Imperceptible Data Hiding in MCLT Domain for Acoustic Data Transmission Using Loudspeaker and Microphone

Kiho Cho, Hwan Sik Yun, Sung Soo Kim and Nam Soo Kim
School of Electrical Engineering and INMC
Seoul National University
P.O.Box 34 Gwanak-no 599 Gwanak-gu Seoul Korea
E-mail: khcho@hi.snu.ac.kr Tel: +82-2-884-1824

***Abstract*—A data hiding technique in a sound wave imperceptibly can be used for a communication system beyond the audio watermarking. The transmitter with a loudspeaker can send the data by playing the sound wave, and the receiver like mobile phone can detect the message using a microphone. This paper proposes an acoustic data transmission system based on the modulated complex lapped transform (MCLT). In the proposed system, data is embedded in an audio wave file by modifying the phases of the MCLT coefficients of the audio wave. The perceived quality of the data-embedded audio wave is almost similar as the original audio. The proposed system is sufficient to deliver short text messages without any additional hardware devices.**

I. Introduction

Acoustic data transmission means a short-range wireless communication in which data is transmitted through the air with the audible sound. Even it provides a low data transmission rate compared to the radio or ultrasonic communications, it has an advantage of using ordinary loudspeakers and microphones without additional hardware infrastructure. Moreover, increasing use of a mobile phone demands the use of sound for data transmission [1].

Embedding data into the audio wave can be viewed as a data hiding technique like audio watermarking [2]. Generally, there are three requirements in data hiding for acoustic data transmission: imperceptibility, robustness, and data rate. Security, also important requirement, is not considered in this application. Imperceptibility means that embedded data signal should not degrade perception of original signal, and robustness represents that the receiver should be able to detect the data even in noisy environment. Data rate means how many bits can be transmitted. 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. Because the current audio watermarking techniques are focused on imperceptibility, they are insufficient to transmit some useful messages through the air in a few seconds [3], [4]. 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.

Recently, the acoustic orthogonal frequency division multiplexing (Acoustic OFDM) was proposed as a reliable communication with reasonable bit rate [5],[6]. In acoustic OFDM, quality degradation of the data-embedded audio wave is inevitable to achieve enough data rate. The data-embedded audio wave produces impulse noise at the first and the last part of each OFDM frame, and it becomes more serious for low powered signals like speech and classical music [7]. This audio quality degradation is caused by several components of the system such as the guard interval (GI), the bandpass filter, and the overlap interval.

In this paper, we propose an acoustic data transmission system using the modulated complex lapped transform (MCLT). In the proposed system, audio signal is transformed by the MCLT and the phases of the coefficients are modified to insert data. 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 audio [8]. 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.

II. Modulated Complex Lapped Transform

The MCLT applies a cosine-modulated filter bank that maps overlapping blocks into complex-valued blocks of transform coefficients [8], and the fast MCLT algorithm [9] makes it usable to many kinds of signal processing applications.

The MCLT generates M coefficients from the $2M$-length frame of input signal $x(n)$. The i-th input frame which is shifted by M samples is denoted by a vector $\mathbf{x}_i = [x(iM), x(iM+1), ..., x(iM+2M-1)]^T$ with T representing the transpose of a vector or matrix, and the MCLT $\mathbf{X}_i$ corresponding to the i-th input frame is given by [8]

$$\mathbf{X}_i = (\mathbf{C} - j\mathbf{S})\mathbf{W}\mathbf{x}_i, \tag{1}$$

where $\mathbf{C}$ and $\mathbf{S}$ denote the $M \times 2M$ cosine and sine modulation matrices whose (k, n)-th elements are defined by

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

$$(\mathbf{S})_{kn} = \sqrt{\frac{2}{M}} \sin\Big[\Big(n + \frac{M+1}{2}\Big)\Big(k + \frac{1}{2}\Big)\frac{\pi}{M}\Big], \quad (3)$$

respectively, and $j = \sqrt{-1}$. The window matrix $\mathbf{W}$ is the $2M \times 2M$ diagonal matrix whose n-th diagonal element is commonly defined as

$$(\mathbf{W})_{nn} = -\sin\Big[\Big(n + \frac{1}{2}\Big)\frac{\pi}{2M}\Big]. \quad (4)$$

The inverse MCLT of $\mathbf{X}_i = \mathbf{X}_{c,i} - j\mathbf{X}_{s,i}$ is derived as

$$\mathbf{y}_i = \mathbf{W}(\beta_c \mathbf{C}^T \mathbf{X}_{c,i} + \beta_s \mathbf{S}^T \mathbf{X}_{s,i}), \quad (5)$$

where β_c and β_s are arbitrary values that satisfy $\beta_c + \beta_s = 1$. In this work, we choose $\beta_c = \beta_s = \frac{1}{2}$. Then, (5) leads to

$$\mathbf{y}_i = \mathbf{W}^2 \mathbf{x}_i, \quad (6)$$

and (6) is perfectly reconstructed by overlapping and adding by M samples with its adjacent frames. Let $\hat{\mathbf{y}}_i$ be the i-th reconstructed frame. Then,

$$\hat{\mathbf{y}}_i = \begin{bmatrix} \mathbf{y}_{2,i-1} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{y}_{1,i} \\ \mathbf{y}_{2,i} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{y}_{1,i+1} \end{bmatrix} \quad (7)$$

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

By (1), (5) and (7), it can be shown that $\hat{\mathbf{Y}}_i$, the MCLT of $\hat{\mathbf{y}}_i$, can be formulated as follows :

$$\begin{aligned} \hat{\mathbf{Y}}_i &= \hat{\mathbf{Y}}_{c,i} - j\hat{\mathbf{Y}}_{s,i} \\ \hat{\mathbf{Y}}_{c,i} &= \frac{1}{2}\mathbf{X}_{c,i} + \frac{1}{2}[\mathbf{A}_{-1}\mathbf{X}_{s,i-1} + \mathbf{A}_0\mathbf{X}_{s,i} + \mathbf{A}_1\mathbf{X}_{s,i+1}] \\ \hat{\mathbf{Y}}_{s,i} &= \frac{1}{2}\mathbf{X}_{s,i} + \frac{1}{2}[\mathbf{B}_{-1}\mathbf{X}_{c,i-1} + \mathbf{B}_0\mathbf{X}_{c,i} + \mathbf{B}_1\mathbf{X}_{c,i+1}] \end{aligned} \quad (8)$$

As can be seen from (8), the interference terms represented as $\frac{1}{2}[\mathbf{A}_{-1}\mathbf{X}_{s,i-1} + \mathbf{A}_0\mathbf{X}_{s,i} + \mathbf{A}_1\mathbf{X}_{s,i+1}]$ and $\frac{1}{2}[\mathbf{B}_{-1}\mathbf{X}_{c,i-1} + \mathbf{B}_0\mathbf{X}_{c,i} + \mathbf{B}_1\mathbf{X}_{c,i+1}]$ are added to $\mathbf{X}_i$ by overlapping and adding the adjacent frames when we apply the MCLT to the reconstructed frame $\hat{\mathbf{y}}_i$ to obtain the MCLT coefficient at the receiver. Due to the interferences, we cannot recover $\mathbf{X}_i$ perfectly from the reconstructed signal $\hat{\mathbf{y}}_i$. The interference terms can be decided by subdividing the MCLT transform matrices as well as the window matrix [10], [11].

III. The Proposed Acoustic Data Transmission System

In the proposed system, data is embedded by modifying the phases of the MCLT coefficients extracted from the audio signal. The modified MCLT coefficients are converted into a time-domain signal frame, and it is overlapped and added with adjacent frames by half of the frame length. This data embedded audio wave is played by a loudspeaker and catched by a microphone at a distance away as shown in Fig. 1. Our strategy is to modify the phases of the MCLT coefficients while the magnitudes are maintained.

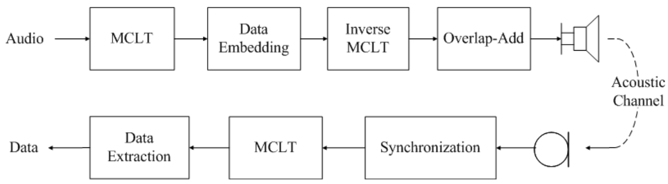

Fig. 1. The proposed acoustic data transmission system.

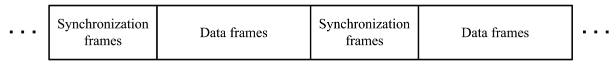

Fig. 2. Structure of synchronization and data frames.

In order to make a precise detection of the data, the receiver has to know the exact location of the analysis interval. In the proposed system, a synchronization frame with a synchronization sequence like M-sequence is inserted to identify the starting point of each analysis frame of the MCLT at the receiver.

The structure of synchronization and data frame is shown in Fig. 2 where each block of synchronization and data 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 data block.

A. The Data Frame

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

$$\hat{X}_i(k) = \begin{cases} |X_i(k)| b_i(k) & \text{if } k \in \mathbb{D} \\ X_i(k) & \text{otherwise,} \end{cases} \quad (9)$$

where $b_i(k) \in \{-1, 1\}$ depending on the input binary data and $\mathbb{D}$ is the set of the coefficient lines corresponding to the target frequency band. At receiver, the received data bit is decided by the phase of the MCLT coefficient. For example, if the phase is closer to 0 than π, received bit is 1. Although the phase of the MCLT coefficient at the receiver is changed due to the interferences by overlapping of each frame, it is still possible to decode the data successfully. From (8), we can see that the real part of the MCLT coefficient is not affected by the overlapping interferences if (9) is applied to all the consecutive frames. As a result, decoding data at the receiver can be performed by checking the sign of the real part of MCLT coefficient on the target frequency lines.

To increase robustness to ambient noise, a spread spectrum technique can be applied. The data bit $b_i(k)$ is spreaded by an L-length vector whose elements are 1 or -1, and a single data bit is embedded in L coefficients. The receiver correlates the received MCLT coefficients with the corresponding spreading sequence to retrieve the original data bits. If the length of spreading vector L becomes larger, the system will be more robust to noise but decrease the data rate.

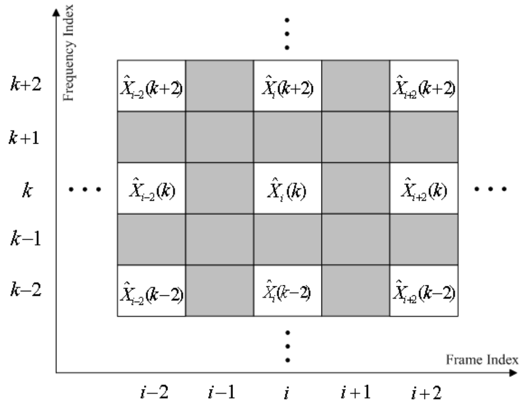

Fig. 3. Embedding of synchronization sequence.

B. The Synchronization Frame

To achieve a good correlation property, the synchronization frame should be generated differently from data frame. If we embed the synchronization sequence as given by (9), the original phases is changed due to interferences described by (8). To prevent this, when embedding the synchronization sequence, (9) is modified in the following way:

$$\begin{aligned}\hat{X}_{c,i}(k) =& |X_i(k)|p(k) - [\mathbf{a}_{-1,k}^T\mathbf{X}_{s,i-1} + \frac{1}{2}X_{s,i}(k-1) \\ & - \frac{1}{2}X_{s,i}(k+1) + \mathbf{a}_{1,k}^T\mathbf{X}_{s,i+1}] \\ \hat{X}_{s,i}(k) =& - [\mathbf{b}_{-1,k}^T\mathbf{X}_{c,i-1} - \frac{1}{2}X_{c,i}(k-1) + \frac{1}{2}X_{c,i}(k+1) \\ & + \mathbf{b}_{1,k}^T\mathbf{X}_{c,i+1}],\end{aligned} \tag{10}$$

where $\mathbf{a}_{l,k}^T$ and $\mathbf{b}_{l,k}^T$ are the k-th row of $\mathbf{A}_l$ and $\mathbf{B}_l$, respectively, and $p(k)$ is the synchronization sequence which should be known to both transmitter and receiver. The phases of received MCLT coefficients are 0 or π at the starting time of the synchronization frames.

From (10), it is noted that when we want to modify $X_i(k)$ to embed the synchronization sequence, the two adjacent frames and two adjacent coefficients $\mathbf{X}_{i-1}$, $\mathbf{X}_{i+1}$, $X_i(k-1)$, and $X_i(k+1)$ (the shaded blocks in Fig. 3) can be modified from their original values to minimize the error between $\hat{X}_i(k)$ and $X_i(k)$. The synchronization sequence is embedded in the every other frame and frequency line as shown in Fig. 3 where each block refers to a MCLT coefficient. By substituting (10) to (8), it can be shown that the interferences are cancelled and the received phase becomes 0 or π as desired.

The receiver computes the phase correlation between the already known synchronization sequence $p(k)$ and the received MCLT coefficients extracted at each possible location of the analysis window for exact synchronization; it is defined as,

$$c(n) = \sum_{k \in \mathbb{S}} \frac{\hat{Y}(k,n)p(k)}{|\hat{Y}(k,n)|} \tag{11}$$

TABLE I
THE PROPOSED SYSTEM PARAMETERS

Sampling frequency	44.1 kHz
Frequency band of data	6400 - 8000 Hz
MCLT frame Size	512 samples
Synchronization block length	12 frames
Data block length	40 frames
Number of subcarriers	36
Spreading length	4
Data rate	0.6 kbps

TABLE II
ACOUSTIC OFDM SYSTEM PARAMETERS

Sampling frequency	44.1 kHz
Frequency band of data	6400 - 8000 Hz
OFDM frame length	2124 samples
Data length	1024 samples
Guard interval	600 samples
Original signal length	500 samples
Synchronization block length	3248 samples
Data block length	12 OFDM frames
Number of subcarriers	36
Data rate	0.6 kbps

where $\hat{Y}(k,n)$ is the received MCLT coefficient on spectral line k at time index n and $\mathbb{S}$ means the set of the coefficient lines corresponding to the synchronization sequences, which represent the white blocks in Fig. 3. We can obtain the exact starting point of the analysis frame from the index which maximize $c(n)$.

IV. EXPERIMENT

To evaluate the performance of the proposed system, we conducted subjective quality tests and transmission performance measurements. To make comparison, we implemented the acoustic OFDM system [5] and set the data rate and the frequency range of the data almost same. The configurations of the proposed system and acoustic OFDM system are listed in Tables I and II.

A. Subjective Quality Test

The perceived quality of the data-embedded audio clips by the proposed system was compared with that by the acoustic OFDM technique through the MUSHRA test [12] and the preference test. 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. The test material consisted of eight audio clips from rock, pop, jazz and classical music genres and nine listeners participated in this test. We included two anchor signals which were low-pass filtered with a cutoff frequency of 3.5 kHz and 7 kHz, respectively and the hidden reference which was the original audio clip. The results are shown in Fig. 4 where the average score and 95 % confidence interval are displayed. From the results, we can conclude that the quality of the data-embedded audio clips by the proposed system was not much degraded than the original audio clips. Furthermore,

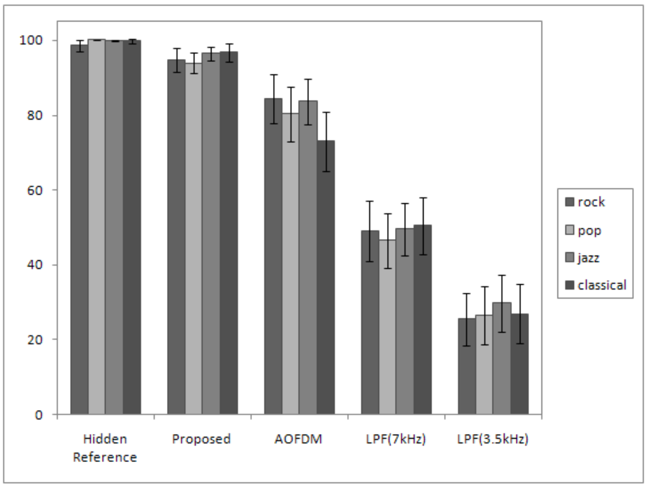

Fig. 4. Subjective quality test results.

TABLE III
THE PREFERENCE RESULT

	Proposed	Acoustic OFDM
Rock	0.25	1.3
Pop	0.2	1.25
Jazz	0.1	1.05
Classical	-0.05	1.75
average	0.13	1.34

the proposed system improved significantly for the classical music clips than the acoustic OFDM system.

The preference test means that the testers give their opinion on the perceptual preference with seven scores : 3 (much better), 2 (better), 1 (slightly better), 0 (about same), -1 (slightly worse), -2 (worse), and -3 (much worse). In this experiment, the quality of each original audio was compared with that of corresponding data-embedded audio. Positive score means that the quality of original audio is better than that of data-embedded audio. The preference test result in Table III shows that the quality degradation of the proposed system was lower than that of the acoustic OFDM. The average score between the acoustic OFDM audio waves and the original ones was 1.34; the listeners distinguished these two and thought that the originals are slightly better. The audio waves from the proposed system, however, were almost indistinguishable; the average test score was only 0.13.

B. Transmission Performance Test

Bit error rate (BER) of the received data was measured at different distances from the loudspeaker to the microphone in a office room. We did not use any channel coding algorithm. In Table IV, the proposed system showed better transmission performance than the acoustic OFDM system except at a distance of 1 m. If we apply forward error correction techniques to the proposed system, the error can be almost perfectly recovered until at a distance of 2 m [6].

TABLE IV
TRANSMISSION PERFORMANCE

Test system	Proposed			Acoustic OFDM		
Distance	1m	2m	3m	1m	2m	3m
Rock	0.0415	0.0760	0.1273	0.0368	0.1110	0.1927
Pop	0.0545	0.0877	0.1369	0.0432	0.1166	0.1933
Jazz	0.0634	0.1021	0.1542	0.0463	0.1176	0.2126
Classical	0.0994	0.1386	0.2061	0.0634	0.1375	0.2002
Average	0.0647	0.1011	0.1561	0.0474	0.1207	0.1997

V. CONCLUSIONS

In this paper, we proposed an acoustic data transmission system based on modifying the phase of the MCLT coefficients. Due to the overlap property of the MCLT, data-embedded audio waves are almost indistinguishable from the original ones. The proposed system shows better performance for imperceptibility than acoustic OFDM, and it is shown by MUSHRA test and preference test. Moreover, the proposed system has better transmission performance than the acoustic OFDM based system at a distance longer than 1 m.

ACKNOWLEDGMENT

This work was supported in part by in part by the Korea Science and Engineering Foundation (KOSEF) Grant funded by the Korean Government (MOST) (No. R0A-2007-000-10022-0) and in part by the Advanced Industrial Technology Development Program funded by the Ministry of Knowledge Economy (No. 10031489).

REFERENCES

[1] C. V. Lopes and P. M. Q. Aguiar, “Aerial acoustic communication,” in *P*roc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Oct. 2001, pp. 219-222.
[2] Ingemar J. Cox, Matthew L. Miller and Jeffrey A. Bloom, *D*igital Watermarking, Academic Press, 2002.
[3] Y. Suzuki, R. Nishimure and H. Tao, ”Audio watermarking enhanced by LDPC coding for air transmission”, *P*roc. Int. Conf. IIH-MSP06, pp.23-26, Dec. 2006.
[4] N. Lazic and P. Aarabi, “Communication over an acoustic channel using data hiding techniques,” *I*EEE Trans. Multimedia, vol. 8, no. 5, pp. 918-924, Oct. 2006.
[5] H. Matsuoka, Y. Nakashima and T. Yoshimura, “Acoustic communication system using mobile terminal microphones,” *N*TT DoCoMo Technical journal, vol. 8, no. 2, pp. 2-12, Sep. 2006.
[6] H. Matsuoka, Y. Nakashima and T. Yoshimura, “Acoustic communication with OFDM signal embedded in Audio,” in *P*roc. Int. Conf. AES 29th , Sep. 2006, pp. 1-6.
[7] K. Cho, H. S. Yun, J. -H. Chang and N. S. Kim, “An analysis on audio quality deterioration of acoustic OFDM,” *T*he Journal of the Acoustical Society of Korea, vol. 28, no. 2, pp. 107-111, Feb. 2009. (in Korean)
[8] H. S. Malvar, “A modulated complex lapped transform and its applications to audio processing,” in *P*roc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, Phoenix, AZ, Mar. 1999, pp. 1421-1424.
[9] H. S. Malvar, “Fast algorithm for the modulated complex lapped transform,” *I*EEE Signal Process. Lett., vol. 10, no. 1, pp. 8-10, Jan. 2003.
[10] F. Kuech and B. Edler, “Aliasing reduction for modified discrete cosine transform domain filtering and its application to speech enhancement,” in *P*roc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Oct. 2007, pp. 131-134.
[11] H. S. Yun, K. Cho and N. S. Kim, “Acoustic data transmission based on modulated complex lapped transform,” unpublished.
[12] ITU-R Recommendation BS. 1534, “Method for the subjective assessment of intermediate quality level of coding systems,” 2001.