission Performance Test

To compare the transmission performance of the SMA algorithm with that of the previous one, the audio clips were recorded at various distance from a loudspeaker and evaluated BER. A mobile phone (Samsung Galaxy S) was used to record the signal. The distance from the loudspeaker to the mobile phone was 1m, 2m, and 3m. The result is given in Table 3.8. As can be seen in this table, the SMA technique enhanced the transmission performance of the proposed audio data hiding method.

Table 3.8 BER of the proposed acoustic data transmission systems with spectral magnitude adjustment algorithm

test system	without SMA			with SMA		
distance	1m	2m	3m	1m	2m	3m
rock1	0.0110	0.0079	0.0372	0.0031	0.0019	0.0182
rock2	0.0122	0.0166	0.0498	0.0039	0.0088	0.0307
pop1	0.0045	0.0193	0.0602	0.0081	0.0098	0.0321
pop2	0.0030	0.0219	0.0692	0.0001	0.0090	0.0573
jazz1	0.1085	0.1495	0.2012	0.0628	0.1037	0.1469
jazz2	0.0024	0.0224	0.0577	0.0005	0.0070	0.0280
classical1	0.0962	0.1331	0.1736	0.0551	0.0917	0.1577
classical2	0.1460	0.1469	0.2269	0.1027	0.1082	0.1545
average	0.0480	0.0647	0.1095	0.0295	0.0425	0.0782

3.6 Summary

In this chapter, a novel acoustic data transmission system based on the MCLT is proposed. In the proposed system, the data is embedded in audio clips by modifying the phases of the MCLT coefficients and transmitted at about 0.6 kbps. To prevent alteration of the received phases, an efficient embedding method to cancel the interferences is devised. Moreover, the techniques to improve robustness of the proposed acoustic data transmission system is introduced. Adopting the masking model, we can increase the audio spectral power without making distinguishable audio quality degradation. Based on the clustering method, we can correct the possible bit errors caused by synchronization mismatch. The spectral magnitude adjustment (SMA) algorithm can reduce the difference between the magnitude of MCLT coefficients and of original and data-embedded audio signals. The subjective quality tests have shown

that the audio quality obtained from the proposed system is better than that from the AOFDM based system. It has also been shown that the proposed system has better transmission performance than the AOFDM based system at a distance which is longer than a certain distance. Moreover, incorporating the masking threshold, clustering based decoding, and adjusting spectral magnitude after data embedding have been found to further improve the performance.

Chapter 4

Robust Acoustic Data Transmission against Reverberant Environments

4.1 Introduction

The audio data hiding method for acoustic data transmission proposed in Chapter 3 have been found to possibly overcome the limitations of the conventional approaches. Main features of these techniques are summarized as follows: First, the phases of the frequency components are modified in a proper way to hide a large amount of data. Second, the lapped transforms, MCLT in this work, which overlap-and-add the adjacent transform windows are applied to reduce the blocking artifacts. The experimental results have demonstrated that the proposed data hiding method in Chapter 3 yields better performance than AOFDM in terms of both the audio quality and transmission range.

Due to the response of the room, the existence of obstacles, and the movement of a receiver, however, the sent communication signals suffer from channel imperfections such as multipath fading, interference, and Doppler shift. The channel imperfec-

tions are severe problem for practical acoustic data transmission system, nevertheless, the acoustic data transmission method proposed in the previous chapter cannot cope with them.

In this chapter, an audio data hiding technique which recomposes and extends the methods proposed in the Chapter 4 to be more suitable for practical acoustic data transmission is presented by adopting techniques used in wireless communication fields. Adopting the wireless communication techniques could be effective to mitigate distortion from the channel effects if the channel model for the radio signal also can be adopted in that for the acoustic signal. The newly proposed features in this chapter are summarized as follows: First, a proper range of MCLT length to cope with reverberant environments is analyzed based on the wireless communication theory. It has been found that the transmission error in reverberant environments would be reduced if a long MCLT length is taken. Second, a channel estimation technique designed based on the Wiener estimator is applied in conjunction with a suitable data packet structure. It enables the receiver to compensate the effect of channel due to the noise, reverberation, and response of the loudspeakers and the microphone.

4.2 Data Embedding

In this section, the procedure for data embedding in the proposed acoustic data transmission system is described. The block diagram of the data embedding procedure is shown in Fig. 4.1. For a reliable transmission of messages, a cyclic redundancy check (CRC) algorithm and a forward error correction (FEC) technique are indispensable for detecting and correcting bit errors. An interleaver permutes the bit sequence in a pseudo-random manner. It can improve the performance of FEC by avoiding errors bursting on certain time or frequency regions and reduce the peak-to-average ratio (PAPR) which possibly degrades the quality of the data-embedded audio signal [34].

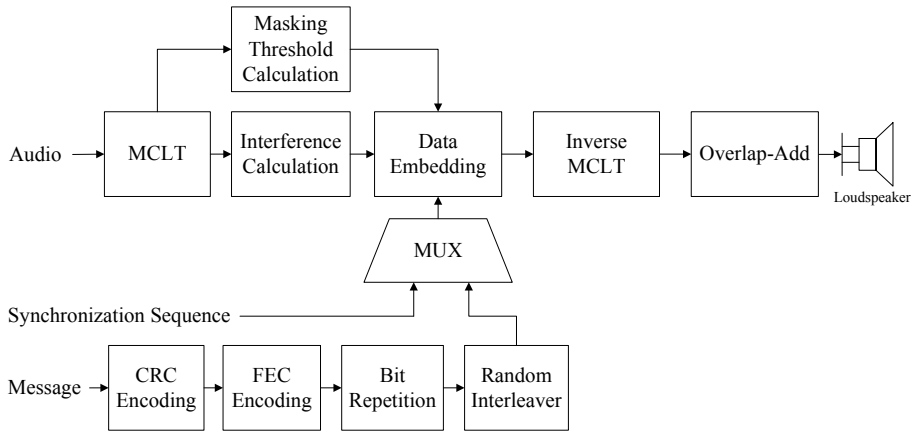


Figure 4.1 Embedding procedure of proposed audio data hiding method for acoustic data transmission system.

4.2.1 Data Embedding

A host audio signal is first divided into consecutive MCLT frames and the data bits are embedded by modifying the MCLT coefficients. The main strategy of data embedding is to modify the MCLT coefficients of the host audio signal in such a way that the phases of MCLT coefficients of the reconstructed frame are decided as either 0 or π. Here without loss of generality, the usage of a binary signaling scheme is presumed. The procedure to obtain the reconstructed MCLT coefficients is displayed in Fig. 3.3.

To obtain the data-embedded MCLT coefficient $X_i^D(k)$ for the k-th frequency element at the i-th input MCLT frame which corresponds the data frame, the target value of the MCLT coefficient of data-embedded audio at the i-th reconstructed MCLT frame $\hat{Y}_i^D(k)$ should be given by

$$\hat{Y}_i^D(k) = \max\Big(|X_i(k)|, M_i(k)\Big) b_i(k), \tag{4.1}$$

where $M_i(k)$ represents the masking threshold [38] and $b_i(k) \in \{-1, 1\}$ means the

data bit. In (4.1), the magnitude of the coefficients is lower bounded by the level of masking threshold for improving the transmission performance while maintaining the audio quality [20].

Because the only difference between deriving $\hat{Y}_i^D(k)$ in (4.1) and $\hat{Y}_i(k)$ in (2.11) is the addition of the data embedding process, $\hat{Y}_i^D(k)$ can be obtained by replacing $X_i(k)$ with $X_i^D(k)$ in (2.11) as follows:

$$\begin{aligned}\hat{Y}_i^D(k) = \frac{1}{2}X_i^D(k) + j\frac{1}{2}\Bigg[&(\mathbf{A}_{-1})_k\vec{\mathbf{X}}_{i-1} \\ &+ \frac{1}{2}X_i(k-1) - \frac{1}{2}X_i(k+1) + (\mathbf{A}_1)_k\vec{\mathbf{X}}_{i+1}\Bigg].\end{aligned} \tag{4.2}$$

Based on (4.1) and (4.2), the data-embedded MCLT coefficient $X_i^D(k)$ should be related to the original MCLT coefficient $X_i(k)$ in the following way:

$$\begin{aligned}X_i^D(k) = 2\max\Big(|X_i(k)|, M_i(k)\Big)b_i(k) - j\Bigg[&(\mathbf{A}_{-1})_k\vec{\mathbf{X}}_{i-1} \\ &+ \frac{1}{2}X_i(k-1) - \frac{1}{2}X_i(k+1) + (\mathbf{A}_1)_k\vec{\mathbf{X}}_{i+1}\Bigg], k \in \mathbb{D},\end{aligned} \tag{4.3}$$

where $\mathbb{D}$ is the set of the frequency indices in which data bit sequences are embedded. The data is embedded in every other frame and frequency line corresponding to transparent windows and white blocks in Fig. 3.4, and other coefficients are not modified from their original values. The modified MCLT coefficients are then converted into a time-domain signal segment by applying inverse MCLT and overlapping with the previous and next MCLT frames. The data embedding procedure is repeatedly performed for the next data frame.

4.2.2 MCLT Length

MCLT frame length is one of the most important parameters in acoustic data transmission. To determine a proper length of an MCLT frame, two parameters of wireless communication should be considered: the maximum (excess) delay τ_{max}, and the maximum Doppler frequency f_{D} [34].

A delay power density spectrum characterizes the frequency selectivity of the mobile radio channel due to the multipath propagation which represents that the received waves arrive from many different directions and delays. If the channel is linear, the delay power density can be modeled as a time-varying filter, which represents the impulse response of the aerial space. The maximum excess delay $\tau_{\max}$, therefore, represents the effective length of the impulse response of the communication channel. To decide the maximum delay $\tau_{\max}$ in practical application, the channel characteristics such as mean delay $\bar{\tau}$ and RMS delay spread σ_τ are obtained from the measured or modeled channel. The mean delay $\bar{\tau}$ and RMS delay spread σ_τ are defined as the expectation and standard variation of the delay power density spectrum, respectively. If the duration of a transmitted symbol, which is identical with the length of the MCLT window, is significantly larger than $\tau_{\max}$, the interference by the previous symbol, which refers to inter-symbol interference (ISI), would be negligible. In this work, although ISI is inevitable because of the overlap property of MCLT explained in Chapter 2, the amount of ISI would be decreased if the length of the MCLT window is quite longer than $\tau_{\max}$.

Similarly, the Doppler density spectrum can be defined the statistical characteristics for the time variance of the channel and the movement of the receiver. The quantitative dispersion of the frequency components is restricted by the maximum Doppler frequency f_{D}. The sub-channel spacing, which can refer to the distance between adjacent MCLT coefficients, should be much larger than f_{D} in order to avoid the performance degradation due to the interference by the adjacent frequency components also known as inter-carrier interference (ICI).

The desired length of each MCLT frame in discrete time domain, therefore, should be chosen such that

$$\frac{\tau_{\max} f_s}{2} \ll M \ll \frac{f_s}{2 f_{\mathrm{D}}} \tag{4.4}$$

where f_s represents the sampling frequency. The embedded data can be extracted

successfully only when $\tau_{\max}$ and f_{D} of the communication channel meet this condition.

4.2.3 Data Packet Structure

The structure of a data packet employed in this work is shown in Fig. 4.2, which is similar to the data structure commonly used in the wireless communication systems [34]. In this figure, the indices for MCLT frame, data frame, pilot (or synchronization) frame, and message frame are specified based on the given data packet structure. Data frame represents the part of the MCLT frames where the data corresponding to the synchronization sequence or message is embedded. From this structure, we can see that a data packet consists of P blocks each of which contains a pilot (or synchronization) frame and $N_t - 1$ message frames where N_t denotes the number of frames in a block.

According to the packet structure, the binary data $b_i(k)$ in (4.3) can be a part of either the synchronization sequence or the message to be transmitted. The synchronization sequence $p(k)$ embedded at every pilot frame should be known a priori at the receiver not only to synchronize the message frame but also to compensate the channel effects.

At message frames, the n-th message bit $d_i(n)$ is repeated L times with L-length spreading sequence to improve robustness of the data-embedded audio signal against channel effect and additive noise [35]. The L-length message sequence is then embedded in L different MCLT coefficients of which the frequency location is respectively decided by the interleaving function. The binary data at the i-th message frame is decided by repeating the n-th message bit denoted by $d_i(n) \in \{-1, 1\}$, which should be extracted at the receiver, by L times as follows:

$$b_i(\Pi(nL + m)) = d_i(n)s(m), \quad nL + m \in \mathbb{D} \tag{4.5}$$

where $s(m) \in \{-1, 1\}$ $(0 \leq m \leq L - 1)$ is the spreading sequence, $\Pi(.)$ refers to

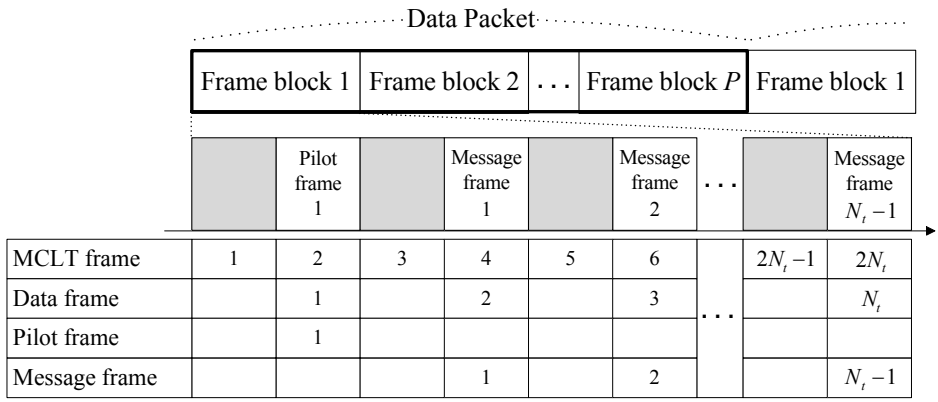


Figure 4.2 Structure and frame indices representation of a data packet. A data packet consists of several blocks containing a pilot (or synchronization) frame and message frames. Data embedding is performed only on the data frames (white blocks).

a random interleaving function, and $\mathbb{D}$ is the set of the frequency indices in which data bit sequences are embedded. For a better robustness to the channel effect, the separation distance of each $\Pi(nL + m)$ according to m should be larger than the inverse of maximum delay spread $1/\tau_{\max}$ [35].

4.3 Data Extraction

The data extraction procedure performed at a receiver is described in this section. The block diagram of this data extraction procedure is shown in Fig. 4.3.

4.3.1 Synchronization

Before extracting data from the audio signal, the received audio signal needs to be synchronized. The receiver exhaustively computes the phase correlation between the known synchronization sequence and the received MCLT coefficients, and finds the time index at which the phase correlation achieves the maximum; this enables identifying the starting time index of the first data frame of a data packet. The starting

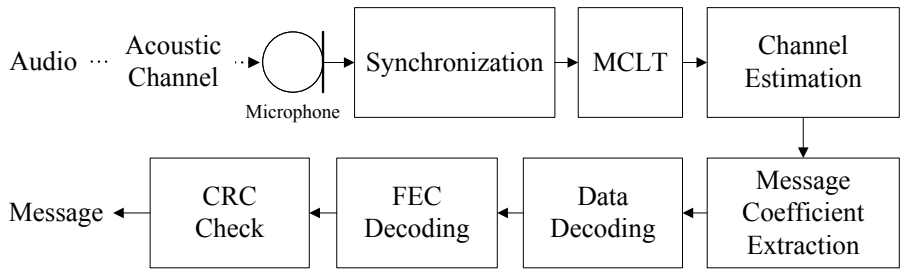


Figure 4.3 Data extraction procedure of proposed audio data hiding method for acoustic data transmission system.

time index $\hat{n}$ is given by

$$\hat{n} = \arg\max_{n} \sum_{k \in \mathbb{D}} \frac{\hat{Y}^R(k,n)p(k)}{|\hat{Y}^R(k,n)|} \tag{4.6}$$

where $p(k)$ is the known synchronization sequence and $\hat{Y}^R(k,n)$ is the k-th MCLT coefficient computed at the receiver when the analysis window starts at time n.

The k-th frequency element of the received MCLT coefficient vector corresponding to the i-th data frame, $\hat{Y}_i(k)$, is now given by

$$\hat{Y}_i^R(k) = \hat{Y}^R(k, \hat{n} + 2M(i-1)), \tag{4.7}$$

where the pilot frames are located at every multiple of N_t-th frame and the other data frames are used as the message frames.

4.3.2 Channel Estimation and Compensation

After synchronization, the received MCLT coefficient vector at the i-th data frame $\hat{Y}_i^R(k)$ can be approximated as follows:

$$\hat{Y}_i^R(k) = H_i(k) \max\Big(|X_i(k)|, M_i(k)\Big) b_i(k) + N_i(k), \tag{4.8}$$

where $H_i(k)$ and $N_i(k)$ denote the channel coefficient and the additive noise modeled by independent and identically distributed Gaussian with mean zero and variance

σ^2, respectively. Similar to a mobile radio channel, an acoustic environment can be modeled as a frequency-selective fading channel because of the commonly observed phenomena of multipath propagation. This channel effect significantly aggravates the data transmission performance.

Due to the presence of the channel coefficients, it is needed to perform channel estimation and compensation to decode the message bits successfully. Without channel estimation and compensation, data decoding is almost impossible because of the phase rotation in (3.6) and other channel effects. Compared to the clustering-based method proposed in Subsection 3.4.2, the algorithm which will be introduced in this subsection has been found more suitable to deal with a long MCLT window.

To estimate the channel coefficient, a proper algorithm to adaptively track the power density spectrum of the channel is usually required in a typical wireless communication system. In contrast to the OFDM technique, however, the proposed system quite often suffers from the inter-symbol and inter-carrier interferences because it does not possess any guard intervals, and consequently it becomes practically difficult to trace the power density spectrum of the channel perfectly. Furthermore, the number of pilot frames is considered insufficient to take full advantage of an adaptive tracking algorithm.

For these reasons, in this work, a non-adaptive Wiener estimation technique for channel equalization is applied [43]. A key idea of the Wiener estimator is first to smooth the channel measurements at the pilot positions and then to interpolate them in order to estimate the channel coefficients at the message positions. In the current work, interpolation is performed along the time axis only because message frames reside between synchronization frames (block-type pilot).

In order to obtain the channel measurements at a certain frequency from the pilot frames which are the elements of $\hat{\vec{\mathbf{H}}}_{\text{Pilot}} = [\hat{H}_{-(K-1)N_t}(k), \hat{H}_{-(K-2)N_t}(k), ..., \hat{H}_{KN_t}(k)]^T$, the received pilot coefficient is multiplied by the known synchronization

sequence $p(k)$ as follows:

$$\hat{H}_{i_p}(k) = \hat{Y}_{i_p}^{R}(k)p(k), \tag{4.9}$$

where i_p denotes the indices of pilot frames. Then, the estimated channel coefficients vector for the message frames $\hat{\vec{\mathbf{H}}}_{\text{Message}} = [\hat{H}_1(k), \hat{H}_2(k), ..., \hat{H}_{N_t-1}(k)]^T$ is interpolated by solving the normal equation, resulting in

$$\hat{\vec{\mathbf{H}}}_{\text{Message}} = \mathbf{R}_{\text{MP}}\mathbf{R}_{\text{PP}}^{-1}\hat{\vec{\mathbf{H}}}_{\text{Pilot}}, \tag{4.10}$$

where $\mathbf{R}_{\text{PP}}$ and $\mathbf{R}_{\text{MP}}$ represents the auto-correlation matrix of $\hat{\vec{\mathbf{H}}}_{\text{Pilot}}$ and the cross-correlation matrix between $\hat{\vec{\mathbf{H}}}_{\text{Message}}$ and $\hat{\vec{\mathbf{H}}}_{\text{Pilot}}$, respectively. The channel estimation process is repeatedly performed at each frequency bin separately.

The auto-correlation matrix $\mathbf{R}_{\text{PP}}$ is given by

$$\mathbf{R}_{\text{PP}} = \begin{bmatrix} R(0)+\sigma^2 & R(N_t) & \cdots & R((2K-1)N_t) \\ R(-N_t) & R(0)+\sigma^2 & \cdots & R(2KN_t) \\ \vdots & \vdots & \ddots & \vdots \\ R(-(2K-1)N_t) & R(-(2K-2)N_t) & \cdots & R(0)+\sigma^2 \end{bmatrix}, \tag{4.11}$$

where σ^2 denotes the assumed noise variance. In the proposed approach, σ^2 is set to a constant since it is practically impossible to estimate the noise spectrum from an unknown audio signal. The cross-correlation matrix $\mathbf{R}_{\text{MP}}$ is given by

$$\mathbf{R}_{\text{MP}} = \begin{bmatrix} R((K-1)N_t+1) & R((K-2)N_t+1) & \cdots & R(-KN_t+1) \\ R((K-1)N_t+2) & R((K-2)N_t+2) & \cdots & R(-KN_t+2) \\ \vdots & \vdots & \ddots & \vdots \\ R(KN_t-1) & R(KN_t) & \cdots & R(-(K-1)N_t-1) \end{bmatrix}. \tag{4.12}$$

Finally, the compensated MCLT coefficients for the k-th frequency element at the i-th message frame $\tilde{Y}_i^R(k)$ are obtained by multiplying the normalized conjugate of the estimated channel coefficient at the corresponding position as follows:

$$\tilde{Y}_i^R(k) = \frac{\hat{Y}_i^R(k)\hat{H}_i^*(k)}{|\hat{H}_i(k)|}, \tag{4.13}$$

where * denotes conjugate of the complex value.

For deriving $\mathbf{R}_{\text{PP}}$ and $\mathbf{R}_{\text{MP}}$, the correlation function $R(l)$ of two channel coefficients with respect to the frame distance l should be defined. It can be defined in

various ways according to the channel characteristics, and the sinc function is employed in this work which is given by

$$R(l) = \text{sinc}\left(\frac{4\pi M f_{\text{D,filter}}}{f_s} l\right) \tag{4.14}$$

where $f_{\text{D,filter}}$ represents the pre-defined filter parameter for the maximum frequency of the Doppler spread of the channel. From the sampling theorem, it can be shown that the distance between two pilot frames N_t satisfies

$$N_t \leq \frac{f_s}{4 f_{\text{D,filter}} M}. \tag{4.15}$$

In this case, the filter coefficients are following a uniform Doppler power density spectrum given by

$$S(f) = \begin{cases} \frac{1}{2 f_{\text{D,filter}}} & |f| < f_{\text{D,filter}} \\ 0 & \text{otherwise.} \end{cases} \tag{4.16}$$

4.3.3 Data Decoding

After the channel is compensated, message bits are extracted from the corresponding message frames through the following steps. First, the sequence of the compensated MCLT coefficients are restored by the de-interleaver performing inverse operation with the interelaver. Next, the message coefficient is obtained by calculating normalized correlation between the MCLT coefficients and the spreading sequence. The n-th message coefficient of the i-th message frame $\hat{d}_i^R(n)$ is calculated as follows:

$$\hat{d}_i^R(n) = \frac{1}{L} \sum_{m=0}^{L-1} s(m) \frac{\tilde{Y}_i^R(\Pi(nL+m))}{|\tilde{Y}_i^R(\Pi(nL+m))|}, \tag{4.17}$$

where $nL + m$ are the elements of $\mathbb{D}$. Finally, the received message bit $d_i^R(n)$ is obtained by examining the sign of real part of $\hat{d}_i^R(n)$ as follows:

$$d_i^R(n) = \begin{cases} 1 & \text{if } Re\{\hat{d}_i^R(n)\} \geq 0 \\ -1 & \text{otherwise,} \end{cases} \tag{4.18}$$

where $Re\{.\}$ indicates the real part. The message bit is decoded successfully if $d_i^R(n)$ is equal to the sent data $d_i(n)$.

After message bits for a data packet are obtained, FEC decoding and CRC checking are carried out to correct and detect errors in the message sequence. The data extraction procedure consisting of synchronization, channel compensation, and data decoding is performed on a packet-by-packet basis.

4.4 Experimental Results

In order to evaluate the performance of the proposed audio data hiding system, experiments concerned with audio quality and data transmission performance were conducted. In the experiments in this chapter, the objective audio quality and BER's when the data-embedded audio signals were convolved with an impulse response of a simulated room and influenced by a Doppler effect were examined while varying the MCLT length. Moreover, the robustness performance against various attacks incorporating typical signal processing and malicious removal attacks were examined. Sixteen stereo audio clips consisting of pop, rock, jazz, classical, and Latin music each with length 30 seconds were used in these experiments and these audio clips are listed in Table 4.1. The sampling frequency was 44.1 kHz. The average power over all the tested audio signals was adjusted to -18 dB in digital domain.

Experimentally selected parameters of the proposed method are summarized in Table 4.2 and some important parameters were selected as follows.

- Frequency band where the data are embedded was chosen from 6.5 kHz to 9.2 kHz which is not sensitive to human auditory system, the loudspeaker characteristics, and speech-like babble noises.

- The number of bit repetition L was set to four which not only shows good performance but also provides a data rate sufficient to send useful text messages.

Table 4.1 Audio clips list

M1	Yellow submarine (Peter Breiner Chamber Orchestra)
M2	Dancing queen (ABBA)
M3	Remember the time (Michael Jackson)
M4	Lucky (Jason Mraz)
M5	Honey, Honey (*Mamma Mia!* OST)
M6	Big girl (MIKA)
M7	Always (Bon Jovi)
M8	Womanizer (Britney Spears)
M9	Straight through my heart (Backstreet Boys)
M10	Hey now! (Oasis)
M11	Love the way you lie (Eminem)
M12	Champagne Supernova (Oasis)
M13	Make it mine (Jason Mraz)
M14	Primavera (Santana)
M15	Happy ending (MIKA)
M16	Bad romance (Lady Gaga)

- Maximum Doppler frequency of the Wiener filter $f_{\text{D,filter}}$ in (4.14) was set to $f_s/16M$ and as a result the maximum allowable N_t in this condition is four; in this experiment, N_t was set to 3.

- The noise variance σ^2 in the auto-correlation matrix $\mathbf{R}_{\text{PP}}$ was adjusted so that SNR would be 3 dB.

- The parameter K for the length of channel measurement $\hat{\vec{\mathbf{H}}}_{\text{Pilot}}$ was set to 1 because the future channel measurements are needed if K is greater than 1 which incurs an algorithmic delay.

Table 4.2 Parameters of system configurations

Parameters	Values
Frequency band	6.5-9.2 kHz
Spreading sequence length (L)	4
Distance between synchronization frames (N_t)	3
Bit rate	231 bps

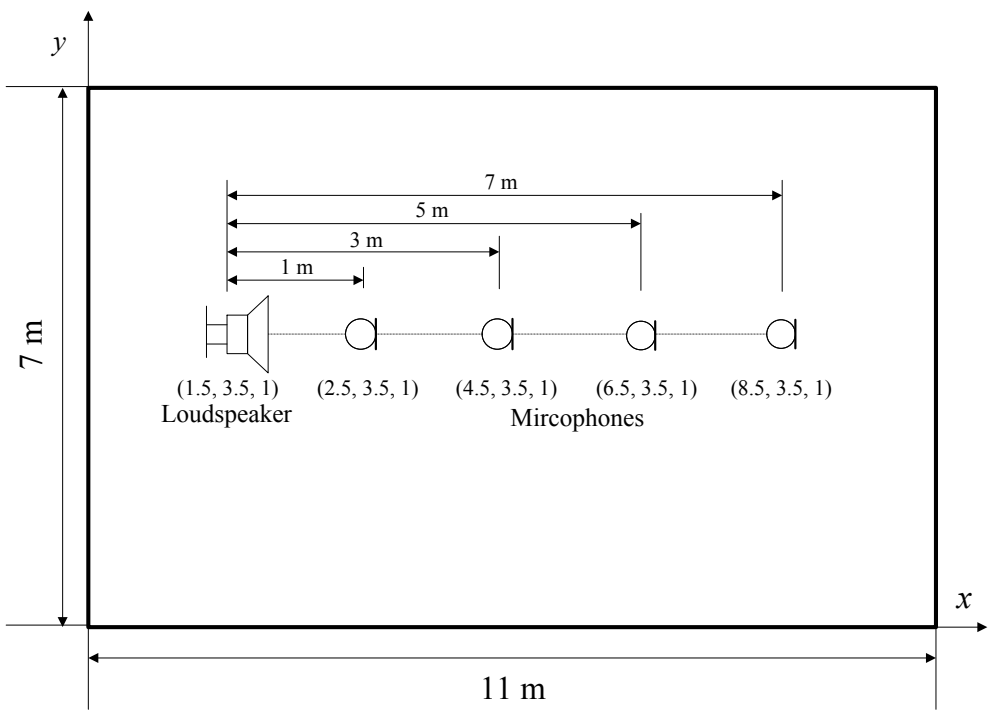


Figure 4.4 Room environment with location of the loudspeaker and microphone. Height of the room is 3 m. At the below of each installation, Cartesian coordinate is written.

4.4.1 Robustness to Reverberation

To examine how the MCLT length has effect on the BER in reverberant environments, the audio signals were convolved with the room impulse response obtained by a simulator based on the image method [45]. The dimension of the simulated room was 11 m $\times$ 7 m $\times$ 3 m and the reflection coefficient was set to -0.9, which presumes the severely reverberant environments. The materials in the room are assumed to reproduce huge acoustic reflection like concrete [46]. The location of the loudspeaker and the microphones is illustrated in Fig. 4.4.

To decide the length of MCLT robust to reverberation, a proper value of $\tau_{\max}$ in (4.4) should be decided. There are various ways to decide $\tau_{\max}$ and they are listed as follows:

- *Method 1*: $\tau_{\max}$ can be estimated as the time instance when the cumulative distribution function of room impulse response becomes greater than a certain threshold, which in this work was set to 0.99.

- *Method 2*: A maximum excess delay can be used which represents the time delay during which multipath energy of power delay profile falls to a certain threshold below the maximum. In this work, the threshold was set to 20 dB.

- *Method 3*: The channel delay τ can be modeled from power delay profile such that it follows the normal distribution $N(\bar{\tau}, \sigma_\tau)$. The maximum delay can be decided as the time instance at which the probability $\Pr(\tau > \tau_{\max})$ is lower than a certain threshold, which was set 0.01 in this work.

The simulated result for estimated $\tau_{\max}$ is shown in Table 4.3. From the simulation, it was found that $\tau_{\max} = 161$ msec was proper. Given this $\tau_{\max}$, the length of MCLT M should be longer than at least 3558 for reliable data transmission. In practical acoustic data transmission, the frame window length cannot exceed dozens of RMS delay spread like wireless communication due to the speed of acoustic wave.

Table 4.3 Estimated maximum delay of simulated channel with various distance

distance	1 m	3 m	5 m	7 m
method 1 (msec)	144	154	157	151
method 2 (msec)	70	140	161	161
method 3 (msec)	112	134	132	132

The BER obtained in the simulated room with various MCLT length is displayed in Fig. 4.5 where each line refers to the result at different distance between the loudspeaker and the receiver. In this figure, we can see the tendency that using longer MCLT window makes the data hiding system more robust to the reverberant environment.

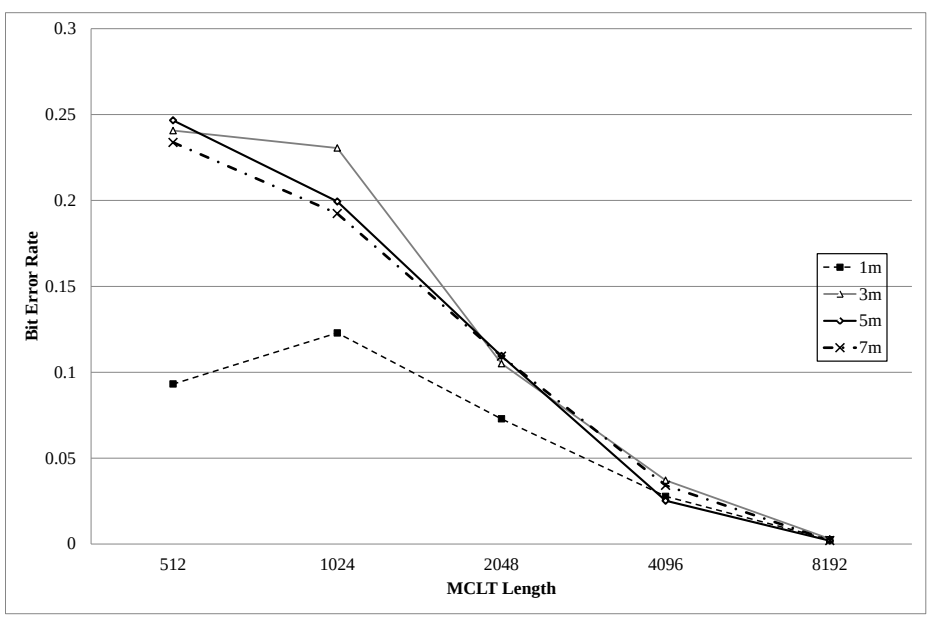


Figure 4.5 BER of the data transmission systems in simulated room with various MCLT length.

4.4.2 Audio Quality

To measure the quality of the data-embedded audio contents with various MCLT length, the objective difference grade (ODG) using the perceptual evaluation of audio quality (PEAQ) method was calculated [47]. The average ODG score obtained from the sixteen test music clips is shown in Table 4.4, which clearly demonstrates that the data-embedded audio is almost indistinguishable from the host audio if the MCLT length is shorter than 1024.

Table 4.4 Objective difference score with various MCLT length

	512	1024	2048	4096	8192
PEAQ (ODG)	-0.308	-0.511	-0.883	-1.097	-1.157

4.4.3 Robustness to Doppler Effect

To examine the robustness against the Doppler effect, the BER under the effect of different Doppler frequencies was evaluated, as given in Table 4.5. The result shown in this table demonstrates that the data can be extracted successfully only when the Doppler frequency f_{D} is much smaller than $f_s/2M$, which is derived from (4.4). Even though the long MCLT case showed a better performance, the data can be fragile to the movement of a receiver or playback speed mismatch.

4.4.4 Robustness to Attacks

In order to evaluate the robustness of the proposed audio data hiding system, the audio quality and transmission performance with various attacks incorporating typical signal processing and malicious removal attacks were examined. The quality of the attacked audio signal was evaluated by an objective measurement using using the perceptual evaluation of audio quality (PEAQ) method [47]. Robustness perfor-

Table 4.5 BER versus the length of MCLT with different Doppler frequency

Doppler frequency (Hz)	512	1024	2048	4096	8192
+2	0.0001	0.0000	0.0000	0.0000	0.0032
+5	0.0003	0.0000	0.0000	0.3204	0.5000
+10	0.0001	0.0001	0.3138	0.5000	0.5000
+20	0.0004	0.3094	0.5000	0.5000	0.5000
+40	0.2630	0.5000	0.5000	0.5000	0.5000

mance was measured in terms of the bit error rate (BER) which refers to the ratio of the number of bits successfully detected to the total number of bits in each audio clip.

Attack Description

The attacks which may impair the detection of data hiding method can be categorized based on the presence of intention to remove data by pirates; in this work, they can be subdivided them into signal processing, malicious attacks, and overwriting attacks. The typical signal processing modules applied in this experiment are listed as follows.

- *Requantization*: Each sample of test audio signal is re-quantized to 8 bits.
- *MP3*: Each test audio is encoded and decoded by using MP3 codec at the bit rates 64 kbps.
- *AWGN*: White Gaussian noise signal is added to each data-embedded audio. The signal-to-noise ratio (SNR) is 10 dB.

The malicious attacks which can remove data intentionally without severe quality degradation of attacked audio signal are listed below.

- *Denoising*: The magnitudes of MCLT coefficients with AWGN at the SNR 10 is multiplied with Wiener gain given by

$$H(k) = \frac{|\hat{X}(k)|^2}{|\hat{X}(k)|^2 + S_{NN}}, \tag{4.19}$$

  where S_{NN} represents the power spectral density of the noise.

- *All-pass filter*: All-pass filter has equal magnitude response for all frequencies, but the phases are altered. The pole of the filter is randomly chosen between 0.05 and 0.95 at every 128 samples, which represents much stronger attack than embedding data using the cochlear delay characteristics based method [55].

Overwriting using audio data hiding methods, which are listed below, could be an effective mean of attack to distort the data without severe quality degradation. Those methods, moreover, are also widely applied to add authorization information in the distribution procedure of audio clips.

- *Spread spectrum* [30]: The magnitudes of MCLT coefficients are amplified or attenuated by 4 dB according to the value of spreading code. Frequency band for embedding data is 6.5-10.5 kHz.

- *Echo hiding* [50]: Time-spread echo kernel is convolved with each test audio. Parameters incorporating amplitude, length, and location of the echo (α, L_{PN}, and Δ in [50]) was set to 0.02, 1023, 44, respectively, which implies strong data hiding condition.

- *Phase coding* [53]: Each test audio is divided into consecutive sequences with length I and subdivided into a series of N segments and then I/N-point FFT is applied to each segment. The phases of the first segment of each sequence are set either to $-\pi/2$ or to $\pi/2$ and those of the other segments are adjusted in order to preserve the relative phase between adjacent segments. In this work,

frequency band within 3.5-6.5 kHz is used, and I and N were set to 512 and 4, respectively.

- *QIM (quantization index modulation)* [54]: The magnitudes of MCLT coefficients are quantized with respect to the input data bit as follows:

$$|\hat{X}'(k)| = \begin{cases} [|\hat{X}(k)|/S]S & \text{if } b(k) = 0 \\ \lfloor |\hat{X}(k)|/S \rfloor S + S/2 & \text{otherwise,} \end{cases} \tag{4.20}$$

where $[.]$, $\lfloor . \rfloor$ respectively refer to the nearest integer, the greatest integer of input and S is quantization step size. This step size was set to

$$S = \sqrt{4M\mathbb{E}_{av}}/32, \tag{4.21}$$

where $\mathbb{E}_{av}$ is the average energy of the test audio clip.

Experimental Result

The experimental result is shown in Table 4.6. In this table, the average PEAQ score of attacked audio signal compared with original (host) and with data-embedded (test) audio signal, and BER of each attack are provided respectively. Note that some of PEAQ scores are not included in this table because PEAQ score is meaningless in the case of AWGN, and the time scaling is not an appropriate attack to define audio quality with PEAQ score. This result proves that the proposed method is as robust as other audio data hiding techniques [50], [51] against compression, adding noise, denoising, all-pass filtering, and overwriting but with comparatively high data rate [56].

Table 4.6 BER of attacked test audio clips with respect to the length of MCLT

Attacks	512	1024	2048	4096	8192
No Attack	0.0000	0.0000	0.0000	0.0000	0.0000
Requantize (8 bit)	0.0011	0.0013	0.0008	0.0006	0.0002
MP3 (64 kbps)	0.0106	0.0166	0.0113	0.0072	0.0047
AWGN (10 dB)	0.0511	0.0600	0.0558	0.0461	0.0386
Denoising (10 dB)	0.0293	0.0277	0.0264	0.0234	0.0190
All-pass filter	0.0115	0.0162	0.0177	0.0158	0.0129
Spread spectrum [30]	0.0010	0.0003	0.0001	0.0000	0.0000
Echo hiding [50]	0.0076	0.0030	0.0002	0.0000	0.0000
Phase coding [53]	0.0015	0.0020	0.0012	0.0008	0.0005
QIM [54]	0.0007	0.0025	0.0018	0.0040	0.0032

4.5 Summary

In this chapter, an audio data hiding technique for acoustic transmission is proposed which incorporates the fundamental schemes of wireless communication into the previous works in Chapter 3. The data bits are embedded by modifying the phases of MCLT coefficients such that they can be detected as pre-defined values at the receiver. From the experimental results, it has been found that the relationship for the MCLT length adopted from wireless communication theory similar tendency in the acoustic data transmission. If the length of MCLT window is quite longer than that of the channel, the acoustic data transmission system can send data reliably through the reverberant aerial space. The lengthy windows, however, usually degrade the quality of the data-embedded audio. The experimental results also have shown that the proposed method is robust against various forms of attacks incorporating signal processing, overwriting, and malicious removal methods.

Chapter 5

Segmental SNR Adjustment for Audio Quality Enhancement

5.1 Introduction

In the acoustic data communication system proposed in the previous chapter, we can see that the transmission performance in reverberant environment would improve as the length of the MCLT frame gets longer. However, it would be highly likely for the phase modification of the audio signal to incur a significant quality degradation if the length of time-frequency transform is very long.

The experiments in [40] and [41] showed the similar tendency for the relationship between the quality of phase-modified signal and the length of analysis window. In these papers, human perception experiments are conducted to measure the intelligibility of speech stimuli (a-Consonant-a context) synthesized either from magnitude or phase spectra. The experiments conclude that the magnitude spectrum plays a dominant role for small window durations (20–40 ms) while the phase spectrum is more important for window durations greater than 200 ms as shown in Fig. 5.1.

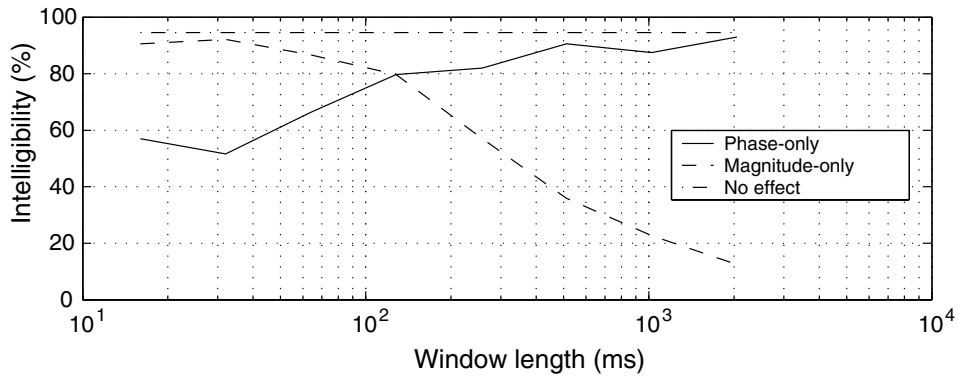


Figure 5.1 Consonant intelligibility as a function of window duration for the magnitude-only and phase-only a-Consonant-a stimuli [40].

The most important problem for the method in the previous chapters would be that it is almost impossible to find a proper window length satisfying both the inaudible distortion and robust data transmission in the reverberant environments. If a percussive sound made by drums, cymbals, or short unvoiced speech exists within an interval of the MCLT window $2M$, pre-echo may frequently occur. The pre-echo is one of the most important causes of quality degradation in data-embedded audio. The length of the MCLT window, however, is usually set longer than the onset time of the percussive sounds in order to get robustness against the reverberation in a typical room environment resulting in pre-echoes. One of the most efficient techniques in audio watermarking is to avoid embedding data at the positions that pre-echo would occur [30], which, however, is considered doubtful to apply because the window length in this work is usually much longer than that in audio watermarking.

In this chapter, therefore, segmental SNR adjustment (SSA) technique [22] with additional trade-off parameter is proposed to further modify the spectral components for attenuating the pre-echo to ameliorate the audio quality degradation resulted from applying a long MCLT length. In order to further investigate the performance based on the results in the previous chapter, moreover, the performance of the five different

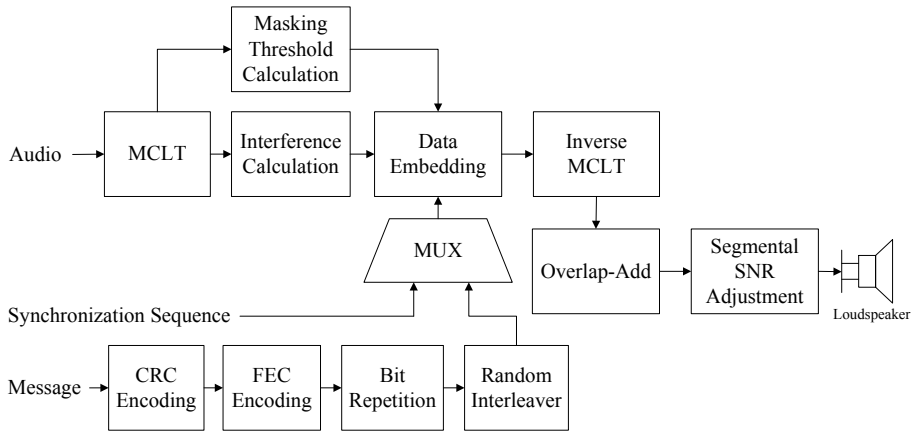


Figure 5.2 Embedding procedure of proposed audio data hiding method with segmental SNR adjustment algorithm.

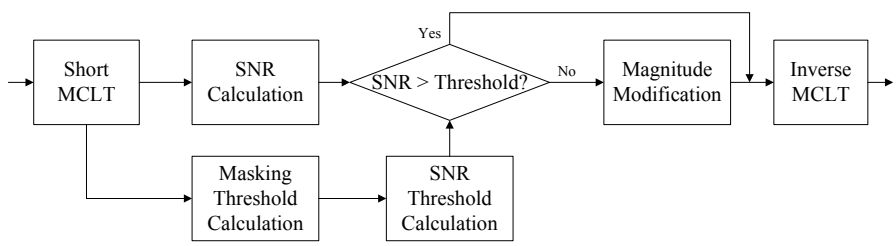


Figure 5.3 Block diagram of segmental SNR adjustment (SSA) algorithm.

system configurations are evaluated in terms of audio quality and transmission performance. The block diagram of the data embedding procedure with proposed SSA algorithm is shown in Fig. 5.2.

5.2 Segmental SNR Adjustment Algorithm

The block diagram of the proposed SSA algorithm is shown in Fig. 5.3. In this algorithm, an MCLT analysis with frame length M_s, which is shorter than the typical onset time of typical percussive sounds, is needed to calculate the segmental SNR of the data-embedded audio signal.

Let $X_{i_s}(k_s)$ and $X_{i_s}^D(k_s)$ denote the MCLT coefficients of the original and data-embedded audio signal, respectively, obtained from the MCLT analysis with frame length M_s. Here, i_s and k_s respectively indicate the frame and frequency indices, which are introduced in order to distinguish them from the long window-based MCLT analysis. The segmental SNR for the i_s-th MCLT frame and k_s-th frequency bin is defined as follows:

$$\text{SNR}_{i_s}(k_s) = 10\log_{10}\left(\frac{|X_{i_s}(k_s)|^2}{(|X_{i_s}(k_s)| - |X_{i_s}^D(k_s)|)^2}\right). \tag{5.1}$$

If $\text{SNR}_{i_s}(k_s)$ is smaller than a target segmental SNR, $\Gamma_{i_s}(k_s)$, the magnitude of $X_{i_s}^D(k_s)$ is modified such that

$$|\tilde{X}_{i_s}^D(k_s)| = \begin{cases} \frac{\Gamma_{i_s}(k_s)+1}{\Gamma_{i_s}(k_s)}|X_{i_s}(k_s)|, & |X_{i_s}(k_s)| < |X_{i_s}^D(k_s)| \\ \frac{\Gamma_{i_s}(k_s)-1}{\Gamma_{i_s}(k_s)}|X_{i_s}(k_s)|, & \text{otherwise,} \end{cases} \tag{5.2}$$

where $\tilde{X}_{i_s}^D(k_s)$ denotes the adjusted MCLT coefficient of data-embedded audio signal and this adjusted MCLT vector is then transformed back to a time domain signal and overlap-added.

The target segmental SNR $\Gamma_{i_s}(k_s)$ can be set to a constant value [22] or adaptively determined utilizing the signal-to-masking ratio (SMR) which represents the ratio between the magnitude of the MCLT coefficient and the masking threshold derived from a psychoacoustic model [38]. In this paper, $\Gamma_{i_s}(k_s)$ is defined as follows:

$$\Gamma_{i_s}(k_s) = \text{SMR}_{i_s}(k_s) + \Delta_{i_s}, \tag{5.3}$$

where the offset Δ_{i_s} is decided by upper-limiting the minimum SNR at the i_s-th frame.

The lower-bound of the SNR at i_s-th MCLT frame can be established when the segmental SNR $\text{SNR}_{i_s}(k_s)$ is smaller than the target segmental SNR, $\Gamma_{i_s}(k_s)$ for all

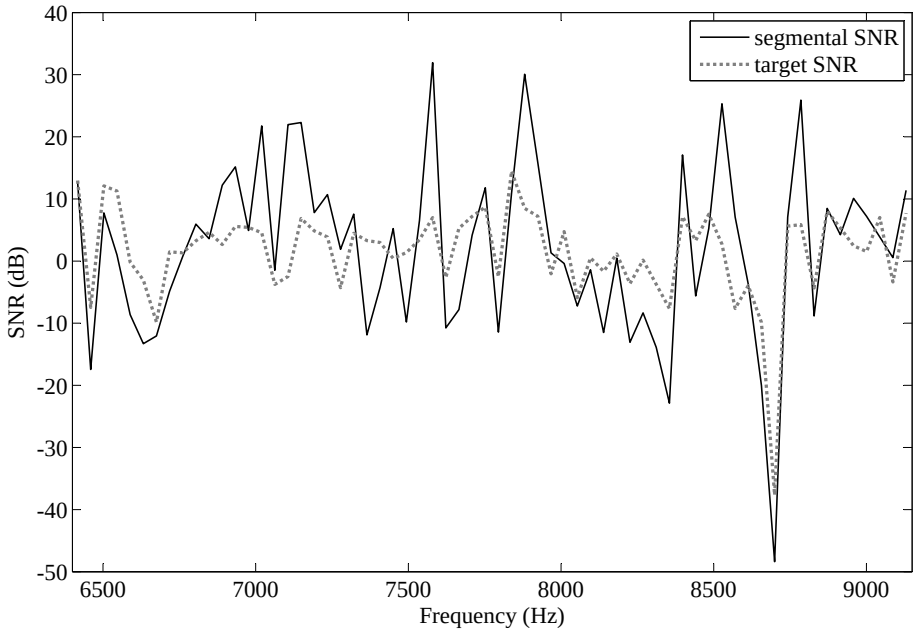


Figure 5.4 Example of segmental SNR and target SNR in MCLT domain. The segmental SNR is upper-limited to the target segmental SNR.

k_s. In this condition, all of $|X_{i_s}^D(k_s)|$ are replaced to $|\tilde{X}_{i_s}^D(k_s)|$ in (5.2) and the lower-bounded SNR can be obtained as follows:

$$\begin{aligned} \text{SNR}_{i_s,\min} &= 20\log_{10}\left(\frac{\sum_{k_s}|X_{i_s}(k_s)|}{\sum_{k_s}(|X_{i_s}(k_s)|10^{-\Gamma_{i_s}(k_s)/20})}\right) \\ &= 20\log_{10}\left(\frac{\sum_{k_s}|X_{i_s}(k_s)|}{\sum_{k_s}(|X_{i_s}(k_s)|10^{-\text{SMR}_{i_s}(k_s)/20})}\right) + \Delta_{i_s}, \end{aligned} \tag{5.4}$$

and the offset Δ_{i_s} is defined by

$$\Delta_{i_s} = S_t - 20\log_{10}\left(\frac{\sum_{k_s}|X_{i_s}(k_s)|}{\sum_{k_s}|X_{i_s}(k_s)|10^{(-\text{SMR}_{i_s}(k_s)/20)}}\right). \tag{5.5}$$

The parameter denoted by S_t is pre-defined to a target value of the minimum SNR for each MCLT frame with length M_s. Hence, S_t plays the role of making a trade-off between the audio quality and the transmission performance which can be determined

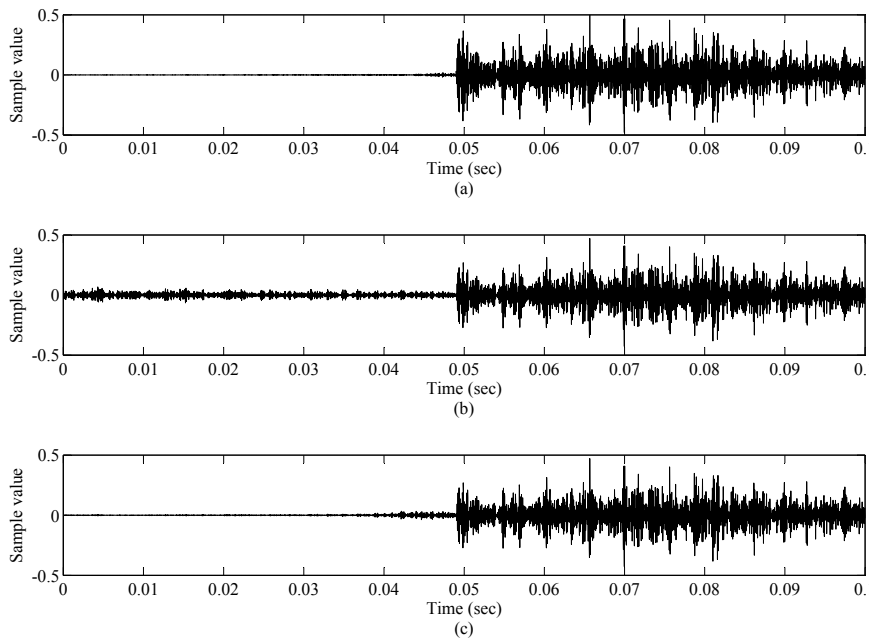


Figure 5.5 Example of pre-echo where the length of MCLT frame is 0.37 second: (a) host signal, (b) data-embedded signal, (c) data-embedded signal with segmental SNR adjustment algorithm.

experimentally; applying higher S_t indicates better audio quality with degraded transmission performance. An example of the segmental SNR and the target value is displayed in Fig. 5.4.

An example of mitigating the pre-echo by applying SSA algorithm is shown in Fig. 5.5. In this figure, comparing host and data-embedded signals, noise-like signal is detected in front of the onset (about 0-50 msec) in data-embedded audio. As can be seen from Fig. 5.5, however, the data-embedded audio signal when adjusted by the proposed SSA algorithm demonstrates attenuated pre-echo. Since the SSA algorithm adjusts the magnitude only, the phase alteration at the receiver is not severe.

Table 5.1 Parameters of system configurations

Parameters	S_{SP}	S_{SC}	S_L	S_{LA1}	S_{LA3}
M	512		8192		
Data frames per packet	16		2		
Pilot frames per packet	8		1		
Channel estimator	X	O			
SSA (S_t)	N/A		X	O (1 dB)	O (3 dB)
Bit rate (bps)	231				

5.3 Experimental Results

In order to evaluate the performance of the proposed SSA algorithm, experiments concerned with audio quality and data transmission performance were conducted. In the experiments in this chapter, the audio quality and BER's in a real classroom were obtained for five system configurations which will be presented in the following subsection. The robustness performance against various attacks incorporating typical signal processing and malicious removal attacks were also examined. The audio clips listed in Table 4.1 were used in the experiments, too. The average power over all the tested audio signals was adjusted to -18 dB in digital domain.

5.3.1 System Configurations

From the results in the previous chapter, it is considered doubtful to find an optimal MCLT length guaranteeing both good audio quality and transmission performance in reverberant condition. In order to further investigate the performance based on the results in the previous chapter, five different system configurations denoted by S_{SP}, S_{SC}, S_L, S_{LA1}, and S_{LA3} are implemented.

S_{SP} refers to the previous system proposed in Chapter 3 with short MCLT length,

and S_{SC} is the same to S_{SP} except that it adopts the packet structure shown in Chapter 4. S_L represents the system using a long MCLT window without the SSA algorithm, and S_{LA1} and S_{LA3} denote the systems using the SSA algorithm setting S_t to 1 dB and 3 dB, respectively. The length of the short MCLT window M_s for the SSA algorithm was set to 512. The system parameters of each configuration are specified in Table 5.1 and other parameters not shown in this table are same with those in Chapter 4.

5.3.2 Audio Quality Test

For subjective audio quality evaluation, MUSHRA test [48] was conducted. Thirteen listeners participated in this test. The results are shown in Fig. 5.6 where the average scores of the test audio clips are displayed in conjunction with 95 % confidence intervals. The scores for two anchor signals are omitted in the figure for a clear display.

As can be seen in Fig. 5.6, the average MUSHRA scores of S_{SP} and S_{SC} were 96.4 and 96.6, respectively. In the case of S_{SLA1} and S_{LA3}, the scores were 92.6 and 93.9, respectively, which still indicate good audio quality even though they are slightly lower than those of S_{SP} and S_{SC}. Comparing between S_{LA1} and S_{LA3}, we can see that using higher S_t usually gives rise to less quality degradation. The average score obtained from S_L was 81.0, which is quite worse than those of S_{LA1} and S_{LA3}. As can be seen in this figure, the average MUSHRA scores of S_{SP} and S_{SC} were slightly higher than those of S_{LA1} and S_{LA3}, which, however, indicate good audio quality except for S_L. Comparing between S_{LA1} and S_{LA3}, we can see that using higher S_t usually gives rise to less quality degradation. The average score obtained from S_L was quite worse than those of other configurations.

Because the average MUSHRA score of the proposed configurations are very high and the difference among them is very little except for S_L as can be seen from Fig. 5.6, the PEAQ test was additionally performed to each music clip. The obtained

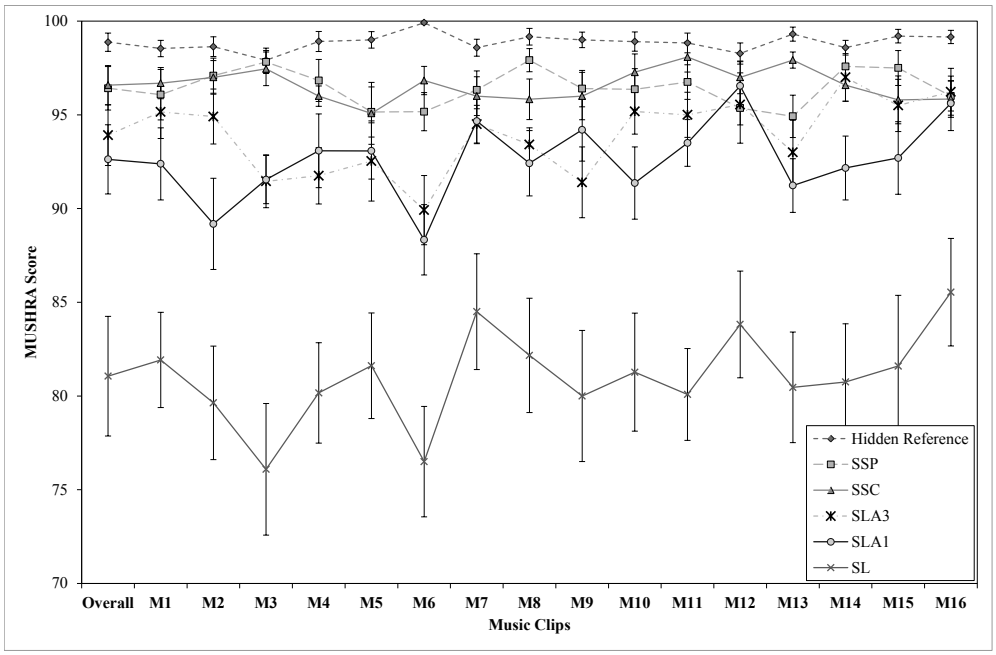


Figure 5.6 MUSHRA test scores with 95% confidence intervals according to test music clips. M1-M16 on horizontal axis represents the name of each music clip.

ODG scores are shown in Table 5.2. From the results, we can see that the ODG scores demonstrate similar tendency to the MUSHRA scores. The ODG scores obtained from S_{SP} and S_{SC} are almost the same since the embedding method is almost the same except for the permutations of the data and pilot blocks. Even though S_{LA1} and S_{LA3} produced slightly worse score than S_{SP} and S_{SC}, the ODG scores were above -0.5 which indicates that some of listeners might distinguish the data-embedded audio signals from the original ones without annoyance. The ODG score for S_L, however, was slightly lower than -1. In summary, the results of quality evaluation tests have shown that the SSA method is very useful for maintaining the quality of data-embedded audio.

Table 5.2 Objective difference score of test audio clips

	S_{SP}	S_{SC}	S_{L}	S_{LA1}	S_{LA3}
M1	-0.309	-0.292	-1.321	-0.477	-0.461
M2	-0.246	-0.267	-0.805	-0.448	-0.423
M3	-0.420	-0.422	-1.992	-0.565	-0.519
M4	-0.557	-0.531	-1.653	-0.636	-0.580
M5	-0.191	-0.239	-0.654	-0.423	-0.381
M6	-0.579	-0.566	-2.350	-0.664	-0.613
M7	-0.259	-0.243	-0.623	-0.383	-0.366
M8	-0.298	-0.326	-1.314	-0.532	-0.491
M9	-0.124	-0.134	-0.609	-0.430	-0.414
M10	-0.206	-0.250	-0.854	-0.492	-0.463
M11	-0.335	-0.352	-1.475	-0.535	-0.495
M12	-0.198	-0.140	-0.477	-0.339	-0.334
M13	-0.353	-0.328	-1.567	-0.600	-0.550
M14	-0.373	-0.389	-1.529	-0.585	-0.532
M15	-0.234	-0.246	-0.803	-0.492	-0.469
M16	-0.158	-0.199	-0.490	-0.361	-0.354
Average	-0.303	-0.308	-1.157	-0.496	-0.465

5.3.3 Robustness to Attacks

The data embedded in the audio signal should be robust to various attacks because the data-embedded audio signal can be distributed after being passed through several signal processing modules such as quantization, compression, and additive noise. The BER of attacked audio clips was measured and the results are shown in Table 5.3. In this table, we can see that S_{LA1} and S_{LA3} showed slightly higher BER's than others

because of the distortion due to the SSA algorithm. S_{LA1} showed slightly lower BER than S_{LA3}, which stands for the adversarial relationship between the audio quality and robustness. Nevertheless, we can see that the all of the configurations of the proposed method are robust to various attacks, which are comparable with other audio data hiding techniques [49], [51].

Table 5.3 BER of attacked test audio clips

	S_{SP}	S_{SC}	S_L	S_{LA1}	S_{LA3}
No Attack	0.0000	0.0000	0.0000	0.0008	0.0013
Requantize (8 bit)	0.0006	0.0011	0.0002	0.0016	0.0021
MP3 (64 kbps)	0.0053	0.0106	0.0047	0.0173	0.0199
AWGN (10 dB)	0.0388	0.0511	0.0386	0.0683	0.0753
Denoising (10 dB)	0.0225	0.0293	0.0190	0.0508	0.0579
All-pass filter	0.0047	0.0115	0.0129	0.0297	0.0334
Spread spectrum [30]	0.0007	0.0010	0.0000	0.0013	0.0019
Echo hiding [50]	0.0029	0.0076	0.0000	0.0012	0.0017
Phase coding [53]	0.0008	0.0015	0.0005	0.0025	0.0031
QIM [54]	0.0001	0.0007	0.0032	0.0140	0.0167

5.3.4 Transmission Performance of Recorded Signals in Indoor Environment

The transmission performance of the proposed approach was evaluated in actual room environments. For the performance measure, BER was used as in the previous experiments. The audio clips which were played back from a loudspeaker were recorded by a mobile phone at various distances in a real reverberant room of dimension 11 m $\times$ 7 m $\times$ 3 m. The loudspeaker is installed at the same position with that in Fig. 4.4. In this room environment, the average measured sound pressure level of

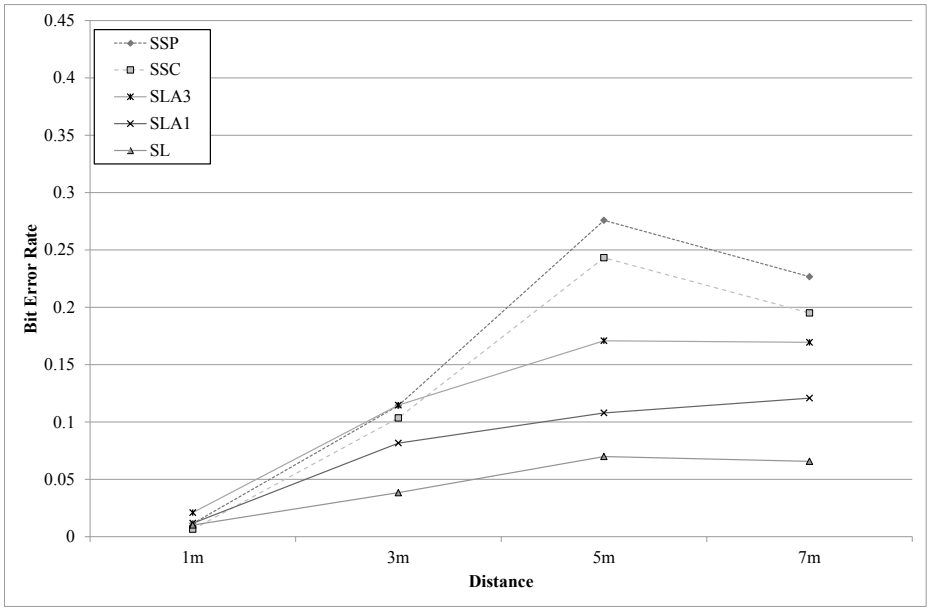


Figure 5.7 BER of the data transmission systems in a real room with respect to the distance between loudspeaker and microphone.

the background noise was 40 dB and that of the audio signal was 65 dB at 1 m from the front of the loudspeaker.

The results obtained from the actual room environments are shown in Fig. 5.7. As can be seen from this figure, all configurations showed good data transmission performance at 1 m distance from the loudspeaker. At 5 m and 7 m, however, S_L, S_{LA1}, and S_{LA3} using long MCLT length showed better data transmission performance than S_{SP} and S_{SC} with short MCLT length.

In the case of S_{SP} and S_{SC}, it is noted that the BER's at 5 m were higher than those at 7 m. This can be accounted for by the fact that the position corresponding to 7 m was close to the opposite wall which might somewhat reduce the reverberation. Also we can see that the channel estimation can reduce the data transmission error when

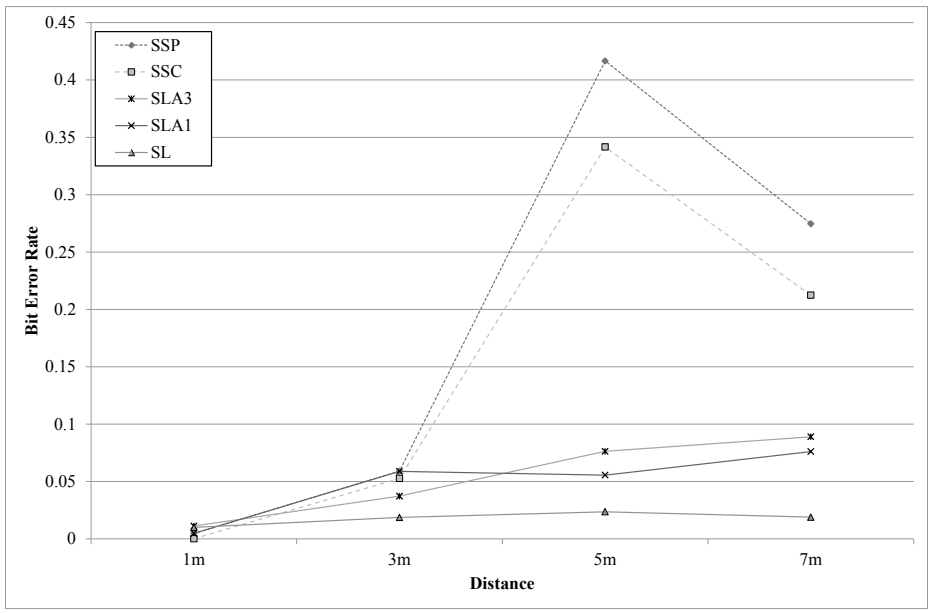


Figure 5.8 BER with 1/3 convolutional coding of the data transmission systems with respect to the distance between loudspeaker and microphone.

comparing the result of S_{SP} with that of S_{SC}. Similar to the previous experiment, S_{LA1} showed better transmission performance than S_{LA3}.

Summarizing the results, it can be said that using long MCLT frame length in conjunction with the SSA algorithm can enhance the transmission performance without significant audio quality degradation. In addition, applying different value of S_t can trade off between the audio quality and the data transmission performance; higher S_t makes the audio quality better at the cost of increasing BER.

5.3.5 Error correction using convolutional coding

The performance of FEC coding usually depends on the error pattern and the length of a data packet. The performance of a FEC coding technique, however, is

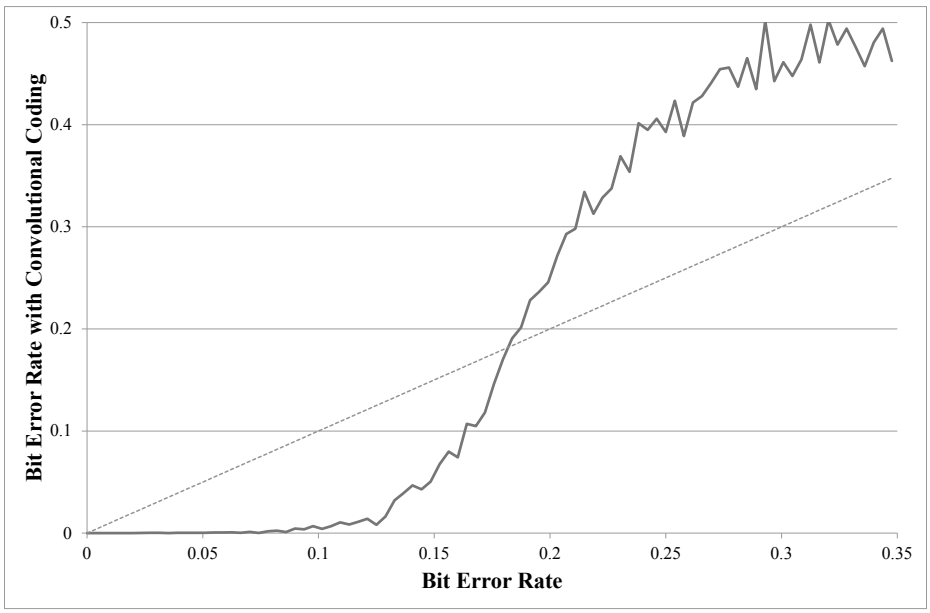


Figure 5.9 BER with 1/3 convolutional coding as a function of BER (solid line). Dotted line denotes the identity line.

usually given in the form of the statistical relationship of BER versus SNR [52], which is however inappropriate for this work since the length of a data packet is much shorter than in typical communication system. For this reason, the BER corrected by convolutional coding with code rate 1/3, one of the most famous FEC coding algorithms, was obtained from the same recorded audio clips used in the previous experiment and the results are displayed in Fig. 5.8.

Comparing with the result in Fig. 5.7, the BER's in Fig. 5.8 are usually decreased, however those obtained from S_{SP} and S_{SC} at 5 m and 7 m become larger than the case without FEC coding. The reason can be accounted for by investigating the relationship between the BER with and without convolutional coding as shown in Fig. 5.9. In this figure, it is observed that the BER is increasing dramatically with the use of the

convolutional coding when it is greater than a certain threshold, which is a common phenomenon of digital communication [52].

5.4 Summary

In this chapter, to cope with the restriction between the audio quality and data transmission performance, the segmental SNR adjustment (SSA) method for the long MCLT window is introduced. The SSA algorithm which further modify the spectral components after data embedding can attenuate the pre-echo effectively. In order to further investigate the experimental results in the previous chapter in real environments, the experiments are conducted for the five different system configurations including previously proposed ones in the previous chapters. The experimental results have been shown that all of the configurations are found to be robust against not only reverberant environment but also signal processing attacks related to typical data hiding applications. The tendency of the audio quality and transmission performance according to the length of MCLT is also shown to be similar in terms of subjective audio quality and bit error rate in a real room, respectively. Especially, using long MCLT window with the SSA algorithm makes the audio data hiding system more suitable for the acoustic data transmission application because it can transmit data in reverberant environments while preserving good audio quality. In addition, a good trade-off between the audio quality and data transmission performance can be achieved by adjusting only a single parameter in the SSA algorithm.

Chapter 6

Multichannel Acoustic Data Transmission

6.1 Introduction

Due to the destructive addition of multipath propagation, the channel usually suffer from heavy attenuation at a certain frequency or time, which is the most severe problem for wireless data transmission. To cope with the heavy attenuation of the channel, the diversity denoting various techniques to transmit or receive duplicates through statistically independent channel are exploited. If each channel is independent, the probability that the heavy attenuation occurs simultaneously is very low and the receiver can detect message more reliably.

The diversity can be subdivided into time diversity, frequency diversity, and spatial diversity [35]. One of the most widely used to achieve time or frequency diversity is to spread same data along time or frequency axis, which have been already adopted for data embedding through the previous chapters. The spatial diversity in the wireless communication represents that each antenna sends messages at the spa-

tially separated position enough to guarantee the statistically independence of the corresponding channel (typically more than ten times of the wavelength) [57]. If the number of antenna is more than one, the spatial diversity could be useful to increase data rate or enhance the transmission reliability. The spatial diversity is playing a major role to realize the fourth-generation wireless communication systems, which each transmitter or receiver is assumed to use multiple antennas. In an acoustic data transmission system, loudspeakers and microphones can be respectively equivalent with antennas at the transmitter and receiver, which, consequently, exploiting the spatial diversity techniques would be available.

In this chapter, the acoustic data transmission technique is extended to multichannel scheme which take advantage of the multi-loudspeakers and multi-microphones. First, the spatial diversity techniques for the multi-loudspeaker and multi-microphone are introduced. Second, combining-based technique is proposed for transmitting data with lower errors. The most noticeable property for combining-based technique is that it provides compatibility with the acoustic data transmission system using a single microphone when each loudspeaker sends same information.

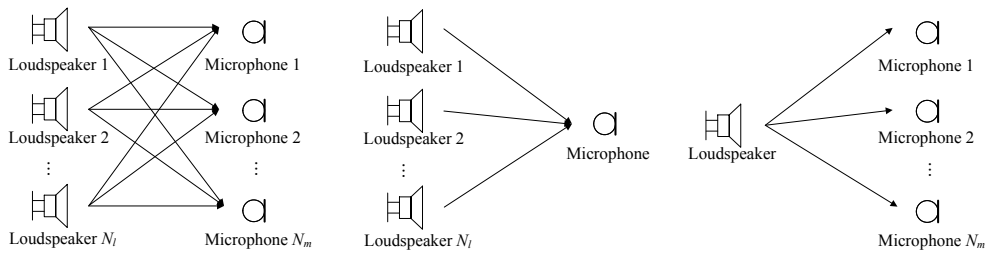
6.2 Multichannel Techniques for Robust Data Transmission

6.2.1 Diversity Techniques for Multichannel System

Multiple Input Multiple Output (MIMO) Channel

In order to exploit the spatial diversity in acoustic data transmission, it should be assumed that more than one loudspeakers and/or microphone is used. In general, the channel in this situation can be referred to multiple input multiple output (MIMO) channel. If N_L loudspeakers and N_M microphones are considered, the channel capacity of MIMO channels C [58] can be simplified by

$$C = \log_2 \left[\det \left(I_{N_M} + \frac{\text{SNR}}{N_L} \mathbf{H}\mathbf{H}^H \right) \right] (\text{bps/Hz}), \tag{6.1}$$



(a) MIMO (b) MISO (c) SIMO

Figure 6.1 Examples of channel configurations.

where $\det(.)$ means determinant, I_{N_M} represents an $N_M \times N_M$ identity matrix, and $\mathbf{H}$ is the channel matrix of dimension $N_M \times N_L$ with ^{H} denoting its Hermitian matrix. In (6.1), we can see that it is theoretically available to transmit more amount of data reliably if more loudspeakers or microphones are installed with sufficient distance achieving statistically independent channel. As a special case, when the number of loudspeakers and microphones are same, which implies $N_L = N_M = N$, then the MIMO channel capacity is maximized to N times of each parallel channel when each channel is orthogonal [57]. If $N_L = 1$, the channel results in single input multiple output (SIMO) channel, and if $N_M = 1$, the channel can be defined as multiple input single output (MISO) channel. The channel configurations are illustrated in Fig. 6.1.

Delay and Phase Diversity

Delay and phase diversity introduce additional constructive signals with different delay or phase to artificially increase the frequency selectivity of the channel [60], [61]. If the frequency selectivity is increased with same power, some of the frequency components would be less attenuated, which enables a better exploitation of the diversity techniques. The delay and phase diversity can be applied both transmitter and receiver.

Table 6.1 Mapping with spcae-frequency block codes for two loudspeakers

MCLT index	Loudspeaker 1	Loudspeaker 2
k	$b(k)$	$-b^*(k+2)$
$k+2$	$b(k+2)$	$b^*(k)$

Space-frequency Block Coding

When the number of microphones is restricted to be equal or more than that of loudspeakers, increasing data rate is available based on multiplexing method. However, this is considered unrealistic. If the number of loudspeakers are known and there is no assumption for the number of microphones, transmit diversity techniques such as a space-time (or space-frequency) coding can be applied to increase the reliability of data transmission. For the multicarrier system which use relatively lengthy frame, the space-frequency block coding (SFBC) would be preferable among the transmit diversity techniques. The SFBC requires only one received data frame for detection, which can retain a flexible data packet structure [59]. The basic idea of SFBC is to provide constructive superposition from the transmitted signal played back at different loudspeakers.

A brief block diagram of the SFBC is shown in Fig. 6.2. For the simplicity, it is assumed that the number of loudspeakers N_L are two and the number of microphones N_M is one (MISO channel). The data bit for k-th frequency at a message frame $b(k)$ is embedded by following the mapping scheme shown in Table 6.1 where * denotes the conjugate operation. Please note that $k+2$ is used instead of $k+1$ because the data should be embedded every other frequency in the proposed method as shown in Fig. 4.2.

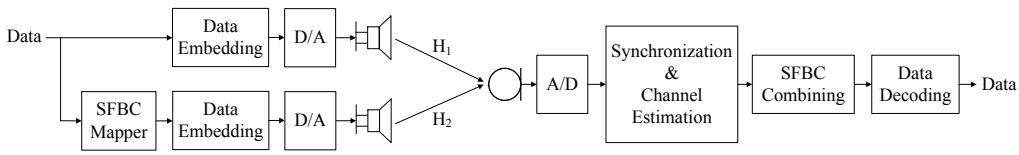


Figure 6.2 Block Diagram of space-frequency block coding (SFBC) for multichannel acoustic data transmission.

The received data coefficient $\hat{b}^R(k)$ and $\hat{b}^R(k+2)$ can be derived as

$$\begin{aligned}\hat{b}^R(k) &= H_1(k)|X_1(k)|b(k) - H_2(k)|X_2(k)|b^*(k+2) + N(k)\\ \hat{b}^R(k+2) &= H_1(k+2)|X_1(k+2)|b(k+2)\\ &\quad + H_2(k+2)|X_2(k+2)|b^*(k) + N(k+2),\end{aligned} \tag{6.2}$$

where $H_l(k)$ and $|X_l(k)|$ refer to the channel coefficient and the magnitude of MCLT coefficient for k-th frequency bin at l-th loudspeaker, and $N(k)$ is the additive noise. If it can be assumed that $H_l(k) = H_l(k+2) = H_l$, the received data is combined as follows:

$$\begin{aligned}R(k) &= \frac{H_1^*}{|H_1|}\hat{b}^R(k) + \frac{H_2}{|H_2|}\hat{b}^{R*}(k+2)\\ R(k+2) &= -\frac{H_2}{|H_2|}\hat{b}^{R*}(k) + \frac{H_1^*}{|H_1|}\hat{b}^R(k+2),\end{aligned} \tag{6.3}$$

where the combined signal result in

$$\begin{aligned}R(k) &= \{|H_1X_1(k)| + |H_2X_2(k+2)|\}b(k) + \tilde{N}(k)\\ R(k+2) &= \{|H_1X_1(k+2)| + |H_2X_2(k)|\}b(k+2) + \tilde{N}(k+2),\end{aligned} \tag{6.4}$$

where $\tilde{N}(k)$ is a residual noise. If H_1 and H_2 are statistically independent, the probability that $R(k)$ or $R(k+2)$ is highly attenuated would be very low and this implies that the SNR would increase.

6.2.2 Combining-based Multichannel Acoustic Data Transmission

In order to estimate each channel coefficient to decode SFBC, the position of the pilot coefficients for each channel should not be duplicated at frequency or time interval. For example, if a pilot is located at a certain MCLT coefficient for a loudspeaker, the MCLT coefficient at that position for other loudspeakers should be zero. Although these vacant components are essential to estimate all channel coefficients, these are not acceptable in the audio data hiding because this components can incur severe audio quality degradation. The SFBC, therefore, could be useful for the acoustic data transmission systems which exploits inaudible frequency higher than 18 kHz. However, it is considered doubtful for the SFBC to apply directly to the acoustic data transmission systems based on the audio data hiding. To exploit the transmit diversity such as the SFBC, furthermore, the number of loudspeaker should be known to the receiver, which results in the restricted range of applications.

In this work, the multichannel acoustic data transmission based on receive diversity is proposed. Consider that it is assumed that the number of loudspeakers is unknown and multi-microphone at a receiver is used (SIMO channel). The channel capacity in SIMO channel increases logarithmically when microphones are installed at the statistically independent channel [57]. In order to achieve the compatibility with the single microphone system, however, increasing data rate is considered doubtful and the multichannel techniques in this chapter should focus on increasing transmission reliability against channel. Therefore, combining-based multichannel method is used to enhance the transmission performance. The most important characteristic for combining-based multichannel method is that embedding procedure is exactly same and it is compatible with the receiver using a single microphone. To guarantee the compatibility, the data signal should be same for all of the loudspeakers.

The block diagram of the combining-based multichannel method is shown in Fig. 6.3. For the simplicity, it is assumed that the number of loudspeakers N_L is one and

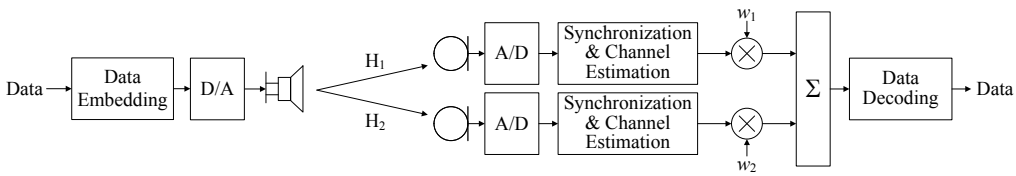


Figure 6.3 Block Diagram of combining-based method for multichannel acoustic data transmission.

the number of microphones N_M are two (SIMO channel). In the combining method, the signals captured at different microphones are combined linearly to increase SNR. After the synchronization at the receiver, the received MCLT coefficient at m-th microphone $\hat{Y}^{R_m}(k)$ can be approximated as follows:

$$\begin{aligned} \hat{Y}^{R_1}(k) &= H_1(k)|X(k)|b(k) + N_1(k) \\ \hat{Y}^{R_2}(k) &= H_2(k)|X(k)|b(k) + N_2(k), \end{aligned} \tag{6.5}$$

where $H_m(k)$ and $N_m(k)$ refer to the channel coefficient and the additive noise for k-th frequency bin at m-th microphone, and $|X(k)|$ is the magnitude of MCLT coefficient.

The combined coefficient can be given by

$$\hat{Y}^{R}(k) = w_1(k)\hat{Y}^{R_1}(k) + w_2(k)\hat{Y}^{R_2}(k), \tag{6.6}$$

and each weighting factor $w_1(k)$ and $w_2(k)$ can be decided by various combining techniques. The most frequently used technique is maximal ratio combining (MRC), selective combining (SC), and equal gain combining (EGC) [36]. MRC is an optimal solution to maximize the SNR after combining. EGC is a special case of MRC where all signals obtained from different microphones are combined with equal amplitudes.

The weighting factor of MRC and EGC are defined by

$$\begin{aligned} w_{m,\text{MRC}}(k) &= H_m^*(k), \\ w_{m,\text{EGC}}(k) &= \frac{H_m^*(k)}{|H_m(k)|}, \end{aligned} \tag{6.7}$$

and weighting factor of SC $w_{m,\text{SC}}(k)$ is same with $w_{m,\text{EGC}}(k)$ for selective channel and zero for the other channels.

The combined coefficient respectively results by

$$\hat{Y}^R(k) = G(k)|X(k)|b(k) + \tilde{N}(k) \tag{6.8}$$

where $G(k)$ and $\tilde{N}(k)$ denote a real-valued gain and a residual noise, respectively. In this work, EGC might be preferable because it is hard to estimate the amplitude of the channel exactly due to the magnitude spectrum of the audio signal.

6.3 Experimental Results

In this chapter, the transmission performance of the combining-based multichannel technique was evaluated for three system configurations S_{SC}, S_{LA1}, and S_L. The system parameters are listed in Table 5.1. In this experiment, BER and relative error reduction rate (RERR) were calculated in a simulated room and in a real reverberant room when single or stereo loudspeakers are installed. The RERR refers to the ratio of the reduced error rates to the baseline error rates. It is worth to evaluate the performance of multichannel acoustic data transmission system with multi-loudspeaker environments because the audio is usually played back with two or more loudspeakers in the practical applications. Moreover, the effect of the distance between two microphones at the receiver was evaluated by calculating RERR while varying parameters. The audio clips listed in Table 4.1 with length 30 seconds were used in the experiments. The average power over all the tested audio signals was adjusted to -18 dB in digital domain.

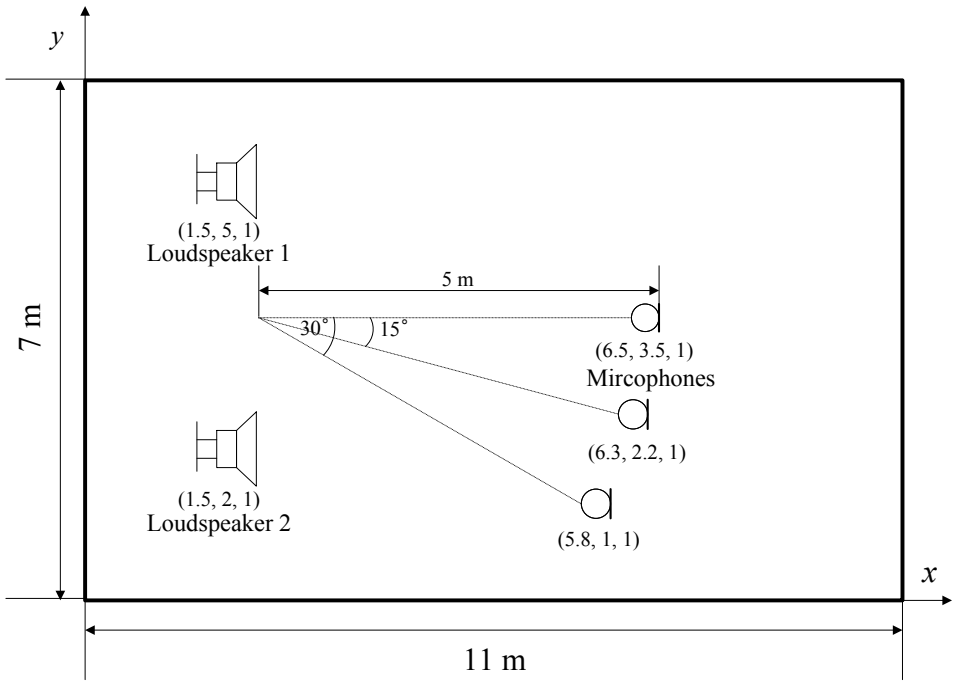


Figure 6.4 Room environment with location of two loudspeakers and microphones. Height of the room is 3 m. At the below of each installation, Cartesian coordinate is written.

6.3.1 Room Environments

In this experiments, the audio signals were convolved with the room impulse response from a simulated room and they were also recorded in a real reverberant room with same dimension and installations. The dimension of the room was 11 m $\times$ 7 m $\times$ 3 m and the reflection coefficient of the simulated room was set to -0.9, which assumes the severely reverberant room. For the single loudspeaker case, the installed positions of the loudspeaker and main microphones were same with the positions shown in Fig. 4.4. In order to evaluate how the multi-loudspeaker environments effect on the performance of the proposed method, stereo loudspeakers are installed at the position shown in Fig. 6.4. As can be seen in this figure, the main microphone is

installed at 5 m from the center of two loudspeakers with degree 0°, 15°, and 30°.

6.3.2 Transmission Performance of Simulated Environments

To examine how the combining multichannel methods have effect in the sever conditions, the audio signals obtained by S_{SC}, S_{LA1}, and S_L were convolved with the simulated room impulse response (RIR) [45]. The auxiliary microphone for combining is installed such that it is at the distance 15 cm along y-axis from the main microphone, which is assumed that the receiver is a mobile phone operating with two microphones. In this simulation, the white noise is added with SNR 20 dB at the receiver. The results obtained form different combining method are shown in Table 6.2. From the results, it is found that EGC reduced the bit errors most. Although MRC is the optimal solution, it showed the worst performance due to the estimation error of the amplitude of channel; it is unlikely to estimate the amplitude of channel correctly because of the audio spectrum. The RERR of S_{SC} is lower than that of the others and it could be explained that S_{SC} is found to be fragile to the reverberation.

In order to examine the effect of combining method more precisely, BER's for EGC were calculated and the results are shown in Fig. 6.5. In this figure, we can see that the combining method can reduce BER in the reverberant environment regardless of the MCLT length. For the stereo loudspeakers case, the BER's without combining method at 15° were higher than those at 0° due to the destructive superposition of acoustic signals radiated from different loudspeakers. The BER's with combining method, however, were almost similar regardless of the location of the microphones except for S_{SC}. As a result, we can see that the combining method is also effective to transmit data with two loudspeakers more reliably.

The statistical independence of the channel between two received signals can result in the further improvement of the transmission performance, and the sufficient distance between receiving microphones would be able to actualize the channel inde-

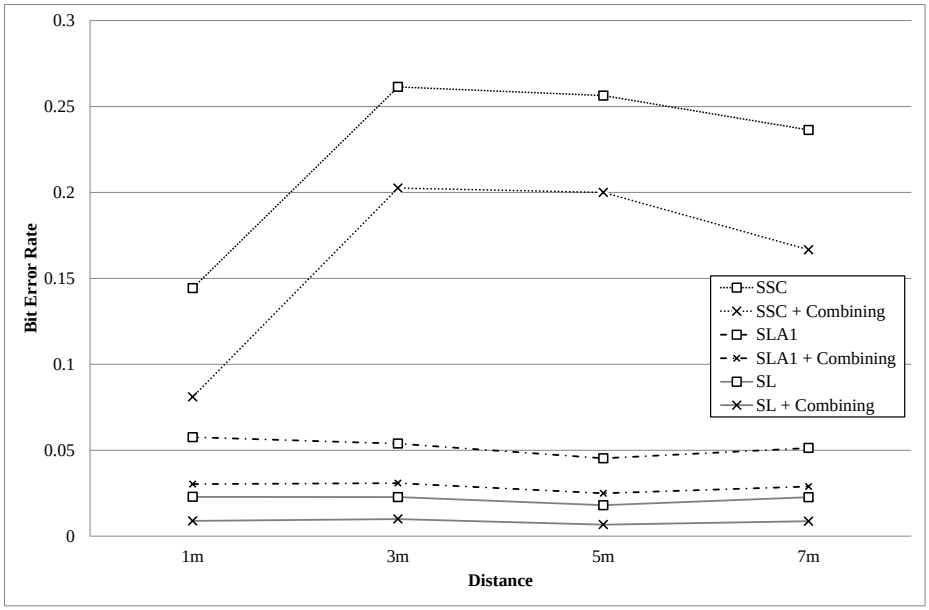
Table 6.2 RERR of the data transmission systems in a simulated room with different combining methods

	single loudspeaker			stereo loudspeakers		
RERR (%)	S_{SC}	S_{LA1}	S_L	S_{SC}	S_{LA1}	S_L
EGC	26.37	44.41	59.17	20.67	42.38	60.18
SC	14.84	27.48	38.33	12.64	25.88	40.24
MRC	7.24	17.47	22.06	0.32	8.49	14.40

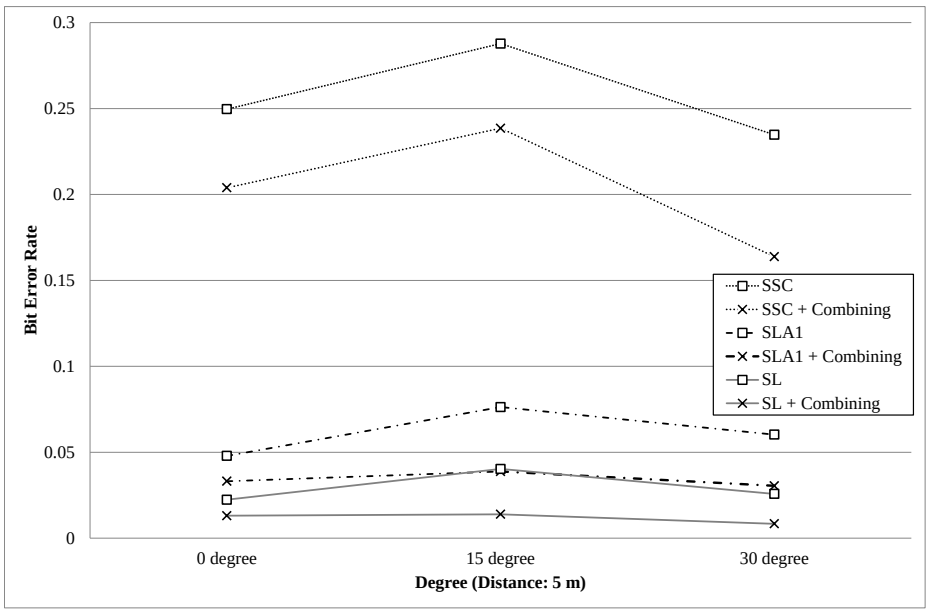
Table 6.3 Average RERR of the data transmission systems in a simulated room with combining method versus distance between two microphones of a receiver

RERR (%)	S_{SC}	S_{LA1}	S_L
15 cm	27.03	44.85	60.37
40 cm	23.73	45.65	60.71

pendence. Most of the applications of the acoustic data transmission is assumed that a receiver is a mobile device, however, there is a upper-limitation for the distance between microphones. In this experiment, therefore, the RERR's were calculated with respect to the distance between microphones 15 cm and 40 cm, respectively. The distance 40 cm presumes the upper-limit for the mobile receiver such as a laptop PC with two microphones. The RERR's were obtained at the distance between the loudspeaker and microphones 1 m, 3 m, 5 m, and 7 m and their averaged values are shown in Table. 6.3. From the result, it is doubtful that significant differences have been found when the RERR's for 15 cm are compared with those for 40 cm.



(a)



(b)

Figure 6.5 BER of the data transmission systems in a simulated room with equal gain combining method: (a) single loudspeaker and (b) stereo loudspeakers.

Table 6.4 RERR of the data transmission systems in a real room with different combining methods

	single loudspeaker			stereo loudspeakers		
RERR (%)	S_{SC}	S_{LA1}	S_L	S_{SC}	S_{LA1}	S_L
EGC	39.85	48.53	69.63	45.75	54.63	77.06
SC	23.79	33.17	49.96	29.95	39.79	57.30
MRC	37.66	45.89	66.62	42.78	52.29	73.62

6.3.3 Transmission Performance of Recorded Signals in Reverberant Environment

In this experiment, the transmission performance of the combining method was evaluated. The audio signals obtained from S_{SC}, S_{LA1}, and S_L were recorded in a real reverberant room of dimension 11 m $\times$ 7 m $\times$ 3 m. The position of loudspeakers and microphones were same with previous experiments and they were shown in Figs. 4.4 and 6.4. The distance of auxiliary microphone for combining was 15 cm along y-axis from the main microphone. The average measured sound pressure level of the background noise was 40 dB and that of the audio signal was 65 dB at 1 m in front of the loudspeaker. The equipment of this experiment were Blitz BR1500 (loudspeakers) and AKG C1000S (microphones).

The RERR's were calculated for different combining method from the recorded audio signals and the results were shown in Table 6.4. From the results, we can see that the combining method can reduce the bit error. The RERR's of MRC were higher than those of SC, which is different from the simulated results shown in Table 6.2. This could proceed from the fact that the magnitude response of the loudspeakers and microphones can be estimated relatively correctly whose ratio would be dominant factor for MRC weighting factor. However, EGC is also preferable in a real condition.

Table 6.5 Average RERR of the data transmission systems in a simulated room with combining method versus distance between two microphones of a receiver

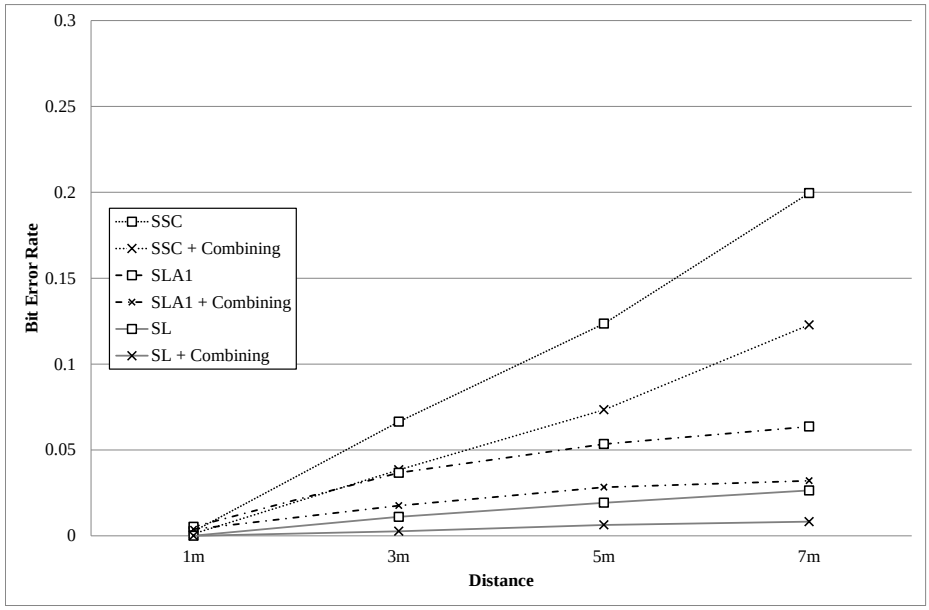
RERR (%)	S_{SC}	S_{LA1}	S_L
15 cm	39.85	48.53	69.62
40 cm	40.61	49.25	70.43

In this experiment, the BER's for EGC and RERR's for different distance between two microphones were also evaluated and they were shown in Fig. 6.6 and Table 6.5. From the results, the combining method is considered effective to reduce the bit errors regardless of the system configurations. The BER's at the 7 m were higher than those at the 5 m, which is different from the results shown in Fig. 5.7. The main reason was that the microphones are cardioid and they cannot record the audio signals reflected by the opposite wall. In Table 6.5, there are only small enhancement when the distance between two microphones is increased to 40 cm. Therefore, it can be concluded that it would be hard to achieve greater statistical independence between the channels when the receivers are restricted to the mobile devices.

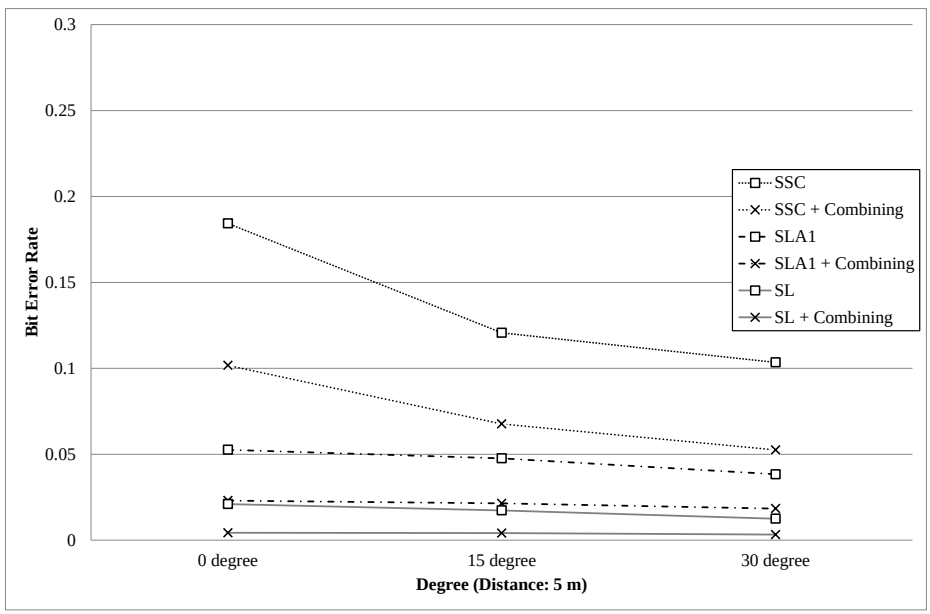
6.4 Summary

In this chapter, multichannel techniques for the acoustic data transmission are delineated and combining-based multichannel method is proposed. The proposed method is the most practical solution under the restriction that the data should be embedded into audio signals inaudibly. In the proposed method, the synchronization and channel estimation are performed each for received signal and the received signals are linearly combined with weighting factor. It is noted that the proposed method is compatible with the single channel systems. From the series of the experiments, the combining-based method can enhance the reliability of the acoustic channel in aerial

space regardless of the MCLT length and the usage of the multi-loudspeaker. However, achieving greater statistical characteristic of the channel is doubtful when the distance between two microphones at the receiver is restricted to the range of typical mobile devices. Despite of the dependency of each channel, the combining-based multichannel method have shown to be useful to apply to the practical application.



(a)



(b)

Figure 6.6 BER of the data transmission systems in a real room with equal gain combining method: (a) single loudspeaker and (b) stereo loudspeakers.

Chapter 7

Conclusions

In this dissertation, audio data hiding methods suitable for acoustic data transmission which communicates data in short-range aerial space between a loudspeaker and a microphone are studied. To overcome the limitations of the previous techniques, a novel audio data hiding method for the acoustic data transmission using MCLT is proposed with additional techniques to improve the performance. Because of this overlapping property, the perceived quality of the data embedded audio signal can be kept almost similar to that of the original audio while transmitting data at several hundreds of bits per second. The experimental results have shown that the audio quality and transmission performance of proposed system are better than those of the AOFDM based system. The proposed audio data hiding method could overcome the limitation of AOFDM which have shown the severe quality degradation for speech or classical music.

Moreover, an audio data hiding technique is extended to be more suitable for acoustic data transmission purpose by incorporating a number of sophisticated techniques widely deployed in wireless communication and audio watermarking. From

the experimental results, it can be found that the length of MCLT window is one of the most important parameters for the practical acoustic data transmission system. Because the length of MCLT window is a fixed parameter known to both the embedder and receiver, it should be determined carefully depending on the target applications. Large MCLT window, for example, can achieve a good transmission performance in a reverberant environment such as living room or cafeteria although the data can be fragile to a possible motion of the receiver or to the playback speed mismatch. For the applications dedicated to very short distance such as device-to-device data transmission, however, short MCLT window might be preferable due to the better audio quality and robustness to the movement of the devices.

The most important problem for the method would be that it is almost impossible to find a proper window length satisfying both the inaudible distortion and robust data transmission in the reverberant environments due to the pre-echo phenomena. To overcome the limitation, segmental SNR adjustment (SSA) technique with trade-off parameter is proposed to further modify the spectral components for attenuating the pre-echo to ameliorate the audio quality degradation resulted from applying a long MCLT length. Using long MCLT window with the SSA algorithm makes the audio data hiding system more suitable for the acoustic data transmission application because it can transmit data in reverberant environments while preserving good audio quality. In addition, a good trade-off between the audio quality and data transmission performance can be achieved by adjusting only a single parameter in the SSA algorithm. As a result, the SSA algorithm is an essential component of the practical acoustic data transmission system to transmit data reliably in reverberant environment while preserving the audio quality.

If number of loudspeaker or microphone is more than one, the diversity technique which takes advantage of transmitting duplicates through statistically independent channel could be useful to increase data throughput or enhance the transmission

reliability. The combining-based multichannel method, which is the most practical solution under the restriction that the data should be embedded into audio signals inaudibly is proposed. The most noticeable property for combining-based technique is that it provides compatibility with the acoustic data transmission system using a single microphone. Because the mobile devices for receiver usually have two microphones, the proposed multichannel method would be more valuable; it can enhance the transmission performance without any additional costs in hardware.

The experimental results have shown that the audio data hiding method proposed through this dissertation can be applied to the practical acoustic data transmission solutions. Although the proposed methods with lengthy window are doubtful to be robust against Doppler shift, the solution of desynchronization attacks in audio watermarking such as time or pitch scaling would be useful to find the exact location in frequency domain exhaustively [30]. Furthermore, the proposed methods have a possibility to be applied to the typical applications of audio data hiding, which will be one of the future directions of this study.

Bibliography

[1] N. Cvejic and T. Seppänen, *Digital Audio Watermarking Tehcniques and Technologies*, IGI Global, 2008.

[2] I. J. Cox, M. L. Miller, J. A. Bloom, J. Fridrich, and T. Kalker, *Digital Watermarking and Steganography*, Elsevier, 2008.

[3] Y. Nakashima, R. Tachibana, and N. Babaguchi, "Watermarked movie soundtrack finds the position of the camcorder in a theater," *IEEE Trans. Multimedia*, vol. 11, no. 3, pp. 443-454, Apr. 2009.

[4] Daniel Gruhl, Anthony Lu and Walter Bender, "Echo hiding," in *Pre- Proceedings: Information Hiding*, pp. 295-316, May 1996.

[5] Y. Suzuki, R. Nishimura, and H. Tao, "Audio watermarking enhanced by LDPC coding for air transmission," in *Proc. IEEE Int. Conf. IIH-MSP'06*, pp. 23-26, Dec. 2006.

[6] N. Lazic and P. Aarabi, "Communication over an acoustic channel using data hiding techniques," *IEEE Trans. Multimedia*, vol. 8, no. 5, pp. 918-924, Oct. 2006.

[7] P.-W. Chen, C.-H. Huang, Y.-C. Shen and J.-L. Wu, "Pushing information over acoustic channels," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 1421-1424, Apr. 2009.

[8] G. D. Galdo, J. Borsum, T. Bliem, A. Carciun, and S. Krägeloh, "Audio watermarking for acoustic propagation in reverberant environments," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 2364-2367, May. 2011.

[9] Y. F. Huang, S. Tang, and J. Yuan, "Steganography in inactive frames of VoIP streams encoded by source codec," *IEEE Trans. Information Forensics and Security*, vol. 6, no. 2, Jun. 2011.

[10] C. V. Lopes and P. M. Q. Aguiar, "Aerial acoustic communication," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 219-222, Oct. 2001.

[11] K. Mizutani, N. Wakatsuki, and K. Mizutani, "Acoustic communication in air using differential biphase shift keying with influence of impulse response and background noise," *Japan Journal of Applied Physics*, vol. 46, no. 7B, pp. 4541-4544, Jul. 2007.

[12] Y. Nakashima, H. Matsuoka, and T. Yoshimura, "Evaluation and demonstration of acoustic OFDM," in *Proc. Fortieth Asilomar Conf. on Signals, Systems and Computers*, pp. 1747-1751, Oct.-Nov. 2006.

[13] H. Matsuoka, Y. Nakashima and T. Yoshimura, "Acoustic communication system using mobile terminal microphones," *NTT DoCoMo Technical journal*, vol. 8, no. 2, pp. 2-12, Sep. 2006.

[14] H. Matsuoka, Y. Nakashima and T. Yoshimura, "Acoustic communication with OFDM signal embedded in Audio," in *Proc. Int. Conf. AES 29th* , pp. 1-6, Sep. 2006.

[15] H. Matsuoka, Y. Nakashima, T. Yoshimura and T. Kawahara, "Acoustic OFDM: Embedding high bit-rate data in audio," in *Proc. Int. Conf. MMM 2008*, pp. 498-507, 2008.

[16] H. Matsuoka, Y. Nakashima, and T. Yoshimura, "Acoustic OFDM system and performance analysis," *IEEE Trans. Fundamentals*, vol. E91-A, no. 7, pp. 1652-1658, Jul. 2008.

[17] H. Matsuoka, Y. Nakashima, and T. Yoshimura, "Acoustic OFDM system and its extention," *Visual Computer*, vol. 12, no. 1, pp. 3-12, Jan. 2009.

[18] K. Cho, H. S. Yun, J.-H. Chang and N. S. Kim, "An analysis on audio quality deterioration of acoustic OFDM," *Journal of the Acoustical Society of Korea*, vol. 28, no. 2, pp. 107-111, Jan. 2010.

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

[20] K. Cho, H. S. Yun, and N. S. Kim, "Robust data hiding for MCLT based acoustic data transmission," *IEEE Signal Processing Letter*, vol. 17, no. 7, pp. 679-682, Jul. 2010.

[21] H. S. Yun, K. Cho, and N. S. Kim, "Spectral magnitude adjustment for MCLT-based acoustic data transmission," *IEICE Transactions on Information and Systems*, vol. E95-D, no. 5, pp. 1523-1526, May. 2012.

[22] K. Cho, J. Choi, Y. G. Jin, and N. S. Kim, "Quality enhancement of audio watermarking for data transmission in aerial space based on segmental SNR adjustment," in *Proc. IEEE Int. Conf. IIH-MSP'12*, pp. 122-125, Jul. 2012.

[23] H. S. Yun, *Data Hiding Techniques for MCLT-based Acoustic Data Transmission*,' Ph. D. Dissertation, EECS Department, Seoul National University, 2011.

[24] S. Shlien, "The modulated lapped transform, its time-varying forms, and applications to audio coding standards," *IEEE Trans. Speech Audio Process.*, vol. 5, pp. 359-366, Jul. 1997.

[25] H. S. Malvar, "A modulated complex lapped transform and its applications to audio processing," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 1421-1424, Mar. 1999.

[26] Q. Dai and X. Chen, "New algorithm for modulated complex lapped transform with symmetrical window function," *IEEE Signal Processing Letter*, vol. 11, pp. 925-928, Dec. 2004.

[27] V. Britanak, "An efficient computing of oddlystacked MDCT/MDST via evenlystacked MDCT/MDST and vice versa," *Signal Processing*, vol. 85, issue. 7, pp. 1353-1374, Jul. 2005.

[28] H. S. Malvar, "Fast algorithm for the modulated complex lapped transform," *IEEE Signal Processing Letter*, vol. 10, no. 1, pp. 8-10, Jan. 2003.

[29] H. S. Malvar, "Modulated complex lapped transform for integrated signal enhancement and coding," U.S. Patent 6 496 795, Dec. 17, 2006.

[30] D. Kirovski and S. Malvar, "Spread-spectrum watermarking of audio signals," *IEEE Trans. Signal Processing*, vol. 51, no. 3, pp. 1020-1033, Apr. 2003.

[31] J. J. Garcia-Hernandez, C. Feregrino-Uribe, R. Cumplido, and C. Reta, "On the implementation of a hardware architecture for an audio data hiding system," *Journal of Signal Processing Systems*, vol. 64, issue. 3, pp. 457-468, Sep. 2011.

[32] F. Kuech and B. Edler, "Aliasing reduction for modified discrete cosine transform domain filtering and its application to speech enhancement," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 131-134, Oct. 2007.

[33] Richard D. Van Nee and Ramjee Prasad, *OFDM for Wireless Multimedia Communications*, Artech House Publishers, 1999.

[34] H. Schulze and C. Lueders, *Theory and Applications of OFDM and CDMA : Wideband Wireless Communications*, Jone Wiley & Sons, 2005.

[35] K. Fazel and S. Kaiser, *Multi-carrier and Spread Spectrum Systems*, Jone Wiley & Sons, 2003.

[36] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels*, Wiley-interscience, 2005.

[37] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing*, Prentice Hall, 2007.

[38] International Standard ISO/IEC 11172-3, "Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5Mbits/s - Part 3: Audio," 1993.

[39] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 4th ed. New York: Academic, 1997.

[40] K. K. Paliwal and L. D. Alsteris, "On the usefulness of STFT phase spectrum in human listening tests," *Speech Communication*, vol. 45, issue. 2, pp. 153-170, Feb. 2005.

[41] K. K. Wójcicki and K. K. Paliwal, "Importance of the dynamic range of an analysis window function for phase-only and magnitude-only reconstruction of

speech," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 729-732, Apr. 2007.

[42] L. Deng and D. O'Shaughnessy, *Speech Processing−A Dynamic and Optimization-Oriented Approach*, New York: Marcel Dekker, 2003.

[43] P. Höher, S. Kaiser, and P. Robertson, "Two-dimensional pilot-symbol-aided channel estimation by Wiener filtering," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 1845-1848, Apr. 1997.

[44] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[45] S. G. McGovern, "Fast image method for impulse response calculators of box-shaped rooms," *Applied Acoustics*, vol. 70, issue. 1, pp. 182-189, Jan. 2009.

[46] L. E. Kinsler, A. R. Frey, A. B. Coppens and J. V. Sanders, *Fundatmentals of Acoustics*, Jone Wiley & Sons, 1999.

[47] P. Kabal, "An examination and interpretation of ITU-R BS.1387: perceptual evaluation of audio quality," *TSP Lab Technical Report*, Dept. Electrical & Computer Engineering, McGill University, May. 2002.

[48] ITU-R Recommendation BS. 1534, "Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems," 2001.

[49] B.-S. Ko, R. Nishimura, and Y. Suzuki, "Robust watermarking based on time-spread echo method with subband decomposition," *IEICE Trans. Fundamentals*, vol. E87-A, no. 6, pp. 1647-1650, Jun. 2004.

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

[51] 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 Trans. Multimedia*, vol. 13, no. 1, pp. 2-13, Feb. 2011.

[52] C. H. Chung, S. H. Cho, S. Kang, and Y. W. Lee, "Performance of convolutional coded and uncoded DS/CDMA system in Nakagami fading channels," in *Proc. IEEE Int. Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 2, pp. 502-506, Sep. 1995.

[53] W. Bendner, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM System Journal*, vol. 35, no. 3&4, pp. 313-336, 1996.

[54] B. Chen, and G. W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE trans. Information Theory*, vol. 47, no. 4, pp. 1423-1443, May. 2001.

[55] H. Daiki, and M. Unoki, "Embedding limitations with audio-watermarking method based on cochlear-delay characteristics," in *Proc. IEEE Int. Conf. IIH-MSP'09*, pp. 82-85, Sep. 2009.

[56] W.-N. Lie, and L.-C. Chang, "Robust and high-quality time-domain audio watermarking based on low-frequency amplitude modification," *IEEE Trans. Multimedia*, vol. 8, no. 1, pp. 46-59, Feb. 2006.

[57] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang, *MIMO-OFDM Wireless Communications with MATLAB*, John Wiley & Sons, 2010.

[58] E. Teletar, "Capacity of multi-antenna Gaussian cahnnels," *European Trans. Telecommunications*, vol. 10, no. 6, pp. 585-595, Dec. 1999.

[59] G. Bauch, "Space-time block codes versus space-frequency block codes," in *Proc. IEEE Conf. Vehicular Technology (VTC)*, pp. 22-25, Apr. 2003.

[60] S. Kaiser, "Spatial transmit diversity techniques for broadband OFDM systems," in *Proc. IEEE GLOBECOM 2000*, pp. 1824-1828, Nov./Dec. 2000.

[61] Y. Li, J. C. Chuang, and N. Rn. Sollenberger, "Transmit diversity for OFDM systems and its impact on high-rate data wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 1233-1243, Jul. 1999.

국문초록

본 논문에서는 음향 데이터 전송을 위한 오디오 정보 은닉 기법에 대한 연구를 수행하였다. 음향 데이터 전송은 전달하고자 하는 메시지를 스피커로 재생하여 전송하고, 공간을 지나 마이크로 녹음하여 검출하는 통신 기법을 의미한다. 오디오 정보 은닉은 전달하고자 하는 메시지를 음악 안에 사람이 구분할 수 없도록 삽입하는 기술이다. 음향 데이터 전송에 적합한 오디오 정보 은닉 기술은 기존 정보 은닉 기술에 비해 높은 강인성과 데이터 전송량을 가져야 한다.

기존 방식 중에 음향 OFDM은 이러한 요구사항을 만족하는 기술이라고 볼 수 있다. 하지만 음향 OFDM 및 다른 기존 방식들은 전송 성능 또는 음질 모두를 만족시킨다고 보기에는 어려움이 있다. 따라서 본 논문에서는 기존 기술이 가지고 있는 한계를 해결하기 위하여 modulated complex lapped transform (MCLT)를 도입하였다. MCLT를 사용하게 되면 음질 저하를 일으키는 블록 효과가 없기 때문에, 음향 OFDM가 가지는 음질 문제를 해결할 수 있다. 또한 본 논문에서는 MCLT를 사용하면서 생기는 간섭 효과에 대해 분석 및 수식화를 진행하였다.

본 논문의 제 3장에서는 MCLT 기반의 오디오 정보 은닉 기술을 제안하였다. 제안된 MCLT 기반의 오디오 정보 은닉 기술은 사람의 귀가 주파수의 크기에 비해 위상에 둔감한 점을 이용하여 데이터 삽입 시 MCLT 계수의 위상 값을 변경하였다. 실험 결과, 제안된 오디오 정보 은닉 기술은 초당 수백 비트를 전송 하면서도 음질이 거의 유지될 수 있었으며, 음향 OFDM보다 음질 및 전송 성능 측면에서 더 나은 성능을 보였다. 더 나아가, 마스킹 문턱값의 이용, 클러스터링 기반 데이터 검출, 그리고 주파수 크기 보정 알고리즘 등의 사용은 제안된 오디오 정보 은닉 기술 성능을 더욱 향상 시킬 수 있었다.

여기에서 더 나아가, 본 논문의 제 4장에서는 반향 환경에 강인한 기법들을 제안하였다. 무선 통신 이론에 기반하여 반향에 강인할 수 있는 MCLT 길이에 대한

논의가 진행되었다. 또한 위너 필터 기반의 채널 추정 기법들과, 채널 추정을 적용하기 위한 새로운 데이터 패킷 구조를 제안하였다. 실험 결과, 길이가 긴 MCLT를 사용하게 되면 반향에 강인해지지만, 그에 따른 음질 저하가 발생함을 알 수 있었다. 또한, 지금까지 제안된 오디오 정보 은닉 기법들은 신호처리, 데이터 재삽입, 의도적 데이터 삭제 등의 공격에 강인함을 보였다.

그러나 반향 환경에서 강인한 성능을 보이면서도 좋은 음질을 만족시키는 MCLT 길이를 찾는 것은 거의 불가능하다. MCLT 길이가 긴 경우, 데이터를 삽입하기 위한 위상의 변경은 음질을 크게 저하시키는 프리에코 현상의 원인이 된다. 따라서 본 논문의 5장에서는 이러한 프리에코를 감쇄할 수 있는 세그멘탈 SNR 보정 (SSA) 알고리즘을 제안하였다. 제안된 SSA 알고리즘은 짧은 길이의 MCLT 변환을 통한 주파수 계수들의 세그멘탈 SNR을 계산하고, 최소 값을 특정 값으로 제한한다. 실험 결과에서 SSA 알고리즘을 사용하면 좋은 음질을 보이면서도 반향 환경에 강인한 성능을 보였다. 또한 하나의 파라미터 값을 조정함으로써 음질과 전송 성능의 트레이드-오프가 가능하다.

마이크의 갯수가 한 개 이상인 경우, 통계적으로 독립적인 채널에 같은 데이터를 전송하는 다이버시티 기법을 적용하여 데이터 전송의 신뢰성을 높일 수 있다. 본 논문의 6장에서는 결합 (combining) 기반의 다중채널 기법을 제안하여 음향 데이터 시스템을 다중 마이크를 사용하도록 확장하였다. 제안된 다중채널 기법에서는, 동기화 및 채널 추정을 각각의 채널에 대해 따로 수행한 후에 그 결과를 SNR이 증가할 수 있도록 선형적으로 결합하였다. 제안된 기법은 기존의 단일 마이크를 사용하는 시스템과 호환이 가능한 것이 가장 큰 특징이다. 실험 결과, 제안된 다중채널 기법은 비록 각각의 채널이 통계적으로 독립이라고 보기는 어렵지만 데이터 전송 성능을 높일 수 있음을 보였다.

주요어: 음향 데이터 전송, 오디오 정보 은닉, modulated complex lapped transform (MCLT), 반향, 세그멘탈 SNR 보정, 다중채널
학번: 2009-30212