QUALITY ENHANCEMENT OF AUDIO WATERMARKING FOR DATA TRANSMISSION IN AERIAL SPACE BASED ON SEGMENTAL SNR ADJUSTMENT

Kiho Cho, Jae Choi, Yu Gwang Jin, and Nam Soo Kim

School of Electrical Engineering and INMC
Seoul National University, Seoul 151-742, Korea
E-mail: {khcho, sjchoi, ygjin}@hi.snu.ac.kr, nkim@snu.ac.kr

ABSTRACT

Audio watermarking techniques can be used not only as a copyright protect system but also as a short-range wireless data transmission system between a loudspeaker and a microphone. To send data through aerial space using acoustic signal, the length of frequency analysis window should be longer than the reverberation time, which, however, can degrade the quality of watermarked audio signal. In this paper, we propose an audio quality enhancement technique for acoustic data transmission application based on segmental SNR adjustment. The acoustic data transmission system which employs the proposed audio watermarking technique with long frequency analysis window showed better transmission performance than the previous system.

Index Terms— Audio watermarking, acoustic data transmission, acoustic communication, data hiding, modulated complex lapped transform

1. INTRODUCTION

Audio watermarking has been widely applied in many areas such as copyright protection and broadcast monitoring. For instance, it embeds data related to the copyright information in a host audio signal and only the owner can extract the data to claim one’s copyright. Beyond the traditional application of the audio watermarking techniques, the audio watermarking method can be used as a fundamental framework for acoustic data transmission. Acoustic data transmission represents a method which sends message signals 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. This technique can be applied to a variety of applications, for example, broadcast monitoring, advertising, or social television, which provides additional information during a TV show.

There have been a number of acoustic data transmission methods based on the audio watermarking and wireless communication techniques: echo hiding with LDPC channel coding [1], spread spectrum [2], and acoustic orthogonal frequency division multiplexing (Acoustic OFDM) technique [3]. Even though conventional audio watermarking techniques are robust to signal processing distortions and malicious attacks, generally, they cannot support sufficient data rates to transmit some useful messages. On the other hand, acoustic OFDM can transmit the data with higher data rate and more reliability than other watermarking methods, but it might produce annoying sound after the embedding process.

In order to overcome the obstacles in employing previous methods as an acoustic transmission, we proposed a novel audio watermarking technique for acoustic data transmission which utilizes the phase of the modulated complex lapped transform (MCLT) [8] coefficients in our previous works [4], [5]. This method shows better results than acoustic OFDM in terms of audio quality and transmission performance. However, the transmission performance needs to be improved while the quality of watermarked audio is maintained.

In this paper, we propose a quality enhancement technique for the watermarked signal named segmental SNR adjustment (SSA) which was incorporated into our previous audio watermarking method for acoustic data transmission [4], [5].

2. DATA EMBEDDING

In this section, we describe the procedure for data embedding in the proposed acoustic data transmission system. The block diagram of the embedding procedure is shown in Fig. 1. A host audio signal is divided into consecutive MCLT frames and watermarked by modifying the phase of the MCLT coef-
coefficients taking advantage of the fact that the phase alternation of signal is hardly distinguishable. 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 MCLT coefficients of the host audio signal to be 0 or π at the receiver. Given an M-point MCLT coefficient vector at the i-th frame $X_i$, $X_i(k)$ which represents the k-th frequency element of $X_i$ is modified depending on the embedded data as follows:

$$
\hat{X}_i(k) = 2|X_i(k)|b_i(k) - j\left(\left(A_{i-1}\right)_kX_{i-1} + \frac{1}{2}X_i(k-1) - \frac{1}{2}X_i(k+1) + \left(A_i\right)_kX_{i+1}\right), \quad k \in \mathbb{D}
$$

(1)

where $b_i \in \{-1, 1\}$ is the binary data and $\mathbb{D}$ is the set of the frequency indices in which data and synchronization sequence are embedded. The l-th element of interference term $(A_i)_l$ is given by

$$
(A_i)_{kl} = \begin{cases} 
(-i)\pi/(2m-1)(2m+1) & \text{if } |l-k| = 2m \\
(-1)^{k-l} & \text{else if } |l-k| = 1 \\
0 & \text{otherwise}
\end{cases}
$$

(2)

The binary data $b_i(k)$ only can be embedded in every other frame and frequency index. In order to avoid interference, this data or synchronization sequence is embedded in every other frame and frequency line. From Eq. (1), one can see that the two adjacent frames, $X_{i-1}$ and $X_{i+1}$, and two adjacent coefficients, $X_i(k-1)$ and $X_i(k+1)$ contribute to interfernece MCLT coefficient $X_i(k)$ and these interference terms are cancelled at the receiver [4], [5].

To increase robustness of watermarked audio signal against additive noise, we take advantage of the frequency diversity technique; a single data bit is embedded in a certain number of MCLT coefficients [6]. As the length of frequency diversity gets larger, the robustness of the system improves accordingly at the cost of reduced bit rate. In this work, the length of frequency diversity was set to four.

The structure of a data packet consists of the synchronization and data frame blocks, spanning over several successive MCLT frames. The synchronization frame can also perform a role of pilot signals which compensate the reverberant channel effects and synchronization timing error.

3. DATA EXTRACTION

The data extraction procedure of the proposed audio watermarking method for acoustic data transmission system is described in this section. The block diagram of this decoding procedure is shown in Fig. 2.

Before extracting data from the watermarked audio signal at the receiver, the received audio signal needs to be synchron-}

![Decoding procedure of proposed audio watermarking method for acoustic data transmission system.](image)

ized. 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.

After the received signal is synchronized in time, the channel estimation and compensation is performed. Like a mobile radio channel, an acoustic channel can be modeled as a frequency-selective fading channel because of multipath propagation. To recover the transmission error by the channel effects and additive noise, the pilot symbol aided channel estimation which can recover the received data symbols is applied [7]. One-dimensional Wiener estimator is adopted in this work. In this work, only time-directional filtering is performed because the synchronization frames which are employed as pilot symbols can interpolate the channel coefficients at the data frames between several synchronization frames. To implement the Wiener estimator, auto-correlation and cross-correlation matrices derived from power density spectrum of the channel are needed. In this work, a robust non-adaptive design which the filter coefficients are fixed by the time correlation function [7].

After the channel is compensated, the despread data coefficients are obtained by correlating the received MCLT coefficients with the corresponding spreading sequence. To make the extraction process robust to phase rotation caused by synchronization error, the clustering based data decoding algorithm is applied [5].

4. SEGMENTAL SNR ADJUSTMENT ALGORITHM FOR QUALITY ENHANCEMENT OF WATERMARKED AUDIO

According to the wireless communication theories, as the length of the MCLT frame gets longer, the transmission performance in reverberant environment would improve [6]. 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 too long [9]. For instance, if a percussive sound made by drums, cymbals, or speech exists in an interval, pre-echo can occur if the length of the MCLT window $2M$ is much longer than the onset time of the percussion as can be seen in Fig. 3 (b). The pre-echo is one of the most important causes of quality degradation in watermarked audio. The length of the MCLT window, however, is usually longer than the onset time of the percusive sounds to cope with the reverberation in a typical room, which may result in pre-echoes.

---

Fig. 2. Decoding procedure of proposed audio watermarking method for acoustic data transmission system.
Fig. 3. Example of pre-echo where the length of MCLT is 371 msec: (a) host signal, (b) watermarked signal, (c) watermarked signal with segmental SNR adjustment algorithm.

Fig. 4. Block diagram of segmental SNR adjustment algorithm (SSA).

For this reason, we propose an adjustment algorithm that attenuates the pre-echo of the watermarked audio signal by limiting the minimum value of the segmental SNR. The procedure of the proposed audio quality enhancement algorithm by adjusting segmental SNR is shown in Fig. 4. In this algorithm, a short MCLT window of length $M_s$, which is shorter than the typical onset time of percussive sounds, is needed to calculate the segmental SNR of the watermarked audio signal. Let $X_{s_i}(k_s)$ and $\hat{X}_{s_i}(k_s)$ denote the MCLT coefficients of the original and watermarked audio signal, respectively, obtained from the MCLT analysis with a shorter window length $M_s$. Here, $s_i$ 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 frame and $k_s$-th frequency bin is defined as follows:

$$\text{SNR}_{i_s}(k_s) = 10 \log_{10} \left( \frac{|X_{i_s}(k_s)|^2}{(|X_{i_s}(k_s)| - |\hat{X}_{i_s}(k_s)|)^2} \right). \quad (3)$$

If the segmental SNR is smaller than predefined target SNR $\gamma$, the magnitude MCLT coefficient of watermarked audio signal is modified by

$$|\hat{X}_{i_s}(k_s)| = \begin{cases} (1 + 10^{-\gamma})|X_{i_s}(k_s)| & |X_{i_s}(k_s)| < |\hat{X}_{i_s}(k_s)| \\ (1 - 10^{-\gamma})|X_{i_s}(k_s)| & \text{otherwise} \end{cases} \quad (4)$$

and this modified MCLT vector is then converted into time domain signal and synthesized by overlapping and adding with previous time domain signal. In this work, $\gamma$ was set to 1dB. As can be seen in Fig. 3, the watermarked signal with the proposed algorithm can attenuate the pre-echo.

5. EXPERIMENTS

In order to evaluate the performance of the proposed audio watermarking system, we conducted experiments concerned with audio quality and data transmission performance. In our experiments, we applied three different system configurations: S1, S2, and S3. S1 represents the system using short MCLT window which is similar to the configuration used in our previous works [4], [5]. S2 stands for the system using a long MCLT window with the proposed SSA algorithm. S3 is the system using a long MCLT window without the SSA algorithm. These three different configurations were applied to investigate the effect of each module of the algorithm individually. The system parameters are specified in Table 1. Six stereo audio clips were used in these experiments and all of the audio clips were 30 seconds long and the sampling frequency was 44.1 kHz. The average energy overall the tested audio signals was set to -18 dB. The length of a packet consisting of data and pilot blocks are 1.1 seconds. The data rate was 231 bps without channel coding.

### Table 1. System parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>512</td>
<td>8192</td>
<td></td>
</tr>
<tr>
<td>$D$</td>
<td>{150, 152, 154, ... , 244}</td>
<td>{1200, 1202, ... , 1952}</td>
<td></td>
</tr>
<tr>
<td>Data block</td>
<td>16 frames</td>
<td>2 frames</td>
<td></td>
</tr>
<tr>
<td>Pilot block</td>
<td>8 frames</td>
<td>1 frame</td>
<td></td>
</tr>
<tr>
<td>SSA</td>
<td>X</td>
<td>O</td>
<td>X</td>
</tr>
</tbody>
</table>

5.1. Transmission Performance Test

The transmission performance was evaluated in terms of the bit error rate (BER) of the received data. The BER was then calculated based on the data extracted from the recorded audio clips. In this experiment, we did not apply any channel coding method. The audio clips which played back from a loudspeaker were recorded by a mobile phone at various distances in a reverberant classroom where the size is 11 m × 7
Table 2. Result of transmission performance test (BER)

<table>
<thead>
<tr>
<th></th>
<th>1 m</th>
<th>3 m</th>
<th>5 m</th>
<th>7 m</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>0.0113</td>
<td>0.0399</td>
<td>0.1480</td>
<td>0.1674</td>
</tr>
<tr>
<td>S2</td>
<td>0.0182</td>
<td>0.0311</td>
<td>0.0959</td>
<td>0.0800</td>
</tr>
<tr>
<td>S3</td>
<td>0.0019</td>
<td>0.0060</td>
<td>0.0363</td>
<td>0.0323</td>
</tr>
</tbody>
</table>

Fig. 5. MUSHRA test scores according to test music clips. M1-M6 on horizontal axis represents the name of each music clip. Numbers on vertical axis represents the MUSHRA score of the test music clips in different experimental conditions.

In this paper, we proposed the segmental SNR adjustment method which improves the quality of watermarked audio modifying the phase of MCLT coefficient for acoustic data transmission application. The proposed method can compensate the quality degradation due to using the large MCLT window to be robust in reverberant environment. The proposed acoustic transmission system with long MCLT window and the SSA method showed similar audio quality to the reference system with short MCLT window, and it showed much better transmission performance than the reference system.

5.2. Subjective Quality Test

MUSHRA test [10] was conducted to compare the perceived quality of the watermarked audio clips generated from three different systems. In the MUSHRA test, each listener compared the reference signal (host audio signal) with six test audio examples: hidden reference, watermarked audio signals by S1-S3, and two anchor signals by low pass filtering. Ten listeners participated in this experiment. The results are shown in Fig. 5 where the average scores of test audio clips are displayed in conjunction with 95% confidence intervals. As can be seen in Fig. 5, the average MUSHRA score of S1 and S2 was higher than 90. S3 showed worse audio quality than S1 and S2 but the score of S3 was still higher than 80. From the results, we can conclude that the SSA method can maintain the audio quality almost indistinguishable from our previous work even though the MCLT frames are lengthened for better transmission performance.

6. CONCLUSIONS

In this paper, we proposed the segmental SNR adjustment method which improves the quality of watermarked audio modifying the phase of MCLT coefficient for acoustic data transmission application. The proposed method can compensate the quality degradation due to using the large MCLT window to be robust in reverberant environment. The proposed acoustic transmission system with long MCLT window and the SSA method showed similar audio quality to the reference system with short MCLT window, and it showed much better transmission performance than the reference system.

7. REFERENCES


