Sound Quality Improvement by Embedding Position Grouping for Audio Watermarking Method Using Smearing Transformation

Akio Ogihara, Takashi Iwamoto, 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, iwamoto_t@ch.cs.osakafu-u.ac.jp

Abstract: The objective of this work is to improve the sound quality of the audio watermarking method using smearing transformation. In this paper, we propose "Embedding Position Grouping" to improve the sound quality of watermarked audio.

1. Introduction

Recently, with the rise of Internet, it has increased that the number of opportunities using many types of digital contents, like video, music and graphics. Because copying digital contents is too easy without deteriorating them, its illegal reproduction can be easily made. So, copyright infringement becomes a social problem such as the illegal reproduction is distributed on the Internet. Hence, digital watermark is proposed in order to protect copyright of digital contents.

In the case of audio, when digital audio is distributed in Internet, it is almost compressed by MP3 [1],[2]. Hence audio watermark systems must be tolerant to MP3 compression. The audio watermarking method using smearing transformation which is one of the audio watermark systems has been proposed as a prevention technique against the copyright infringement [1]. However, this method tends to generate noise, because the strength of embedding depends on the embedding position's amplitude of original audio signal. And if we select the sample which have small amplitude as the embedding position, the strength of embedding is large and noise is generated too. Hence, we aim to reduce the strength of embedding. As a result, we aim to improve the sound quality of watermarked audio in this paper.

In the conventional method, one impulse signal is embedded as watermark signal in one block which is group of samples of original audio signal as shown in Fig.1. When we embed the value of watermark information '1', we embed the impulse signal at the 1st sample of the block. In this case, if the amplitude of the 1st sample is small, we must embed a large impulse as a watermark signal. To avoid this problem, we divide block into some groups, and we embed any sample of the 1st group when we embed the value of watermark information '1'. So we can select the sample which has large absolute amplitude, and we can avoid the surplus strength of embedding. Hence, we can improve the sound quality of the conventional method.

We explain the smearing transformation and the desmearing transformation in chapter 2 and the embedding process of the conventional method in chapter 3. In chapter 4, we introduce the embedding process of the proposed method. In chapter 5, we describe experimental result. Finally, we present a brief conclusion in chapter 6.

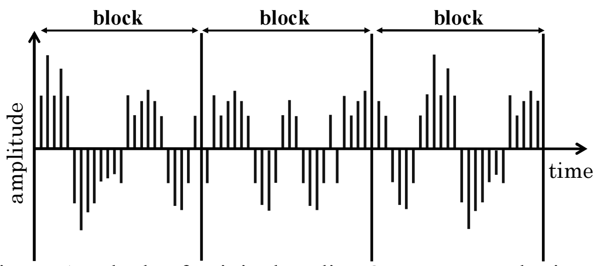

Figure 1. Block of original audio. One watermark signal is embedded in one block.

2. Smearing Transformation and Desmearing Transformation

In this section, smearing transformation and desmearing transformation are described.

Smearing transformation is a transformation which spreads each sample into the whole of data sequence. In audio signal, it is equal to changing its Fourier phase spectrum without changing Fourier amplitude spectrum. Smearing transformation and desmearing transformation are described by the following equations.

$$\tilde{X} = S \cdot X \qquad \text{(Smearing transformation)} \qquad (1)$$

$$X = D \cdot \tilde{X} \qquad \text{(Desmearing transformation)} \qquad (2)$$

where $\tilde{X}$ is a spread sequence of audio obtained by smearing transformation, X is a sequence of original audio, S is a matrix of smearing transformation, and D is a matrix of desmearing transformation.

Now matrix of smearing transformation S is defined by the following equation.

$$S = F^{-1} \begin{pmatrix} e^{-j\theta_0} & 0 & \ldots & 0 \\ 0 & e^{-j\theta_1} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & e^{-j\theta_{N-1}} \end{pmatrix} F \qquad (3)$$

where F is the matrix of Fourier transformation. θ_k is phase characteristic parameter, and is defined by the following equation.

$$\theta_k = \theta\left(\frac{2\pi}{T} \cdot \frac{k}{N}\right) \qquad (k = 0,1,2,\cdots,N-1) \qquad (4)$$

where $\theta(\omega)$ is periodical function with $2\pi / T$ period, N is the length of sequence, T is obtained from period of θ_k, and it has $\theta'(\omega)$ in Eq.(5) as a fundamental function.

$$\theta'(\omega) = \int_0^{\omega} \tau(\omega')\,d\omega' \quad \left(\frac{-\pi}{T} < \omega < \frac{\pi}{T}\right) \tag{5}$$

where $\tau(\omega)$ is delay function. Equation (6) is used as the delay function in the conventional method.

$$\tau(\omega) = \alpha \frac{N}{\pi} T^2 |\omega| \tag{6}$$

where α is the spreading coefficient which is the parameter to decide the ability of spreading.

On the other hand, the matrix of desmearing transformation is obtained by calculating inverse matrix of smearing transformation as the following equation.

$$D = S^{-1} \tag{7}$$

3. Conventional Audio Watermarking Method Using Smearing Transformation

In this section, the outline of the conventional audio watermarking method using smearing transformation [1] is described. Process of the conventional method is shown in Fig.2

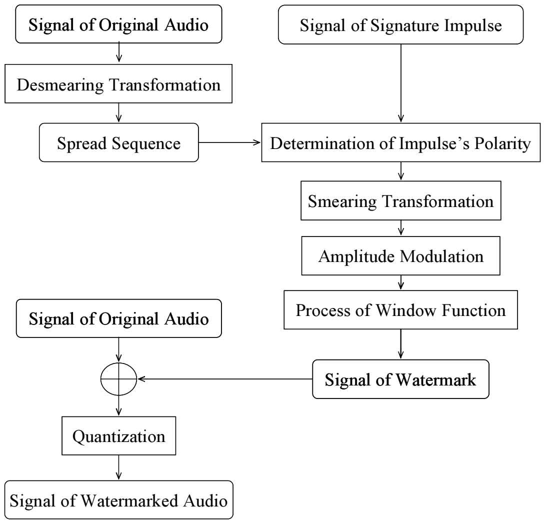

Figure 2. Embedding process of conventional method.

3. 1 Embedding process

In the conventional method, original audio signal is divided into consecutive N-length blocks, and each block contains some M-length modulation-blocks. Impulse signal is embedded in each block as shown in Fig.3. Impulse signal in the first block is called "time correct impulse signal", and impulse signals in the others are called "signature impulse signal". The distance between "time correct impulse signal" and "signature impulse signal" is used as watermark information. Time correct impulse is used for revising the time-shift processing like as MP3. In this section, we explain embedding process of the conventional method.

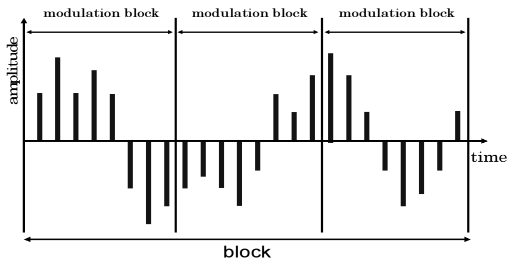

Figure 3. Modulation block. Block is divided into the some modulation blocks. In this figure, block is divided into 3 blocks.

3. 1. 1 Determination of impulse's polarity and spreading impulse

Desmearing transformation applies to a block of original audio signal G, and the spread sequence $G' = (g'_0, g'_1, \cdots, g'_{N-1})$ is obtained, and we calculate the average of spread sequence A. When the value of sample at embedding position in spread sequence is larger than A, we make positive impulse. Otherwise we make negative impulse. Here we define s as embedding position, and define $W^s = (w^s(0), w^s(1), \cdots, w^s(N-1))$ as impulse sequence. We express the determination of impulse polarity as the following equation.

$$w^s(k) = \begin{cases} h & (k = s \text{ and } g'_s > A) \\ -h & (k = s \text{ and } g'_s \le A) \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

Finally, smearing transformation applies to W^s, and spread signature impulse sequence $\tilde{W}^s$ is obtained.

3. 1. 2 Amplitude modulation

Spread impulse sequence $\tilde{W}^s$ is spread uniformly all over the sequence. Hence, when the volume of original audio is small partly, the noise is generated by watermarking. So according to "the modulation block's power of the original audio signal" and "the average of block's power", amplitude modulation applies to the spread signature impulse sequence.

Original audio signal G and spread signature impulse sequence $\tilde{W}^s$ are divided into l parts of M-length modulation blocks, $G_i = (g_i(0), g_i(1), \cdots, g_i(m-1))$ $\tilde{W}^s = (\tilde{w}_i^s(0), \tilde{w}_i^s(1), \cdots, \tilde{w}_i^s(m-1))$ $(0 \le i \le l-1)$ respectively. We define E as average of energy of a block. We define $E' = (e'_0, e'_1, \cdots, e'_{l-1})$ as the average energy sequence of modulation blocks which is obtained by the following equation.

$$e'_i = \frac{1}{m} \sum_{p=0}^{m-1} \left(g_i(p)\right)^2 \tag{9}$$

Hence, spread signature impulse sequence after amplitude modulation is obtained by the following equation.

$$\tilde{w}'^s_i(p) = \begin{cases} \tilde{w}_i^s(p) \times a & (e'_i \le E) \\ \tilde{w}_i^s(p) \times b & (e'_i > E) \end{cases} \quad (i = 0, 1, \cdots, l-1) \tag{10}$$

where we define $0 < a < 1$ and $1 \le b$.

3. 1. 3 Window function

After the amplitude modulation, the difference of amplitude between the adjacent modulation blocks which are an amplified modulation block and an attenuated modulation block is generated. This difference generates the noise. Hence, we eliminate this difference by using window function. In the conventional method, the following equation is used for window function.

$$M(p)=1-\left(1-\frac{a}{b}\right)\cos\left(\frac{2\pi p}{m-1}\right) \tag{11}$$

Window function applies to amplified modulation block. And the modulation block after applying the window function is obtained by the following equation.

$$y_i^p=\begin{cases}\tilde{x}_i^{p\,\prime}\times 1 & (e_i'\le E)\\ \tilde{x}_i^{p\,\prime}\times M(p) & (e_i'>E)\end{cases} \tag{12}$$

Finally, adjacent modulation blocks are linked smoothly, and the final watermark signal sequence is obtained.

3. 1. 4 Generating watermarked audio

Watermarked audio data $U'=(u_0',u_1',\cdots,u_{N-1}')$ is obtained by adding watermark signal and quantization. In the conventional method, we use the following equation for quantization.

$$u_k'=\begin{cases}-32768 & \left(\left\lfloor u_k+\frac{1}{2}\right\rfloor<-32768\right)\\ \left\lfloor u_k+\frac{1}{2}\right\rfloor & \left(-32768\le\left\lfloor u_k+\frac{1}{2}\right\rfloor\le 32767\right)\\ 32767 & \left(\left\lfloor u_k+\frac{1}{2}\right\rfloor>32767\right)\end{cases} \tag{13}$$

3. 2 Extraction process

First, watermarked audio is divided into blocks. Second, desmearing transformation applies to these blocks, and spread sequences are obtained. Third, Laplacian filter applies to spread sequences in order to emphasize the edge, and the edge emphasized spread sequence $\tilde{X}'$ is obtained. And the average A' of $\tilde{X}'$ is calculated. Finally, the embedded position s' is obtained by the following equation.

$$s'=\left\{k\mid\max_{0\le k\le N-1}\left|\tilde{x}_k'-A'\right|\right\} \tag{14}$$

4. Improvement of Sound Quality by Embedding Position Grouping

In the conventional method mentioned in Sect.3, sound quality often declines in case that the embedding position's amplitude of spread sequence of original audio signal is small. In this case, large impulse intensity is needed in order to make the embedding position have the largest absolute amplitude in the block because extracting the impulses correctly is needed in extracting phase. As above, sound quality depends on the embedding position's amplitude of spread sequence of original audio signal. Hence, we proposed embedding the impulse not into one decided sample but into the arbitrary sample of one decided group. Figure 4 shows the process of proposed method. The processes written by bold-face in Fig.4 are additional processes in proposed method. The embedding watermark is performed by the following process.

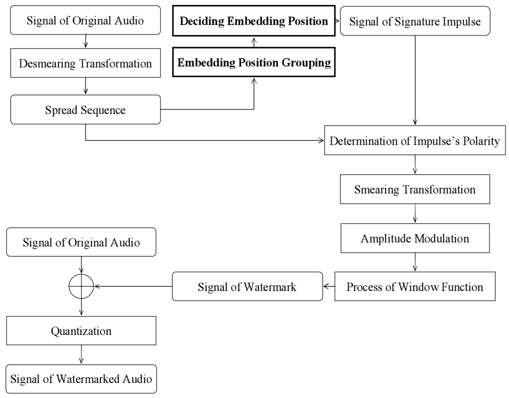

Figure 4. Embedding process of proposed method.

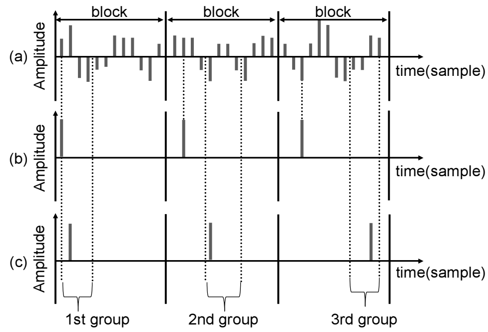

Figure 5. Embedding position of impulse and embedding position grouping (4 samples in a group): (a) spread sequence obtained from original audio signal by desmearing transformation, (b) embedding position in conventional method, (c) embedding position in proposed method.

[step1] Spread sequence is divided into some L-length groups. When the value of watermark information is 1, we can select the embedding position from L samples in 1st-group (i.e. 1st-4th sample shown in Fig.5).

[step2] Determining the sample as embedding position, we select the sample which has the largest absolute amplitude in the block as the embedding position. Because, impulse intensity is made smaller when embedding position's amplitude is large. Selected sample is obtained by the following equation.

$$q=\left\{n\,\middle|\,\max_{0\le n\le \mathrm{gs}-1}\left(\mathrm{abs}(\mathrm{group}(n))\right)\right\} \tag{15}$$

where group(n) is nth samples of the group, smp is the number of samples in the group, gs is the number of samples, and q is embedding position in the group.

Embedding same watermark information, the difference of embedding position between the conventional method and the proposed method is shown in Fig.5 When the value of watermark information is 1 in the 1st block, the 1st sample of the block is selected in the conventional method, but the 2nd sample in the 1st group of the block in the proposed method. The difference of impulse intensity between the conventional method and the proposed method is shown in Fig.6. We can confirm that the impulse intensity is attenuated by the proposed method.

5. Experiment

5. 1 Experimental condition

In this experiment, 15 music data selected from "RWC music database: music genre" were used. These music data were five types of music data, including Pops, Rock, Dance Jazz, and Classical music, each of 30 sec duration, at a 44.1 kHz sampling rate, with stereo channel. We defined the number of sample in block as 65536 samples. The other parameters are defined in order to extract all impulses correctly before MP3 compression.

Segmental SNR was used as objective evaluation of sound quality. It is obtained by the following equations.

$$\mathrm{SNRseg}\ (\mathrm{dB}) = \frac{1}{\mathrm{seg}} \sum_{n_1=0}^{\mathrm{seg}-1} \mathrm{SNR}_{n_1} \tag{16}$$

$$\mathrm{SNR}\ (\mathrm{dB}) = 20 \log_{10} \frac{\sqrt{\sum_{n_2=0}^{\mathrm{smp}-1} \left(\mathrm{org}_{n_2}\right)^2}}{\sqrt{\sum_{n_2=0}^{\mathrm{smp}-1} \left(\mathrm{org}_{n_2} - \mathrm{emb}_{n_2}\right)^2}} \tag{17}$$

where seg is the number of segments, smp is the number of samples in the segment, org is original audio signal, and emb is audio signal after embedding.

5. 2 Experimental result

Table 1 shows experimental result of segmental SNR and the intensity of signature impulse and time correct impulse under the condition of watermark recovery rate=100% before MP3 compression. In Music no.13 and no.14, time correct impulse intensity and signature impulse intensity are decreased greatly by the proposed method. However, the impulses' intensity in these two music is much larger than impulse intensity in the others. The reason of this problem is that impulse intensity depends to the volume of music. So we have to propose the new method in which impulse intensity is independent to the volume of original audio. In the other music, impulses' intensity is decreased by the proposed method. And, the proposed method increases segmental SNR by 2.1dB. As a consequence, we confirmed that the proposed method is effective.

6. Conclusion

We are attention to the impulse intensity which depends to embedding position in the conventional method. We have proposed *embedding position grouping* to improve the sound quality of audio watermarking method using smearing transformation. This proposed method is selecting an arbitrary sample from some samples in group, in order to eliminate the impulse intensity which depends to embedding position. We can attenuate the impulse intensity of all music. As a consequence, it has been confirmed from this result that the proposed method has increased segmental SNR by 2.1dB. We think that the research of tolerance to MP3 compression and eliminating the dependence on volume of original audio are subjects of future investigation.

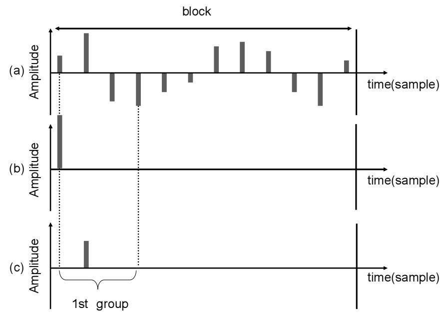

Figure 6. Difference of impulse intensity: (a) spread sequence obtained from original audio signal by desmearing transformation, (b) amplitude of impulse in conventional method, (c) amplitude of impulse in proposed method.

Table 1. Time correct impulse intensity and segmental SNR of conventional method and proposed method under the condition of watermark recovery rate=100% before MP3 compression.

music no.	conventional method			proposed method		
	time correct impulse intensity	signature impulse intensity	SNRseg (dB)	time correct impulse intensity	signature impulse intensity	SNRseg (dB)
1	1400	2250	43.3	1200	1750	45.5
2	2050	5050	41.8	500	4300	43.7
3	1450	3400	50.7	900	2800	52.5
7	3250	4800	45.5	2700	3600	48.0
8	1000	5850	38.5	250	4800	38.8
9	4200	6350	44.2	2450	5150	46.2
13	7900	11500	40.2	7300	8550	42.7
14	6350	10050	40.8	5750	6800	44.1
15	400	3050	49.8	250	2400	51.9
28	150	1000	59.0	150	600	63.2
29	1750	2650	46.6	700	2050	49.1
30	800	2050	54.3	550	1400	57.6
49	250	500	59.3	150	500	59.5
50	350	250	58.9	250	250	59.0
51	400	500	62.5	350	400	64.4
average	2113	3950	49.0	1563	3023	51.1

References

[1] M.Wakaki, T.Okahisa, K.Ohue, "Digital audio watermarking method with MP3 tolerance using smearing-desmearing transformation," *Technical report of IEICE*, CAS2001-103, 2002.

[2] Kirovski, D, Malvar, H.S,"Spread-spectrum watermarking of audio signals," IEEE Transactions on Signal Processing vol.51, no.4, pp.1020-1033, April 2003.