The Watermarking of Digital Sound by Band Division and Spectral Spreading

(帯域分割と直接拡散を用いた音響データへの電子透かし)

Munetoshi Iwakiri (member)[†], Kineo Matsui (member)[†]

Abstract This paper presents a technique for the digital watermarking of musical sound. It consists of two main steps: the first is the division of the audio-frequency band into several sub-bands and the second is the embedding of a watermark in one of these bands by the spectral spreading method. The proposed scheme produces a watermark that is robust against data compression and ordinary attacks, while retaining high quality for the sound because we randomly select the specific band for the embedding of the watermark among the many sub-bands. For example, our experimental results have shown that the quality of watermarked music is good in about 45dB and that the watermark remains detectable after MPEG1 Audio LayerIII compression and jitter attacks.

Key words: Watermark, Spread spectrum, Sub-band, MPEG, Jitter attack.

1. Introduction

Recent development of computer technology and networks brings us more utilization of high quality digital contents. However, a serious problem has arisen in copyright protection of the digital contents from the networked worlds, because the contents are easily copied out of any permission and widely distributed without the degradation. As one of the solution for this problem, a watermarking technique is introduced recently to digital image¹⁾. This is to embed some copyright information into the image so that any user may not perceive it in advance while the content is circulated, and is to activate the watermark if an illegal use would be detected in the network.

Watermarking techniques proposed now are most applied for visual contents such as pictures and videos²⁾. There are a few studies for music sound up to date, but severe conditions for the sound are imposed on the watermark in practical use. One of them is a watermarking scheme which uses the spectrum spreading algorithm in communication³⁾. It is robust for attacks because each watermark needs different random sequences. It is, especially, reported that the direct sequence scheme of spectrum spreading can generate a watermark which is

resistive to data compression⁴⁾.

On the other hand, if the band-width to embed the watermark is too narrow, it brings us a small jarring noise in the music sound. The direct sequence scheme is less resistive to a jitter attack because it changes the sampling rate of sound in time. In this paper, we try to improve this scheme by using the sub-band division technique so that the embedded watermark becomes robust under MPEG compression and jitter attack.

First, a series of sound data are classified into two sets of component in high and low frequency ranges based on sub-band technique⁴⁾. This process is recurrently applied to each low band and it is divided into the specified sub-bands. Next, a band is chosen from them and a watermark is embedded into the band. This results in decreasing the amount of processing because the band-width to embed the watermark is less than the one in the original scheme⁵⁾. Moreover, the spectrum components of watermark are spread onto the components in low frequency domain, and it generates only a bit of noise which we may not find. This paper provides another new tool to resist digital processing in amplitude.

2. Band Division

We first define several symbols in this section; let $l_0(t)$, $l_1(t)$ and $h_1(t)$ be sampled sequences of component L_0 in the original band, L_1 in the low frequency band and H_1 in high frequency band at time $t = 0, 1, 2, \dots$, respectively, where $l_1(t)$ and $h_1(t)$ are written by $l_0(t)$

Received October 29, 2001; Revised January 30, 2002; Accepted March 7, 2002

[†]Department of Computer Science, National Defense Academy
(1-10-20 Hashirimizu, Yokosuka-shi, 239-8686, Japan)

as

$$l_1(t_1) = \frac{l_0(2 \cdot t_1) + l_0(2 \cdot t_1 + 1)}{2} \quad (1)$$

$$h_1(t_0) = l_0(t_0) - l_1(\lfloor \frac{t_0}{2} \rfloor) \quad (2)$$

where $t_1 = t_0/2$. Equations (1) and (2) show that L_1 becomes a half of L_0 , which is a low frequency band, and H_1 consists of high-frequency components as well as those in L_0 . Repeating these frequency divisions, we have the following sequences step by step as

$$\begin{aligned} L_0 &\rightarrow \{L_1, H_1\} \\ L_1 &\rightarrow \{L_2, H_2\} \\ &\vdots \\ L_{j-2} &\rightarrow \{L_{j-1}, H_{j-1}\} \\ L_{j-1} &\rightarrow \{L_j, H_j\} \end{aligned} \quad (3)$$

This processing is shown in **Fig. 1**. From the disjoint bands of L_1 and H_1 , inversely, we have the original band L_0 by

$$l_0(t_0) = l_1(\lfloor \frac{t_0}{2} \rfloor) + h_1(t_0) \quad (4)$$

where $t_0 = 0, 1, 2, \dots$. In the same way, we restore the upper bound L_{i-1} from both the sub-band L_i and H_i .

3. Proposal Technique

3.1 Watermarking Method

A schematic diagram of our watermarking method is shown in **Fig. 2**. First, a source data S is a sequence of signals $s(t)$ for $t = 0, 1, 2, \dots$, where S is equal to L_0 in this paper, and it is divided into sub-bands L_j and $\{H_1, H_2, \dots, H_j\}$ through each filter shown in Fig. 1. Next, we apply the direct sequence scheme (i.e., DSS) of spectrum spreading to these bands so as to diffuse a watermark signal into a specified band.

Let $g(t_j)$ be the value of random PN sequence at time t_j which consists of +1 or -1, and k_s be the key for generating the PN sequence, where t_j is a sampling time

Fig. 1 Basic processing of the band division.

Fig. 2 The processing block for embedding.

on L_j . Then we have a diffused signal $x(t_j)$, i.e.,

$$x(t_j) = l_j(t_j) \cdot g(t_j) \quad (5)$$

which is spread in the broad band. The signal $x(t_j)$ is, next, converted in the frequency area by the modified discrete cosine transformation (i.e., MDCT). This transformation requires $2M$ samples at t_j and generates M coefficients in spectrum as shown in **Fig. 3**. Let $X_i(k)$ be an MDCT coefficient at i -th frame, i.e.,

$$X_i(k) = \frac{2}{M} \sum_{n=0}^{2M-1} w(n) c(k, n) x(n + iM) \quad (6)$$

where $0 \leq k \leq M - 1$ and $0 \leq n \leq 2M - 1$. In Eq.(6), $w(n)$ and $c(k, n)$ are defined as

$$w(n) = \sin\left(\frac{\pi(2n+1)}{4M}\right) \quad (7)$$

and

$$c(k, n) = \cos\left(\frac{\pi(2k+1)(2n+M+1)}{4M}\right) \quad (8)$$

where $0 \leq k \leq M - 1$ and $0 \leq n \leq 2M - 1$.

Now, we define a random number r_i that specifies a frequency component to embed a watermarking bit b_i , and a positive number p that controls the watermarked level, where p is experimentally estimated under considering the sound quality and the robustness of watermark. Let R stand for a set of $\{r_i | i = 0, 1, \dots, M - 2\}$.

Fig. 3 The window function.

It will be a key for restoring the embedded signals to the watermark. Next $X_i(r_i)$ and $X_i(r_i + 1)$ are selected as shown in **Fig. 4** and are compared with the absolute value of them, where it is denoted by $|X|$, and our watermark can resist the level modulation attack (see Section 4.2) than the one in other techniques using the fixed quantization parameter⁵⁾. Then the following algorithm is applied:

Embedding algorithm

(a) To embed $b_i = 0$:

if

$$|X_i(r_i + 1)| - |X_i(r_i)| \geq p, \quad (9)$$

then $X_i(r_i)$ and $X_i(r_i + 1)$ remain as these are, otherwise $X_i(r_i + 1)$ is changed with

$$|X_i(r_i + 1)| \leftarrow |X_i(r_i)| + p. \quad (10)$$

(b) To embed $b_i = 1$:

if

$$|X_i(r_i)| - |X_i(r_i + 1)| \geq p, \quad (11)$$

then $X_i(r_i)$ and $X_i(r_i + 1)$ remain as these are, otherwise $X_i(r_i)$ is changed with

$$|X_i(r_i)| \leftarrow |X_i(r_i + 1)| + p. \quad (12)$$

This completes the embedding process when $|X_i(r_i)|$ and $|X_i(r_i + 1)|$ are replaced with the corresponding absolute values of MDCT coefficients, respectively. Thus, let $X'_i(k)$ be the watermarked $X_i(k)$. By using $X'_i(k)$ where $k = 0, 1, \dots, M - 1$, we have

$$x'_i(n) = w(n) \sum_{k=0}^{M-1} c(k, n) X'_i(k) \quad (13)$$

where $0 \leq k \leq M - 1$ and $0 \leq n \leq 2M - 1$, and these become wave elements $x'_i(n)$ at frame i . For every i , we add the wave elements as

$$x'(n + iM) = x'_{i-1}(n + M) + x'_i(n) \quad (14)$$

where $0 \leq n \leq M - 1$, then it generates $x'(t_j)$ at time t_j in which the watermark is embedded. Since $\{g(t_j)\}^2 = 1$, we have

$$\begin{aligned} x'(t_j) \cdot g(t) &= l'_j(t_j) \cdot \{g(t_j)\}^2 \\ &= l'_j(t_j). \end{aligned} \quad (15)$$

This processing is applied to the component L_j . Then we have the watermarked audio data S' from L'_j and $\{H_1, H_2, \dots, H_j\}$. Using the sampling rate B of original sound and embedding parameter (j, M) , the watermarking bit rate β is calculated by

$$\beta = \frac{B}{M \cdot 2^j}. \quad (16)$$

For example, when $B = 44,100[\text{Hz}]$, $j = 10$ and

(a) $b_i = 0$

(b) $b_i = 1$

Fig. 4 The embedding control.

$M = 16$, we say that the embedded rate is $2.69 (= 44,100 / (16 \times 2^{10}))[\text{bit/s}]$. When the bit rate for embedding the watermark is constant ($\beta = 2.69[\text{bit/s}]$), the processing time changes in our experiment as shown in **Fig. 5**. It suggests that the size of frame L_j to embed a watermark becomes smaller as the number of band division (j) is larger, and thus our algorithm can reduce the CPU time for watermarking than the one in other spectrum spreading schemes⁵⁾.

Fig. 5 Processing time $j(2.69[\text{bit/s}])$

Table 1 The experimental sounds			
Name	Samples(R)	Samples(L)	Sec.
Classic	440,832	440,832	10
Jazz	440,832	440,832	10
Dance	440,832	440,832	10

3.2 Decoding method

First, let us divide S' into sub-bands L'_j and $\{H'_1, H'_2, \dots, H'_j\}$. We get $l'_j(t_j)$ from L'_j and $x'(t_j)$ as $x'(t_j) = l'_j(t_j) \cdot g(t_j)$. Then, $X'_i(r_i)$ and $X'_i(r_i + 1)$ are checked and we estimate the embedded signal by the following step ; i.e., if

$$|X'_i(r_i)| < |X'_i(r_i + 1)| \quad (17)$$

then we have $b'_i = 0$, otherwise $b'_i = 1$. Note that nobody can restore the watermark unless $g(t_j)$ for decoding is identical to the one for spreading in Eq.(5) and R is legal for the scheme.

4. Experimental Results

It is desirable that the watermark remains in the signal even if the data are compressed efficiently and that the music sounds well. The data listed in Table 1 are used for our experiments. We study the effect of watermark on music quality and that of digital processing such as popular compression and some attacks on the watermark. The experimental data are digitized at the rate of 44.1 kHz and are expressed in 16 bits per each sample. Ordinarily, these data are constituted in stereo-typed format, but we use only a single channel(right side) of stereo sounds in order to simplify the discussion. We introduce the 32ms segmental signal-to-noise ratio⁶⁾, i.e., $SNR_{seg}(16 \sim 32ms)$, corresponding to subjective evaluation, approximately. A detection rate is defined with the rate of which the decoded bits in the watermark sequence coincide with the embedded ones. For example, when a single bit is not matched in the sequence of 25 bits, we say that the detection rate is $0.96(= 24/25)$.

4.1 Sound Quality

The sound qualities by embedding under L_j for $j = 3, 4, \dots, 10$ are shown in Fig. 6 when the rate of bits embedded and the watermarked level p are constant, respectively. It is shown that the sound quality SNR_{seg} is not less than about 40dB when $j \geq 6$. Since the watermarking bits are embedded to L_j component, it seems to be able to hide the watermark inaudible. The sound qualities SNR_{seg} to p are also shown in Fig. 7 when $j = 10$ and $M = 16$, and those are deteriorated as p is greater. For example, we have about 45dB in

Fig. 6 SNR_{seg} to j ($p = 25, 2.69[bit/s]$)

Fig. 7 SNR_{seg} to p ($j = 10, 2.69[bit/s]$)

the sound level for $p = 25$, but we could not detect any noise on the music.

4.2 Robustness for level modulation attack

The proposed scheme embeds a watermark based on the relative relationship between two frequency components $X_i(r_i)$ and $X_i(r_i + 1)$. Thus, this is robust for attacks such as increasing or decreasing the sound levels. Therefore, we checked if the watermarks embedded could be restored from the spectrum of the sounds shown in Table 1, where the sample was changed with the coefficients of 0.5 to 1.5 under $j = 10$, $p = 25$ and $M = 16$. Then we got a high detection rate close to 1.0. It tells us that our watermarking scheme is robust for level modulation of sound waves.

The watermark is included in the sound that is modulated in band by using the so-called graphical equalization tool so as to keep the low-band components, and then we may detect it easily, also.

4.3 Robustness for channel mixing

We check a watermark against the channel mixing modulation. First, we embedded watermarks to both channels for each data shown in Table 1, where the same parameters were used as $j = 10$, $p = 25$ and

Table 2 Influence on watermark by amplification.

Amplification	0.50	0.75	1.25	1.50
Classic	1.00	1.00	1.00	1.00
Jazz	1.00	1.00	1.00	0.96
Dance	0.96	0.96	0.96	0.96

$M = 16$, and these were made into a single channel by averaging the corresponding watermarked signals on both the channels. We tried to detect the watermarks from the unified channel. Then, the results of this experiment are shown in **Table 3**. It is understood that higher detection rate may be attained in two samples, but the other is not so much. The latter will be caused as the waveform spectrums between both the channels are more different than the formers.

4.4 Robustness for white noise addition attack

We check a watermark against the white noise addition attack. First, we superimposed a white noise consisting of the dynamic range with about 10% to 40% onto a single channel of the watermarked sound under the parameters of $j = 10$, $p = 25$ and $M = 16$. Next, we tried to detect the embedded watermark from this deteriorated sounds, and then the high detection rates shown in **Table 4** are attained. Thus, it tells us that our watermarking scheme is robust for the white noise addition attack.

4.5 Robustness for re-sampling attack

This scheme is also based on the band division with the re-sampling technique. Thus, the watermarking signals are robust for up-or-down sampling ratio attack. Therefore, we checked if the embedded watermark could be restored from the data when the sampling rate is changed under $j = 10$, $p = 25$ and $M = 16$. The experimental results are shown in **Table 5**. It tells us that the watermark by our scheme is robust for the re-sampling attacks, also. But, only the result in Dance deteriorates as compared to the others, and it seems to be greatly affected by the re-sampling attacks, because the frequency component in Dance are widely distributed

Table 3 Influence on watermark by channel mixing.

j	10	9	8	7
Classic	0.80	0.68	0.64	0.80
Jazz	1.00	1.00	1.00	0.96
Dance	0.96	0.72	0.96	0.92

Table 4 Influence on watermark by noise addition.

Noise	10%	20%	30%	40%
Classic	1.00	1.00	1.00	1.00
Jazz	1.00	1.00	1.00	1.00
Dance	0.96	0.96	0.96	0.96

Table 5 Influence on watermark by re-sampling.

Sampling rate	48,000	22,050	11,025	8,000
Classic	1.00	1.00	1.00	1.00
Jazz	1.00	1.00	1.00	1.00
Dance	1.00	1.00	0.79	1.00

Table 6 The effect by the band pass filter.

Band[Hz]	Classic	Jazz	Dance	Average
0 ~ 20,000	1.00	1.00	1.00	1.00
10 ~ 20,000	0.71	0.67	0.67	0.68
20 ~ 20,000	0.54	0.50	0.46	0.50
0 ~ 20	0.83	0.79	0.50	0.71
0 ~ 100	0.92	0.83	0.75	0.83
100 ~ 200	0.50	0.75	0.50	0.58

on all bands.

4.6 Robustness for bandpass filtering attack

This scheme is also based on the direct sequence method of spectrum spreading. Thus, the watermarking signals are spread to a wide frequency band and may be omitted through the band-pass filter. Therefore, we checked how much the signals were omitted through the band-pass filter. The results are shown in **Table 6**, where the samples are transformed to the frequency components by DFT method and the watermarked signals are embedded under $j = 10$, $p = 25$ and $M = 16$, respectively. From the experimental results, we may estimate that the watermarked signals are concentrated to lower frequency bands. Thus, it is considered that we may not decode the watermark correctly when the lower frequency components are reduced through the filter. For this attack, it is necessary that the watermarks must be also embedded into higher frequency band.

4.7 Robustness for jitter attack

We check a watermark against the jitter attack⁷⁾. It is a popular attack technique that increases or decreases the frequency band of sound about 0.1% at random. First, the watermark is embedded into the sound data for some p under $j = 10$ and $M = 16$. Then, we execute the jitter attack on the watermarked sound, and get the deformed one. These results are shown in **Fig. 8**, which tells us that the robustness for this attack is rising trend subject to p . For example, we have the embedded watermark in more than 90% under $p \geq 20$.

From the above discussions, we conclude that our scheme attains a good sound in high quality of about 45dB and favorable robustness for popular attacks.

When the watermarked sound is attacked with pitch-shifting processing, however, we could not restore almost the watermarks if the shifted rate was greater than 1% in pitch. The reason is estimated as the synchronization between coder and decoder was broken.

Fig. 8 The effect of Jitter Attack(Embedding: 2.69[bit/s]).

Fig. 9 The effect of the MPEG processing(Embedding: 2.69[bit/s]).

4.8 Robustness for data compression

We check our scheme for MPEG Audio Layer III, i.e., so called MP3, data compression method under the compression rate of 1/10. The experimental results are shown in Fig. 9, where the watermark signals are embedded for p into each sample under $j = 10$ and $M = 16$. Then, the samples are compressed by the MP3 algorithm. Our results show higher detection rate over 90% in $p \geq 25$, and the detection rate is usually satisfactory for copy-protection with watermarking.

5. Conclusion

This paper proposes a watermarking scheme to music sound in high quality. It is robust for popular attacks because the watermarks are spread and concealed from illegal spectrum analysis. Moreover, the sub-banding scheme brings us some reduction of computation time for watermarking. It is important in our scheme that we have to select a set of control parameters $\{j, p, M\}$ in optimal under preliminary tests. Increasing the detection rate of watermark will be required into practical

use with the incorporation of error-correcting technique and adaptive control method. It will be available for detecting an illegal copy under the internet environment. Furthermore, we must study a protection scheme in this method against the band-pass filtering and resampling attacks, and it is important to try to estimate the watermarked sound quality by introducing the mean-opinion-score⁶⁾ as subjective evaluation.

[References]

- 1) K.Matsui, "Elementary digital watermarking - New protection technique for multimedia," Chap.7, Morikita Pub. Co., 1998(in japanese).
- 2) "The research report on the watermark technology," JEITA, 1999(in japanese).
- 3) Y.Miyaji, "The spread spectrum communication," Tokyo denki university press, 1994(in japanese).
- 4) K.R.Rao, J.J.Hwang, "Techniques and Standards for Image," Video and Audio Coding, Prentice Hall, 1996.
- 5) M.Iwakiri, K.Matsui, "Digital watermark scheme for high quality audio data by spectrum spreading and modified discrete cosine transform," Trans.IPS.Japan, 39, 9, pp. 2631-2637, 1998(in japanese).
- 6) K.Ozawa, "The high efficiency sound coding techniques for digital mobile-communication," Triceps Pub. Co., 1992(in japanese).
- 7) F.A.P.Petitcolas, R.J.Anderson, and M.G.Kuhn, "Attacks on copyright marking systems," Second workshop on information hiding, pp. 218-238, 1998.

Munetoshi Iwakiri received B.E and M.E. degrees from National Defense Academy, Yokosuka, Japan in 1993 and 1998. He is currently a research associate in department of computer science of National Defense Academy. His research interest is information security, watermarking, multimedia processing, and their applications. He is a member of IPSJ, IEICE and ASJ.

Kineo Matsui received M.E. degree in electrical engineering from Kyushu University, Fukuoka, Japan in 1965 and Dr. Eng. degree in electrical engineering from Keio University, Tokyo, Japan, in 1975. He is currently a professor in department of computer science of National Defense Academy. His research interest is information security, watermarking, multimedia processing, and their applications. He is a member of IPSJ, IEEE, and IEICE.