An Efficient Audio Watermark Extraction in Time Domain

Heawon Kang*, and Sung-Hwan Jung*

Abstract: In this paper, we propose an audio extraction method to decrease the influence of the original signal by modifying the watermarking detection system proposed by P. Bassia et al. In the extraction of the watermark, we employ a simple mean filter to remove the influence of the original signal as a preprocessing of extraction and the repetitive insertion of the watermark. As the result of the experiment, for which we used about 20 kinds of actual audio data, we obtain a watermark detection rate of about 95% and a good performance even after the various signal processing attacks.

Keywords: audio watermarking, blind detection Wavelet

1. Introduction

The watermark is the copyright information that is embedded into the multimedia content in order to protect it from being illegally copied and distributed. Requirements of the watermark depend on the purpose of its application. A watermark has various features, among which the most important are imperceptibility and robustness, which can conflict with each other. Thus, a compromise is needed [1-3].

In order to satisfy the imperceptibility of the watermark, most of the watermark is embedded into multimedia content as a "noise" both in the time domain and in the frequency domain. Therefore, the energy of the original signal is relatively much stronger than the energy of the watermark.

The watermarking detection system proposed by P. Bassia et al. is a blind detection system based on the assumption that the frame size is sufficiently large [4]. In its practical application, however, the frame size is not large enough for the original signal and the watermark to be uncorrelated [5]. Consequently, the detection result based on the system of P. Bassia et al. is affected significantly by the original signal in the practical application.

This paper presents a method to reduce the influence of the original signal by employing a simple high-pass filtering using a mean filter. In order to increase robustness, we add the repetitive insertion of the watermark to the embedding system of P. Bassia et al. The work presented here significantly improves the efficiency of watermark detection in the time domain.

2. Previous Audio Watermarking System

2.1 Previous Audio Watermarking Systems

P. Bassia et al. proposed a watermarking system in the time domain with a blind detection [4]. For embedding, they modulate the watermark, the streams of {-1, 1}, according to the amplitude of the original signal. This "signal-dependent watermark" is embedded into the original signal in order to increase the robustness of the watermark while satisfying its imperceptibility.

The audio signal is divided into N_s frames of L samples. $x_k(i)$ denotes audio samples of a k-th frame. If the watermark $w_k(i)$ is modulated into the signal-dependent watermark $w_k'(i)$, the obtained watermarked signal $y_k(i)$ is,

$$y_k(i) = x_k(i) + w_k'(i) \quad i = 0, 1, ..., N_s - 1 \tag{1}$$

For detecting, they use the correlation between the watermark and the subject signal. If the subject signal for detection is $y_k(i)$, the correlation between the watermark and the signal y, denoted by S_k, is,

$$S_k = \frac{1}{L}\sum_{i=1}^{L} y_k(i)w(i) \cdot \tag{2}$$

$y_k(i)$ can be rewritten as $x_k(i) + w'_k(i)$. So, the equation (2) is rewritten as,

$$S_k = \frac{1}{L}\left(\sum_{i=1}^{L} x_k(i)w(i) + \sum_{i=1}^{L} w_k'(i)w(i)\right) \tag{3}$$

They divide the watermark pattern $w(i)$ into two subsets, a set A whose sum is zero and a set B whose sum is not zero [4].

So, they rewrite S_k as follows:

$$S_k = \frac{1}{L}\left(\sum_{i \in A} x_k(i)w(i) + \sum_{i \in B} x_k(i)w(i) + \sum_{i=1}^{L} w_k'(i)w(i)\right) \tag{4}$$

Manuscript received February 10, 2006; accepted March 3, 2006.
Corresponding author: Sung-Hwan Jung
* Dept. of Computer Engineering, Changwon National University, Korea (sjung@changwon.ac.kr)

To simplify the above equation, they define its parts as:

$$T_{1,k} = \frac{1}{L}\sum_{i \in A} x_k(i)w(i) \tag{5}$$

$$T_{2,k} = \frac{1}{L}\sum_{i \in B} x_k(i)w(i) \tag{6}$$

$$T_{3,k} = \frac{1}{L}\sum w'(i)w(i) \tag{7}$$

They say that the expected value of $T_{1,k}$ is zero, and that if the number of subset A is sufficiently large, $T_{1,k}$ is close to zero. So,

$$S_k = (T_{1,k} + T_{2,k} + T_{3,k}) \cong (T_{2,k} + T_{3,k}). \tag{8}$$

If not watermarked, because $T_{3,k}$ is zero, $S_k = T_{2,k}$.

If watermarked, $S_k = T_{2,k} + T_{3,k}$.

Therefore, they propose the following ratio, r_k, to detect a watermark [4].

$$r_k \cong (S_k - T_{2,k}) / T_{3,k} \tag{9}$$

If r_k is bigger than a threshold, it is declared as being watermarked.

2.2 Problems with the previous system

In probability, if a set X and a set Y are independent random variables, the expectation value of the average product of each element of X and Y is

$$E(X \cdot Y) = E(X) \cdot E(Y) \tag{10}$$

where $E()$ denotes the expected value operator. If either $E(X)$ or $E(Y)$ is zero, $E(X \cdot Y)$ becomes zero [6].

Going back to the extracting system, P. Bassia et al. assume that $T_{1,k}$ is zero based on the assumption that

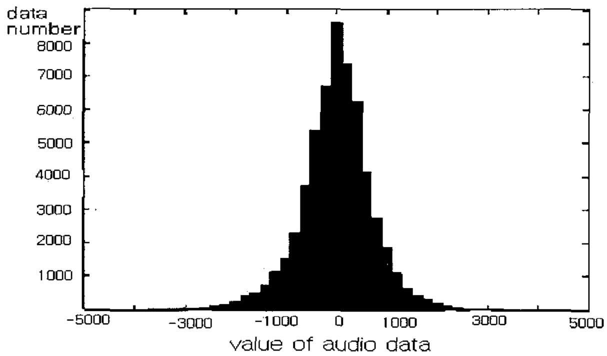
$$T_{1,k} = \frac{1}{L}\sum_{i \in A} x_k(i)w(i) = \frac{1}{L}\sum x_k(i) \cdot \frac{1}{L}\sum_{i \in A} w(i)$$
$$= \frac{1}{L}\sum x_k(i) \cdot 0 = 0 \tag{11}$$

Also, P. Bassia et al. choose the frame length $L = 2^{17} = 131072$ in their experiments. However, the frame length L is not sufficiently long enough to provide the independency between the original signal and the watermark in the practical audio watermarking application. Therefore, if the subset A of the original signal $x_k(i)$ and the watermark pattern $w_k(i)$ are not independent, the above assumption cannot be applied.

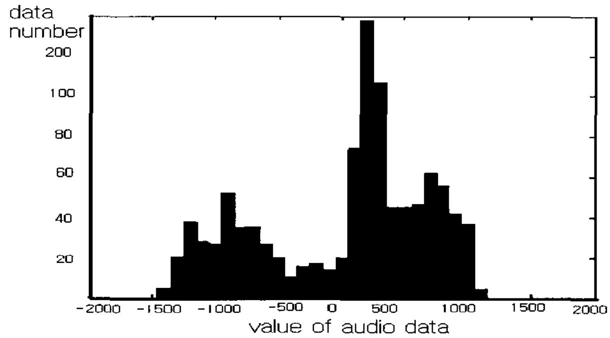
In addition, the energy of the watermark is very weak compared with that of the original signal, because the watermark is embedded into the original signal as a noise for imperceptibility. Therefore, the correlation between them is affected significantly by the original signal when the watermarked signal and the original signal are not independent.

Fig. 1 shows an example of the distribution of audio samples in one frame according to the frame size. Fig. 1(a) shows the distribution when the frame length is two to the power of seventeen and Fig. 1(b) shows the distribution when the frame length is two to the power of ten.

D. Kirovski and H. S. Malvar applied a pre-process before embedding a watermark in order to make a Gaussian distribution with a zero mean artificially and increase the reliability of the experiments [7]. However, this process not only brings about an irreversible change of the audio data but also degrades the quality of the audio so that it is inadequate for the original purpose of the audio.



(a) The frame length $L = 2^{17}$



(b) The frame length $L = 2^{10}$

Fig. 1. The distribution of one frame sample data according to different frame sizes.

For these reasons, we present a method for removing the influence of the original signal in the extracting process. It employs a simple mean filter to significantly reduce the influence of the original signal and obtain the high correlation of the watermark pattern and the watermarked signal, i.e. the subject signal. Also, we improve the accuracy of the extraction by using the repetitive embedding of the watermark. This is possible because we use a relatively smaller frame length compared with that of the system proposed by P. Bassia et al.

3. Proposed Audio Watermarking System

3.1 Embedding Process

Fig. 2 shows the embedding process. The message data is the stream of {-1, 1} that has copyright information.

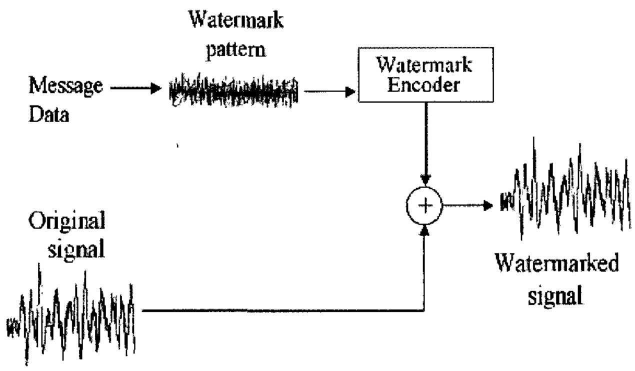


Fig. 2. The embedding process of a generic audio watermarking system

The watermark pattern or reference pattern is the predefined pattern that is the stream of {-1, 1} and its length is the frame length. For this system, we have two patterns. One is for the message data -1, and the other is for the message data 1. Each message data corresponds to a specific frame. If the message data for the corresponding frame is -1, then the modulated watermark pattern for the message data -1 is embedded into the audio signal of the frame, and vice versa. We refer to this process as mapping a message data into a watermark pattern.

Assuming an audio signal of T samples: $x(i),\ i=0,1,...,T-1$. The audio signal is divided into N_s frames of L samples. $x_k(i)$ denotes audio samples of a k th frame.

$$x_k(i) = x(kL+i), \quad i = 0,1,...L-1, \quad k = 0,1,...N_s-1 \tag{12}$$

This bipolar array of {-1,1} is used as a watermark pattern $w(i)$.

$$w(i) \in \{-1,1\}, i = 0,1,...L-1 \tag{13}$$

The watermark pattern is modulated according to the amplitude of the original signal so as to produce the "signal-dependent" watermark pattern $w_k^{'}(i)$.

$$w_k^{'}(i) = \alpha \cdot |x_k(i)| \cdot w(i), \quad i = 0,1,...L-1, \quad k = 0,1,....N_s-1 \tag{14}$$

Here, α is a constant to control the amplitude of the watermark pattern. Finally, the watermarked signal $y_k(i)$ is obtained by adding the modulated watermark pattern to the original signal.

$$y_k(i) = x_k(i) + w_k^{'}(i) \tag{15}$$

We can employ the repetitive insertion of the message data to increase robustness. This helps in extracting more reliable data through attacks. If the length of message data is M and the number of overall frames is N_s, the message data is repeated N_s / M times. If the repetition number N increases, the extracted message data becomes more reliable.

3.2 Extracting Process

For extracting, we employ a high-pass filtering as a pre-process. Then we apply the pre-processed signal to the detecting system used by P. Bassia et al.

We utilize the high-pass filtering in order to reduce the influence of the original signal by removing a signal similar to the original signal from the subject signal. First, this system employs a simple four point mean filter. It is a low-pass filter [8]. If the subject signal is y and the filtered signal through the four point mean filter is y', we obtain signal z by subtracting y' from y.

$$z = y - y^{'} \tag{16}$$

The signal z is divided into N_s frames of L samples

$$z_k(i) = z(kL+i), \quad i = 0,1,...L-1, \quad k = 0,1,...N_s-1 \tag{17}$$

In k th frame, $w_{-1,k}(i)$ denotes the watermark pattern for the message data -1 and $w_{1,k}(i)$ for the message data 1 The correlation between z_k and each watermark pattern, $w_{-1}(i)$ and $w_1(i)$ is denoted by S_k.

$$S_k = \frac{1}{L}\sum_{i=1}^{L} z_k(i) w_{j,k}(i), \quad k = 0,1,....N_s-1, \quad j \in \{-1,1\} \tag{18}$$

Because the influence of the original signal is reduced significantly, we use the watermark detection ratio as follows:

$$r_{j,k} \cong (S_{j,k} - T_{j2,k}) / T_{j3,k}, j \in \{-1,1\} \tag{19}$$

$T_{j2,k}$ can be ignored because the value is very small. So, it can be rewritten as

$$r_{j,k} \cong S_{j,k} / T_{j3,k}. \tag{20}$$

The inserted message data can be determined in this way:

$$\begin{aligned} r_{-1,k} > r_{1,k}, &\quad message\ \ data = -1 \\ r_{-1,k} \le r_{1,k}, &\quad message\ \ data = 1 \end{aligned} \tag{21}$$

4. Experiment

The proposed system is evaluated by watermark extraction efficiency, compared with the system proposed by P. Bassia et al. In order to make a similar experimental environment, two watermark patterns are used for mapping a message data for both the proposed system and the previous system.

Every audio test data is a data with a 44,100 sampling rate per second, mono-channel, and 16 bits per sampling. The 20 kinds of test data from the various music genres are extracted from music CDs. The frame size used for the experiment is 1,024 because it is one of the most widely used frame sizes [5].

Fig. 3 shows the extraction ratio of the watermark without the signal processing attacks. The average extraction ratio in the proposed system is 95.0%. The average extraction ratio in the system of P. Bassia et al. is 2.3%.

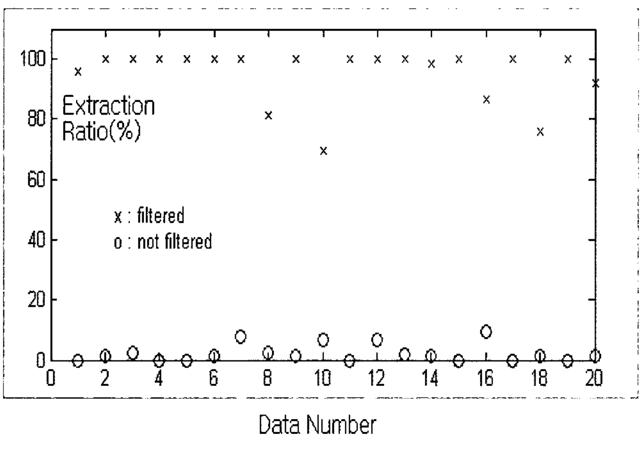


Fig. 3. Extraction ratios without attacks

For the various signal processing attacks, 20 kinds of test music data are down-sampled, down-quantized, and compressed. They are down-sampled from 44,100 Hz to 32,000 Hz. The resolution of test music data is lowered from 16 bits per sampling to 8 bits per sampling. The compression used for the attack here is MPEG-layer3 [9].

Table 1. Comparison of the average extraction ratio

Signal processing	Extraction ratio of proposed system (%)	Extraction ratio of Bassia et al. system (%)
Without attacks	95.00	2.31
Down-sampling	51.53	13.40
Down-quantization	63.20	19.27
Compression-mp3	55.00	9.20
Average	66.18	11.05

Table 1 shows the average of the extraction ratios of the 20 kinds of test music data in the case of various signal processing attacks. The averaged extraction ratio of the proposed system is improved by about 55% using the proposed approach.

5. Conclusion

The blind detection system proposed by P. Bassia et al. is significantly affected by the original signal when the frame size is not sufficiently large. Because the watermark is embedded as a type of noise in order to be unobtrusive, 1) its energy is much smaller than the energy of the original signal. Also, in the practical application, 2) the frame size is not sufficiently large to provide independency between the original signal and the watermark. These two facts affect the correlation between the watermark and the subject signal for detection and make watermark extraction difficult.

In this paper, we propose a method to reduce the influence of the original signal by employing high-pass filtering. By using a simple four point mean filter, we remove a signal similar to the original signal from the subject signal. Also, we increase robustness through the repetitive insertion of the watermark.

In experiments, the extraction ratio is improved by about 55% by the proposed system. The proposed system also shows better performance after various signal processing attacks when compared with the previous system.

References

[1] I. J. Cox, M. L. Miller, and J. A. Bloom, Digital Watermarking, Morgan Kaufmann Pub., 2002.

[2] I. J. Cox, J. Kilian, T. Leight and T. Shamoon, "A Secure, Robust Watermark for Multimedia," in Workshop on Information Hiding, Newton Institute, Univ. of Cambridge, pp. 147-158, May 1996.

[3] I. J. Cox, J. Kilia, T. Leighton and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia," in IEEE Trans. on Image Processing, vol. 6, no. 12, pp 1673-1687, 1997.

[4] P. Bassia, I. Pitas and N. Nikolaidis, "Robust Audio Watermarking in the Time Domain," in IEEE Trans. on Multimedia, vol. 3, no. 2, pp. 232-241, Jun. 2001.
[5] Hui Suk Lee, An Audio Watermark Method Using the Attribute of the Tonal Masker, a doctoral dissertation in Changwon National University, July 2003.
[6] Alberto, Leon-Garcia, Probability and Random Process for Electrical Engineering, Addison Wesley Pub. 1994.
[7] D. Kirovski and H. Malvar, "Spread Spectrum Watermarking of Audio Signals," in IEEE Transactions on Signal Processing, vol. 51, no. 4, pp. 1020-1033, Apr. 2003.
[8] James H. McClellan, Ronald W. Schafer, and Mark A. Yoder, DSP First a Multimedia Approach, Prentice Hall Pub. 1999.
[9] K. Brandenburg and H. Popp, An Introduction to MPEG Layer-3, EBU Technical Review, Jun. 2000.

Haewon Kang
She received her BS and MS in Electrical Engineering and Computer Engineering from Changwon National University, Korea, in 2002 and 2004 respectively. She received a MS in Media Arts and Technology from the University of California (UCSB) in 2006. Her research interests are in Audio Watermarking, and Multimedia Information Processing.

Sung-Hwan Jung
He received a BS, MS, and Ph.D. in Electronics Engineering (Information and Communication & Image Processing) from Kyungpook National University, Korea, in 1979, 1983, and 1988, respectively. From 1983 to 1985, he was with the Electronics and Telecommunications Research Institute in Korea as a member of the research staff, where he developed a portable computer. In 1988, he joined the faculty of the Department of Computer Engineering at Changwon National University in Korea, where he is currently working as a full professor. From 1992 to 1994, he was on the postdoctoral research staff of the Department of Electrical and Computer Engineering at the University of California at Santa Barbara (UCSB). From 1999 to 2000, he also worked for the Colorado School of Mines (CSM) in Golden as a visiting professor. He is an Information System Auditor and P.E. in the area of Information Processing and Electronic Computers. His research interests are in Content-based Image Retrieval, Image Steganography and Watermarking, Internet-based Remote Monitoring, Text Mining, and Bioinformatics.