An Acoustic Communication Technique Based on Audio Data Hiding Utilizing Artificial Flowing Water Sounds

Tetsuya Kojima*, Naoyuki Muraoka† and Raito Matsuzaki*
* Department of Computer Science, National Institute of Technology, Tokyo College, Tokyo, Japan
E-mail: {kojt, matsuzaki}@tokyo-ct.ac.jp
† Advanced Course of Mechanical and Computer Systems Engineering,
National Institute of Technology, Tokyo College, Hachioji, Tokyo, Japan
E-mail: s21610@tokyo.kosen-ac.jp

***Abstract*—Data hiding can be utilized as a communication medium as well as digital watermarking or fingerprinting applications which can protect intellectual property rights or avoid illegal copying of digital media. In the previous studies, audio data hiding techniques have been applied to disaster prevention broadcasting systems and a technique called Tone Code that can embed the secret messages by utilizing the characteristics of musical pieces, such as melody, harmony and rhythm. In this study, an audio data hiding scheme utilizing artificial flowing water sounds is proposed and applied it into an acoustic communication technique in a small chamber such as a restroom stall. Such flowing water sounds have characteristics quite similar to white Gaussian noise. Therefore, it is expected that the deterioration of the stego sound quality can be ignored after embedding secret messages in those sounds. It is shown that the proposed scheme can transmit the secret messages with almost no errors through the experiments under the analog transmission scenario using loud speakers and mobile terminals in a small chamber. Some problems to be tackled in our future study are also discussed.**

I. Introduction

Data hiding is a technique to embed secret messages into digital media such as images, audio and video signals without being recognized by human sensory systems[1]. Data hiding techniques are usually applied into digital watermarking or digital fingerprinting, which is used to protect intellectual property rights or to avoid illegal copies of the original works. On the other hand, it has been pointed out that data hiding can be utilized as a communication medium. In other words, messages are secretly embedded into digital media that can be recorded by receivers such as smartphones and tablets. The extracted messages can be displayed on the screens of such receivers. The authors have proposed data hiding techniques for such communications purposes, for example, disaster prevention broadcasting systems using audio data hiding[2], [3], an audio data hiding called 'Tone Code' utilizing musical components[4], a visible video data hiding using visual effects[5].

In this study, the audio broadcasting scheme[2], [3] is applied into artificial flowing water sounds [1]. In the proposed scheme, secret messages are embedded in artificial flowing water sounds. The stego water sounds, that is, the audio signals conveying secret messages are broadcast from a loud speaker and recorded by the receiver such as smartphones or tablets. Some restroom facilities have functions that play a melody or flushing sounds. The stego water sounds generated by the proposed scheme can be used for such purposes as well as broadcasting of advertising messages, emergency purposes, and so on.

In this paper, the detailed method of the proposed scheme and the results of experiments under realistic scenario where the spatial acoustic communication is realized using a loud speaker and a receiver in a small chamber like a restroom stall. It is shown that the messages can be received without almost no errors through such experiments.

This paper is constructed as follows. Sect.II describes the overview of the data hiding method of the previous and the proposed scheme. We developed an Android application operated at the receivers such as smartphones or mobile tablets. Sect.III describes the details of the developed Android application. In Sect.IV, the performance of the proposed method is evaluated through the realistic experiments employing a loud speaker and the developed application. Sect.V gives some discussions on the problems to be tackled in the future study. Sect.VI summarizes our proposal.

Main contribution: We proposed an acoustic communication scheme based on audio data hiding using artificial flowing water sounds. It has been shown that the embedded messages can be successfully extracted at the receiver with almost no errors in a small chamber such as a restroom stall. As far as the authors know, there are no proposals of audio data hiding schemes or acoustic communication methods based on such artificial flowing water sounds.

[1] A part of this paper was presented at IEICE Technical Conference on Enriched Multimedia as "An audio data hiding scheme utilizing artificial flowing water sounds (in Japanese)," *IEICE Tech. Rep.*, vol.121, no.29, EMM2021-7, pp.37–41, May 2021.

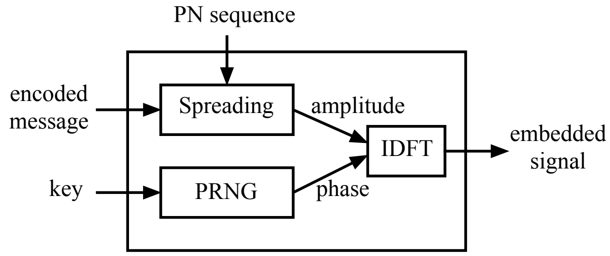


Fig. 1. Generator of the embedded signal (watermark) of the previous method.

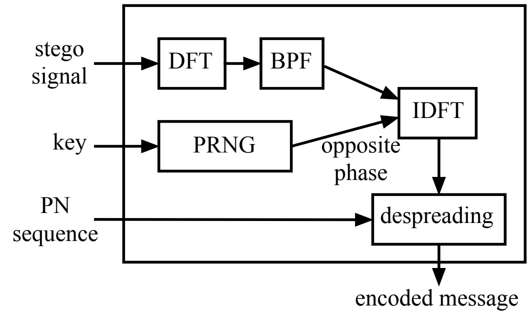


Fig. 2. Detector of the embedded signal (watermark) of the previous method.

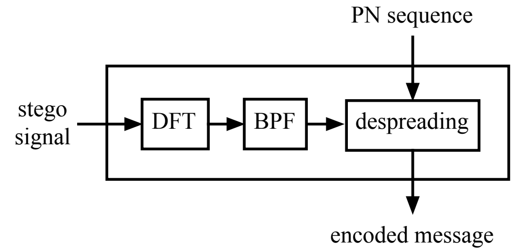


Fig. 3. Detector of the embedded signal (watermark) of the proposed method.

II. Proposed Data Hiding Scheme

A. *Difference from the Previous Method*

In the authors' previous studies, disaster prevention broadcasting systems based on audio data hiding techniques have been proposed. In this system, disaster information is embedded into an original siren sound and broadcast from a loud speaker. The broadcast siren sound conveying the embedded information is recorded by a microphone of the receiver, such as smartphones or mobile tablets. The transmitted disaster information is extracted from the received signal, and displayed on the screen of the receiver. In this sense, such a system is considered as an acoustic communication scheme employing siren sounds instead of electro-magnetic waves. In this study, we propose a similar communication scheme employing artificial flowing water sounds.

The previous studies[3] employ a variation of direct spread spectrum modulation to embed and extract the secret messages. In this study, we take the almost same method in embedding and extracting the secret messages.

In the preliminary experiments, it has been shown that the bit error performances deteriorate in the small chamber, especially near the walls. To solve this problem, the embedding and extracting method in [3] are modified and applied into the proposed scheme. The detailed embedding and extracting methods of the previous scheme[3] and the proposed scheme are described as follows.

In the following, an original signal that does not carry any secret information is called a *cover* signal while the signal where the secret messages are embedded is called the *stego* signal.

1) Previous Method: Fig.1 shows the block diagram of the embedded signal generator employed in the previous method, that is, the watermark generator in the digital watermarking terminology. The secret message is encoded by an error detecting code in advance, and input into the generator. The encoded message is modulated by a spread spectrum (SS) method using a pseudo random noise (PN) sequence. On the other hand, a real-valued random sequence is generated by a pseudo random number generator (PRNG) by a random seed (key). A complex-valued signal is generated by taking the SS-modulated signal and the real-valued random sequences as the amplitude and the phase, respectively. and transformed into a real-valued time series signal by inverse discrete Fourier transform (IDFT). The output of IDFT is considered as a signal embedded into the cover audio signal.

Fig.2 shows the block diagram of the embedded signal detector used in the previous method. At the receiver, the received signal is first synchronized by the synchronization (sync) signal, and input in to the detector. The input signal is transformed into a complex-valued signal by discrete Fourier transform (DFT), and input into a band-pass filter (BPF) to reduce the noise components. On the other hand, PRNG generates the real-valued random sequence identical to the one used in the embedded signal generator from the key shared with the embedding side. The random sequence is transformed into a complex-valued signal which has the opposite phase to the complex signal generated in the embedding process, and multiplied to the output signal of BPF. Then, this complex valued signal is transformed into a real-valued time series signal by IDFT, and demodulated by the same PN sequences as the one used at the embedding side. The output signal is expected to be almost identical to the input signal for the embedded signal generator[3].

2) Proposed Method: The previous method requires the complex-valued signal which has the opposite phase as the signal generated in the embedding process as described above. If the phase components of the received signal are completely identical to the one generated at the embedding side, the output signal of the detector is almost same as the encoded signal input into the embedded signal generator. However, the phase components of the received signal are altered from the original ones since it is affected by the noise or echo effects during the acoustic transmission. In this study, we modify the embedding

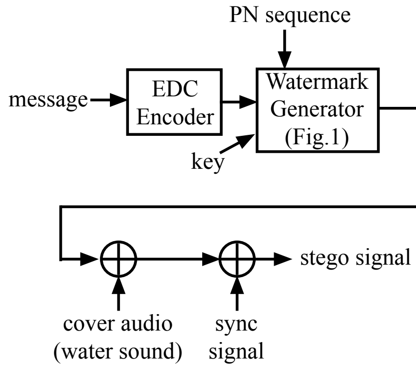


Fig. 4. Block diagram of the embedding process.

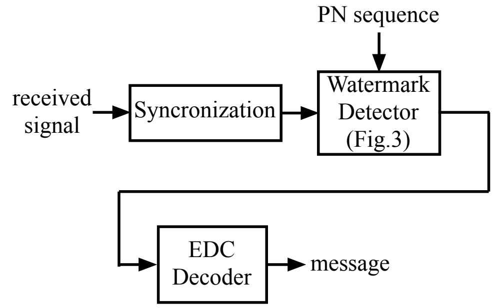


Fig. 5. Block diagram of the extracting process.

and extracting method a bit, and propose a method that does not employ the phase component of the signals at all.

In the previous method, the signal obtained by spread spectrum modulation is considered as a bi-phase signal on $\{-1, +1\}$. In the proposed signal, the bi-phase, that is, $\{-1, +1\}$-valued modulated signal is added by a constant '1', such that the obtained signal is a binary signal on $\{0, 2\}$. The other processes at the embedding side are same as those in the previous method. In other words, the embedded signal is obtained as IDFT of the binary modulated signal.

Fig.3 shows the block diagram of the embedded signal detector in the proposed method. In the proposed method, it is not necessary to consider the phase components of the signals at all since it is considered that the modulated signal can be obtained by digital amplitude modulation or on-off keying. It makes the signal detecting process quite simple, such that the output signal from BPF is directly demodulated by the PN sequence without IDFT shown in Fig.3. It is expected that the bit-error-rate (BER) performances of the extracted secret messages can be improved. BER performances of the proposed method is evaluated in Sect.IV.

B. Overview of the Embedding Process

Fig.4 shows the block diagram of the embedding process. The only difference between the previous and the proposed method can be found in the watermark generator, that is, the embedded signal generator shown in Fig.1. Therefore, the almost same embedding method shown in Fig.4 is employed in both of the previous and the proposed method. As described above, the binary secret message is encoded by an error detecting code (EDC). We employ Adler-32 checksum algorithm as EDC. In the previous method[3], the output of EDC is compressed by Deflate algorithm based on LZ77 and Huffman codes. In the proposed method, however, we don't consider any data compression since quite short messages such as URL are considered as secret messages. The output of EDC encoder is input into the watermark generator shown in Fig.1, which output the embedded signal obtained by employing a PN sequence and a random seed (key). The stego audio signal is obtained by adding the embedded signal by a cover audio signal and a synchronization (sync) signal. In this study, artificial flowing water sounds are employed as cover audio signals.

C. Overview of the Extracting Process

Fig.5 shows the block diagram of the extracting process. In a similar way as the embedding process, the only difference between the previous and the proposed method is the process in the watermark detector, that is, the embedded signal detectors shown in Figs.2 and 3.

The signal received by a microphone is first synchronized by a sync signal which is identical to the one used in the embedding process. The location of the embedded signal is identified by the synchronization. In this study, a method called cross power spectrum phase analysis (CSP)[6], [7] is employed as the synchronization method. The synchronized signal is input into the watermark detector shown in Fig.3. Finally, the message can be obtained by decoding the output of the detector at EDC decoder.

III. Receiver Application

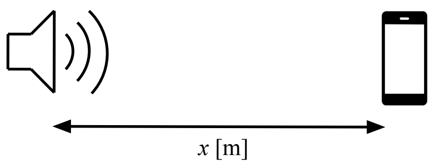
We developed an Android application that can extract and display the secret messages extracted by the proposed method described in the previous section. The developed application is used in the experiments described in Sect.IV.

The operations of the application are described as follows.

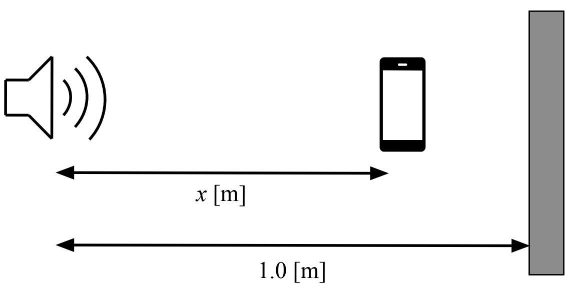
1) Pressing the central button starts to record the stego audio signal with the built-in microphone.
2) The recording is stopped after a predetermined period of time, or if the central button is pressed again.
3) The message is extracted from the recorded stego audio signal.
4) If no error is detected at EDC decoder, the extracted message is displayed on the screen. Otherwise, the error message is displayed.

The screen shot of the developed Android application is shown in Fig.6. In this screen, the number of experimental trials, the maximal number of bit-errors in all of the experimental trials and the averaged bit error rate are displayed for the verification purposes of the application. The averaged bit error rate is defined as the accumulated sum of the number of

Fig. 6. A screen shot of the developed Android application.



(a) Situation without any walls around the receiver.



(b) Situation with a wall near the receiver.

Fig. 7. Scenarios for performance evaluation experiments.

bit errors divided by the sum of the message lengths used in all of the experimental trials.

IV. Experimental Results

A. Comparison of the Previous and Proposed Methods

In this section, we compare the bit error rate (BER) performances of the proposed method with those of the previous method through realistic acoustic communication experiments. In the experiments, the stego flowing water sounds broadcast from a loud speaker are recorded by the receivers. In these experiments, the embedded messages are not extracted by the mobile terminals, but by PCs.

TABLE I
BER performances in the situation without any walls.

Distance x[m]	0.10	0.20	0.40
BER[%] for the previous method	0.000	0.010	0.010
BER[%] for the proposed method	0.000	0.000	0.093

TABLE II
BER performances in the situation with a wall.

Distance x[m]	0.10	0.20	0.40
BER[%] for the previous method	27.642	39.574	37.726
BER[%] for the proposed method	0.000	0.001	0.003

The experiments are executed under the two different situations shown in Fig.7. One is the situation where there are not any walls around the receiver (Fig.7(a)). Another one is the situation where a loud speaker locates 1.0[m] away from a wall. On the other hand, the receiver locates x[m] away from a loud speaker, where $x < 1.0$[m] (Fig.7(b)). The distance x between the loud speaker and the receiver is set to 0.10[m], 0.20[m], or 0.40[m].

The results for the situation without any walls and with a wall are shown in Tables I and II, respectively. Table I shows that BER performances for the situation where there are not any walls are quite good and less than 0.1[%] in both of the previous and the proposed method. On the other hand, as shown in Table II, in the situation where there is a wall behind the receiver, BER performances for the previous method deteriorate while BER values for the proposed method are almost zero. It implies that the proposed extracting method without employing phase component of the signals works for this situation. From these two tables, the proposed method causes almost no errors at the receiver in both of the two situations.

B. Performance Evaluation by Android Application

In this section, acoustic communication experiments are executed using the Android receiver application described in Sect.III. In a similar way as above, the two different situations without any walls and with walls are considered. Flowing water sounds used in this study are frequently used as artificial flushing sounds in restroom stalls. Therefore, we consider a space like a small chamber shown in Fig.8 as the situation with walls. The distance x between the loud speaker and the receiver is set to 0.30[m], 0.50[m], 0.70[m] and 1.00[m]. In the situation without any walls, the same scenario shown in Fig.7(a) is employed. In these experiments, the embedded messages are recorded and extracted on the receiver mobile terminals by the developed Android application.

The experimental results are shown in Table III. In this table, it is shown that BER performances are less than 0.1[%] in almost all cases. It implies that the proposed method achieves successful acoustic communications with almost no bit errors

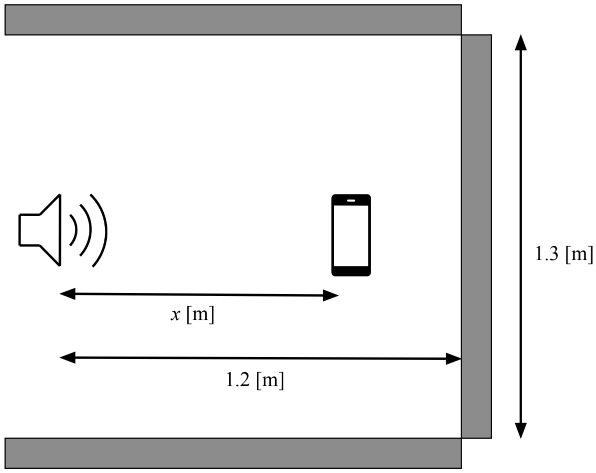


Fig. 8. Situations like a small chamber such as a restroom stall.

TABLE III
BER[%] PERFORMANCES IN THE EXPERIMENTS USING THE DEVELOPED ANDROID APPLICATION.

Distance x[m]	0.30	0.50	0.70	1.00
Without any walls	0.000	0.045	0.045	0.257
In a small chamber	0.000	0.078	0.067	0.078

in the extracted messages.

V. Discussions

As described in Sect.IV, the proposed scheme has quite good BER performances. However, there are some problems to be tackled in our future study in order to improve the proposed scheme. In this section, we describe some of these problems.

One problem is that the computational cost is high in the extraction process. In the Android application described in Sect.III, it takes over 30[sec] to display the extracted messages after the recording of the stego audio signal is finished. The primal reason for this problem is that it takes quite a long time for synchronization. Especially, the computation for the cross-correlation functions required for CSP method[6], [7] has a high computational cost. On the other hand, a recent study[8] has proposed a fast computation method for the cross-correlation functions used in CSP synchronization for the previous study[3] by employing fast Fourier transform (FFT). Therefore, it is expected that the proposed scheme can be also accelerated by employing FFT.

Another problem is to improve BER performances. In the experiments using Android application described in Sect.IV, we have acoustic communication trials 160 times as a whole. The number of trials where any bit errors occur is only seven out of 160 trials. In other words, no errors occur at all in almost all trials, but there are many errors found in only these seven trials. In this situation, it is possible to improve BER performances, for example, if we transmit the same messages successively several times and demodulate and extract the embedded messages by majority decision.

The next problem is to perform the experiments using compressed stego audio signals. In this paper, we use a loud speaker with relatively wide dynamic range. However, the dynamic range of the built-in speaker on the restroom facilities is not very wide in general, which implies that the broadcast stego audio signal can be largely compressed. Therefore. more realistic performance evaluations are required to bring the proposed scheme in a practical use.

The last problem is to improve the user interface of the Android application. It is difficult to alter the stego sound signals frequently in the restroom facilities. Therefore, we consider that the possible embedded messages should be URLs for a website since it is easy to alter the contents of the web page shown by a constant URL. Current Android application only shows the extracted messages, but it should be modified to show the webpages shown by the extracted URL.

VI. Summary

In this study, we proposed an audio data hiding scheme using artificial flowing water sounds and its application to acoustic communications from a built-in speaker to the mobile terminals in the small chambers such as restroom stalls. We also developed an Android application operated in mobile terminals used as the receivers. It is shown that the proposed scheme has quite good BER performances through the experiments employing the developed Android application. In our future study, it is important to tackle the problems discussed in Sect.V.

Acknowledgment

A part of this work was supported by JSPS KAKENHI Grand Number 18K11609. The authors sincerely thank to Prof. Akira Tanaka, National Institute of Technology, Tokyo College, for his fruitful comments and discussions.

References

[1] F.A.P. Petitcolas, R.J. Anderson and M.G. Kuhn, “Information hiding — a survay,” *Proc. of IEEE*, vol. 87, no.7, pp.1062–1078, July 1999.

[2] T. Kojima, T. Tachikawa, A. Oizumi, Y. Yamaguchi and U. Parampalli, “A disaster prevention broadcasting based on data hiding scheme using complete complementary codes,” *Proc. of ISITA 2014, Melbourne, Australia*, October 2014, pp.45–49.

[3] Y. Sada and T. Kojima, “Improvement of emergency broadcasting system based on audio data hiding (in Japanese),” *IEICE Tech. Rep.*, vol.117, no.476, EMM2017-88, pp.55–60, March 2018.

[4] K. Kamada, T. Kojima and U. Parampalli, “Tone Code : a novel method of covert communications based on musical components,” *Proc. of ISITA 2016, Monterey, CA, USA*, November 2016, pp.335–339.

[5] T. Kojima and K. Akimoto, “Visible video data hiding techniques based on visual effects utilizing barcodes” *Proc. of ISITA 2020, Kapolei, HI, USA (Virtual)*, October, 2020, pp.514–518.

[6] M. Omologo, “Acoustic source location in noisy and reverberant environment using CSP analysis,” *Proc. of ICASSP ’96, Atlanta, GA, USA*, May 1996, pp.921–924.

[7] T. Nishiura, T. Yamada, S. Nakamura and K. Shikano, “Localization of multiple sound sources based on CSP analysis with a microphone array (in Japanese),” *IEICE Trans. Inf & Syst.*, vol.J83-D-II, no.8, pp.1713–1721, August 2000.

[8] T. Otsuka, T. Kojima, K. Ikeguchi and R. Matsuzaki, “Improvements on the receiving functions of disaster prevention broadcasting systems based on audio data hiding (in Japanese),” *IEICE Tech. Rep.*, vol.121, no.29, EMM2021-8, pp.42–47, May 2021.