Multiple Embedding for Time-Domain Audio Watermarking Based on Low-Frequency Amplitude Modification

Harumi Murata, Akio Ogihara, Motoi Iwata and Akira Shiozaki
Department of Computer Sciences and Intelligent Systems,
Graduate School of Engineering, Osaka Prefecture University
1-1 Gakuen-cho, Naka-ku, Sakai, Osaka, 599-8531, Japan
E-mail: ogi@cs.osakafu-u.ac.jp

Abstract: The objective of this work is to increase the capacity of watermark information in the audio watermarking method based on low-frequency amplitude modification. We increase the capacity of watermark information by embedding multiple watermarks in individual data channels.

1. Introduction

Because digital audio data is not deteriorated by copying, its illegal reproduction can be easily made. Recently, copyright infringement becomes a social problem such as the illegal reproduction is distributed on the Internet. Hence, the audio watermarking methods, techniques to embed proprietary data into digital audio data, have attached attention as a prevention technique against the copyright infringement. The most representative of the audio watermarking methods add their watermarks in the time domain [1]-[3], do so in the subband domain [4] and do so in the Fourier domain [5]. W.N.Lie *et al*. have proposed the audio watermarking method based on amplitude modification [1]. This conventional method maintains high sound quality and is highly robust to pirate attacks, including MP3 compression, low-pass filtering, time scaling, digital-to-analog/analog-to-digital reacquisition and cropping. The watermark information is embedded into audio signals in the time domain. One-bit watermark information is embedded by modifying the differences of average-of-absolute-amplitude from three sections in a GOS(Group of Samples). In the conventional method, it is necessary to embed watermark information in short GOS length in order to increase the capacity of watermark information. However, if the watermark information is embedded in too short GOS length, the robustness to MP3 compression become low. And hence, the embedded watermark informatioin may not extracted correctly after MP3 compression.

In the conventional method, the capacity of watermark information is not enough. And hence, it is desirable that the capacity of watermark information is increased. In this paper, we aim to increase the capacity of watermark information in the audio watermarking method based on amplitude modification by multiple embedding. By multiple embedding, the watermark informatoin can be embed in different channels independently. We called this channel **data channel** in this paper. For example, we consider that both copyright information and image are embedded into audio signals. The former data has small capacity but it requires correct recovery rate. And the latter data has large capacity but it can be recognized even if it is not possible to extract it perfectly. In the conventional method, since it has only one data channel, GOS length must be set to be a certain value. If the watermark information is embedded in long GOS length that extracted copyright information correctly, the embedded watermark information can be extracted correctly but it may not be secured the capacity for image. On the other hand, if the watermark information is embedded in short GOS length that secured the capacity for image, the capacity is increased but it may not to extract copyright information correctly. Hence, in the conventional method, it is difficult to embed both copyright information and image.

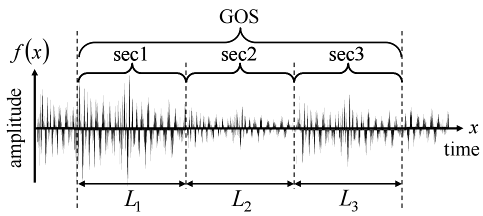

Figure 1. GOS and three sections in the conventional method proposed by W.N.Lie, et al.

In the proposed method, it has a lot of data channel, so it is possible to embed watermark information by selecting proper data channel according to the capacity of data or extraction accuracy. Hence, it can be secured the capacity for image and copyright information is embedded in high accuracy.

2. Conventional Audio Watermarking Method Based on Low-Frequency Amplitude Modification

In this section, 'embedding of watermark information' and 'extraction of watermark information' of the conventional audio watermarking method [1] are described.

2. 1 Embedding of Watermark Information

Original audio signal $f(x)$ is divided into consecutive L-length GOSs (Group of Samples) as shown in Fig.1. Each GOS contains three nonoverlapping sections (sec_1, sec_2 and sec_3), and the lengths of these three sections are L_1, L_2 and L_3, respectively. And hence $L = L_1 + L_2 + L_3$.

AOAAs (Average of Absolute Amplitudes) are calculated from the three sections according to the following equations.

$$E_{i1} = \frac{1}{L_1} \cdot \sum_{x=0}^{L_1-1} \left| f\left(L \cdot i + x\right) \right| \qquad (1)$$

$$E_{i2} = \frac{1}{L_2} \cdot \sum_{x=L_1}^{L_1+L_2-1} |f(L \cdot i + x)| \tag{2}$$

$$E_{i3} = \frac{1}{L_3} \cdot \sum_{x=L_1+L_2}^{L_1+L_2+L_3-1} |f(L \cdot i + x)| \tag{3}$$

where i is the GOS index, i = 0, 1, 2, …

E_{i1}, E_{i2} and E_{i3} are sorted in descending order, and they are renamed as $E_{\max}$, E_{mid} and $E_{\min}$, respectively. The differences of them are calculated by Eqs.(4) and (5) as shown in Fig.2.

$$A = E_{\max} - E_{\mathrm{mid}} \tag{4}$$

$$B = E_{\mathrm{mid}} - E_{\min} \tag{5}$$

The relationship $A < B$ means state "0", and $A \geq B$ means state "1". And hence one binary bit can be embedded in one GOS by modifying the original audio signal.

The embedding scheme is based on the following rules.

・To embed watermark bit "1"

If ($A - B \geq$ Thd1), then no operation is performed. Else increase $E_{\max}$ and decrease E_{mid} by the same amount so that the above condition is satisfied.

・To embed watermark bit "0"

If ($B - A \geq$ Thd1), then no operation is performed. Else increase E_{mid} and decrease $E_{\min}$ by the same amount so that the above condition is satisfied.

The threshold Thd1 is calculated by Eq.(6)

$$\mathrm{Thd1} = (E_{\max} + 2E_{\mathrm{mid}} + E_{\min}) \cdot d \tag{6}$$

where d is the parameter that adjusts a threshold.

2. 2 Extraction of Watermark Information

Assuming that the start point of data embedding has been recognized and the section lengths are known, every three consecutive sections of samples are grouped as a GOS and examined to extract the watermark. AOAAs are calculated for the ith GOS by Eqs.(7)-(9).

$$E'_{i1} = \frac{1}{L_1} \cdot \sum_{x=0}^{L_1-1} |f'(L \cdot i + x)| \tag{7}$$

$$E'_{i2} = \frac{1}{L_2} \cdot \sum_{x=L_1}^{L_1+L_2-1} |f'(L \cdot i + x)| \tag{8}$$

$$E'_{i3} = \frac{1}{L_3} \cdot \sum_{x=L_1+L_2}^{L_1+L_2+L_3-1} |f'(L \cdot i + x)| \tag{9}$$

where $f'(x)$ is the watermarked signal. E'_{i1}, E'_{i2} and E'_{i3} are ordered to yield $E'_{\max}$, E'_{mid} and $E'_{\min}$. The differences of them are calculated by Eqs.(10) and (11).

$$A' = E'_{\max} - E'_{\mathrm{mid}} \tag{10}$$

$$B' = E'_{\mathrm{mid}} - E'_{\min} \tag{11}$$

Comparing A' and B' yield the retrieved bit "1" if $A' \geq B'$ and "0" if $B' > A'$. This process is repeated for every GOS to extract the entire embedded bits.

3. Proposed Multiple Embedding Method

In order to increase the capacity of watermark information,

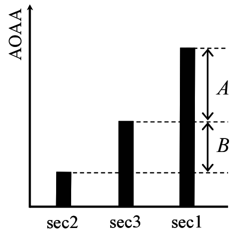

Figure 2. Differences of AOAAs calculated from three sections.

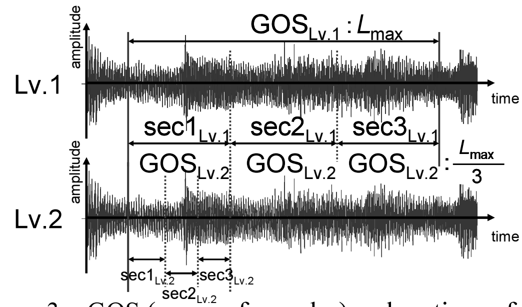

Figure 3. GOS (group of samples) and sections of each level in the proposed multipule embedding method.

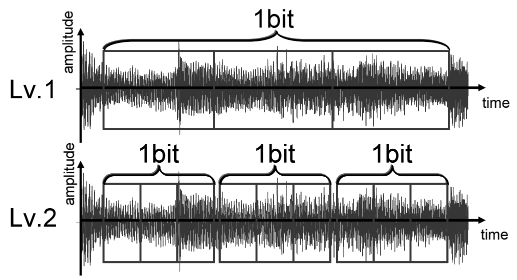

Figure 4. Capacity of watermark information in the proposed multipule embedding method.

we embed multiple watermarks in individual data channels. The proposed method has plural data channels: Level 1, Level 2,..., Level K. The larger the number of data channel level, the smaller the length of the data channel is. And hence, the larger the number of data channel level, the more the capacity of watermark information is. However, the larger the number of data channel level, the weaker the robustness to MP3 compression. In other words, there is the tradeoff between capacity and robustness.

First, 1st-watermark is embedded in GOSs in the same manner as the conventional method, and we call this embedding process "Level 1." And then, 2nd-watermark is embedded in sections of Level 1, and this embedding process is called "Level 2." In other words, we regard the section of Level 1 as the GOS of Level 2, as shown in Fig.3. According to the number of level, the above processing is repeated recursively. For the example, in the case of K = 2, the proposed method can embed four binary bits in the length of one GOS of Level 1 as shown in Fig.4. In the recursive embedding process, in order to extract correct watermark information from each data channel level, it is important that AOAA of Level N is unchanged by the embedding process of Level N+1.

The procedure of multiple embedding until Level K is shown below.

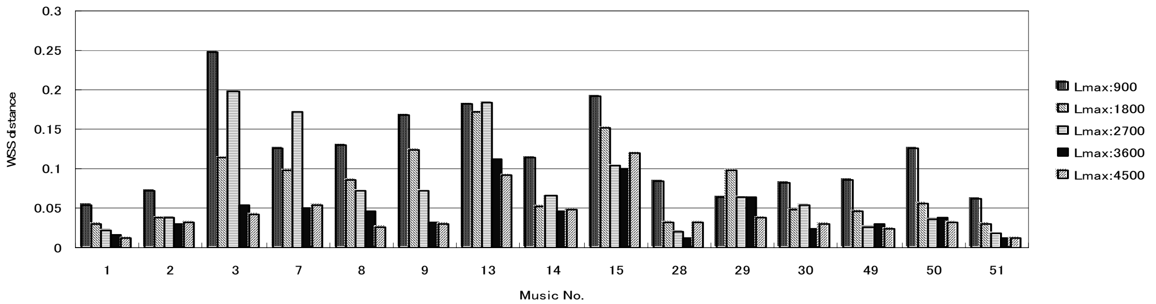

Figure 5. Objective evaluation of sound quality by WSS distance to each music data. (conventional method)

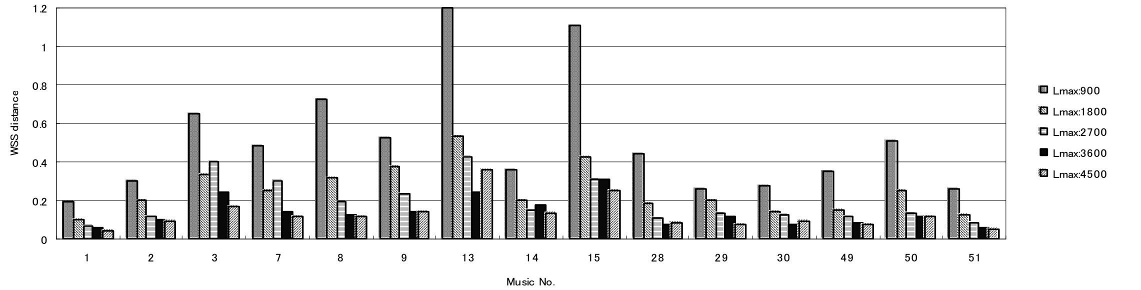

Figure 6. Objective evaluation of sound quality by WSS distance to each music data. (proposed method)

Step 1. $N \leftarrow 1$.
Step 2. GOS length of Level N is calculated by the following equation, where $L_{\max}$ is GOS length of Level 1.

$$L_{\mathrm{Lv}.N} = \frac{L_{\max}}{3^{N-1}} \tag{12}$$

Step 3. The watermark information is embedded in Level N data channel in the same manner as the conventional method.
Step 4. $N \leftarrow N+1$. If $N \leq K$ then go to Step.2, else terminate the multiple embedding process.

4. Experimental Results

Five types of music, including pops (Music No.1-3), rock (Music No.7-9), dance (Music No.13-15), jazz (Music No.28-30) , classic (Music No.49-51), each of 30 seconds duration, at a 44.1 kHz sampling rate, with stereo channels, were used for testing. These music data were selected from "RWC music database: music genre". We embedded the pseudo random sequence as watermark information in the left cannel of the music data. In the proposed method, maximum embedding level is Lv.2 (K = 2), and GOS length of Lv.1 is set to $L_{\max}$ = 900, 1800, 2700, 3600, 4500, respectively. In the conventional method, GOS length L is set to be equal to $L_{\max}$.

First, we made a comparison of sound quality between the conventional method and the proposed method after embedding the watermark information. We use WSS (Weighted Spectral Slope) distance [6] as objective evaluation of sound quality. WSS distance is close to a value evaluated subjectively, and low WSS distance means good sound quality. Table 1 shows the comparison result of the average of watermark recovery rate after MP3 compression between the conventional method and the proposed method. Detail experimental results to each music data are shown in Figs.5 and 6. In these results, the larger the value of $L_{\max}$ became, the smaller WSS distance became. Because, according to the increase of the value of $L_{\max}$, the

Table 1. Average of objective evaluation of sound quality by WSS distance.

$L_{\max}$	WSS distance (conv.)	WSS distance (prop.)
900	0.120	0.508
1800	0.0789	0.253
2700	0.0770	0.192
3600	0.0448	0.137
4500	0.0420	0.128

amount of change per one sample of music data becomes small. There is a little increase of WSS distance in the proposed method in comparison with the conventional method. However, it is not a problem on practical use and it seems to be under the limit of ordinary human's perception, since it is hard to perceive the noise in subjective evaluation.

Second experiment is about the robustness to MP3 compression. Lame3.93.1 [7] is used for MP3 compression, and bit rate is set to be 128 kbps. Table 2 shows the experimental result concerned with the average of watermark recovery rate after MP3 compression. Detail experimental results to each music data are shown in Figs.7 and 8. The watermark recovery rate R is defined as the following equation

$$R = \frac{\sum_{k=1}^{H} B\left(w_k = \tilde{w}_k\right)}{H} \tag{13}$$

where w_k is the watermark bit embedded in kth GOS, $\tilde{w}_k$ is the watermark bit extracted from kth GOS, H is the length of watermark information, and $B()$ is a Boolean function that returns B(True) = 1 and B(False) = 0. The watermark recovery rate of Level 1 in the proposed method is almost equal to that of the conventional method, and hence Level 1 is able to provide good robustness as the conventional method. In exchange for an increase of the capacity of watermark information, the watermark recovery rate of the Level 2 decreased slightly in comparison with Level 1 and the conventional method. We think that Level 2 is used as

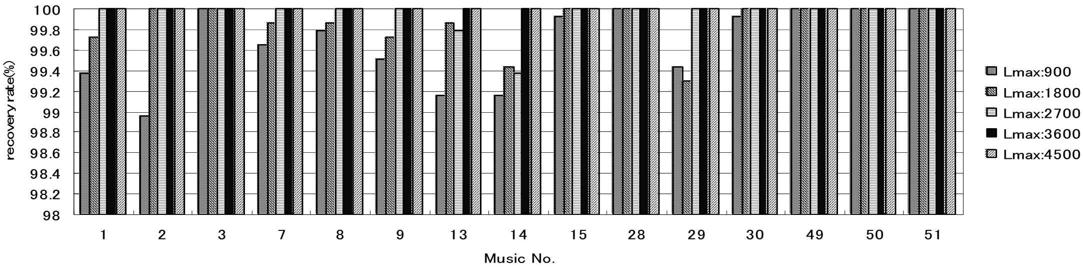

Figure 7. Watermark recovery rate after MP3 compression to each music data. (conventional method)

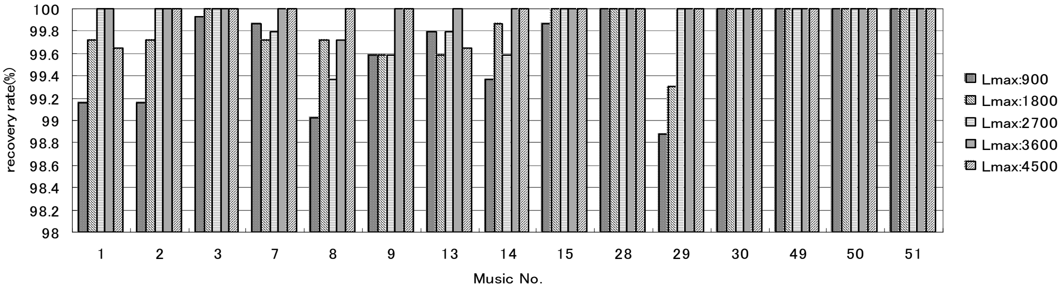

Figure 8. Watermark recovery rate after MP3 compression to each music data. (proposed method, Lv.1)

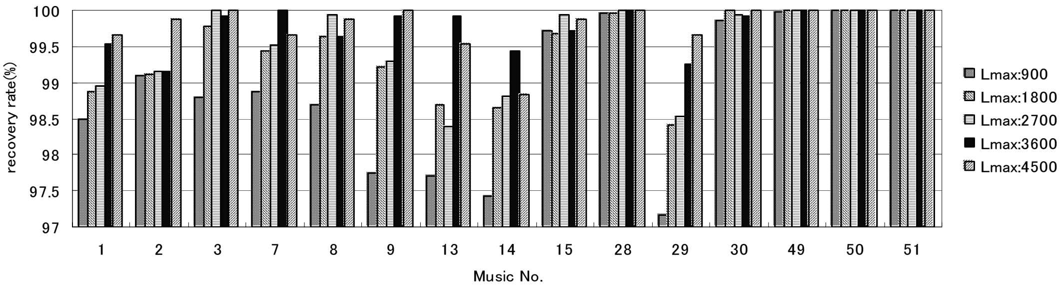

Figure 9. Watermark recovery rate after MP3 compression to each music data. (proposed method, Lv.2)

the data channel for the multimedia data (sound or image data) which can permit a little error.

5. Conclusion

This paper has proposed the method to increase the capacity of watermark information by embedding multiple watermarks in individual data channel. From the experimental results, it is confirmed that the capacity of watermark information increases without deterioration of sound quality in comparison with the conventional method. We have made a trial of our multiple embedding technique in the basic condition of $K = 2$ in this paper. As the future works, it is necessary to investigate the measures to control the deterioration of sound quality when the number of embedding level is extended.

Table 2. Average of watermark recovery rate after MP3 compression.

$L_{\max}$	recovery rate (conv.)	recovery rate (prop.)	
		Lv.1	Lv.2
900	99.66 %	99.64 %	98.90 %
1800	99.85 %	99.81 %	99.43 %
2700	99.94 %	99.87 %	99.50 %
3600	100 %	99.98 %	99.76 %
4500	100 %	99.95 %	99.80 %

References

[1] W.N.Lie, L.C.Chang, "Robust and high-quality time-domain audio watermarking based on low-frequency amplitude modification," *IEEE Trans. Multimedia*, vol. 8, no. 1, pp.46-59, 2006.

[2] P.Bassia, I.Pitas, and N.Nikolaidis, "Robust audio watermarking in the time domain," *IEEE Trans. Multimedia*, Vol.3, no.2, pp.232-241, Jun.2001.

[3] S.Shin, O.Kim, J.Kim, and J.Choil, "A robust audio watermarking algorithm using pitch scaling," *Proc. IEEE 14th Int. Conf. Digital Signal Processing*, 2002, pp.701-704.

[4] X.Li and H.H.Yu, "Transparent and robust audio data hiding in subband domain," *Proc. IEEE Int. Conf. Information Technology: Coding and Computing*, 2000, pp.74-79.

[5] S.-S.Kuo, J.D.Johnston, W.Turin, and S.R. Quackenbush, "Covert audio watermarking using perceptually tuned signal independent multi-band phase modulation," *Proc. IEEE. Int. Conf. Acoustics, Speech, and Signal Processing*, vol.II, 2002, pp.1753-1756.

[6] J.R.Deller, J.G.Proakis, and J.H.L.Hansen, *Discrete-Time Processing of Speech Signals*, Macmillan, 1993.

[7] http://lame.sourceforge.net/