A Robust Zero-Watermarking Scheme for Digital Audio

Shu-Min Tsai

***Abstract*—As the Internet rapidly grows, networks carry large amounts of multimedia data. Protecting the copyright of digital audio is becoming an increasingly important issue. This paper presents a zero-watermarking algorithm for audio based on the energy of each frame. The characteristic of energy is utilized to retrieve the watermark after various attacks. To measure the robustness and security of the proposed method, the accuracies of extracting watermark from an audio signal following an attack by various signal processing schemes are evaluated. Simulation results demonstrate that the proposed algorithm has greater robustness than the methods in the literature and it exhibits good security. The proposed scheme can be used to protect rights of intellectual property in audio signals.**

***Index Terms*—Intellectual property right, information security, zero-watermark, robustness, framed energy.**

I. Introduction

With the fast development of digital media and the applications over the Internet, digital watermarking technique is playing an increasingly important role in the protection of copyright and the authentication of digital signals. To protect digital works against illegal use and tampering, the digital watermark was developed to protect copyright and authenticate content integrity. An extensively investigated scheme for protecting digital audio copyright is based on digital watermarking technology. Although many preliminary studies of watermarks have focused on both image and video, audio watermarks are causing more and more attention in recent years.

Various watermarking schemes have been presented but they may slightly distort the original signal and cannot provide a perfect balance between robustness and imperceptibility. The technique of zero-watermark was developed to be robust and imperceptible in digital media [1].

Recently, zero-watermarking schemes have been proposed to analyze the security of audio signals. [2] presented a method for mapping the approximate coefficients of the wavelet transform of an audio segment into a binary matrix. A blind zero-watermarking algorithm, based on the Discrete Wavelet Transform (DWT), has been used to construct secrete keys [3]. [4] used the Lifting Wavelet Transform (LWT) technique on audio aggregation zero-watermark. DWT and Discrete Cosine Transformation (DCT) have been combined to generate the watermarking sequences [5]-[7] employed an audio zero-watermarking algorithm that combined DCT and Zernike moments. [8] discovered that the entropy of an audio signal on different scales yielded its statistical features. [9] combined the DWT algorithm and Linear Prediction Cepstrum Coefficients (LPCC) to construct the zero-watermarking image. The method of Modified Discrete Cosine Transform (MDCT) is applied in the zero-watermark of audio is proposed by [10]. The concept of the zero-watermark based on LPCC with a weighting function was also proposed to protect audio signals [11], [12].

This work presents audio verification using the zero-watermark. The novel approach creates an audio watermark that represents a favorable balance of robustness and transparency, and resists the attack by various signal processes.

The rest of the paper is organized as follows. Section II, introduces the proposed zero-watermark constructing algorithm in detail. Section III presents the detection of the watermark. Section IV presents and analyzes experimental results to verify the robustness and security of the proposed technique. Finally, Section V draws conclusions.

II. Method for Constructing Zero-Watermark

The scheme for generating a zero-watermark consists of two parts, which are described as follows.

A. Preprocess

First, a watermarking image **I** = $\{I(i, j), 1 \le i \le X, 1 \le j \le Y\}$ of $X \times Y$ size is selected. This image is converted to a one-dimensional signal **V**, **V** $= \{v(k), 1 \le k \le X \times Y\}$ for the host audio. Fig. 1 displays the watermarking image.

Second, a high-pass filter is applied to the audio signal before the embedding process. The high-pass filter eliminates undesired low-frequency components. After the audio signal is filtered through the high-pass filter, it is framed by a Hamming window into $X \times Y$ segments.

Fig. 1. Watermarking image.

B. Constructing Process

Fig. 2 shows a block diagram of the zero-watermarking construction as follows.

1) Generate a random permutation that is used as a key (denoted as random seed) in the watermark detecting process. Then **V** is scrambled into **W**. By this way, we can decrease the relationship of pixel spaces in the

Manuscript received July 1, 2014; revised September 3, 2014. The author would like to thank the National Science Council of the Republic of China, Taiwan, for financially supporting this research under Contract No. NSC 101-2221-E-346 -012.

S. M. Tsai is with the Department of Communication Engineering of National Penghu University of Science and Technology Penghu, Taiwan (e-mail: smtsai@npu.edu.tw).

watermarking image, and improve the security of the system.

$$\mathbf{W} = \{w(k), 1 \le k \le X \times Y\}, \tag{1}$$

where $X \times Y$ is the size of image, and $w(k)$ represents the scrambled pixel.

2) The zero-watermarking sequences should be perceptually significant features. In this paper, the energy of each frame is taken as a characteristic value of audio signal.

E_{ave} is the average energy of all frames of the audio signal.

$$E_{ave} = \frac{\sum [x(t)]^2}{M}, \tag{2}$$

where $x(t)$ is the audio signal, and M is the number of frame. $FE(k)$ is defined the energy of frame, and will be called a framed energy.

$$FE(k) = \sum_{t=k}^{k+N} [x(t)]^2, \ k = 1, \ldots, M, \tag{3}$$

where k is a frame index, and N is the size of each frame.

3) A binary pattern **E**, is defined as

$$\mathbf{E} = \{e(k), 1 \le k \le X \times Y\}, \tag{4}$$

An element of the binary pattern **E** is obtained by comparing with E_{ave}, as follows.

$$e(k) = \begin{cases} 1, & if \ FE(k) > E_{ave}, \\ 0, & otherwise. \end{cases} \tag{5}$$

4) The zero-watermark ZW is generated as follows.

$$zw(k) = e(k) \otimes w(k), \tag{6}$$

where $\otimes$ is the XOR operation. The zero-watermark image is shown in Fig. 3.

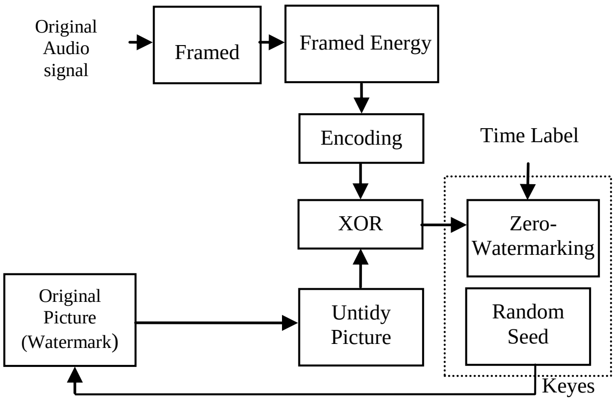

Fig. 2. Block diagram of watermark construction.

III. Watermark Detection

The watermark can be recovered. Fig. 5 shows the block diagram of the watermark-detecting scheme. The extraction processes are as follows.

1) The test audio is divided into $X \times Y$ frames.

2) The framed energy (*FE*) of the test audio is calculated.

3) The same procedure that is described as an embedding process in Section II-B is performed to estimate the binary pattern $\tilde{\mathbf{E}}$.

$$\tilde{\mathbf{E}} = \{\tilde{e}(k), 1 \le k \le X \times Y\}, \tag{7}$$

4) The XOR operation is applied between $\tilde{\mathbf{E}}$ and $zw(k)$, that is,

$$zw(k) = e(k) \otimes w(k), \tag{8}$$

where $\tilde{\mathbf{W}}$ is a sequence of the scrambled image.

5) The random seed is used to obtain the extracted watermark from $\tilde{\mathbf{W}}$.

Fig. 3. Zero-Watermark image of music signal.

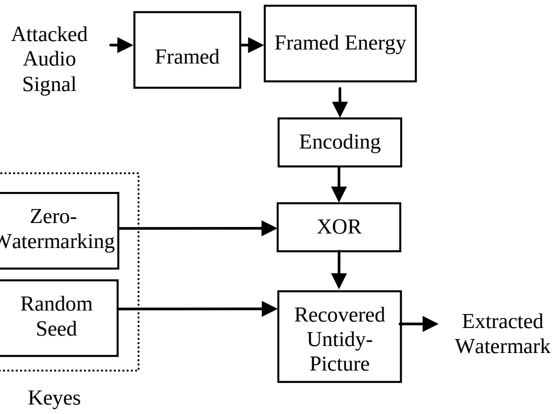

Fig. 4. Block diagram of watermark detection.

IV. Simulations and Analysis

This section analyzes the performance of the proposed algorithm. The bit error rate (BER), defined in Eq. (9), is used to measure the reliability of the proposed zero-watermarking technique.

$$BER = \frac{no.of \ error \ bits}{no.of \ total \ bits} \times 100\%, \tag{9}$$

where the error bits represent the difference between the retrieved watermark and original watermarking image. Hence, the watermarking method is highly robust and a trade-off must be made between the time required for the corresponding calculation and the quality of the digital watermark. The proposed algorithm provides a resolution.

The normalized cross-correlation (NC), defined in Eq. (10), is used to evaluate the similarity between the extracted and the original watermark.

$$NC(\mathbf{I},\tilde{\mathbf{I}}) = \frac{\sum_{i=1}^{X}\sum_{j=1}^{Y} i(i,j)\tilde{i}(i,j)}{\sqrt{\sum_{i=1}^{X}\sum_{j=1}^{Y} i^2(i,j)}\sqrt{\sum_{i=1}^{X}\sum_{j=1}^{Y} \tilde{i}^2(i,j)}}, \quad (10)$$

where $\mathbf{I}$ and $\tilde{\mathbf{I}}$ denote the original watermark and the extracted watermark, respectively.

The watermarking image for the audio signal is an 88×88 bit binary image, so the size $X \times Y$ is 7744.The original audio signal that is used in the experiment is 16 bit PCM with a sampling frequency of 44.1 kHz. There are ten audio signal included five song and five music. The languages of the songs are Japanese, Taiwanese, English, Chinese and Cantonese. The types of music are Classical, Rock, Electronic, Hip hop and Jazz. The average duration of each song signal was approximately 290 seconds, and that of each music signal was about 237 seconds.

Table I presents the attack methods including re-sampling, re-quantization, low-pass filter with a cut-off frequency of 11 kHz, the additive white Gaussian noise with a power of -20 dB and -10 dB, and MP3 compression with a bit rate of 128 kbps, 64 kbps and 32 kbps.

TABLE I: ATTACKED METHODS

TYPE	ATTACKED METHODS
1	Re-sampling 44.1→22.05→44.1 (kHz)
2	Re-quantization 16→8→16 (bit)
3	Low-pass filter (11 kHz)
4	AWGN (-20 dB)
5	AWGN (-10 dB)
6	MP3 compression (128 kbps)
7	MP3 compression (64 kbps)
8	MP3 compression (32 kbps)

A. Robustness Test

To verify the reliability, the bit error rates of the FE and DWT plus LPCC approaches are computed. Table II and Table III summarize the results of the simulation for song signal and music signal, respectively. The simulation indicates that the proposed algorithm is highly robust for audio signal, and can resist various attacks by signal processes. Moreover, the proposed method outperforms that of [9], which is based on DWT plus LPCC. The right-hand columns in Table II and III show watermarks extracted by the FE method with the highest BER. These extracted watermarks indicate that the robustness of the proposed method even for high BER. As shown in Fig. 5, the difference between the bit error rate of FE and that of DWT plus LPCC indicates that the proposed method is more reliable than that of [9] under every attack. Moreover, the FE method has a much lower bit error rate under the MP3 compression attack at with a bit rate of 32 kbps. The corresponding values are 12.39% and 8.702% for the song signal and the music signal, respectively. Fig. 6 and Fig. 7 plot the standard deviation of BER for the audio signal, revealing that the result indicates the proposed algorithm has a smaller standard deviation than that of the DWT plus LPCC, so the average BER obtained is closer to represent that by various audio signal. Restated, the FE approach is more accurate than that of [9].

TABLE II: THE AVERAGE OF BER FOR SONG SIGNAL BY VARIANT ATTACK

Attack	FE (Proposed)	DWT+LPCC	Extracted Watermark from FE (highest BER)
Re-sampling 8-4-8 (kHz)	0.208%	0.228%	 0.58%
Re-quantization 16-8-16 (bit)	0.068%	1.572%	 0.13%
LPF 11 kHz	0.200%	3.172%	 0.45%
AWGN -20 dB	0.054%	1.384%	 0.09%
AWGN -10 dB	0.434%	6.830%	 0.46%
MP3 128 kbps	0.198%	3.722%	 0.26%
MP3 64 kbps	1.694%	7.818%	 1.87%
MP3 32 kbps	5.282%	17.672%	 6.48%

TABLE III: THE AVERAGE OF BER FOR MUSIC SIGNAL BY VARIANT ATTACK

Attack	FE (Proposed)	DWT+LPCC	Extracted Watermark from FE (highest BER)
Re-sampling 8-4-8 (kHz)	0.196	0.342	 0.31%
Re-quantization 16-8-16 (bit)	0.14	0.726	 0.3%
LPF 11 kHz	0.378	3.374	 0.62%
AWGN -20 dB	0.06	0.636	 0.09%
AWGN -10 dB	0.882	3.812	 1.69%
MP3 128 kbps	0.436	2.738	 0.76%
MP3 64 kbps	2.858	6.482	 4.65%
MP3 32 kbps	12.086	20.788	 17.39%

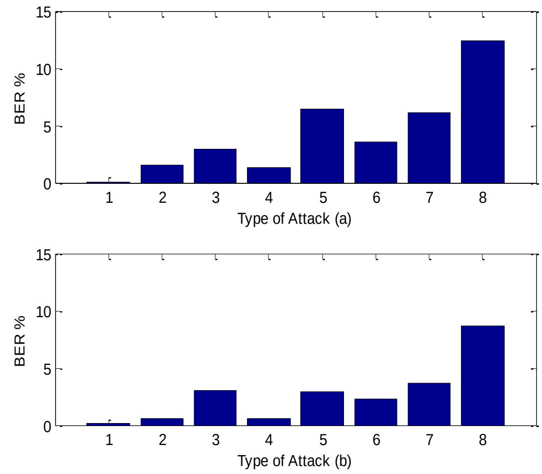

Fig. 5. The difference of bit error rate between FE and DWT+LPCC (a) song signal (b) music signal.

B. Security Test

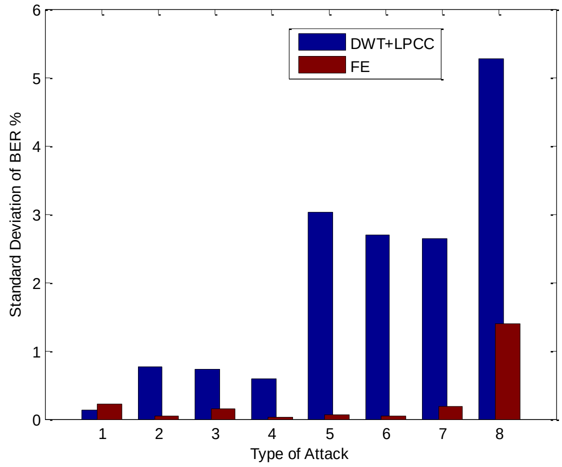

Fig. 6. The standard deviation of BER for song signal.

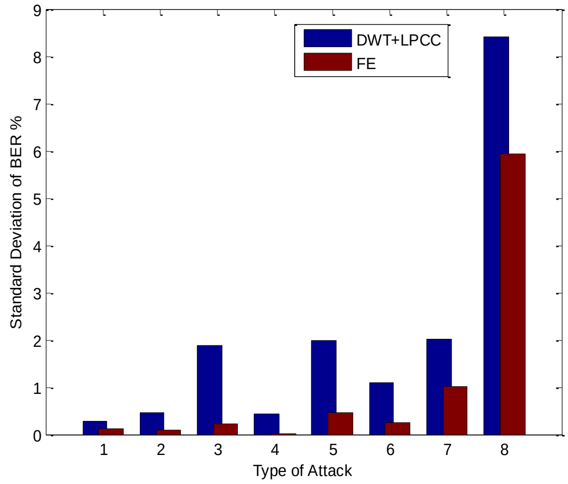

Fig. 7. The standard deviation of BER for music signal.

To evaluate the security of the proposed scheme, the watermark is retrieved from different audio by attack. Normalized cross-correlation (NC) is applied to evaluate the similarity between the extracted and the original watermarks, and is given in Table IV. When the original signal and the attacked signal are the same audio, the extracted watermark is correct; otherwise the wrong watermark has been obtained. In the detection process, the proposed scheme avoids false watermark extracted and it provides security.

TABLE IV: EXTRACTED WATERMARK WITH CORRECT AUDIO AND INCORRECT AUDIO IN DETECTING PROCESS

Original Audio	Attacked Audio	Extracted Watermark	NC
Audio A	Audio A		99.93%
Audio A	Audio B		19.56%

V. CONCLUSION

In this paper, we have presented a new algorithm in constructing and detecting processes for audio zero-watermark. The characteristic parameter of the audio signal is evaluated from the framed energy (FE). In the proposed scheme, the watermark is not embedded into the host audio, so the audio quality is maintained; the method protects intellectual property rights and achieves imperceptibility. Experimental results herein demonstrate that the proposed method provides the lowest BER compared with previous research in the literature, for various types of audio signal under various attacks. Furthermore, the presented method does not extract an inaccurate watermark retrieved in the security test. Simulation results demonstrate that the proposed zero-watermark method, based on the FE scheme, performs excellently in terms of security and robustness.

REFERENCES

[1] Q. Wen, T. F. Sun, and S. X. Wang, "Concept and application of zero-watermark," *Acta Electronica Sinica*, vol. 31, no. 2, pp. 214-216, Feb. 2003.

[2] X. Li and G. He, "A new audio zero-watermark algorithm for copyright protection based on audio segmentation and wavelet coefficients mapping," in *Proc. International Conference on Digital Content, Multimedia Technology and its Applications (IDCTA)*, pp. 22-26, 2011.

[3] X. Zhong, X. Tang, and H. Yuc, "Zero-watermark scheme based on audio's statistical character," in *Proc. International Symposium on Microwave, Antenna, Propagation, and EMC Technologies for Wireless Communications*, pp. 1227-1230, 2007.

[4] J. Li and R. Wang, "Audio aggregation zero-watermark algorithm based on (k, n/k=n)," in *Proc. International Conference on Signal Processing Systems (ICSPS)*, vol. 3, pp. 644-648, 2010.

[5] R. W. Ciptasari, F. A. Yulianto, A. Fajar， and K. Sakurai, "An efficient key generation method in audio zero-Watermarking," in *Proc. International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, pp. 336-339, 2011.

[6] H. L. Dai and D. He, "An efficient and robust zero-watermarking scheme for audio based on DWT and DCT," in *Proc. Asia Pacific Conference on IEEE Microelectronics & Electronics*, pp. 233-236, 2009.

[7] Y. Xiong and R. Wang, "An audio zero-watermark algorithm combined DCT with zernike moments," in *Proc. International Conference on Cyberworlds*, pp. 11-15, 2008.

[8] G. Chen and Q. Hu, "Zero-watermark of audio based on multi-scale entropy vectors," in *Proc. IEEE International Conference on Computer Science and Information Technology (ICCSIT)*, pp. 131-135, 2010.

[9] Y. C. Lu, "Cepstral coefficients-based zero-watermark scheme for digital audio," M.S. Thesis, Institute of Communication Engineering, Tatung University, Taipei, Taiwan, 2008.

[10] M. L. Wang, H. X. Lin, and M. T. Lee, "Robust audio watermarking based on MDCT coefficients," in *Proc. the Sixth International Conference on Genetic and Evolutionary Computing*, pp. 372-375, 2012.
[11] S. M. Tsai, "A robust zero-watermarking algorithm for audio based on LPCC," in *Proc. IEEE International Conference on Orange Technologies (ICOT)*, pp. 63-66, 2013.
[12] S. M. Tsai, "An efficient and robust zero-watermarking scheme for digital audio," in *Proc. IEEE International Conference on Circuits and Systems (ICCAS)*, pp. 51-54, 2013.

Shu-Min Tsai is now an associate professor in the department of Communication Engineering of National Penghu University of Science and Technology, Taiwan. She received the B. S. degree from the National Taiwan Ocean University, Keelung, Taiwan, in 1992, and the M. S. degree in1994, respectively, and the Ph. D. degree from the National Cheng Kung University, Tainan, Taiwan, in 2007, all in Electrical Engineering Department. Her research interests include audio signal compression and digital signal processing. She is a member of IEEE.