CAPTURING-RESISTANT AUDIO WATERMARKING BASED ON DISCRETE WAVELET TRANSFORM

Seungjae Lee, Sang-Kwang Lee, Young-Ho Seo, and Chang D. Yoo[1]

Digital Contents Research Division, ETRI, Daejeon,Korea
[1] Dept. of EECS, Div. of EE, KAIST, Daejon, Korea
{seungjlee, sklee, syh}@etri.re.kr and [1] cdyoo@ee.kaist.ac.kr

ABSTRACT

In this paper, we propose a wavelet-based audio watermarking algorithm that is robust against capturing attack. With a commercial capturing tool, it is possible to capture various audio contents without noticeable degradation, and thus can potentially facilitate the illegal distribution of the audio content. By adjusting the mean value of the lowest subband coefficients of the discrete wavelet transform (DWT) of the audio, the proposed watermark can survive capturing attack including sampling rate conversion, random cropping and compression. By incorporating a simple human auditory model, the inaudibility of the watermark achieved, and the detection probability is improved based on the difference information of extracted values. This is confirmed by experimental results.

1. INTRODUCTION

Illegal distribution of music files has been on the rise in recent years with the development of the Internet and the appearance of portable MP3 players. In proportion to the amount of illegal distribution, the profit of music industry has been declining and this has brought about an urgent need to protect copyrighted music files against piracy.

Many algorithms [1]-[10] have been proposed to protect music files from piracy. The Digital right management (DRM) system, based on traditional cryptographic algorithms, can deliver music files secretly and grant authorization to certain users for the music files in accordance with a proper policy. The DRM system works quite well and it seems to be a good protection method; however, it has a security problem in that it can not prevent malicious users from accessing raw data while a music file is being rendered [4].

For example, when music files are played in a local computer, raw data can be captured by using a commercial capturing tool [11] from sound devices. Once raw data are obtained and shared on the Internet, illegal distribution cannot be blocked by the DRM system. During the capturing process, if information about the user who is capturing can be embedded and extracted, then the illegal distributor can be traced. Accordingly, the DRM system can trace back to the illegal distributor.

Fingerprinting, which is closely related to watermarking, has been proposed for such applications [2]-[4]. Unlike traditional watermarking algorithms, user information is embedded; moreover, to discriminate colluders from illegally distributed files, collusion-resistant fingerprinting code [5],[6] may be included.

To adopt fingerprinting into the DRM system as shown in Fig. 1, capturing-resistant watermarking algorithm is an essential prerequisite. The capturing attack consists of various attacks such as compression, cropping and so on; therefore, capturing-resistant watermarking algorithm is to be robust against each attack and combination of them. Because cropping occurs in every capturing process, the ability to self-synchronize itself is also necessary. A few literatures in audio watermarking have investigated the robustness issues associated with these attacks. In [7]-[10], cropping and shifting were examined and simple kind of multiple attacks were considered, but real audio capturing was not considered.

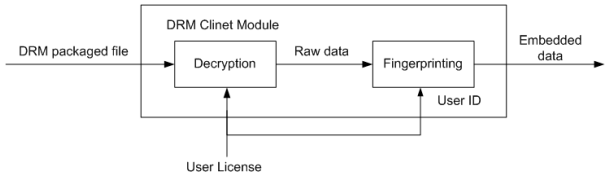

Fig. 1. DRM system with fingerprinting.

In this paper, we propose a capturing-resistant audio watermarking algorithm. The proposed method embeds the watermark by adjusting the mean value of the lowest subband coefficients in DWT domain. By modifying the mean value of low frequency coefficients, the proposed method is robust against attacks such as compression and random noise. By letting the mean values of the subband to be embedded zeros periodically, the proposed watermark can easily synchronize and be resistant against cropping. By adjusting the embedding power depending on a simple human auditory system (HAS) model, the proposed watermark is transparent. The propose method uses the difference of the embedded values which is relatively stable during random perturbation so that the detection probability can be improved.

Experimental results show that the proposed method is robust against capturing attack with various conditions such as different compression ratio, compression algorithm and sampling rate. We also apply the proposed method to DRM client module, and examine the robustness of capturing attack. The audibility of the proposed watermark is tested by a preference test.

This paper is organized as follows. Section 2 presents audio capturing attack and its characteristics. Section 3 explains the proposed method. Section 4 gives the experimental results and discussion. Section 5 summaries.

2. AUDIO CAPTURING ATTACK

Audio capturing is a process that holds and stores audio signal from a playback device. Commercial capturing tools enable us to save sound that we hear on the device, and an audio capturing attack includes the following processes:

- D/A and A/D conversion: When music files are played, D/A

conversion occurs. After then, A/D conversion is performed in order to store the captured audio data.

- Random Cropping: Audio capturing can crop any random area in the music files.
- Sampling rate conversion: When the captured audio data are stored, an arbitrary sampling rate can be chosen.
- Compression: When the captured audio is stored, an arbitrary compression ratio and any kind of an audio file format can be selected.
- Random perturbation: When the captured audio data are stored, random noise may be added or some samples can be unpredictably dropped according to the capturing environments.

After D/A conversion, audio data including random noise are captured at the random position and saved with compression and resample. Therefore, to be capturing-resistant, all the mentioned above attacks should be considered in the design of the watermark.

3. THE PROPOSED METHOD

In the proposed method, a watermark is embedded in the form of a low frequency signal by modifying the mean value of the lowest subband coefficients of the DWT of the input audio. This leads to the robustness against compression and random noise. To avoid degradation in audio quality, the proposed method adjusts the strength of the watermark with a simple HAS model. To be robust against random cropping and perturbation, the synchronization process is introduced and the difference between average mean values of adjacent frames are used to improve the detection probability.

3.1. Embedding Method

Fig. 2 shows the embedding process in the proposed method. Each audio file is divided into 40ms frames with 50% overlap which are windowed by Hanning window. Then, the DWT is performed and the encoded watermark is embedded by modifying the mean value of the lowest subband coefficients in DWT domain. After the embedding process, each frame is overlap-added to smooth the variation of the embedded frame.

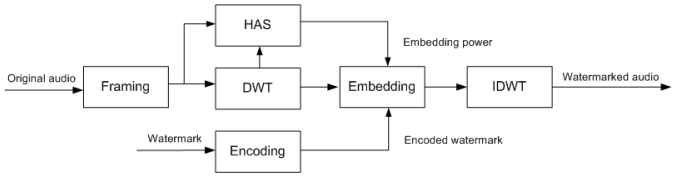

Fig. 2. The embedding process in the proposed method.

3.1.1. Simple HAS

The proposed method adjusts the strength of the watermark to achieve the transparency. Instead of directly applying the psychoacoustic model used in the MPEG, we use the relationship between the absolute sum of the samples of input frame (S_1) and the absolute sum of the high frequency coefficients in the DWT of input frame (S_2). From equal loudness curves and masking effect, the embedding power, which is roughly proportion to S_1 and S_2, is calculated by the experiment. The embedding process uses embedding power tables that give the appropriate strength for a particular S_1 and S_2.

3.1.2. Watermark Encoding

The proposed method uses -1 and 1 to represent the watermark. The encoding process converts the watermark into the encoded watermark as shown in Fig. 3.

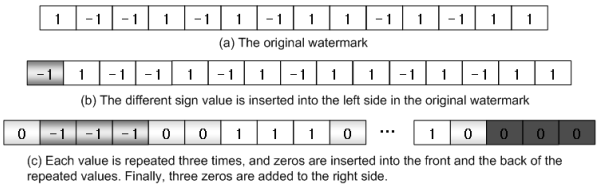

Fig. 3. The watermark encoding process in the proposed method.

Step 1 Insert the opposite sign value of the left side in the watermark into the front of it.

Step 2 Each value is repeated three times, and zeros are inserted into the front and the back of the repeated values.

Step 3 Three successive zeros are added to the right side.

Then, the basic embedding block is generated to be used for embedding, repeatedly.

3.1.3. Watermark Embedding

According to the encoded watermark (w_e), the mean value of the lowest subband coefficients in DWT domain is modified as follows:

- If w_e is 1 or -1, the mean value of the lowest subband coefficients in DWT domain is modified to positive value or negative value, respectively.
- If w_e is 0, the mean value of the lowest subband coefficients in DWT domain is modified to zero.

Each n*th* coefficient of the lowest subband coefficients in DWT domain at the k*th* frame $C_o^k[n]$ is modified by the following:

$$C_m^k[n] = \begin{cases} C_o^k[n] - (2m - P_a^k) \cdot H[n] & if\ w_e = 1 \\ C_o^k[n] - (2m + P_a^k) \cdot H[n] & if\ w_e = -1 \\ C_o^k[n] - 2m \cdot H[n] & if\ w_e = 0 \end{cases} \quad (1)$$

where $C_m^k[n], m, P_a^k$ and $H[n]$ are the n*th* modified coefficient of the lowest subband coefficients in DWT domain at the k*th* frame, the mean of the lowest subband coefficients in DWT domain at the kth frame, the embedding power calculated by the simple HAS, and Hanning window, respectively.

3.2. Extraction Method

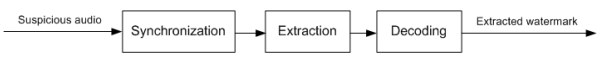

Fig. 4. The extraction process in the proposed method.

Fig. 4 shows the entire extraction process in the proposed method. At first, the suspicious audio is framed and the DWT is performed the same way as the embedding method did. To extract the embedded watermark, the synchronization is performed by the statistical

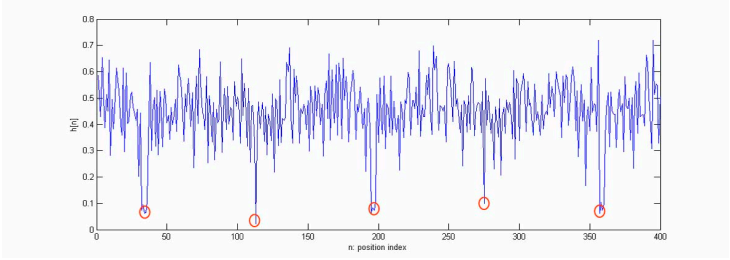

Fig. 5. Each red circle indicates the minimum position of $h[n]$.

property of DWT coefficients which were intentionally introduced. Then, the embedded watermark is extracted from the difference between adjacent averaged mean values of the lowest subband coefficients.

3.2.1. Synchronization

For synchronization, the proposed method inserted successive zeros into the encoded watermark. Synchronization is conducted by searching the minimum position of the following equation:

$$h[n] = \alpha \cdot f[n] + \beta \cdot g[n] \tag{2}$$

where $f[n]$, $g[n]$, α and β are the mean of N adjacent means of the lowest subband coefficients in DWT domain, the variance of them and weighting factors, respectively.

Fig. 5 shows that $h[n]$ can indicate a synchronization position when α and β are 0.6 and 0.4, respectively.

To improve the synchronization ability, the proposed method finds a minimum position among the sum of N $h[n]$ values at the period position of the encoded watermark as follows.

$$\arg\min_{n} \sum_{k=0}^{N-1} h[n - kP] \tag{3}$$

where P is the period of the encoded watermark.

This reduces the synchronization error considerably. With Eq. (3), the approximate synchronization position is determined by shifting a sample position with a proper step size. By using a full search, we can find the exact minimum position; however, a coarse synchronization position is enough to extract the embedded watermark because the propose method uses a small frame size and the repeated values are embedded.

3.2.2. Watermark Extraction and Decoding

Once the synchronization position is determined, each mean value at the middle of every five successive values is normalized by the estimated embedding power at that present frame. Then, the normalized values are calculated, and they are approximately either p or $-p$. For example, p is 0.5 when the watermarked audio itself is an input to the extraction process.

Then, the difference (-2p, 0, 2p) between the previous normalized value and the present one is calculated in order to improve the detection probability because the difference is relatively stable under random noise and an amplitude level shifting. Moreover, the averaged difference values of several blocks are more stable.

The difference and the embedded watermark are modelled as a simple state machine, and the distribution of the difference value is

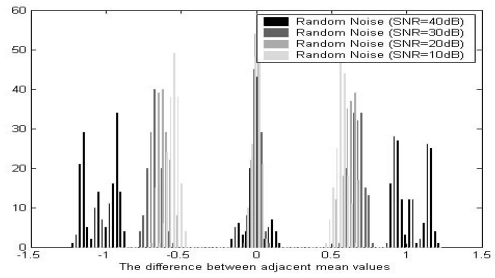

Fig. 6. Histograms of the differences (-2p, 0, 2p) between adjacent mean values.

presented as shown in Fig. 6. Therefore, each difference value is converted into observation $w_o^{'}$ by the following equation.

$$w_o^{'} = \begin{cases} a & d \geq p \\ b & -p \leq d < p \\ c & d < -p \end{cases} \tag{4}$$

where d is the difference value.

To decode the converted differences, we use a simple state diagram as shown in Fig. 7. When the first difference is b, the decoding does not start. To avoid this case, we insert the opposite sign watermark bit into the front of the left side in the watermark as mentioned in Section 3.1.2. The state diagram may fail when the normalized values are corrupted by random noise. But, the averaged difference values reduces the possibility of this case.

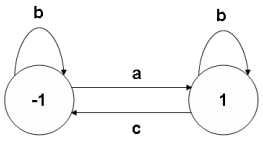

Fig. 7. The state machine for decoding process

4. EXPERIMENTAL RESULTS

In this section, the performance of the proposed method is evaluated. In the experiments, 9/7 biorthogonal wavelet filter was used for DWT, and the decomposition was performed five times. We collected 16 music files of four different genres and each music file is two channels with 44100Hz wave format with 100 second length. We also downloaded five DRM packaged music files. The embedded watermark is a string of eight byte characters.

4.1. Robustness Test

The robustness test was conducted under the following conditions.

1. Attacks on the watermarked file
2. Capturing attack on the watermarked file
3. Capturing attack on the DRM packaged file

In first two experiments, 16 music files were grouped into four genres and each experiment was performed four times. Table 1 shows results of the robustness test. There is no error in the first

experiment, but two errors occurred when a rock music file was captured. After examining that file, we found that the unintentional silence was inserted during music performance. Therefore, synchronization process failed.

Table 1. The Robustness Test I

	cropping[1]	compression[2]	resampling[3]	capturing[4]
classic	100%	100%	100%	100%
ballad	100%	100%	100%	100%
dance	100%	100%	100%	100%
rock	100%	100%	100%	87.5%
avg.	100%	100%	100%	96.8%

[1] 60 sec length at a random position. [2] MP3 128kbps. [3] 32kHz resampling. [4] 60 sec length capturing with MP3 128kbps 32kHz.

For five DRM packaged music files, we performed capturing attack where DRM system including the proposed method as a COM filter [12] was installed. In the DRM packaged files, a few detection error occurred as shown in Table 2. However, in most cases, the embedded watermark is accurately extracted.

In cases where an error occurs, we found that some samples were dropped due to the interruption of capturing process by other process such as the decryption and the execution of other applications.

Table 2. The Robustness Test II

	capturing[1]	capturing[2]	capturing[3]	capturing[4]
DRM 1	94.4%	100%	100%	100%
DRM 2	100%	99.4%	94.4%	100%
DRM 3	100%	100%	100%	94.4%
DRM 4	100%	99.4%	100%	100%
DRM 5	100%	100%	88.9%	100%

[1] 60 sec length capturing with MP3 128kbps 44.1kHz, [2] 60 sec length capturing with MP3 128kbps 32kHz, [3] 60 sec length capturing with MP3 128kbps 48kHz, [4] 60 sec length capturing with WMA 96kbps 44.1kHz

All capturing attacks cropped music files about 60 second length at a random position, then nine encoded watermark blocks are included in a captured file when a string of eight byte characters is used as a watermark. Therefore, the averaged mean was calculated by nine mean values to extract the watermark.

4.2. Sound Quality Test

The sound quality test was conducted by a preference test. The listener listens to both the watermarked audio and the original audio alternating between the two twice, then he or she chooses good one. The order is random. We cropped 20 second from four different genres music files, and six individuals participated in the experiments. Table 3 shows that the listeners can not discriminate the difference between the two. Therefore, the transparency of the watermarked audio is achieved.

5. CONCLUSIONS

In this paper, we proposed a capturing-resistant audio watermarking algorithm in DWT domain. Audio capturing attack includes random cropping and various attacks for common in watermarking and DRM system. By modifying the statistical property of the low frequency coefficients in DWT domain based on a simple HAS, the robustness against audio capturing attack and the transparency of the watermark signal are obtained. By exploiting the averaged difference of the extracted mean values, the detection probability is improved.

Table 3. The Sound Quality Test

	$A > B$	$A = B$	$A < B$
classic	33.3%	50%	16.7%
ballad	50%	50%	0%
dace	16.7%	66.6%	16.7%
rock	50%	33.3%	16.7%
avg.	37.5%	50 %	12.5%

A and B are the original audio and the watermarked audio, respectively.

In audio capturing attack, as each attacker captures at a different position, a collusion attack seems to be more difficult, but it is still possible. To extend the proposed method into fingerprinting, collusion-resilient codes will be considered, and the security of the proposed method will be improved. We leave it as a further work.

6. REFERENCES

[1] The Secure Digital Music Initiative. (http://www.sdmi.org/)

[2] M. Veen, A. N. Lemma, and T. Kalker, "Watermarking and fingerprinting for electronic music delivery," in *Proc. Security, Steganography, and Watermarking of Multimedia Contents VI*, vol. 5306, pp.200 - 210, Jan. 2004.

[3] J. Lee, and S. O. Hwang, S. Jeong, K. S. Yoon, C. S. Park, and J. Ryou, "A DRM Framework for Distributing Digital Contents through the Internet," *ETRI Jornal*,vol. 25, no. 6, pp. 423-436, Dec. 2003.

[4] D. Schonberg and D. Kirovski, "Fingerprinting and Forensic Analysis of Multimedia," *ACM Multimedia*, pp.788-795, 2004.

[5] W. Trappe, M. Wu, Z.J. Wang, and K.J.R. Liu, " Anti-collusion fingerprinting for multimedia," *IEEE Trans. on Signal Processing*, vol. 51, pp. 1069-1087, Apr. 2003.

[6] D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data," *IEEE Trans. on Information Theory*, vol. 44, pp. 1897-1905, Sep. 1998.

[7] S. Wu, J. Huang, D. Huang, and Y.Q. Shi, "Self-synchronized audio watermark in DWT domain ," in *Proc. IEEE Int. Sym. on Circuits and Systems*, vol. 5, pp. 712-715, May 2004.

[8] P. Bassia, I. Pitas, and N. Nikolaidis, "Robust audio watermarking in the time domain," *IEEE Trans. on Multimedia*, vol. 3, pp. 232-241, June 2001.

[9] Z. Liu, and A. Inoue, "Audio watermarking techniques using sinusoidal patterns based on pseudorandom sequences," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, pp. 801-812, Aug. 2003.

[10] N. Cvejic, and T.Seppanen, "Robust audio watermarking in wavelet domain using frequency hopping and patchwork method," in *Proc. IEEE Int. Sym. Image and Signal Processing and Analysis*, vol. 1, pp. 251 - 255, Sep. 2003.

[11] Capture Solution. (http://www.sprosoft.com/)

[12] A. Troelsen, *Developer's Workshop to COM and ATL 3.0*, Wordware Publishing Inc., 2000.