# Audible secret keying for Time-spread-echo based Audio watermarking

Tatsuya MATSUMOTO
Department of Engineering,
Nagasaki University,
Nagasaki, Japan

Kotaro SONODA
Division of Electrical Engineering and Computer Science,
Graduate School of Engineering, Nagasaki University
Nagasaki, Japan,
sonoda-iihmsp15@cis.nagasaki-u.ac.jp

***Abstract*—This paper deals with the pseudo noise (PN) generating method for digital audio watermarking using time-spread echo hiding. In time-spread echo based audio watermarking, the secret payload is embedded in the form of multiple echoes spread by a pseudo noise sequence and the pseudo noise sequence is used as secret key. Generally, the pseudo noise sequence is required to be uncorrelated with other sequence and therefore typically generated by an M-sequence generator. In this paper, we propose a key sequence generating method which generates a key sequence from an audio signal. By using the new key sequence generated from an audio signal instead of an M-sequence, the key information users have to remember is their secret audio signal but not a complicated random PN key. Therefore, the availability of digital audio watermarking would be improved. After introducing the new key generation method, we evaluated the proposed key sequence. The result shows it can detect a secret payload similar to use of a conventional M-sequence and does not deteriorate inaudibility of the embedded watermark.**

***Index Terms*—time-spread echo hiding; audio watermarking;**

## I. Introduction

In the secure watermarking process of embedding and detection, key information which is shared only with limited entity is required. The time-spread echo based audio watermarking technique[1] represents the watermarked secret bit information as added multiple echoes of a host signal and pseudo noise (PN) is used in the echo generation kernel. The pseudo noise sequence used is required to share the process of embedding and detection as the key information. To detect the embedded secret payload from the stego signal, the time-spread echo method distinguishes the watermark bit payload from a correlation between the echo components in the cepstrum of the stego signal and the shared pseudo noise sequence. Increased the randomness of the pseudo noise sequence increases robustness against disturbances and increases security against a fake key sequence. Conventionally, therefore, pseudo noise that has a low correlation with the other in certain period length, e.g. M-sequence and Gold sequence, being used as key sequences. However, such pseudo noise is difficult to remember.

In this paper, we use an audio signal as a watermarking key sequence instead of pseudo random noise sequences. The audio signal used as the watermarking key is available for two reasons. Firstly, what users have to remember is only which audio signal is used as the watermarking key. And secondly, the key is difficult for pirates to be noticed as the watermarking key.

In section II, the conventional time-spread echo method is introduced. The proposed method for generating watermarking key by using an audio signal is described in section III. Then we evaluate the message detection property in the case of using the proposed watermarking key in sections IV and V. Section VI presents the resulting conclusion.

## II. Time-spread echo hiding

The diagram of the time-spread echo hiding is basically similar to the echo hiding[2]; adding echoes started at time delay $\tau_w$ corresponded with a watermark bit, $w_i$, to the segmented host signal $x_i(n)$. The echo adding kernel, $k_w(n)$, $w \in \{0, 1\}$, is defined as following equation (1), and the watermarked stego signal segment $y_i(n)$ is produced by convolution of $x_i(n)$ with $k_w(n)$.

$$k_w(n) = \delta(n) + \beta p(n - \tau_w) \tag{1}$$

In equation (1), $\delta(n)$ is Kronecker delta $\delta_{n,0}$, $\beta$ is the gain of echoes, and $p(n)$ is the echo transfer function. $\tau_w$ is the delay time where adding echoes started and it is corresponded with watermark bit $w$. In the case of conventional echo hiding, the $p(n)$ is also Kronecker delta $\delta_{n,0}$. In the case of the spread-echo hiding, the $p(n)$ is made of random binary sequence (Pseudo noise sequence; PN sequence) and a convolution process with the $p(n)$ works to spread single-echo to multi-echoes. During the watermark detection process, it takes advantage of that the cepstrum of observed stego signal $\hat{y}(n)$ repesents the echo kernel in lower qefrencies. Its cross-correlation with $p(n)$ has a peak at certain lag point. By comparing the lag point with the delay $\tau_w$, the corresponding watermark bit $w$ is identified.

## III. PN-like key generation from audible signal

In spread-echo hiding, the process to cross correlate between the cepstrum of the stego signal $\hat{y}(n)$ and spread-echoes $p(n)$ is equivalent to despreading of the echo kernel. Therefore, $p(n)$ is required have a low correlation with other sequences of certain lengthes and therefore M-sequence or Gold-sequence is often used as $p(n)$. However, such pseudo noise is difficult to remember. In this paper, we use an audio signal as a watermarking key sequence instead of such pseudo random noise sequences.

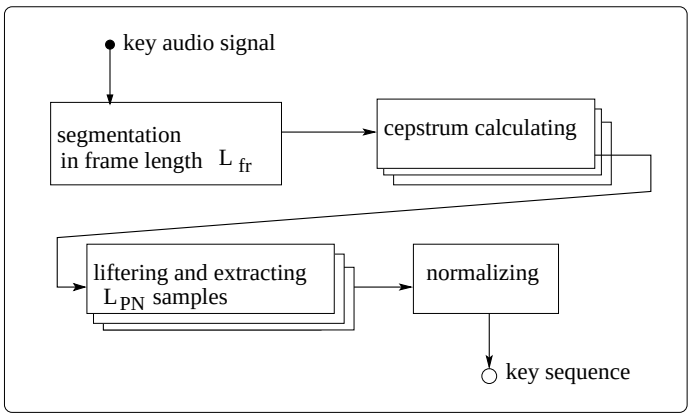


Fig. 1. Blockdiagram for key generation procedure from secret key audio signal

The watermarking key sequence we must generate is a key sequence of length $L_{PN}$ made from an audio signal $s(n)$ of length $L_{sig}$. In this paper, as the simplest case, we generate the key sequence by the cepstrum. A key sequence $p(n)$ is made by averaging the cepstrum coefficients of $L_{fr}$ samples with a window shifting $L_{shift}$ samples on an audio signal $s(n)$, extracting the $L_{PN}$ coefficients from the low quefrency range and making the standard deviation of them to one.

The cepstrum (power cepstrum) $c(\tau)$ is defined by

$$c(\tau) = \mathcal{F}^{-1}[\log S_\omega] \tag{2}$$

where $S_\omega$ is the amplitude spectrum of the time-domain signal $s(n)$ given by

$$S_\omega = |\mathcal{F}[s(n)]| \tag{3}$$

In Fig. 1, the blockdiagram for the proposed key generation procedure from key audio signal is shown.

We compute totally $N = \frac{L_{shift}}{L_{sig}-L_{fr}}$ cepstra from an audio signal $s(n)$ of length $L_{sig}$ with a frame length $L_{fr}$ and a shifting length $L_{shift}$. Then our watermarking key sequence $p(n)$ is generated by

$$p(n) = Norm\left[\frac{1}{N}\sum_{m=1}^{N} c_m(n - q_{th})\right] \tag{4}$$

where $q_{th}$ is cut-off delay and the function $Norm[\cdot]$ makes standard deviation of the sequence to one.

Figure 2 shows an example of the generated key sequence and the base music signal.

## IV. Evaluation of generated key

### A. embeddable bits

The generated key sequences are evaluated in terms of the embeddable bits. Both the audio signals utilized in generating the key sequence and the host signals are picked from SQAM[3] which is provided by EBU. Table I shows the audio file list for generating the key sequence and table II shows that for host signal. All of them are sampled in 44.1 kHz, 16 bit-PCM stereo. The key audio signals are the 10 second's length from 5 seconds of each signals. To generate 1023 key length sequence from 10 seconds (441000 samples), we set a frame length $L_{fr}$ of 4092 samples. The parameters of the watermarking process are summarized in Table III.

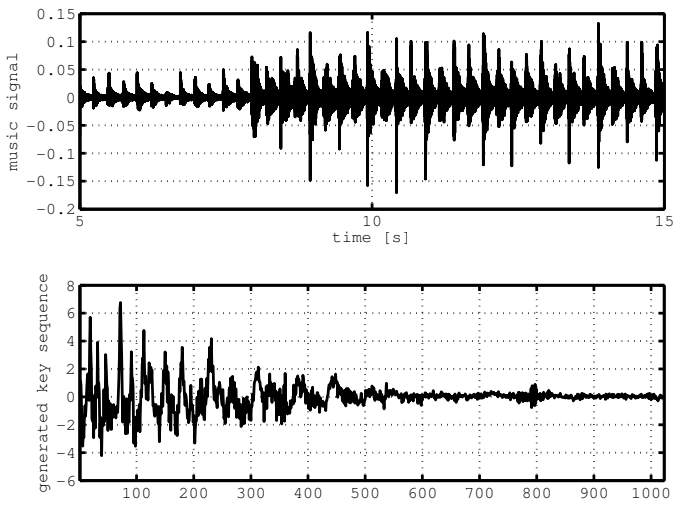


Fig. 2. Generated key sequence and the base music signal

TABLE I
KEY SIGNALS

| Track | Genre |
|---|---|
| 39 | Grand Piano |
| 48 | Quartet |
| 55 | Trumpet |
| 60 | Piano |
| 61 | Soprano |

TABLE II
HOST SIGNALS

| Track | Genre |
|---|---|
| 27 | Castanets |
| 32 | Triangles |
| 35 | Glockenspiel |
| 40 | Harpsichord |
| 65 | Orchestra |
| 66 | Wind Orchestra |
| 69 | ABBA |
| 70 | Eddie Rabbit |

As a result of the evaluation, Figure 3 shows the ratio of the successfully embedded bits against host signals for the tested key signals and a conventional PN sequence.

From Fig. 3, all stego signals watermarked by five key signals are embedded in almost the same ratios compared with a case of watermarked by a conventional PN sequence.

The results of some tested host signal show the low bit ratios of successfully embedded in both case of our keys and the conventional key, because the host signal have many silent frames or much pre-echoed frames.

### B. Detection by fake key signal

In this subsection, we tried to detect the watermark by using fake key signals. The host signal is #40. After generating stego signals by using the true key signal, the watermark bits were detected by using fake key signals. In Fig. 4, bit error rates in case of using fake key signals are depicted. Counted amount

TABLE III
PARAMETER SETUPS FOR WATERMARKING

| | |
|---|---|
| Echo gain $\beta$ | 0.006 |
| Length of PN sequence $L_{PN}$ | 1023 |
| Delay of Echo $\tau$ | 1 ms(44 points)<br>2 ms(88 points) |
| Embedding Bit rate | 5.38 bps |

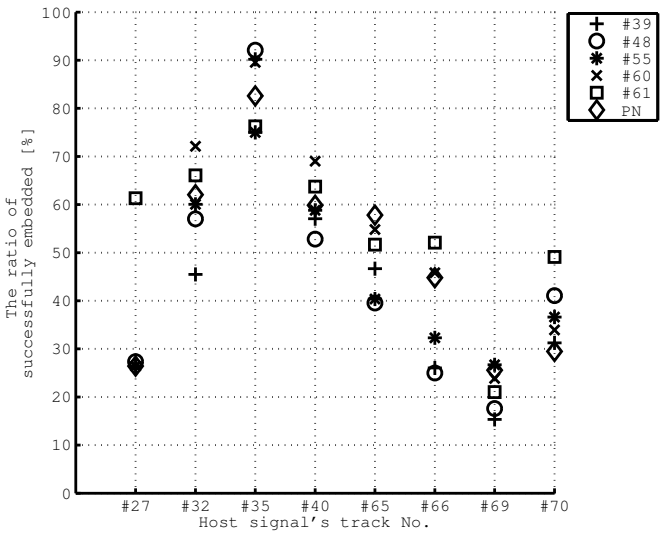


Fig. 3. Bit ratios of successfully embedded

of bits are number of successfully embedded bits in previous subsection.

Figure 4 shows that the cases using fake key signals results in higher than 50% BER. Then the proposed key sequence generated from a music signal is sufficiently secure compared with a conventional PN sequence.

## V. Evaluation of time-spread echo method using audible secret key

To evaluate the detection robustness and the sound quality of the watermarked signal with the key generated from music signal, we carried out two experiments. In our experiments, eight host signals from a SQAM database (Track No. 27, 32, 35, 40, 65, 66, 69, and 70) were used. The all host signals had a 44.1 kHz sampling rate, 16 bit quantization, and monaural. The key is generated by using the 10 seconds length of SQAM track No.60. The watermark embedding settings are the same as before and are shown in Table. III. Their evaluation criteria, attacking manipulations and host signal sources are based on Information Hiding Criteria version.3 [4].

Our proposed method is almost same watermarking technique with original time-spread echo watermarking[1] except for the introduction of the key sequence generated from music signal.

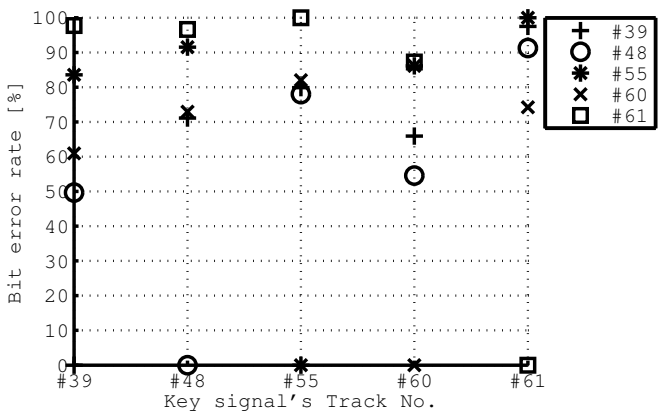


Fig. 4. Bit error rates: Detection by fake keys

### A. Robustness against attacks

First test evaluates the robustness against signal manipulation attacks. Watermarked stego signals were manipulated by attacks listed in Table. IV. Their listed attacks are the recommended manipurations which the stego files should be tested about the robustness by Information Hiding Criteria Committee[4]. The results of the robustness tests are shown in Fig. 5.

From the results in Fig. 5, while the BERs on detection for Track No.27 and 66 are high (about 20%), the BERs on almost other tracks are lower than 10 % against the attacks.

The result shows that the watermarking method introducted our new music key sequence has the robustness against the typical manipulations.

### B. Objective evaluation of audio quality

The second test evaluates the sound quality of the watermarked stego signals. In our experiment, we conducted objective tests using PEAQ (the perceptual evaluation of audio quality)[5]. The PEAQ measures the deterioration of the signal from another signal and scores the deterioration on a scale called ODG (objective difference grade), from -4 (Very annoying) to 0 (Imperceptible) as shown in Table V.

The scored ODG of the stego signals are shown in Table. VI. The results show that the stego signals are scored higher than -1.7 and the deterioration was perceptible, but not annoying.

## VI. Conclusion

This paper presented the evaluation of the key generated from a secret music signal for the well-known time-spread echo based audio watermarking method. In this paper, the key sequence is generated by taking the lower quefrency range in averaged cepstrum of the secret music signal. From our experiment, the generated key sequence has almost as same detection performance as the conventional PN sequence. Furthermore, the stego signal watermarked with our key sequence is robust against many attacks and satisfies inaudibility.

TABLE IV
ATTACK CONDITIONS FOR EVALUATION [4]

| attack | conditions | abbreviation at Fig 5 |
|---|---|---|
| MP3 compression | 128 kbps const. rates | mp3o |
| noise addition | S/N = 36dB | wgn0 |
| bandpass filtering | 100 Hz $\sim$ 6 kHz, -12 dB/oct. | bapf |
| pitch stretching w/o invariant duration (1) | +4% | tsmp |
| pitch stretching w/o invariant duration (2) | -4% | tsmm |
| time stretching (1) | +10% | spep |
| time stretching (2) | -10% | spem |
| echo addition | 100 ms, -6 dB | echo |
| two times MP3 compression | 128kbps const. rates | mp3t |

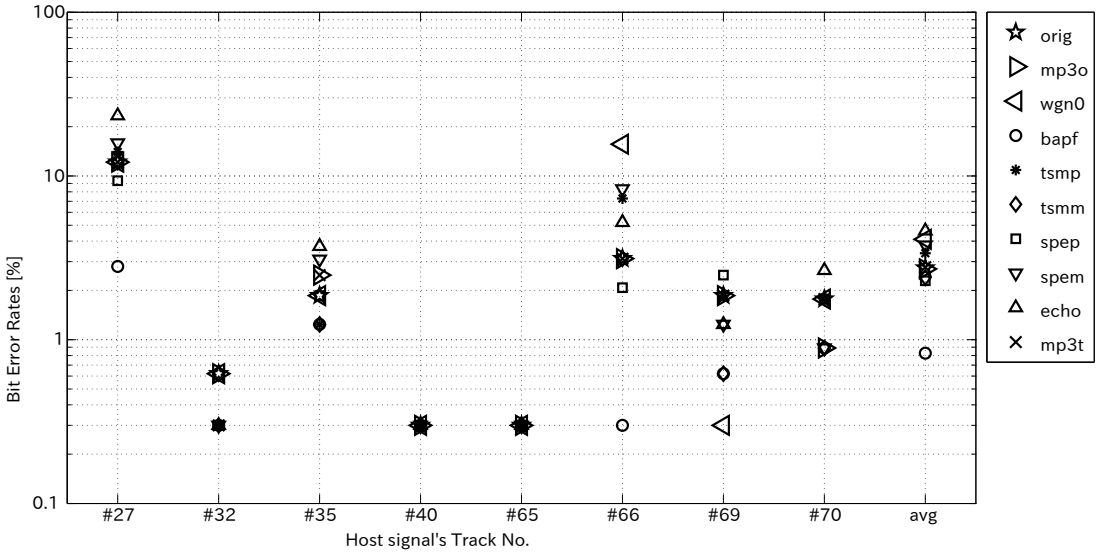


Fig. 5. results of the robustness tests: Bit error rates of stego signals against manipulation attacks

TABLE V
ODG SCORES AND DESCRIPTION IN PEAQ

| ODG score | description |
|---|---|
| 0 | Imperceptible |
| -1 | Perceptible, but not annoying |
| -2 | Slightly annoying |
| -3 | Annoying |
| -4 | Very annoying |

TABLE VI
SCORED ODGS FOR THE STEGO SIGNALS

| Stego signal's Track No | Scored ODG |
|---|---|
| No.27 | -1.56 |
| No.32 | -0.86 |
| No.35 | -0.97 |
| No.40 | -1.20 |
| No.65 | -1.74 |
| No.66 | -1.71 |
| No.69 | -1.58 |
| No.70 | -1.51 |
| Average | -1.39 |

As furture work, we aim to utilize a personal voice as the key. A method to generate the key from a voice securely and conviniently is required.

## REFERENCES


[1] B.-S. Ko, R. Nishimura, and Y. Suzuki, "Time-spread echo method for digital audio watermarking," *IEEE Transactions on Multimedia*, vol. 7, no. 2, pp. 212–221, 2005. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1407894

[2] D. Gruhl, A. Lu, and W. R. Bender, "Echo hiding," in *Information Hiding*, 1996, pp. 293—-315.

[3] Europian Broadcasting Union (EBU). (2008) Sound quality assessment material. [Online]. Available: http://tech.ebu.ch/publications/sqamcd

[4] Information Hiding Committee Audio Group, "IHC Evaluation Criteria and Competition: Watermark Criteria for Audio (ver. 3)," 2014. [Online]. Available: http://www.ieice.org/iss/emm/ihc/IHC_criteriaVer3.pdf

[5] International Telecommunication Union, "ITU Recommendation: Method for Objective Measurements of Perceived Audio Quality (PEAQ)," 2001, no. ITU-R BS 1387-1.