A Method of Increasing Watermark Information Capacity with Five Sections for Audio Watermarking Method Based on Amplitude Modification

Akio Ogihara, Naoya Inoue, 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, inoue_n@ch.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 amplitude modification. We increase the capacity of watermark information by dividing a GOS (Group of Samples) into five sections.

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 dagital 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. The principle of amplitude modification is embployed to scale amplitudes in selected sections of samples so that time-domain waveform envelope can be almost preserved.

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. In the conventional method, one-bit watermark information is embedded into a GOS with three sections. And hence, the capacity of watermark informarion per section is 1/3 bits. If two-bit watermark information is embedded into a GOS with five sections, it is 2/5 bits and more than the conventional method. Hence, We propose a method of dividing a GOS into five sections in order to embed two watermark bits into a GOS.

The capacity of watermark informarion per section in our proposed method is more than the capacity of watermark information per section in the conventional method. This means the proposed method can make the section longer than conventional method. Our test shows that using a long
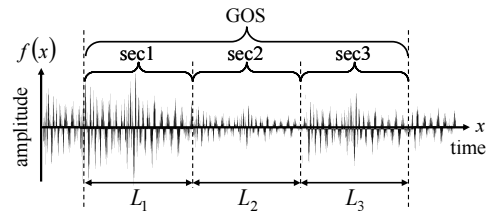

Figure 1. GOS and three sections in conventional method.

section almost preserves the audio envelope and can achieve good sound quality. And hence, the proposed method can achieve an improvement in sound quality.

2. Conventional Audio Watermarking Method Based on 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 non-overlapping sections (sec1, sec2 and sec3), and the lengths of these three sections are L_1, L_2 and L_3, respectively. And hence $L = L_1 + L_2 + L_3$. The section-length parameters (L_1, L_2, L_3) can be determined by using the secret key in order to prevent statistically detecting the watermark information. AOAAs (Average of Absolute Amplitudes) are calculated from the three sections according to the following equations

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

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

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

where i is the GOS index, $i = 0, 1, 2, \ldots$

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.

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

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

The relationship $A \geq B$ means state "1", and $A < B$ means state "0". 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)

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

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)

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

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

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

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).

$$A' = E'_{max} - E'_{mid} \qquad (10)$$

$$B' = E'_{mid} - E'_{min} \qquad (11)$$

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 Method of Increasing Watermark Information Capacity

In this section, 'embedding of watermark information' and 'extraction of watermark information' of the proposed method are described.

3. 1 Embedding of Watermark Information

In order to increase the capacity of watermark information, we divide a GOS into five sections (sec1, sec2, sec3, sec4 and sec5) as shown in Fig.3, and the lengths of these sections are L_1, L_2, L_3, L_4 and L_5, respectively. AOAAs are calculated from the five sections according to the following equations.
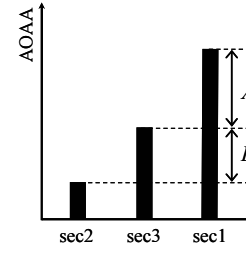

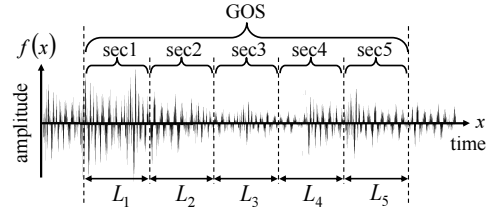Figure 2. Differences of AOAA in conventional method.

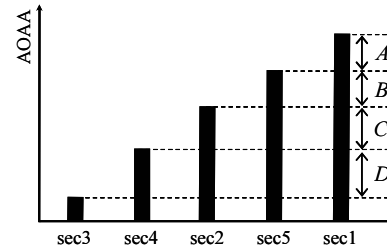Figure 3. GOS and three sections in proposed method.

Figure 4. Differences of AOAA in proposed method.

$$E_{i1} = \frac{1}{L_1} \cdot \sum_{x=0}^{L_1-1} \left| f(L \cdot i + x) \right| \qquad (12)$$

$$E_{i2} = \frac{1}{L_2} \cdot \sum_{x=L_1}^{L_1+L_2-1} \left| f(L \cdot i + x) \right| \qquad (13)$$

$$E_{i3} = \frac{1}{L_3} \cdot \sum_{x=L_1+L_2}^{L_1+L_2+L_3-1} \left| f(L \cdot i + x) \right| \qquad (14)$$

$$E_{i4} = \frac{1}{L_4} \cdot \sum_{x=L_1+L_2+L_3}^{L_1+L_2+L_3+L_4-1} \left| f(L \cdot i + x) \right| \qquad (15)$$

$$E_{i5} = \frac{1}{L_5} \cdot \sum_{x=L-L_5}^{L-1} \left| f(L \cdot i + x) \right| \qquad (16)$$

E_{i1}, E_{i2}, E_{i3}, E_{i4} and E_{i5} are sorted in ascending order, and they are renamed as E_{1st}, E_{2nd}, E_{3rd}, E_{4th}, and E_{5th}, respectively. The differences of them are calculated by Eqs.(17)-(20) as shown in Fig.4.

$$A = E_{5th} - E_{4th} \qquad (17)$$

$$B = E_{4th} - E_{3rd} \qquad (18)$$

$$C = E_{3rd} - E_{2nd} \qquad (19)$$

$$D = E_{2nd} - E_{1st} \qquad (20)$$

In the proposed method, one binary bit is embedded in the relationship between A and B, and another binary bit is embedded in the relationship between C and D. As the result, the proposed method can embed two watermark bits in one GOS.

The relationship $A \geq B$ means state "1", and $A < B$ means state "0". In addition, the relationship $C \geq D$ means state "1", and $C < D$ means state "0". The embedding scheme is based on the following rules.

・ To embed watermark bit "1" in the relationship between A and B

If ($A - B \geq$ Thd1), then no operation is performed. Else increase E_{5th} and decrease E_{4th} by the same amount so that the above condition is satisfied.

・ To embed watermark bit "0" in the relationship between A and B

If ($B - A \geq$ Thd1), then no operation is performed. Else increase E_{4th} and decrease E_{5th} by the same amount so that the above condition is satisfied.

・ To embed watermark bit "1" in the relationship between C and D

If ($C - D \geq$ Thd2), then no operation is performed. Else increase E_{1st} and decrease E_{2nd} by the same amount so that the above condition is satisfied.

・ To embed watermark bit "0" in the relationship between C and D

If ($D - C \geq$ Thd2), then no operation is performed. Else increase E_{2nd} and decrease E_{1st} by the same amount so that the above condition is satisfied.

Thd1 and Thd2 are calculated by Eqs.(21) and (22).

$$\text{Thd1} = \left(E_{3rd} + 2E_{4th} + E_{5th}\right) \cdot d \tag{21}$$

$$\text{Thd2} = \left(E_{1st} + 2E_{2nd} + E_{3rd}\right) \cdot d \tag{22}$$

where d is the parameter that adjusts a threshold.

Fig.5 shows the differences of AOAA in case a watermark bit "0" is embedded in the relationship between A and B and a watermark bit "1" is embedded in the relationship between C and D.

3. 2 Extraction of Watermark Information

Every five consecutive sections of samples are grouped as a GOS and AOAAs are calculated for the ith GOS by Eqs.(23)-(27).

$$E'_{i1} = \frac{1}{L_1} \cdot \sum_{x=0}^{L_1 - 1} \left|f'\left(L \cdot i + x\right)\right| \tag{23}$$

$$E'_{i2} = \frac{1}{L_2} \cdot \sum_{x=L_1}^{L_1 + L_2 - 1} \left|f'\left(L \cdot i + x\right)\right| \tag{24}$$

$$E'_{i3} = \frac{1}{L_3} \cdot \sum_{x=L_1 + L_2}^{L_1 + L_2 + L_3 - 1} \left|f'\left(L \cdot i + x\right)\right| \tag{25}$$

$$E'_{i4} = \frac{1}{L_4} \cdot \sum_{x=L_1 + L_2 + L_3}^{L_1 + L_2 + L_3 + L_4 - 1} \left|f'\left(L \cdot i + x\right)\right| \tag{26}$$

$$E'_{i5} = \frac{1}{L_5} \cdot \sum_{x=L - L_5}^{L - 1} \left|f'\left(L \cdot i + x\right)\right| \tag{27}$$
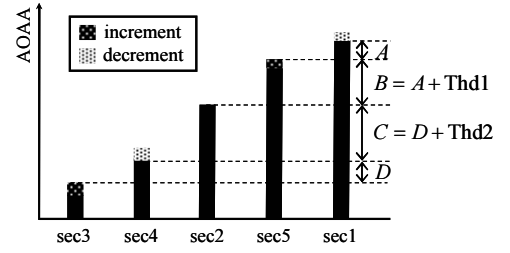

Figure 5. Differences of AOAA after embedding watermark bits "01" in proposed method.

E'_{i1}, E'_{i2}, E'_{i3}, E'_{i4} and E'_{i5} are ordered to yield E'_{1st}, E'_{2nd}, E'_{3rd}, E'_{4th} and E'_{5th}. The differences of them are calculated by Eqs.(28)-(31).

$$A' = E'_{5th} - E'_{4th} \tag{28}$$

$$B' = E'_{4th} - E'_{3rd} \tag{29}$$

$$C' = E'_{3rd} - E'_{2nd} \tag{30}$$

$$D' = E'_{2nd} - E'_{1st} \tag{31}$$

Comparing A' and B' yield the retrieved bit "1" if $A' \geq B'$ and "0" if $B' > A'$. In addition, Comparing C' and D' yield the retrieved bit "1" if $C' \geq D'$ and "0" if $D' > C'$. And hence, as the result, two watermark bits are extracted from one GOS.

4. Experimental Result

In the experiment, 15 music data selected from "RWC music database: music genre" were used. These music data were five types of music, including Pops, Rock, Dance, Jazz and Classic, each of 10s duration, at a 44.1kHz sampling rate, with stereo channels.

In case the section length is set to be 340, the watermark capacity of the conventional method is 43 bps. In contrast, the watermark capacity of the proposed method is 51 bps. And hence, the proposed method increases the watermark capacity by 8 bps.

The experiment for comparing the conventional method and the proposed method was performed on the condition that both watermark capacities were 43 bps. And hence, the section length of the proposed method was set to be 408. A fixed pattern "1010…" was selected in the conventional method and a pattern "11001100…" was selected in the proposed method, as the watermark information.

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 Slop) 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. Secondly, we made a comparison of the watermark robustness between the conventional method and the proposed method. MP3 compression was used as an attack to demonstrate the robustness of embedded

Table 1. Experimental results of WSS distance and watermark recovery rate after MP3 compression.

Music	WSS distance		Recovery rate (%)	
No.	conventional	proposed	conventional	proposed
No.1	0.0278	0.0439	97.5	98.8
No.2	0.0434	0.0322	89.8	92.4
No.3	0.1766	0.1073	91.7	94.7
No.7	0.0611	0.0313	89.6	91.9
No.8	0.0689	0.0369	94.2	91.5
No.9	0.0608	0.0454	86.1	91.4
No.13	0.2459	0.052	89.4	92.4
No.14	0.051	0.0463	86.4	83.5
No.15	0.1638	0.048	99.8	99.5
No.28	0.0645	0.0411	98.2	97.7
No.29	0.0399	0.0634	92.8	94
No.30	0.0409	0.0314	95.2	95.4
No.49	0.0596	0.0492	97.9	99.1
No.50	0.0746	0.0317	98.2	99.8
No.51	0.0277	0.0253	96.3	96.3
average	0.0807	0.0453	93.5	94.6

watermark information. MP3 compression was performed using Lame3.96.1 [7] and its compression bit rate was chosen 128 kbps. We calculated the watermark recovery rate R after MP3 compression. It is defined as the following equation

$$R = \frac{\sum_{k=1}^{H} B(w_k = \widetilde{w}_k)}{H} \qquad (32)$$

where w_k is the watermark bit embedded in kth GOS, $\widetilde{w}_k$ is the watermark bit extracted from kth GOS, H is the length of watermark information, and $B()$ is a Boolean function that returns $B(\text{True}) = 1$ and $B(\text{False}) = 0$.

Table 1 shows experimental results of WSS distance and the watermark recovery rate after MP3 compression. In Music No.1 and No.29, the conventional method indicated lower WSS distance than the proposed method. However, these two music data were hardly degraded and they were acceptable in subjective evaluation of sound quality. In all other music, the proposed method indicated lower WSS distance than the conventional method. Especially in Music No.3, No.13 and No.15, there were great differences of WSS distance between conventional method and proposed method. Additionally, in subjective evaluation of Music No.3 and No.9, the proposed method reduced perceptible noises. This agreed with the results of objective evaluation by WSS distance. The described above results mean the proposed method attained good results in parts of music, and we confirmed that the proposed method improved the sound quality. The average of the watermark recovery rate increased slightly. This means the proposed method has a high robustness to MP3 compression as the conventional method.

5. Conclusion

We have proposed a method of increasing the capacity of watermark information by dividing a GOS into five sections. We modify AOAAs calculated from the five sections in order to embed two watermark bits into a GOS. The proposed method is able to increase the capacity of watermark information, and to improve the sound quality, in the audio watermarking method based on amplitude modification.

References

[1] W.N.Lie, L.C.Chang, "Robust and high-quality time-domain audio watermarking based on low-frequency amplitude modification," *IEEE Trans. on 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. on 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.2, 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/