LETTER

# Spectral Magnitude Adjustment for MCLT-Based Acoustic Data Transmission

**Hwan Sik YUN**[†a)], **Kiho CHO**[†b)], ***Nonmembers*, *and* Nam Soo KIM**[†c)], ***Member***

**SUMMARY** Acoustic data transmission is a technique which embeds data in a sound wave imperceptibly and detects it at a receiver. The data are embedded in an original audio signal and transmitted through the air by playing back the data-embedded audio using a loudspeaker. At the receiver, the data are extracted from the received audio signal captured by a microphone. In our previous work, we proposed an acoustic data transmission system designed based on phase modification of the modulated complex lapped transform (MCLT) coefficients. In this paper, we propose the spectral magnitude adjustment (SMA) technique which not only enhances the quality of the data-embedded audio signal but also improves the transmission performance of the system.

***key words:*** *acoustic data transmission, acoustic communication, data hiding, modulated complex lapped transform*

## 1. Introduction

Acoustic data transmission is a technique for short-range wireless communication using an audible sound [1]–[4]. A transmitter embeds a data stream into an original audio signal and broadcasts the data through the air by playing back the data-embedded sound using loudspeakers. A receiver picks up the sound signal and extracts the hidden message. Although it provides a low data rate and a limited range compared to radio or ultrasonic communications, it has an advantage of using ordinary loudspeakers and microphones without additional hardware infrastructure. Hence, it can be a convenient and low-cost method to transfer information among devices. Possible applications for acoustic data transmission are delivering flight schedules in airports, providing products information in shopping malls, and transferring advertisement [1].

Several acoustic data transmission techniques have been developed. One of these techniques employs an echo hiding method in which data are embedded by adding echoes with different delay times [2]. However, the echo hiding method is not appropriate for acoustic data transmission because the transmitted audio signal suffers from various acoustic reflections which disturb the detection of the embedded echoes. In [1], an acoustic data transmission technique using the spread spectrum watermarking in discrete Fourier transform (DFT) domain was proposed. Even though it provides a good quality of the embedded audio signal with the application of a frequency masking model, it hardly shows reliable transmission performance and sufficient data rate. The acoustic orthogonal frequency-division multiplexing (AOFDM) approach is considered a candidate technique for short-range acoustic data transmission [3]. Although it successfully establishes a reliable communication at a reasonable bit rate, some amount of audio distortion is inevitable due to several building blocks of the system such as the guard interval (GI) and the bandpass filter.

To overcome these limitations, we previously proposed a novel acoustic data transmission approach based on the modulated complex lapped transform (MCLT) [4]. Because each MCLT frame overlaps half of the adjacent frames, the MCLT-based approach reduces the blocking artifacts which degrade the quality of the data-embedded audio signal [5]. In our previous work [6], we also proposed techniques that made the system more robust to ambient noise, synchronization failures, and low signal power. These techniques have improved transmission performance without any audio quality degradation.

In our previous studies [4], [6], the data are embedded by modifying the phases of the MCLT coefficients. Although the magnitude of an MCLT coefficient is mostly preserved when the phase is modified, it has been discovered that the overall magnitude spectrum of a data-embedded audio signal is altered compared with the original audio signal. This altered magnitude of the data-embedded audio signal causes a degraded transmission performance. In addition, this magnitude difference may degrade the perceived quality of the data-embedded audio. In this paper, we propose the spectral magnitude adjustment (SMA) technique to maintain the magnitude of the data-embedded audio signal to the level of the original audio signal. A block diagram of the acoustic data transmission system proposed in this work is shown in Fig. 1.

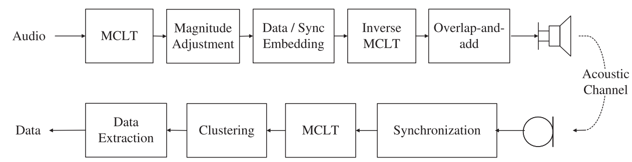


**Fig. 1** Block diagram of proposed acoustic data transmission system.

Manuscript received February 22, 2011.
Manuscript revised December 5, 2011.

[†]The authors are with the School of Electrical Engineering and the Institute of New Media and Communications, Seoul National University, Seoul 151–742, Korea.
a) E-mail: hsyun@hi.snu.ac.kr
b) E-mail: khcho@hi.snu.ac.kr
c) E-mail: nkim@snu.ac.kr

DOI: 10.1587/transinf.E95.D.1523

Copyright © 2012 The Institute of Electronics, Information and Communication Engineers

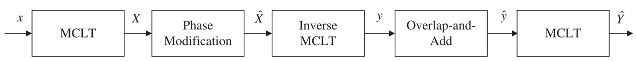


**Fig. 2** Procedure to extract ideally received MCLT coefficient.

## 2. Acoustic Data Transmission System Based on MCLT

In this section, we describe the data embedding and extracting procedures of the acoustic data transmission system based on MCLT as proposed in [4] and [6].

### 2.1 Data and Synchronization Sequence Embedding

For data embedding, 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 the overlap-and-add method.

To recover the desired phase at the receiver, we should consider the interference when the MCLT coefficient is modified to embed the data. Let $\mathbf{X}_i = [X_i(0), X_i(1), \ldots, X_i(M-1)]^T$ be an MCLT coefficient vector obtained from the original audio signal at the $i$-th frame and $\hat{\mathbf{Y}}_i = [\hat{Y}_i(0), \hat{Y}_i(1), \ldots, \hat{Y}_i(M-1)]^T$ be the MCLT coefficients vector obtained at the receiver by following the procedure shown in Fig. 2. Then, $\hat{Y}_i(k)$ is given by

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

where $j = \sqrt{-1}$ and $\mathbf{a}_{i,k} = [\mathbf{a}_{i,k}(0), \mathbf{a}_{i,k}(1), \ldots, \mathbf{a}_{i,k}(M-1)]^T$ is a constant vector which can be derived from the transformation matrices of MCLT [4], [7]. As shown in (1), the phase of $\hat{X}_i(k)$ which is the MCLT coefficient modified so as to embed the data can be altered by the interference terms $j\frac{1}{2}[\mathbf{a}_{-1,k}^T\mathbf{X}_{i-1} + \frac{1}{2}X_i(k-1) - \frac{1}{2}X_i(k+1) + \mathbf{a}_{1,k}^T\mathbf{X}_{i+1}]$. To preserve the phase of $\hat{X}_i(k)$, data embedding is modified such that

$$\begin{aligned}\hat{X}_i(k) =& |X_i(k)|d_i(k) - j\Big[\mathbf{a}_{-1,k}^T\mathbf{X}_{i-1} + \frac{1}{2}X_i(k-1) \\ & - \frac{1}{2}X_i(k+1) + \mathbf{a}_{1,k}^T\mathbf{X}_{i+1}\Big], \quad k \in \mathbb{D}\end{aligned} \tag{2}$$

where $d_i(k) \in \{-1, 1\}$ is a binary data bit and $\mathbb{D}$ denotes the set of coefficient indices corresponding to the target frequency band for data transmission. By substituting (2) to (1), it can be shown that the interferences are canceled and the received MCLT coefficient becomes $\hat{Y}_i(k) = \frac{1}{2}|X_i(k)|d_i(k)$. From (2), it is noted that the two adjacent frames and two adjacent coefficients $\mathbf{X}_{i-1}$, $\mathbf{X}_{i+1}$, $X_i(k-1)$, and $X_i(k+1)$ contribute to modifying the target MCLT coefficient $X_i(k)$. Hence, the data are embedded in every other frame and coefficient.

To improve robustness in a noisy environment, each single data bit is spread by an $L$-length spreading sequence. With larger value of $L$, higher robustness can be achieved at the cost of a decrease in the data rate.

A known synchronization sequence should also be embedded in front of the data frames to identify the exact starting point of each MCLT analysis frame. Given a synchronization sequence $p(k) \in \{-1, 1\}$, it is embedded in the frequency band $\mathbb{S}$ in the same manner as the data embedding method described by (2).

### 2.2 Synchronization and Data Extraction

At the receiver, we compute the phase correlation between the synchronization sequence and the MCLT coefficients extracted at each possible location of the analysis window. The window location which yields the maximum correlation value is identified as the starting time of the MCLT analysis frame. The estimated starting time $\hat{n}$ is given by

$$\hat{n} = \underset{n}{\operatorname{argmax}} \sum_{k \in \mathbb{S}} \frac{\hat{Y}(n,k)p(k)}{|\hat{Y}(n,k)|} \tag{3}$$

where $\hat{Y}(n,k)$ is the $k$-th MCLT coefficient when the analysis window starts at time $n$ and $\mathbb{S}$ indicates the set of the coefficient indices for the synchronization sequence.

Once the received signal is synchronized, data extraction is performed using the despread coefficients obtained by correlating the received MCLT coefficients with the corresponding spreading sequence. The data bit then is recovered by making a decision on whether the phase of the despread coefficient is closer to 0 or $\pi$. To make the extraction process robust to phase distortion, the k-means clustering algorithm can be applied [6]. With 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.

## 3. Spectral Magnitude Adjustment

From a number of experiments, we have observed that the data-embedded audio signal has different magnitude spectra from the original audio signal. This difference can be investigated by comparing $\hat{Y}_i(k)$ and $X_i(k)$ which are the MCLT coefficients of the data-embedded and the original audio signal, respectively. For the embedding method in (2), the MCLT coefficient $X_i(k)$ is modified to cancel the interference. The received MCLT coefficient corresponding to the modified coefficient can then be represented as $\hat{Y}_i(k) = \frac{1}{2}|X_i(k)|d_i(k)$. Compared to $X_i(k)$, the received MCLT coefficient $\hat{Y}_i(k)$ has half the magnitude of $X_i(k)$. This decreased magnitude degrades the transmission performance and the quality of the data-embedded audio signal.

To recover the magnitude of the received coefficient and obtain $\hat{Y}_i(k) = |X_i(k)|d_i(k)$, we can merely double $X_i(k)$ in the data embedding process, i.e., replace $|X_i(k)|d_i(k)$ with $2|X_i(k)|d_i(k)$ in (2). However, it still degrades the quality of the data-embedded audio signal even though it contributes to enhance the transmission performance. This is because that the MCLT coefficients adjacent to the data-embedded coefficient may have different magnitudes from their original values.

To alleviate the quality degradation while improving the performance, we propose the SMA approach. By adjusting the magnitude of the MCLT coefficients, the difference between the original and data-embedded audio signals can be reduced. Since the magnitude of an MCLT coefficient can be altered by interferences, we should consider the effect of the interferences among adjacent MCLT coefficients when adjusting the magnitude. The SMA approach is an iterative algorithm 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, $\tilde{X}_i^{(v)}(k) = \alpha_{i,k}^{(v)} X_i(k)$, where $v$ denotes the iteration number. The scaling factor is applied to not only the coefficients to be modified for data embedding but also the coefficients adjacent to them.
3. Embed the data in the scaled MCLT coefficient $\tilde{X}_i^{(v)}(k)$ using (2).
4. Derive the recovered MCLT coefficient $\hat{Y}_i^{(v)}(k)$ from $\tilde{X}_i^{(v)}(k)$ using (1).
5. Given $\hat{Y}_i^{(v)}(k)$ and the original MCLT coefficient $X_i(k)$, compute the magnitude ratio, $\gamma_{i,k}^{(v)} = |X_i(k)|/|\hat{Y}_i^{(v)}(k)|$.
6. Update the scaling factor as $\alpha_{i,k}^{(v+1)} = \gamma_{i,k}^{(v)} \alpha_{i,k}^{(v)}$.
7. Repeat Steps 2 to 6 until the ratio $\gamma_{i,k}^{(v)}$ approaches close to 1.

Each step of the above process is applied for several frames and coefficients simultaneously because adjacent coefficients can affect each other. After completing the above process, we obtain the final scaling factor $\alpha_{i,k}^{*}$ which makes the magnitude difference between $\hat{Y}_i(k)$ and $X_i(k)$ more reduced. The scaling factor $\alpha_{i,k}^{*}$ is applied to each coefficient to obtain the scaled MCLT coefficient $\tilde{X}_i(k) = \alpha_{i,k}^{*} X_i(k)$ and the data embedding process is performed with the scaled MCLT coefficient $\tilde{X}_i(k)$. By generating the data-embedded audio signal using $\tilde{X}_i(k)$ instead of $X_i(k)$, the spectral magnitude difference between the data-embedded and original audio signals can be reduced.

To enhance the computational efficiency of the SMA approach, it is preferred to cluster the coefficients into a number of groups and apply a common scaling factor to the coefficients belonging to the same group. In that case, the scaled MCLT coefficient at Step 2 of the SMA approach is obtained by

$$\tilde{X}_i^{(v)}(k) = \alpha_m^{(v)} X_i(k), \quad (i,k) \in \mathbb{G}_m \tag{4}$$

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 approach is calculated from the averaged magnitude of the coefficients in each group as follows:

$$\gamma_m^{(v)} = \frac{\sum_{(i,k)\in\mathbb{G}_m} |X_i(k)|}{\sum_{(i,k)\in\mathbb{G}_m} |\hat{Y}_i^{(v)}(k)|} \tag{5}$$

Comparing with the scheme which assigns an individual scaling factor to each coefficient, applying a common scaling factor may let the SMA approach less effective for recovering the spectral magnitude. However, the SMA approach with common scaling factors can still achieve a good audio quality by adopting an appropriate grouping method.

## 4. Experimental Results

We conducted subjective quality and transmission performance tests to evaluate the performance of the proposed SMA technique. The acoustic data transmission system parameters are listed in Table 1. The SMA approach was applied for both the synchronization and data blocks. The coefficients to be modified by the SMA approach were clustered into four groups as shown in Fig. 3 where each rectangular bin refers to an individual MCLT coefficient and the white bin denotes the data-embedded coefficient. The number on each rectangular bin in Fig. 3 represents the identity of the group in which the coefficient is included. In this experiment, the data-embedded coefficients were clustered into the same group because all of them have the relation $\hat{Y}_i(k) = \frac{1}{2}|X_i(k)|d_i(k)$. The other coefficients were empirically clustered according to their relative positions to the corresponding data-embedded coefficients. For the $m$-th group, the magnitude ratio $\gamma_m$ is calculated at each iteration and $\gamma_m$ approaches close to 1 as the number of iteration increases. An example of the variation in $\gamma_m$ along the iteration number is plotted in Fig. 4.

First, the perceived quality of the data-embedded audio

**Table 1** 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 |

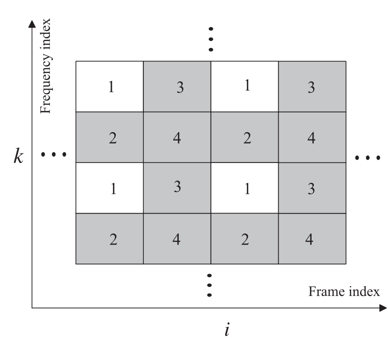


**Fig. 3** Example of grouping coefficients in the SMA approach.

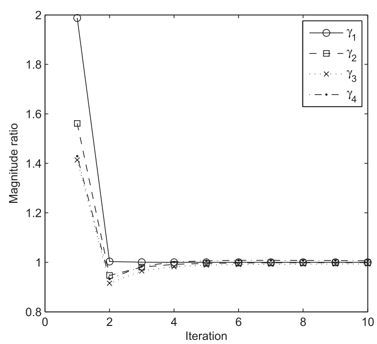


**Fig. 4** Magnitude ratio of each group at each iteration.

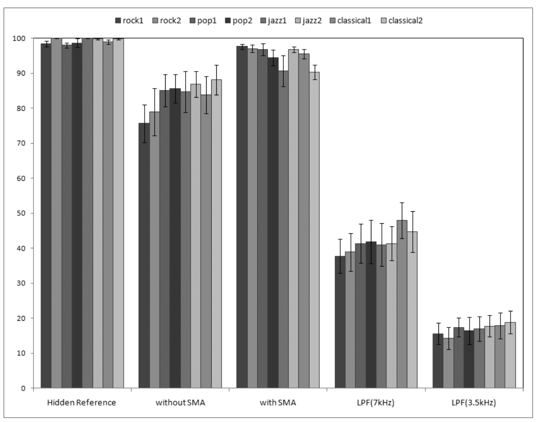


**Fig. 5** Result of subjective quality test.

clips with and without the proposed technique were compared through the MUSHRA test [8]. In the MUSHRA test, each listener compares the test sounds, the hidden reference and anchor signals with the reference and assigns a score between 0 and 100 depending on the perceived quality. We included two low-pass filtered anchor signals, the first with a cutoff frequency of 3.5 kHz and the second with a cut-off frequency of 7 kHz, along with the hidden reference which was the original audio clip. The test sounds consisted of the data-embedded signals with and without the SMA approach. The test materials consisted of eight audio clips from rock, pop, jazz, and classical music genres and nine listeners participated in this test. The results are shown in Fig. 5 where the average scores and their 95% confidence intervals are displayed. From the results, we can conclude that the quality of the data-embedded audio clips is enhanced by the proposed approach.

Next, the transmission performance was evaluated in terms of the bit error rate (BER) of the received data. The data-embedded audio clips were played back by a loudspeaker and recorded by a microphone located at different distances in a quiet office environment. The BER was then calculated based on the data extracted from the recorded audio clips. In this experiment, we did not apply any FEC code. The values of the BER averaged over each genre are given in Table 2. The result shows that the proposed technique improves the data transmission performance.

**Table 2** Result of transmission performance test (BER).

| test system | without SMA | | | with SMA | | |
|---|---|---|---|---|---|---|
| distance | 1 m | 2 m | 3 m | 1 m | 2 m | 3 m |
| 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 |

## 5. Conclusions

In this paper, we have proposed the spectral magnitude adjustment (SMA) approach for the MCLT based acoustic data transmission system. Using the proposed SMA technique, the spectral magnitude difference between the original and data-embedded audio signal can be reduced. The experimental results showed that not only the perceived quality of the data-embedded audio signal but also the transmission performance are improved by the proposed SMA technique.

## Acknowledgments

This research was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2011-0020407) and 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-2012-H0301-12-2005).

### References

[1] 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.

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

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

[4] 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.67–70, Jan. 2010.

[5] H.S. Malvar, "A modulated complex lapped transform and its applications to audio processing," Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, pp.1421–1424, Phoenix, AZ, March 1999.

[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, July 2010.

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

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