Spectral Magnitude Adjusted Data Hiding in MCLT Domain of Audio Signal for Robust Acoustic Data Transmission

Kiho Cho, Jae Choi, Hwan Sik Yun and Nam Soo Kim
Seoul National University, Seoul, Korea
E-mail: {khcho, sjchoi, hsyun}@hi.snu.ac.kr, nkim@snu.ac.kr Tel/Fax: +82-2-884-1824

***Abstract*—Data hiding in audio signal enables acoustic data transmission by playing back the data-embedded audio and by recording the signal and extracting the data at receiver like mobile devices. In our previous work, we proposed an audio data hiding system for acoustic data transmission, of which the audio quality and transmission performance are suitable, based on phase modification of the modulated complex lapped transform (MCLT) coefficients. In this paper, we propose the spectral magnitude adjustment (SMA) technique for this data hiding system which enhances both the quality of the data-embedded audio signal and the transmission performance.**

I. Introduction

Data hiding in audio contents is a widely used approach for audio watermarking, broadcast monitoring, steganography, and data transmission over public audio channel [1]. The embedded data can be short messages, internet URL addresses, or song information according to applications. This data should be extracted under the attack like adding noise, filtering, and MP3 encoding. Moreover, the quality of the data-embedded audio contents should not be distinguished from that of the original ones. In addition, for some applications like audio watermarking, the embedded data should be indelible even though an attacker, who wants to erase the data, knows the whole embedding and extracting procedures except secret keys.

Acoustic data transmission can be one of the applications of data hiding technique. It is a procedure or system to send hidden data through the aerial space to mobile receiver containing microphone by playing the data-embedded audio contents. Most of the data hiding techniques, however, are not suitable because the received audio clips usually suffer from reverberation and heavy noise.

One of the data hiding techniques which can be used for acoustic data transmission is echo hiding [2]. It could hide data without audio quality degradation, but transmission performance is still questionable. Spread spectrum [3] is a basically robust technique to noisy environment. It shows a promising result compared with echo hiding, but the effective transmittible area might be small because it cannot cope with reverberation sufficiently. Acoustic orthogonal frequency-division multiplexing (acoustic OFDM) [4] could be a candidate for acoustic data transmission system because most of the mobile communication methods which should also be robust to reverberation and noise are based on OFDM. The transmission performance of acoustic OFDM is much better than that of the previous techniques, but its audio quality is degraded especially for classical music or speech.

To overcome this weakness, we have proposed a data hiding technique for acoustic data transmission [5], [6] based on the phase modification of the modulated complex lapped transform (MCLT) coefficients [7]. Each MCLT frame is overlapped with half of the adjacent frames, thus the MCLT-based approach reduces the blocking artifacts which degrade the quality of the data embedded audio signal. Although this system have shown better performance than the acoustic OFDM system, additional improvements are needed to use it to practical applications.

In this paper, we propose a spectral magnitude adjustment (SMA) technique in embedding process which can enhance the audio quality. On the other hand, the broader frequency band could be used to increase the transmission performance or data transmission rate with similar audio quality. The overall block diagram of the proposed data hiding system for acoustic data transmission is shown in Fig. 1.

II. Data Hiding Algorithm for Acoustic Data Transmission

A. Data Embedding

To embed the data, an original audio signal is divided into consecutive MCLT frames and the data are embedded by modifying the phases of MCLT coefficients. The modified MCLT coefficients are converted into a time-domain signal by applying inverse MCLT and by overlapping by half of the previous frame.

Data embedding strategy is to modify the phase of the MCLT coefficients of the audio signal to be 0 or π at the receiver. Due to the overlap property of MCLT, interference terms among the adjacent coefficients exist at receiver [5]. Given an M-point MCLT coefficient vector at the i-th frame $\mathbf{X}_i = [X_i(0), X_i(1), ..., X_i(M-1)]^T$, $\hat{X}_i(k)$ which represents

This research was supported in part by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST) (No. 20110020407) and by the MKE(The Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency) (NIPA-2011-C1090-1121-0007).

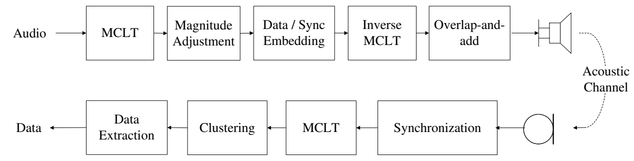

Fig. 1. Block diagram of proposed data hiding system for acoustic data transmission.

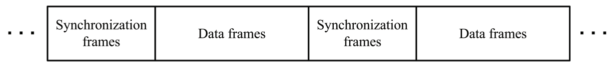

Fig. 2. Structure of synchronization and data frame blocks.

the k-th frequency element of $\mathbf{X}_i$ is modified depending on the embedded data as follows:

$$\begin{aligned}\hat{X}_i(k) &= \max\{|X_i(k)|, M(k)\}b_i(k) - 2j[\mathbf{z}_{-1,k}^T\mathbf{X}_{i-1} \\ &+ \frac{1}{4}X_i(k-1) - \frac{1}{4}X_i(k+1) + \mathbf{z}_{1,k}^T(k)\mathbf{X}_{i+1}], k \in \mathbb{D}, \mathbb{S}\end{aligned} \tag{1}$$

where $b_i \in \{-1, 1\}$ is the binary data, $\mathbf{z}_{i,k}$, means the interference weighting vector, $M(k)$ represents the masking threshold [8], and $\mathbb{D}$, $\mathbb{S}$ are the set of the frequency indices in which data and synchronization sequence are embedded. The binary data $b(k)$ can be either the data to be transmitted or synchronize sequences. The interference weighting vector $\mathbf{z}_{i,k} = [z_{i,k}(0), z_{i,k}(1), ..., z_{i,k}(M-1)]^T$ to compensate the interference term and the l-th element of $\mathbf{z}_{i,k}$ is defined as follows [6]:

$$z_{i,k}(l) = \begin{cases} (-i)\frac{(-1)^{l+m}}{2\pi(2m-1)(2m+1)} & \text{if } |l-k| = 2m \\ \frac{(-1)^l}{8} & \text{else if } |l-k| = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The structure of synchronization and data frame block is shown in Fig. 2 consisting of several successive frames. To synchronize the received signal, the synchronization frame blocks are inserted in the interval of the data frame blocks.

To increase robustness against additive noise, we take advantage of the frequency diversity technique by using L-length spreading sequence in data frames, i.e., a single data bit is embedded in L MCLT coefficients. The binary data in data frames are defined by

$$b(kL+m) = d_i(k)s(m), \quad kL+m \in \mathbb{D} \tag{3}$$

where $d_i(k), s(m) \in \{-1, 1\}$ are the data to transmit at i-th data frame, the L-length spreading sequence, respectively.

B. Synchronization

Before data extraction at the receiver, 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 index at which the phase correlation achieves the maximum; it represents the starting time index of the synchronization frame block. The starting time index $\hat{n}$ is given by

$$\hat{n} = \arg\max_n \sum_{k\in\mathbb{S}} \frac{\hat{Y}(k,n)p(k)}{|\hat{Y}(k,n)|} \tag{4}$$

where $p(k) \in \{-1, 1\}$ is the synchronization sequence and $\hat{Y}(k,n)$ is the k-th MCLT coefficient at the receiver when the analysis window starts at time n. The starting point of the data frame block, then, is defined as the following time index by synchronization frame block length from $\hat{n}$.

In real environment, synchronization timing error could occur because exact timing synchronization is difficult in acoustic data transmission because of variety acoustic interference sources. This error 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

$$\phi(k) = 2\pi\frac{k+0.5}{2M}\tau. \tag{5}$$

C. Data Extraction

After the received signal is timely synchronized, data are extracted from corresponding data frames according to following procedures. At first, despread data coefficients are obtained by correlating the received MCLT coefficients with the corresponding spreading sequence. The k-th despread data coefficient of the i-th data frame $\hat{d}_i(k)$ which represents the mean value of the normalized MCLT coefficients is calculated as follows:

$$\hat{d}_i(k) = \frac{1}{L}\sum_{m=0}^{L-1} s(m)\frac{\hat{Y}_i(kL+m)}{|\hat{Y}_i(kL+m)|}, \quad kL+m \in \mathbb{D}, \tag{6}$$

where $\hat{Y}_i(k)$ means received MCLT coefficient. To make the extraction process robust to phase rotation caused by synchronization error, the clustering based data decoding algorithm is applied [6]. Using the k-means clustering algorithm, each despread coefficient is assigned to one of two clusters which represent the data assigned to bit 1 and -1, respectively. The initial mean value of each cluster is defined by utilizing the normalized MCLT coefficients in the synchronization frames.

III. 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 $\hat{Y}_i(k)$ which represent the MCLT coefficients of the original audio signal and the ideally received data-embedded one following the procedure shown in Fig. 3, respectively. The relation between $\hat{Y}_i(k)$, $\hat{X}_i(k)$ and $X_i(k)$ is defined by

$$\begin{aligned}\hat{Y}_i(k) &= \frac{1}{2}\hat{X}_i(k) + j[\mathbf{z}_{-1,k}^T\mathbf{X}_{i-1} \\ &+ \frac{1}{4}X_i(k-1) - \frac{1}{4}X_i(k+1) + \mathbf{z}_{1,k}^T(k)\mathbf{X}_{i+1}].\end{aligned} \tag{7}$$

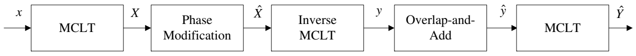

Fig. 3. Procedure to extract ideally received data-embedded MCLT coefficient.

To cope with the audio quality degradation while improving the performance, we propose a spectral magnitude adjustment (SMA) algorithm in embedding process. 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 (1).
4) Derive the ideally received MCLT coefficient $\hat{Y}_i^{(\nu)}(k)$ from $\tilde{X}_i^{(\nu)}(k)$ using (7).
5) Given $\hat{Y}_i^{(\nu)}(k)$ and the original MCLT coefficient $X_i(k)$, compute the magnitude ratio, $\gamma_{i,k}^{(\nu)} = |X_i(k)|/|\hat{Y}_i^{(\nu)}(k)|$.
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

$$\tilde{X}_i^{(\nu)}(k) = \alpha_m^{(\nu)} X_i(k), \quad (i,k) \in \mathbb{G}_m \tag{8}$$

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:

$$\gamma_m^{(\nu)} = \frac{\sum_{(i,k)\in\mathbb{G}_m} |X_i(k)|}{\sum_{(i,k)\in\mathbb{G}_m} |\hat{Y}_i^{(\nu)}(k)|}. \tag{9}$$

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. 4 where 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. 4 represents the the group number m in which the coefficient is included.

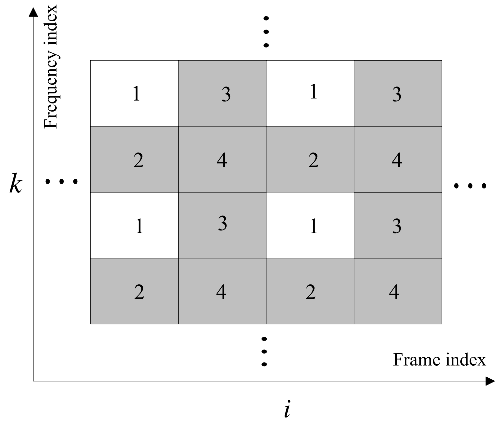

Fig. 4. Grouping example of MCLT coefficients in the SMA algorithm.

TABLE I
SYSTEM PARAMETERS

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
Data block length	80 frames
Data bits per frame	15 bits
Spreading length	4
Data rate	561 bps

IV. Experimental Results

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

A. Subjective and Objective Quality Test

To examine the quality of data-embedded audio contents between the proposed SMA algorithm and the previous one, we conducted the MUSHRA test [9]. In this test, the listeners gave an absolute score, which is from 0 to 100, of each data-embedded audio clip by comparing the original one. The anchor signals (hidden reference, 3.5 kHz and 7 kHz low-pass filtered) were added to meet the experimental condition of the MUSHRA test. Eight audio clips consisting of four genres, which are rock, pop, jazz, and classical music, were used and ten listeners participated in this test.

To measure the objective quality of data-embedded audio contents, perceptual evaluation of audio quality (PEAQ) test was conducted [10]. The PEAQ test score ranges from -4 to 0, which means very annoying, annoying, slightly annoying, perceptible but not annoying, and imperceptible; thus, score 0 indicates perceptually the same between the embedded audio clips and the original ones. The same audio clips with the subjective test were used in this test.

TABLE II
RESULT OF SUBJECTIVE AND OBJECTIVE QUALITY TEST

test method	Subjective test (MUSHRA)			Objective test (PEAQ)		
test system	without SMA	with SMA	diff. score	without SMA	with SMA	diff. score
rock1	71.3 ± 5.5	95.2 ± 1.9	23.9 ± 5.5	-1.46	-0.63	0.83
rock2	74.2 ± 7.0	96.6 ± 1.2	22.4 ± 6.5	-1.41	-0.54	0.86
pop1	79.9 ± 5.3	96.1 ± 1.9	16.2 ± 4.3	-1.54	-0.86	0.68
pop2	83.9 ± 4.0	91.9 ± 2.8	8.0 ± 3.7	-1.22	-0.61	0.61
jazz1	84.7 ± 5.2	84.8 ± 5.7	0.1 ± 4.7	-1.62	-1.43	0.19
jazz2	81.6 ± 5.0	94.5 ± 1.6	12.9 ± 4.3	-1.48	-0.78	0.70
classical1	80.6 ± 5.2	95.4 ± 1.4	14.8 ± 5.4	-1.27	-0.87	0.40
classical2	86.2 ± 4.1	90.7 ± 2.1	4.5 ± 4.2	-1.44	-1.05	0.38
average	80.3 ± 1.9	93.2 ± 1.0	12.9 ± 1.9	-1.43	-0.85	0.58

The result of the subjective and objective audio quality test is shown in Table. II. At the subjective column in this table, the absolute quality scores and 95% confidential intervals of each music clip are listed. As can be seen, the audio 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. At the objective column in Table. II, the PEAQ test scores of each music clip are listed. As can be seen, the proposed method shows better perceptible characteristic. Therefore, the perceptual quality of audio contents produced by the proposed method can be concluded to be better than that by the previous one.

B. Transmission Performance Test

To compare the transmission performance of the proposed system with that of the previous one, we recorded the audio clips at various distance from a loudspeaker and evaluated bit error rate (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 same audio clips for audio quality test were also used. The result is given in Table III. As can be seen in this table, the proposed technique averagely showed better transmission performance than that of the previous one.

V. CONCLUSIONS

The audio data hiding system for acoustic transmission which takes advantage on the phase of the MCLT coefficients shows reasonable audio quality and transmission performance compared with acoustic OFDM or the other audio watermarking systems. In this paper, we proposed the spectral magnitude adjustment (SMA) algorithm to enhance the perceptual quality of data-embedded audio clips. The goal of this algorithm is to reduce the difference between the magnitude of MCLT coefficients and of original and data-embedded audio signals.

TABLE III
RESULT OF TRANSMISSION PERFORMANCE TEST (BER)

test system	without SMA			with SMA		
distance	1m	2m	3m	1m	2m	3m
rock1	0.011	0.008	0.037	0.003	0.002	0.018
rock2	0.012	0.017	0.045	0.004	0.009	0.031
pop1	0.005	0.019	0.060	0.008	0.010	0.032
pop2	0.003	0.022	0.069	0.000	0.009	0.057
jazz1	0.109	0.150	0.201	0.063	0.104	0.147
jazz2	0.002	0.022	0.058	0.001	0.007	0.028
classical1	0.096	0.133	0.174	0.055	0.092	0.158
classical2	0.146	0.147	0.227	0.103	0.108	0.155
average	0.048	0.065	0.109	0.030	0.043	0.078

The subjective and objective experimental results support that the perceptual quality of data-embedded audio clips using SMA algorithm is improved. The BER test results indicate that the SMA algorithm enhances the transmission performance.

REFERENCES

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

[2] Y. Suzuki, R. Nishimure and H. Tao, "Audio watermarking enhanced by LDPC coding for air transmission", *Proc. Int. Conf. IIH-MSP06*, pp.23-26, Dec. 2006.

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

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

[5] H. S. Yun, K. Cho and N. S. Kim, "Acoustic data transmission based on modulated complex lapped transform", *IEEE Signal Process. Lett.*, vol. 17, no. 1, pp. 78-70, Jan. 2010.

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

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

[8] 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.

[9] ITU-R Recommendation BS. 1534, "Method for the subjective assessment of intermediate quality level of coding systems," 2001.

[10] 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.