AUDIO DATA HIDING BY USE OF BAND-LIMITED RANDOM SEQUENCES

Mikio IKEDA†, Kazuya TAKEDA† and Fumitada ITAKURA‡

†Graduate School of Engineering, ‡Center for Information Media Studies, Nagoya University
Furo-Cho, Chikusa-ku, Nagoya, Aichi 464-8603, JAPAN
E-mail:ikeda@nuee.nagoya-u.ac.jp

ABSTRACT

This paper proposes the use of band-limited random sequences to introduce further flexibility in the spread spectrum based audio data hiding. To realize the sub-band data hiding, a systematic method is developed in order to generate band-limited and orthonormal random sequences of any length. In experiments, we evaluated the selective use of frequency channels to be used for information embedding, and the robustness against the MPEG1 layer 3 encoding and decoding. From the results, it is clarified that the proposed method is robust against more than 160 kbps MPEG1 coding and decoding when the center frequency of the sub-band is lower than 11 kHz.

1. INTRODUCTION

The data hiding or digital watermarking technique has been required and developed for the copyright protection and authentication of multimedia contents. In the evaluation of the data hiding performance, not only imperceptibility of watermarks but also the robustness against signal processing such as the AD/DA conversion, the encoding and decoding of source signals is required.

Among the data hiding techniques for audio signals, the use of the spread spectrum using pseudo random sequence[1], phase coding using all-pass filtering [4] and echo hiding[5] have been investigated. In the data hiding technique using spread spectrum, digital information is encoded by superimposing random sequences. If the power level of the random sequences is below the perception level throughout the signal band, the distortion is expected to be imperceptible. From the viewpoint of decoding, on the other hand, there is a trade-off between the length of random sequences and the relative power level to the source signal. In general, long random sequences, which limit the data rate of hiding, are required to keep the distortions imperceptible.

In order to improve the data rate of hiding, it seems to be an effective approach to shape the spectrum of random sequence up to the imperceptive level of the source signal. The method using the masking characteristics of the auditory system[2][3] is one of the effective approaches on this direction.

In this paper, we propose the use of band-limited random sequences to introduce further flexibility in the spread spectrum based audio data hiding. To realize the sub-band data hiding, a systematic method is developed in order to generate band-limited and orthonormal random sequences of any length. Besides, the random sequences can be used for overlapping frame processing. Typically, as an extension of shaping random sequence, we will evaluate the selective use of frequency channels to be used for information embedding. Through the development of the algorithm, we assume that the data rate of hiding is about 50 bits per second and the decoding does not require the original source signal.

In the rest part of paper, we will describe the detail algorithm of the sub-band data-hiding method and random sequence generation. Furthermore, the experimental evaluation using signal-to-deviation ratio (SDR) and the robustness against the MPEG1 layer-3 encoding and decoding are investigated for several information rates.

2. ALGORITHM

2.1. Basic Scheme

The Proposed data hiding method is an extension of spread spectrum method [1] because of the use of band-limited random sequences. The basic scheme of the hiding can be summarized as below.

Given a sequence of code symbols, $S(p), p = 0, 1, \cdots,$ $P-1$, data imposed signal $y(n)$ of the p-th frame is calculated as follows:

$$\begin{aligned} y(n+pB) &= x(n+pB) + u(p)h_{2B}(n; S(p)) \\ &\quad + u(p-1)h_{2B}(n+B; S(p-1)) \\ &\quad (n = 0, 1, \cdots B-1). \end{aligned} \tag{1}$$

where, $x(n)$ is the original signal, $u(p)$ is a modulation index, and $h_{2B}(n; i)$ is the i-th basis function of a set of $2B$-

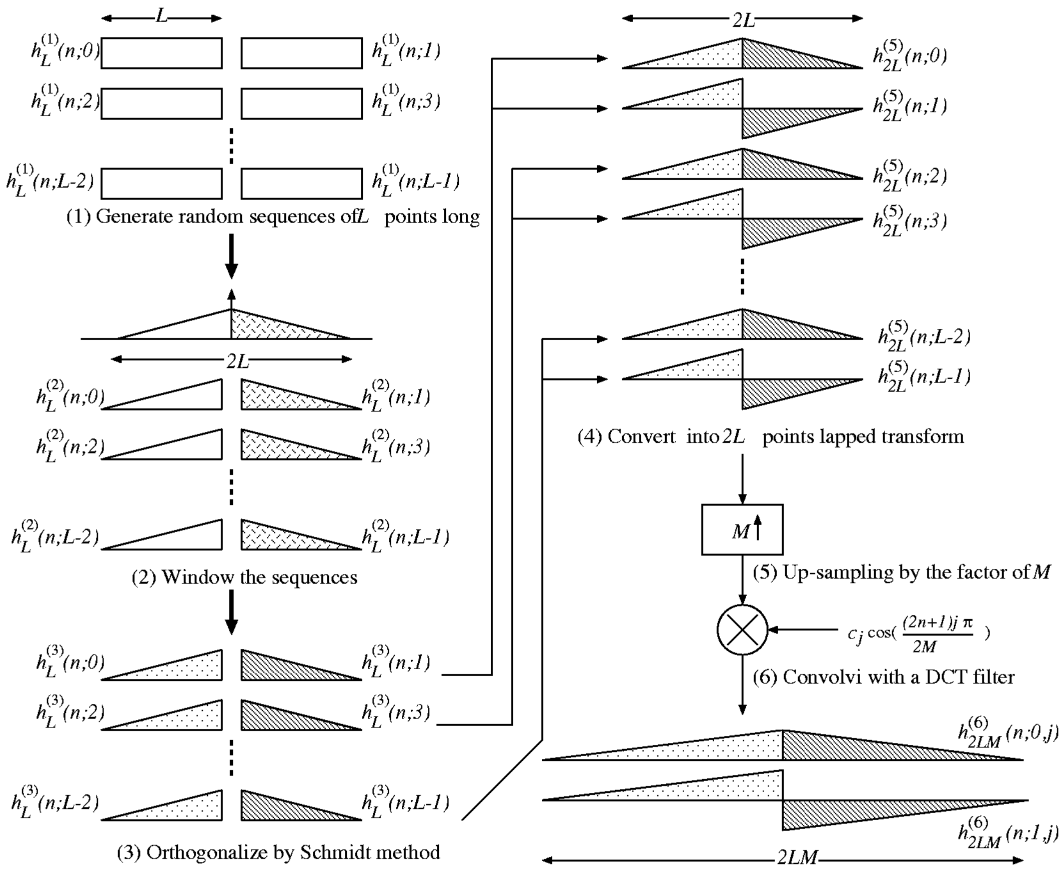

Figure 1: Procedures of designing random sequences.

point length orthonormal random sequence, i.e.,

$$\sum_{n=0}^{2B-1} h_{2B}(n;i)h_{2B}(n;j) = \delta(i-j), \tag{2}$$

$$\sum_{n=0}^{B-1} h_{2B}(n;i)h_{2B}(n+B;j) = 0. \tag{3}$$

As shown in the equation, adjacent two code symbols, $S(p)$ and $S(p-1)$ are superimposed simultaneously through two different (orthogonal) bases for overlapping processing. The design of the basis will be detailed in the following sections. The modulation index $u(p)$ is calculated as the cross-correlation between the original signal and the basis sequences:

$$u(m) = \alpha \,\mathrm{sgn}(r(p, S(p))) \max_k |r(p,k)|, \tag{4}$$

$$r(p,j) = \sum_{n=0}^{n=2B-1} x(n+pB)h(n,j), \tag{5}$$

where $\mathrm{sgn}(z)$ is a sign function given by

$$\mathrm{sgn}(z) = \begin{cases} 1 & (z > 0) \\ -1 & (z \le 0). \end{cases} \tag{6}$$

Extracting symbol sequence is performed by taking cross-correlation or matched filtering as follows:

$$\hat{S}(p) = \underset{k}{\mathrm{argmax}} \left| \sum_{n=0}^{2B-1} y(n+pB)h_{2B}(n;k) \right|. \tag{7}$$

In eqn. (4), α controls gain of the random sequence to be added. If there is no error or distortion in transmission, α which is greater than 1.0 guarantees error-free decoding. In general, greater α improves the robustness of correct decoding. However, the distortion of the music source is more likely perceived. Therefore, the trade-off between distortion and robustness is one of the center issues in the algorithm development.

2.2. Designing Random Sequences

The three issues concerned with designing random sequence are the key technology in the proposed method:

1. orthogonality among overlapped blocks for avoiding block distortion,

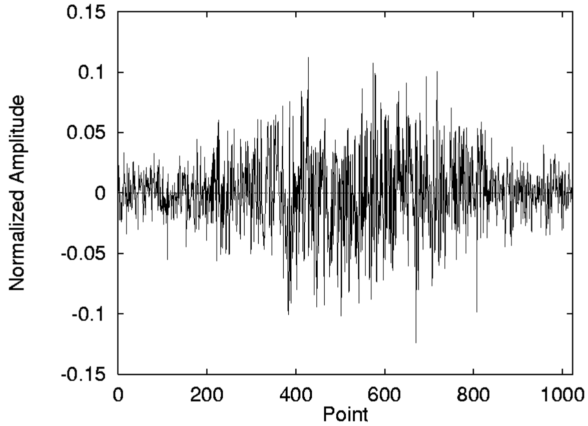

Figure 2: Example of generated random sequence ($2LM = 1024$).

2. band limitation for flexible use of frequency channels, and
3. effective algorithm for generating long orthogonal sequences.

The developed algorithm is depicted in Figure 1. The procedure of the designing random sequence follows.

1. Generate L random sequence of L point length; $h_L^{(1)}(n;i)$, $n = 0,1,2,\cdots L-1, i = 0,1,2\cdots L-1$.
2. Shape each sequence by multiplying a first and last half of a $2L$ point hamming window to even and odd numbered random sequences, respectively.

$$h_L^{(2)}(n;i) = \begin{cases} w(n)h_L^{(1)}(n;i) & (i : \text{even}) \\ w(n+L)h_L^{(1)}(n;i) & (i : \text{odd}) \end{cases} \tag{8}$$

$$w(n) = 0.54 - 0.46\cos(n\pi/L) \tag{9}$$

3. Orthogonalize the L shaped random sequences by Schmidt method,

$$h_L^{(2)}(n;i) \rightarrow h_L^{(3)}(n;i)$$

so that

$$\sum_{n=0}^{L-1} h_L^{(3)}(n;i)h_L^{(3)}(n;j) = \delta(i-j).$$

4. Concatenate even and odd numbered sequences to form L random sequences of $2L$ point length.

$$h_{2L}^{(4)}(n;2i) = \begin{cases} \frac{1}{\sqrt{2}}h_L^{(3)}(n;2i) \\ \quad (n = 0,1,\cdots,L-1) \\ \frac{1}{\sqrt{2}}h_L^{(3)}(n-L;2i+1) \\ \quad (n = L, L+1,\cdots 2L-1) \end{cases} \tag{10}$$

$$h_{2L}^{(4)}(n;2i+1) = \begin{cases} \frac{1}{\sqrt{2}}h_L^{(3)}(n;2i) \\ \quad (n = 0,1,\cdots,L-1) \\ -\frac{1}{\sqrt{2}}h_L^{(3)}(n-L;2i+1) \\ \quad (n = L, L+1,\cdots 2L-1) \end{cases} \tag{11}$$

5. Extend the length of each sequence by up-sampling by the factor of M

$$h_{2LM}^{(5)}(n;i) = \begin{cases} h_{2L}^{(4)}(n/M;i) & (n \bmod M = 0) \\ 0 & (\text{otherwise}). \end{cases} \tag{12}$$

6. Band-limit onto j-th frequency channel by convolving with a DCT filter whose center frequency is indexed by j

$$h_{2LM}^{(6)}(n;i,j) = h_{2LM}^{(5)}(n;i)c_j\cos(\frac{(2n+1)j\pi}{2M}), \tag{13}$$

$$c_j = \begin{cases} \sqrt{2/M} & (j=0) \\ \sqrt{1/M} & (\text{otherwise}). \end{cases} \tag{14}$$

Finally, L orthonormal random sequences of $2LM$ point long are generated for M frequency channels.

3. EXPERIMENTS

As described in the previous section, the use of DCT as the sub-band dividing filter enables us to embed given information in a specific sub-band. In this section, we investigate the best selection of sub-band in which the given information is embedded. Furthermore, we evaluate the robustness against the MPEG1 layer-3 coding and decoding for several information rates.

3.1. Sub-band Selection

In this experiment, we examined the number of sub-band and the center frequency of the sub-band. The experimental conditions are shown in Table 1. The signal-to-deviation ratio (SDR) between the original and data-hidden signals is used to evaluate the performance of the proposed method.

The resultant SDRs for 8, 16 and 32 bands are shown in Figure 3 as a function of the center frequency of sub-band. As shown in the figure, the distortions introduced by the proposed method are almost the same regardless of the number of sub-band. In addition, higher SDR are attained, as the center frequency becomes higher. Since the components of music signal in higher frequency bands are small, this result represents the basic characteristics that the proposed method introduces the distortion proportional to the component of signal included in the sub-band.

Table 1: Experimental Conditions

Music material	Symphony No.6 Mahler (initial 120 s)
Sampling Frequency	44,100 Hz
Channels	2
Frame Shift LM	4096 samples (92.9 ms)
Number of sub-band M	8, 16, 32
Gain control factor α	2.0
Number of symbols	8 (3 bits for each frame)
Data rate	64.6bps

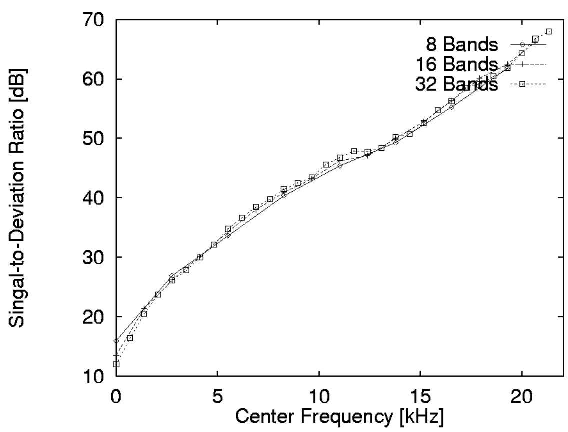

Figure 3: Signal-to-Deviation Ratio with respect to center frequency and number of sub-band.

3.2. Robustness against MPEG1 layer-3 coding and decoding

The selected information rates were 320 kbps, 160 kbps and 80 kbps. In the experiments, the number of sub-band division M was fixed to be eight. The other experimental conditions were the same in the previous experiments as shown in Table 1. The experimental results measured by symbol error rate are shown in Table 2. The results can be summarized by the following three points: (1) there is no error in the case of 320 kbps information rate and the center frequencies below 11,025 Hz, (2) the symbol error rate becomes large as the center frequency is higher, (3) the hidden information can hardly extract for 80 kbps information rate. It should be noted that the distortion introduced by the MPEG1 coding is perceptible. Hence, we can conclude that the proposed method is robust against more than 160 kbps MPEG1 coding and decoding when the center frequency of the sub-band is lower than 11 kHz.

Table 2: Symbol error rates: case of MPEG1 layer 3 coding and decoding

Center Frequency	Symbol error rate (%)		
(Hz)	320 kbps	160 kbps	80 kbps
0	0.0	0.0	15.4
2,756	0.0	0.0	55.5
5,513	0.0	0.0	69.8
8,269	0.0	2.3	78.2
11,025	0.0	5.5	84.9
13,781	2.3	23.4	85.4
16,538	31.9	67.4	87.4
19,294	57.6	86.7	87.2

4. CONCLUSIONS

This paper proposes the use of band-limited random sequences to introduce further flexibility in the spread spectrum based audio data hiding. We evaluated the selective use of frequency channels to be used for information embedding. From the results, the distortions introduced by the proposed method are almost the same regardless of the number of sub-band. In addition, we evaluated the robustness against the MPEG1 layer 3 encoding and decoding . From the results, it is clarified that the proposed method is robust against more than 160 kbps MPEG1 coding and decoding when the center frequency of the sub-band is lower than 11 kHz.

5. REFERENCES

[1] Ingemar J. Cox, Joe Kilian, Tom Leighton, and Talal Shamoon, “Secure spread spectrum watermarking for images, audio and video,” ICIP’96, Vol III, pp. 234–246 (1996)

[2] Laurence Boney, Ahmed H. Tewfik, and Khaled N. Hamdy, “Digital Watarmarks for Audio Signals,” IEEE Intl. Conf. on Multimedia Computing and Systems, Hiroshima, pp. 473–480 (1996)

[3] Laurence Boney, Ahamed H. Tewfik, and Khaled N Hamdy, “Digital Watamasks for Audio Signals,” European Signal Proc. Conf., Trieste, Italy, September 1996.

[4] Yasemin Yardimci, et al, “Data hiding in speech using phase coding,” ECSA. Eurospeech97, pp. 1679–1683 (1997).

[5] D. Gruhl, A. Lu, and W. Wender, “Echo hiding,” Lecture Notes in Computer Science, Vol. 1174, Information Hiding, pp. 295–315, Springer (1996).