Robust Data Hiding for MCLT Based Acoustic Data Transmission

Kiho Cho, Hwan Sik Yun, and Nam Soo Kim, Member, IEEE

Abstract—Acoustic data transmission enables a short-range wireless communication between the loudspeaker and microphone. A transmitter embeds a data stream into a base audio signal such as music or commercial advertisement and broadcasts the data through the air by playing back the data-embedded sound using a loudspeaker. A receiver picks up the sound signal using a microphone and extracts the hidden message. In our previous work, we proposed an acoustic data transmission system which takes advantage of the phase modification of the modulated complex lapped transform (MCLT) coefficients. In this letter, we propose several techniques to realize a more robust communication system in real noisy environments.

Index Terms—Acoustic communication, acoustic data transmission, data hiding, modulated complex lapped transform.

I. INTRODUCTION

A COUSTIC data transmission systems can send short messages through the air using loudspeaker and microphone without additional communication infrastructures devices like “Bluetooth” or “ZigBee.” In the acoustic data transmission system, it is very important that the data-embedded sound should be perceptually indistinguishable from the base audio signal and the receiver should be able to extract the data from the received sound even in noisy environments. A number of previously proposed acoustic data transmission techniques based on audio watermarking [1]–[3] and acoustic orthogonal frequency division multiplexing (Acoustic OFDM) technique [4]–[6], however, could not meet all of requirements for acoustic data transmission systems simultaneously.

To overcome this limitation, in our previous work [7], we proposed a novel acoustic data transmission approach which modifies the phases of the modulated complex lapped transform (MCLT) coefficients and results in both good transmission performance and audio quality. Since each MCLT frame overlaps by half with the adjacent ones, the MCLT-based approach reduces the blocking artifacts which degrade the quality of the data-embedded audio signal [8].

In this letter, we propose several techniques to improve the performance of our previous acoustic data transmission system.

The overall block diagram of the acoustic data transmission system proposed in this letter is shown in Fig. 1. 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, we propose a method to adaptively select the frames for data embedding depending on the power variation.

II. ACoustic DATA TRANSMISSION BASED ON MclT

In this section, we describe the data embedding and extracting procedures of the MCLT-based acoustic data transmission system proposed in [7] briefly.

A. Data Embedding

Data embedding is performed by modifying the phase of the MCLT coefficients of the base audio to 0 or π. To increase robustness against additive noise, a single data bit is spread by an L-length spreading sequence, i.e., 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. Given a MCLT coefficient vector at the th frame, \( \mathbf{x}_i = [x_i(0), x_i(1), \ldots, x_i(M-1)]^T \), it is modified to \( \hat{x}_i = [\tilde{x}_i(0), \tilde{x}_i(1), \ldots, \tilde{x}_i(M-1)]^T \) depending on the embedded data as follows:

\[
\tilde{x}_i(kL + m) = \frac{1}{2} x_i(kL + m) d_m(k) s(m), \quad 0 \leq m \leq L - 1, \quad kL + m \in \mathbb{D}
\]  

(1)

where \( d_m(k), s(m) \in \{-1, 1\} \), and \( \mathbb{D} \) are the input binary data, a L-length spreading sequence, and the set of the coefficients in which the data are embedded, respectively. The modified MCLT coefficients are converted into a time-domain waveform by applying inverse MCLT.

B. Synchronization Sequence Embedding

A known synchronization sequence should be also embedded in front of the data frames to identify the exact starting point of each MCLT analysis frame. Let \( \hat{y}_i = [\hat{y}_i(0), \hat{y}_i(1), \ldots, \hat{y}_i(M-1)]^T \) be the MCLT coefficients vector obtained at the receiver by following the procedure in Fig. 2. Then, it can be derived that

\[
\hat{y}_i(k) = \frac{1}{2} \hat{x}_i(k) + j \left[ \mathbf{z}_{i,k}^T \mathbf{x}_{i-1} + \frac{1}{4} x_i(k-1) \right] - \frac{1}{4} x_i(k+1) + \mathbf{z}_{i,k}^T \mathbf{x}_{i+k+1}
\]  

(2)

where \( \mathbf{z}_{i,k} = [z_{i,k}(0), z_{i,k}(1), \ldots, z_{i,k}(M-1)]^T \) and the th element of \( \mathbf{z}_{i,k} \) is defined as follows [7], [9]:

Manuscript received February 23, 2010; revised April 21, 2010; accepted May 05, 2010. Date of publication May 24, 2010; date of current version June 14, 2010. This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2009-0083044) and SK Telecom. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Rudolf Rabenstein.

The authors are with the School of Electrical Engineering and INMC, Seoul National University, Seoul 151-742, Korea (e-mail: khcho@hi.snu.ac.kr, hsyun@hi.snu.ac.kr, nkim@snu.ac.kr).

Digital Object Identifier 10.1109/LSP.2010.2051174
be 0 or \( \pi \). By applying (4) to contribute is the synchronization sequence known, in the following way:

\[
\mathcal{S} = \{0, 1, -1\}
\]

is the set of the co-
samples is given by

\[
z_{i,k}(l) = \begin{cases} (-1)^{\frac{l}{2(2n+1)}} & \text{if } l = k = 2m \\ \frac{1}{2}, & \text{if } l = k = 1 \\ 0, & \text{otherwise} \end{cases}
\]

For a good correlation property, it is beneficial to make the phase of each \( \hat{Y}_i(k) \) be 0 or \( \pi \). To set the phase of ideally received MCLT coefficients to 0 or \( \pi \), embedding of the synchronization sequence is modified as follows:

\[
\hat{X}_i(k) = |X_i(k)|p(k) - 2J\left[ z_{i-1,k}^T X_{i-1} + \frac{1}{4}X_i(k-1) - \frac{1}{4}X_i(k+1) + z_{i,k}^T X_{i+1} \right],
\]

\[k \in \mathcal{S}\]

where \( p(k) \in \{-1, 1\} \) is the synchronization sequence known to both the transmitter and receiver, and \( \mathcal{S} \) is the set of the coefficients for synchronization. The synchronization sequence is embedded in every other frame and frequency line \[7\]. From (4), we can see that, the two adjacent frames and two adjacent coefficients \( X_{i-1}, X_{i+1}, X_i(k-1) \), and \( X_i(k+1) \) contribute to modifying the MCLT coefficient \( X_i(k) \). By applying (4) to (2), it can be shown that the interferences are cancelled and the ideally received MCLT coefficient becomes \((1/2)|X_i(k)|p(k)\).

C. Synchronization and 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

\[
\hat{n} = \arg \max_n \sum_{k \in \mathcal{S}} \frac{\hat{Y}(k, n)p(k)}{|\hat{Y}(k, n)|}
\]

where \( \hat{Y}(k, n) \) is the \( k \)th MCLT coefficient when the analysis window starts at time \( n \).

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

III. NEW TECHNIQUES FOR PERFORMANCE IMPROVEMENT

In this section, we propose three novel techniques to make our previous system more robust to ambient noise, synchronization failure, and weak base signal power. These techniques enable an improved transmission performance without any audio quality degradation.

A. 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 \[10\]. 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 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:

\[
[X_i(k)] = \max \{ |X_i(k)|, M(k) \}
\]

where \( M(k) \) represents the masking threshold \[11\] in the \( k \)th frequency bin.

B. Clustering

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 noninteger 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 \( \tau \) samples is given by

\[
\phi(k) = \frac{2\pi k + 0.5}{2M}\tau.
\]

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, we propose a bit decoding algorithm that applies the \( k \)-means clustering technique [12]. Despread data coefficient of the \( i \)th input frame \( \hat{d}_i(k) \) is calculated by

\[
\hat{d}_i(k) = \frac{1}{L} \sum_{m=0}^{L-1} s(m) \frac{\hat{Y}_i(kL+m)}{\hat{Y}_i(kL+m)} \quad kL + m \in \mathbb{D}
\]

which represents the mean value of the normalized MCLT coefficients. Each received data coefficient \( \hat{d}_i(k) \) 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 codebook is initialized by utilizing the normalized MCLT coefficients obtained in the synchronization frames. Let \( \mu(I) \) denote the initial codeword for the synchronization bit \( I \). Then

\[
\mu(I) = \frac{1}{|L|} \sum_{k \in L} \frac{\hat{Y}_i(k)}{|\hat{Y}_i(k)|}
\]

where \( L = \{ k | p(k) = I, k \in \mathbb{S} \} \) and \( |L| \) is the number of elements in \( L \). 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.

C. Frame Selective Embedding

The power variation appears very rapid in some audio signals, and this brings about a dramatic change in decoding performance. Sometimes, there even exist silence periods during which it is impossible to embed additional data. A desirable approach to alleviate this problem would be to change the bit rate adaptively in accordance with the power variation of the base audio signal.

To find the proper time region to embed the data, we count the number of coefficients which have more power than a fixed threshold in a candidate region as follows:

\[
Q = \sum_{k \in \mathbb{D} \cap \mathbb{L}} I (|X_i(k)|^2)
\]

\[
I(x) = \begin{cases} 
1 & \text{if } x > T_1, \\
0 & \text{otherwise}
\end{cases}
\]

where \( \mathbb{L} \) is the set of the frame indices corresponding to the candidate region which covers a packet of synchronization and data frames. If the number \( Q \) is bigger than a fixed threshold \( T_2 \), then this region is selected for data embedding. Otherwise, the candidate region is shifted by an integer multiple of the frame length \( T_3 M \) and \( Q \) is recomputed. \( T_1, T_2 \) and \( T_3 \) are determined experimentally.

IV. EXPERIMENTS

The performance of the proposed system was compared with the previous technique [7] through a number of subjective quality and transmission performance tests. The implementation parameters of the system are specified in Table I. Two audio clips from each of the four genres (rock, pop, jazz, and classical music) were used for performance evaluation.

A. Subjective Quality Test

MUSHRA test [13] 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. In the MUSHRA test, each listener compared the reference signal (base audio) with five test audio clips: hidden reference, data-embedded signals with and without magnitude modification, and two anchor signals. Each listener gave a score from 0 to 100 depending on the perceived quality. Nine listeners participated in this experiment. The results are shown in Fig. 3 where the average scores of test audio clips are displayed in conjunction with 95% confidence intervals. In Fig. 4, we show the score difference between the magnitude modified audio clips and those obtained without magnitude modification. From the results, we can conclude that the magnitude modification approach based on frequency masking does not make a serious degradation in audio quality.

B. 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. In this test, we did not apply any channel coding algorithm in order to compare the transmission performance more fairly. The results are shown in Table II in which each method indicates the combination of our previous systems.
technique with the following approaches: magnitude modification (MM), clustering decoding (CLS) and frequency selective embedding (FS). In Table II, we can see that the clustering decoding and magnitude modification techniques are efficient in reducing the BER.

V. CONCLUSIONS

In this letter, we have proposed new techniques to improve robustness of the acoustic data transmission system proposed in [7]. 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. We can also improve the transmission performance using frame selective embedding especially for weak audio signals. The experimental results have shown that the proposed techniques can improve the robustness of our previous acoustic data transmission system.

REFERENCES


