Sound Quality Improvement by Eliminating High Frequency Component from Signature Impulse for Audio Watermarking Method Using Smearing Transformation

Akio Ogihara, Ryota Sawada, 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, sawada_r@ch.cs.osakafu-u.ac.jp

Abstract: In this paper, we aim to improve the sound quality of the audio watermarking method using smearing transformation, which is proposed by M.Wakai et al. Since noise is caused by the signature impulse, we aim to reduce the power of the signature impulse. In our method, the signature impulse is filtered by low-pass filter, and then the filtered impulse is embedded.

1. Introduction

Recently, with the rise of Internet, it has increased that the number of opportunities using many types of digital content, 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, it is compressed by MP3. Hence audio watermark systems must be tolerant to MP3 compression. The audio watermarking method using smearing transformation that is one of the audio watermark systems tolerant to MP3 compression has been proposed as a prevention technique against the copyright infringement [1]. However, if we embed the watermark with enough tolerance for MP3 compression by using the conventional method, embedding strength becomes larger and the amount of the change of the original audio becomes larger. So there are some music data from which can perceive the noise. Then in this paper, we aim to improve sound quality of reference [1].

In MP3 compression process of music data, MP3 tends to eliminate the high frequency component which is hard to perceive by human's auditory sense. Therefore, when watermarked music data is compressed by MP3, the high frequency component of the embedded signature impulse is eliminated. In this paper, we eliminate the high frequency component from the signature impulse by low-pass filter in advance, and then the band-limited signature impulse is used as the embedded watermark signal. Because the band-limited signature impulse does not contain high frequency component, its power is less than the power of the signature impulse which is used in the conventional method. The less the power of watermark signal is, the less the change from the original music data is. As the result, by using the proposed method, it is expected that the deterioration of sound quality caused by watermarking is reduced without the decline of watermark recovery rate from the music data

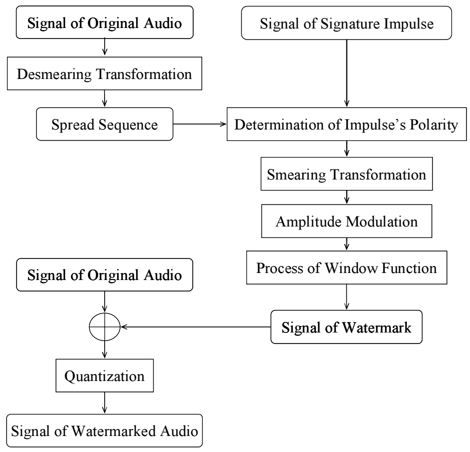

Figure 1. Embedding process of conventional method.

compressed by MP3. In this paper, we attempt two control methods for apposite cutoff frequency. The first method is the identical cutoff frequency is used to all blocks in watermark embedding process. The second method is the different cutoff frequency is used to each block in watermark embedding process.

In chapter 2, we introduce the embedding process of the conventional method. In chapter 3, we explain the optimizing of signature impulse in the proposed method. In chapter 4, we show the experiment. Finally, we describe the conclusion.

2. 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.1

2. 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.2. 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

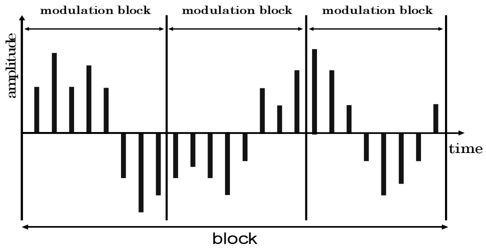

Figure 2. Modulation block. Block is divided into the some modulation blocks. In this figure, block is divided into 3 blocks.

time-shift processing like as MP3. In this section, we explain embedding process of the conventional method.

2. 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.

$$w^s(k) = \begin{cases} h & (k = s \text{ and } g'_s > A) \\ -h & (k = s \text{ and } g'_s \leq A) \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

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

2. 1. 2 Amplitude modulation

Spread impulse sequence $\widetilde{W}^s$ 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 $\widetilde{W}^s$ are divided into l parts of M-length modulation blocks, $G_i = (g_i(0), g_i(1), \cdots, g_i(m-1))$ $\widetilde{W}^s = (\widetilde{w}_i^s(0), \widetilde{w}_i^s(1), \cdots, \widetilde{w}_i^s(m-1))$ $(0 \leq i \leq l-1)$ 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.

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

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

$$\widetilde{w}_i'^s(p) = \begin{cases} \widetilde{w}_i^s(p) \times a & (e'_i \leq E) \\ \widetilde{w}_i^s(p) \times b & (e'_i > E) \end{cases} \quad (i = 0, 1, \cdots, l-1) \tag{3}$$

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

2. 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.

$$M(p) = 1 - \left(1 - \frac{a}{b}\right) \cos\left(\frac{2\pi p}{m-1}\right) \tag{4}$$

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

$$y_i^p = \begin{cases} \widetilde{x}_i^{p'} \times 1 & (e'_i \leq E) \\ \widetilde{x}_i^{p'} \times M(p) & (e'_i > E) \end{cases} \tag{5}$$

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

2. 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.

$$u'_k = \begin{cases} -32768 & \left(\left\lfloor u_k + \frac{1}{2} \right\rfloor < -32768\right) \\ \left\lfloor u_k + \frac{1}{2} \right\rfloor & \left(-32768 \leq \left\lfloor u_k + \frac{1}{2} \right\rfloor \leq 32767\right) \\ 32767 & \left(\left\lfloor u_k + \frac{1}{2} \right\rfloor > 32767\right) \end{cases} \tag{6}$$

2. 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 $\widetilde{X}'$ is obtained. And the average A' of $\widetilde{X}'$ is calculated. Finally, the embedded position s' is obtained by the following equation.

$$s' = \left\{ k \mid \max_{0 \leq k \leq N-1} \left| \widetilde{x}'_k - A' \right| \right\} \tag{7}$$

3. Sound Quality Improvement by Eliminating High Frequency Component from Signature Impulse

3. 1 Basic idea for sound quality improvement

In MP3 compression process of music data, MP3 tends to eliminate the high frequency component which is hard to perceive by human's auditory sense. Therefore, when watermarked music data is compressed by MP3, the high frequency component of the embedded signature impulse is eliminated. In this paper, we eliminate the high frequency component from the signature impulse by low-pass filter in

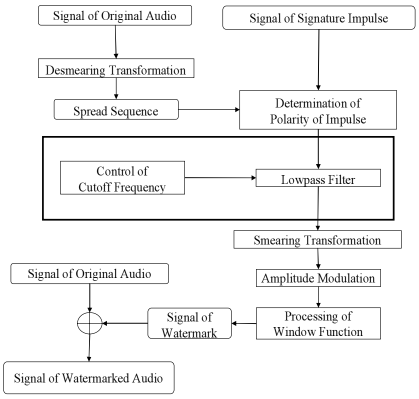

Figure 3. Flowchart of embedding procedure with the band-limited signature impulse in proposed method.

advance, and then the band-limited signature impulse is used as the embedded watermark signal. Because the band-limited signature impulse does not contain high frequency component, its power is less than the power of the signature impulse which is used in the conventional method. The less the power of watermark signal is, the less the change from the original music data is. As the result, by using the proposed method, it is expected that the deterioration of sound quality caused by watermarking is reduced without the decline of watermark recovery rate from the music data compressed by MP3. The flowchart of the embedding procedure of the proposed method with the band-limited signature impulse is shown in Fig.3. The part enclosed in heavy line is the additional processes to the conventional method.

3. 2 Eliminating high frequency component from signature impulse

In the proposed method, in order to eliminate high frequency component from the signature impulse, the filtering technique with Fourier transform is used as low-pass filter. Figure 4 shows the process of low-pass filtering with Fourier transform. First, the original impulse signal is transformed to frequency domain by FFT (Fast Fourier Transform). Then the spectra at high frequency are removed, and finally the band-limited signature impulse signal is obtained by reconstructing the spectra by Inverse-FFT.

The band-limited signature impulse has the side lobes which are occurred by removing high frequency component as shown in Fig.5. The intensity of these side lobes depends on the cutoff frequency of the low-pass filtering process. The smaller the cutoff frequency is, the larger the intensity of side lobes is. If the intensity of side lobes is too large, the correct position of main lobe cannot be detected in the watermark extracting process, and the position of side lobe may be detected wrong. On the other hand, the smaller the cutoff frequency is, the smaller the power of the band-limited signature impulse is. And hence, it is necessary that the cutoff frequency is set to be apposite value.

In this paper, we attempt two control methods for apposite cutoff frequency. In **control method (I)**, the identical cutoff frequency is used to all blocks in watermark

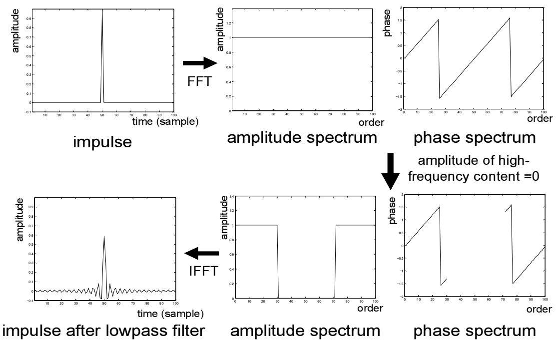

Figure 4. Process of low-pass filtering with Fourier transform.

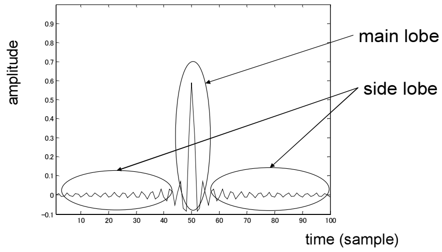

Figure 5. Spectra of the band-limited signature impulse with the side lobes which are occurred by removing high frequency component.

embedding process. After embedding watermark, it is confirmed whether all watermark information is correctly extracted from the watermarked audio data. The confirmation is repeated while the cutoff frequency is decreased, and we find the minimum cutoff frequency which can extract all watermark information correctly. In **control method (II)**, the different cutoff frequency is used to each block in watermark embedding process. The confirmation and the minimization of cutoff frequency are individually performed in each block.

4. Experiments

4. 1 Experimental condition

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 10 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 band-limited signature impulse at the position of the 100th sample in each block. Experiments are performed under the following conditions: the coefficients of amplitude modulation a=0.8 and b=1.0, the length of block is 4096 samples, the length of modulation block is 1024 samples, the strength of embedding is the value shown in Table 1. Lame3.93.1 [2] is used for MP3 compression, and bit rate is set to be 128 kbps.

In this paper, the robustness of watermarking method is evaluated by watermark recovery rate. The watermark recovery rate R is defined as the following equation

Table 1. Minimum of watermark strength with recovery rate 100% of watermark before MP3 compression.

music No.	genre	strength	
		time correct impulse	signature impulse
01	pops	2500	2500
02		1000	5000
03		5000	7000
07	rock	6000	9000
08		2000	8500
09		8000	10000
13	dance	23000	24000
14		11000	16000
15		1000	2500
28	jazz	1000	3000
29		4000	7000
30		5000	3500
49	classic	1000	1000
50		1000	2000
51		2000	2000

$$R = \frac{\sum_{k=1}^{H} B\left(w_k = \widetilde{w}_k\right)}{H} \tag{8}$$

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

The sound quality of watermarked music data is evaluated by the segmental signal-to-noise ratio SNR_{seg} defined by the following equations.

$$\text{SNR}_{\text{seg}}\left(\text{dB}\right) = \frac{1}{\text{seg}} \sum_{n_1=0}^{\text{seg}-1} \text{SNR}_{n_1} \tag{9}$$

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

where org_i is ith sample of original music data, emb_i is ith sample of embedded music data, smp is the length of segment, seg is the number of segments which is included in original music data. In this experiment, smp is set to be 4096.

4. 2 Experimental results

Table 2 shows experimental results of SNR_{seg} and the watermark recovery rate after MP3 compression. In every music data, the conventional method indicated lower SNR_{seg} than the control method (I), and the control method (I) indicated lower SNR_{seg} than the control method (II). Additionally, about the average of recovery rate of all music, the conventional method indicated lower recovery rate than the control method (I). But in every music data, the control method (II) indicated lower recovery rate than the control method (I). In music No.15, the recovery rate is hardly low. Because if we determined the cutoff frequency for each block, the control method (I) would have little high frequency component which could be eliminated in each block. Then the cutoff frequency cannot get hardly low. The above results mean that the control method (I) attained good result in both SNR_{seg} and recovery rate. The control method (II) attained good result in SNR_{seg}, but bad result in recovery rate. If there is a possibility to use for MP3 compression, the recovery rate of the control method (II) is too bad. Then we use the control method (I) in this case. If there isn't a possibility to use for MP3 compression, we should use the control method (II) which can provide high sound quality.

Table 2. Experimental results of SNR_{seg} and watermark recovery rate after MP3 compression.

music No.	genre	SNR_{seg} (dB)			recovery rate (%)		
		conventional	control (I)	control (II)	conventional	control (I)	control (II)
01	pops	27.4	28.2	34.1	90.1	100	28.3
02		28.5	28.5	44.4	95.3	100	40.1
03		29.6	29.7	39.5	99.1	99.1	39.6
07	rock	27.1	27.5	39.6	97.2	97.2	38.2
08		22.5	22.6	40.3	98.1	98.1	40.1
09		26.6	28.0	33.2	92.0	93.9	33.5
13	dance	18.0	18.3	30.8	93.9	100	41.0
14		22.1	23.1	30.6	97.6	99.5	34.4
15		30.3	30.7	47.3	100	100	76.9
28	jazz	37.9	39.3	51.2	97.6	100	42.0
29		17.1	17.4	28.9	98.1	100	42.5
30		34.8	35.2	44.4	100	92.5	39.6
49	classic	38.0	38.6	48.2	92.9	92.9	32.6
50		34.2	35.6	49.5	97.2	97.6	42.0
51		39.5	40.6	50.5	93.4	94.8	36.8
average		28.9	29.6	40.8	96.2	97.7	40.5

5. Conclusion

This paper has proposed the method to improve the sound quality of watermarked audio data by reducing the power of the signature impulse. In our method, the signature impulse is filtered by low-pass filter, and then the band-limited signature impulse is embedded in audio data. From the experimental results, it is confirmed that the segmental SNR of watermarked audio data is improved. But the robustness to MP3 compression in the control method (II) has not been improved. We have made a trial of the basic reducing technique by using low-pass filtering in this paper. As the future works, it is necessary to investigate the better reducing method which adapts to original audio data.

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] http://lame.sourceforge.net/