An Efficient and Robust Zero-Watermarking Scheme for Audio Based on DWT and DCT

Hua-liang Dai and Di He
Department of Electronic Engineering
Shanghai Jiao Tong University
Shanghai, 200240, P.R.China
Email: {dagun,dihe}@sjtu.edu.cn

***Abstract*— A zero-watermarking scheme for audio based on the steady sign of certain DWT-DCT coefficients with maximal absolute value, is proposed. Compared with traditional watermarking schemes, the proposed scheme can solve the contradiction between robustness and imperceptibility perfectly without introducing audio quality degradation as occurred usually in traditional watermarking schemes. The multiresolution characteristic of discrete wavelet transform (DWT), the energy compression characteristic of discrete cosine transform (DCT) and the steady sign of certain DWT-DCT coefficients are combined to extract important features from the host audio signal, which are used to construct the secret key with the watermark image. Simulation results show that the proposed scheme has stronger robustness as well as security, as compared to the schemes in [5] and [8].**

I. Introduction

With the fast development and application of the Internet, digital watermarking technique is now playing a more and more important role in copyright protection and authentication of digital media [1]. So far, many watermarking schemes have been proposed, however, these traditional schemes may distort the host signal to some extent and can not achieve the balance between robustness and imperceptibility perfectly. Then, the technique of zero-watermark [2] was presented to solve the contradiction between robustness and imperceptibility.

Recently, some zero-watermarking algorithms for audio have been proposed [3] [4] [5] [6] [7]. In [3], a part of samples and low-frequency discrete cosine transform (DCT) coefficients of the same volume are combined to perform low-order Zernike transform, and then the watermark is constructed based on Zernike moments. In [4], the watermark is constructed in lifting wavelet domain based on chaotic modulation. In [5], the audio's statistical character is used to construct the watermark sequence. In [6], vector quantization of linear prediction coefficients is applied to construct the secret key with the watermark image. And in [7], the characteristics of DCT, discrete wavelet transform (DWT) and higher-order cumulant are combined to construct the secret key with the watermark image.

To make full use of the auditory masking effect, the multiresolution characteristic of DWT and the energy compression characteristic of DCT [8] [9], a new zero-watermarking scheme based on the steady sign of certain DWT-DCT [1] coefficients with maximal absolute value, is proposed in this paper. The essential features are firstly extracted from the host audio signal and then an exclusive or (XOR) operation is used between the extracted features and the original watermark to construct the secret key, which is used for watermark recovery in the detecting scheme. Simulation results demonstrate the efficiency of the proposed algorithm.

The paper proceeds as follows: Section II introduces the proposed embedding algorithm detailedly. Section III presents the watermark detecting scheme. In Section IV, robustness analysis is given. The simulation results are shown in Section V. Finally, conclusion will be given in Section VI.

II. Watermark Embedding Scheme

Ref. [10] suggests that the watermark should be embedded in the perceptually significant components of the original audio. Consequently, the host audio signal is firstly segmented into segments, and then the segments with high energy [7] are selected in the proposed scheme. For each selected segment, H -level wavelet decomposition is applied to get the coarse signal, which is the perceptually significant part of the host audio. After that the coarse signal is cut into frames, on which DCT is performed to get the DWT-DCT coefficient which has the maximal absolute value. Finally, the binary pattern is generated based on the sign of the selected DWT-DCT coefficients and XOR operation is used between the binary pattern and the original watermark image to construct the secret key as used for watermark recovery. The block diagram of embedding process of the proposed scheme is shown in Fig.1.

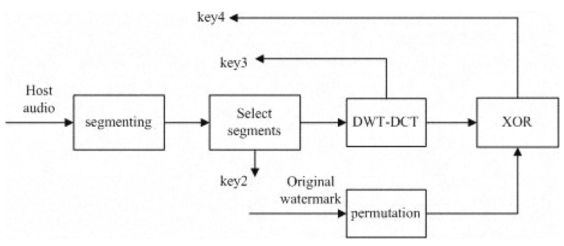

Fig. 1. Block Diagram of Watermark Embedding Scheme

[1]DCT performed on DWT coefficients

A. Preprocessing

Firstly, an image$W = \{w(i,j)\}(1 \le i \le M1, 1 \le j \le M2)$, with size $M1 \times M2$ is chosen. It needs to be converted to one-dimension signal and the converted watermark image U can be expressed as:

$$U = \{u(k) = w(i,j), 1 \le i \le M1, \\ 1 \le j \le M2, k = (i-1) \cdot M2 + j\} \quad (1)$$

that is, the pixel $w(i,j)$ becomes $u(k)$ in U.

Secondly, generate a random permutation which is used as a key (denoted as $key1$) in the watermark detecting process. Then, U is scrambled into $V = \{v(k), 1 \le k \le M1 \cdot M2\}$. In this way, we can decrease the pixel space relationship of the binary watermark image, and improve the security of the system.

B. Embedding Process

Let $\boldsymbol{S} = \{s(i)|i = 0, 1, \ldots, L_s\}$ be the host audio signal and be segmented into l segments, with q samples in each segment. The embedding steps are described as follows:

1)The energy value of each segment is calculated and all the segments are rearranged in order of decreasing energy value. The first $M2$ segments are selected to embed the watermark. The indices of the selected segments is saved as a security key, that is,

$$key2 = \{i(k)|i(k) \in (0, 1, \cdots l-1) \\ k = 0, 1, \cdots M2 - 1\}. \quad (2)$$

2)$H-$level DWT is performed on each selected audio segment, thus we can get the coarse signal $A^H_{i(k)}$ and detailed signal $D^H_{i(k)}, D^{H-1}_{i(k)}, \cdots, D^1_{i(k)}$.To take advantage of low frequency coefficient which has a higher energy value and robust against various signal processing manipulations [8], $A^H_{i(k)}$ is firstly cut into $M1$ frames, denoted as $F^{i(k)} = \{f^{i(k)}_j, j = 0, 1, \cdots, M1-1\}$, with m samples in each frame, and then DCT is performed on each frame to get DWT-DCT coefficients. That is,

$$f^{i(k)c}_j = DCT(f^{i(k)}_j) \\ = \{f^{i(k)c}_j(n)|n = 0, 1, \cdots, m-1\}, \quad (3)$$

where $j = 0, 1, \cdots, M1 - 1$.

Let

$$T = \{t(k,j)|t(k,j) = f^{i(k)c}_j(u), \\ u \in (0, 1, \cdots m-1)\}, \quad (4)$$

where $k = 0, 1, \cdots, M2-1$, $j = 0, 1, \cdots, M1-1$, and $|f^{i(k)c}_j(u)| = max(|f^{i(k)c}_j(0)|, |f^{i(k)c}_j(1)|, \cdots, |f^{i(k)c}_j(m-1)|)$.

Let

$$key3 = \{z(k,j)|k = 0, 1, \cdots M2 - 1, \\ j = 0, 1, \cdots M1 - 1\}, \quad (5)$$

where $z(k,j)$ denotes the index of $f^{i(k)c}_j(u)$ in $f^{i(k)c}_j$ and $f^{i(k)c}_j(u) \in T$.

3) A binary pattern, denoted as C

$$C = \{c(i)|i = 0, 1, \cdots, M1 \times M2 - 1\}, \quad (6)$$

is generated. Before binary pattern generation, T should be converted to one-dimensional signal G, $G = \{g(i)|i = 0, 1, \cdots, M1 \times M2 - 1\}$.

Thus C can be obtained as follows:

$$c(i) = \begin{cases} 1, if\ g(i) > 0 \\ 0, otherwise. \end{cases} \quad (7)$$

4) The secret key, denoted as $key4$, is generated as follows:

$$key4 = C \oplus V, \quad (8)$$

where $\oplus$ is the XOR operation.

In the proposed scheme, L_s should not be smaller than $m \cdot M1 \cdot 2^H \cdot M2$. Besides, H should be chosen properly to balance the robustness of the watermark and the calculation cost.

III. Watermark Detecting Scheme

The watermark can be recovered blindly and Fig.2 shows the block diagram of the watermark detecting scheme.

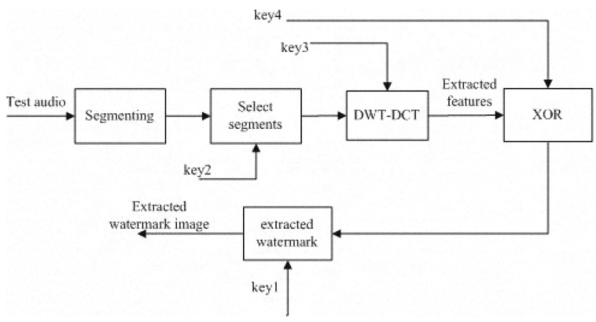

Fig. 2. Block Diagram of Watermark Detecting Scheme

The detailed steps are described as follows:

1)The test audio $\tilde{\boldsymbol{S}} = \{\tilde{s}(i)|i = 0, 1, \ldots, L_s\}$ is segmented into l segments, with q samples in each segment.

2) Use $key2$ to choose the segments, and then $H-$level DWT is performed on each selected segment to get the coarse signal $\tilde{A}^H_{i(k)}$, $(i(k) \in \{0, 1, \cdots, l-1\}, k = 0, 1, \cdots, M2-1)$.

3) $\tilde{A}^H_{i(k)}$ is cut into $M1$ frames, and DCT is performed on each frame. Then, the same procedure in Section Embedding Process is performed to get the estimated binary pattern $\tilde{C}$.

4) XOR operation is used between $\tilde{C}$ and $key4$, that is,

$$\tilde{V} = key4 \oplus \tilde{C}, \quad (9)$$

where $\tilde{V}$ is the estimated watermark, and then $key1$ is used to get the extracted watermark image.

IV. Robustness Analysis

The proposed scheme makes full use of the auditory masking effect and extracts the essential features from DWT and DCT domain of the audio segments with high energy, which possesses strong ability of resisting common attacks according to the characteristics of human auditory system (HAS) [11].

Besides, observing DCT formula:

$$F(k) = c(k) \sum_{n=0}^{N-1} f(n) \cos \frac{(2n+1)k\pi}{2N}, \qquad (10)$$

where f is the time sequence, F is the DCT-coefficient with $0 \le n, k \le N$ and $c(m) = \begin{cases} \frac{1}{\sqrt{N}}, m = 0 \\ \frac{2}{\sqrt{N}}, m \neq 0. \end{cases}$

Assuming

$$|F(m)| = max(|F(0)|, |F(1)|, \cdots, |F(N-1)|), \qquad (11)$$

then it is easily seen that, compared to the sign of $F(m)$, the signs of $F(k), k \neq m$, are more easily to be changed, implying that the strength of $F(m)$ against attack is stronger than that of all the other $F(k)s$. Consequently, extracting features according to the sign of $F(m)$ can obtain better robustness.

V. Computer Simulations and Analysis

In the computer simulations, a piece of 20-second mono music and a piece of 20-second mono speech, with 16bits signed and sampled at 44.1 KHz are used as the host audio signals. A 50×100 bits binary image is taken as the watermark for the host audios. In the simulations, we set $H = 3, m = 12$ and the Haar wavelet basis is used.

Moreover, the signal-to-noise ration (SNR) defined in (12) is used to measure the imperceptibility of the proposed scheme:

$$SNR(\mathbf{s}, \tilde{\mathbf{s}}) = 10 \log_{10} \frac{\sum_{i=0}^{L-1} s^2(i)}{\sum_{i=0}^{L-1} [s(i) - \tilde{s}(i)]^2}, \qquad (12)$$

where L is the length of the audio signal, $\mathbf{s}$ and $\tilde{\mathbf{s}}$ are the original audio signal and watermarked audio signal, respectively.The normalized cross-correlation (NC) defined in (13) is used to evaluate the similarity between the extracted watermark and the original one:

$NC(\mathbf{w}, \bar{\mathbf{w}}) =$

$$\frac{\sum_{i=1}^{M1} \sum_{j=1}^{M2} w(i,j) \cdot \bar{w}(i,j)}{\sqrt{\sum_{i=1}^{M1} \sum_{j=1}^{M2} w^2(i,j)} \cdot \sqrt{\sum_{i=1}^{M1} \sum_{j=1}^{M2} \bar{w}^2(i,j)}}, \qquad (13)$$

where $\mathbf{w}$ and $\bar{\mathbf{w}}$ are the original watermark and the detected watermark, respectively. And the bit error rate (BER) defined in (14) is used to test the reliability of the proposed scheme:

$$BER = \frac{D}{N} \times 100\% \qquad (14)$$

where D denotes the number of error bits, N means the number of total bits.

A. Imperceptibility Test

In the proposed scheme, the watermark is not embedded into the host audio in fact, so the host audio will not be distorted and can achieve the imperceptibility of the scheme naturally, that is, the SNR of the proposed scheme is infinite.

B. Security Test

To test the security of the scheme, try to extract the watermark from the watermarked audio with wrong keys, and extract watermark from non-watermarked audio signals with the right keys ($key1$, $key2$,$key3$ and $key4$ in Fig.2). In addition to the watermarked music/speech, another 30 pieces of mono music signals and 30 pieces of mono speech signals are used in the simulations. The simulation results are shown in Fig.3 and Fig.4, respectively. The peaks in Fig.4 correspond to the watermarked music/speech. So the keys used in the detecting scheme can avoid false watermark detecting and ensure the security of the proposed scheme.

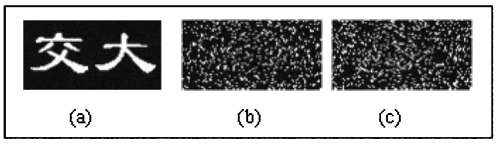

Fig. 3. Watermark Images(a)Original Watermark Image;(b) With Wrong keys (music); (c)With Wrong keys (speech).

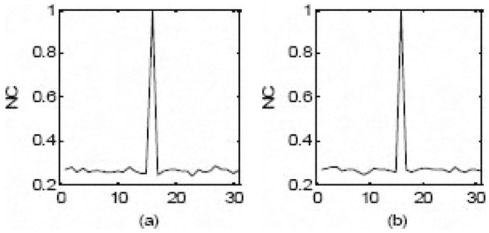

Fig. 4. Security test results of 31 signals (a) music; (b) speech.

C. Robustness Test

The robustness of the proposed scheme is tested by "Stirmark for Audio v02" and Cool Edit. To verify the robustness of the proposed scheme, the proposed scheme is compared with the schemes in [5] and [8]. The simulation results (NC) are summarized in Table 1. The BER of the proposed scheme are shown in Fig.5. The simulation results prove that the proposed scheme has strong robustness, and can resist the common attacks.

Besides, the performances of the proposed scheme are better than those of [8], which is also based on DWT and DCT. In [8], the quantization step should be controlled to balance the robustness and imperceptibility, while the proposed scheme can achieve it naturally.

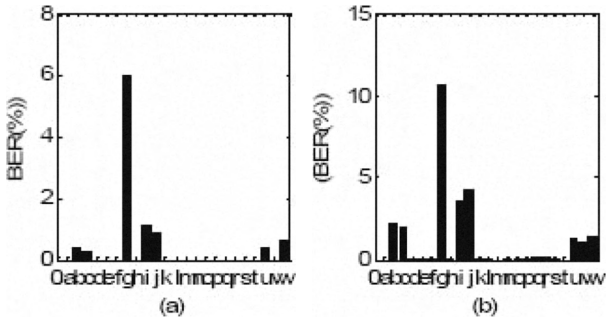

Fig. 5. BER of the proposed scheme; (a) music (b) speech. The description of (a)-(w)are given in Table 1.

Table 1. Simulation Results

attack	NC(proposed)		NC([5])		NC([8])	
	music	speech	music	speech	music	speech
(a) no attack	1	1	1	1	1	1
(b) addnoise_900	0.9856	0.9248	0.5054	0.3316	0.5721	0.5762
(c) addsinus	0.9913	0.9348	0.6219	0.4851	0.7849	0.7862
(d) amplify	1	0.9993	0.9993	0.9857	0.2696	0.245
(e) compressor	1	0.9993	1	0.9985	1	0.9993
(f) normalize	1	0.9993	1	0.9978	0.1911	0.2391
(g) rc_highpass	0.819	0.7188	0.52	0.5144	0.3587	0.2897
(h) rc_lowpass	1	0.9993	0.9705	0.9364	0.8168	0.4606
(i)addbrumm_1100	0.9627	0.8869	0.4502	0.3891	0.8922	0.8922
(j) zerocross	0.969	0.8669	0.7458	0.6745	0.2631	0.2629
(k) flippsample	1	0.9993	0.9150	0.9347	0.2266	0.2846
(l) stat1	1	0.9985	0.9446	0.9331	0.7677	0.3763
(m) stat2	1	1	0.9956	0.9828	0.9971	0.9935
(n) extrastereo_70	1	0.9993	1	0.9993	1	1
(o) fft_real_reverse	1	0.9985	0.9828	0.7767	0.9964	0.9856
(p) exchange	1	0.9993	0.9806	0.9596	1	1
(q) smooth	1	0.9935	0.9753	0.9473	0.9667	0.5388
(r) smooth2	1	0.9957	0.8672	0.8205	0.878	0.489
(s) dynnoise	1	0.9993	0.7202	0.7842	0.6125	0.3051
(t) lsbzero	1	1	0.9964	0.9549	1	1
(u) Echo(100ms, 20%)	0.9849	0.956	0.9213	0.6586	0.5735	0.3084
(v)Re-quantization (16→ 8→ 16 bit)	1	0.9621	0.9677	0.8099	0.9928	0.9836
(w) Delay(200ms,15%)	0.9779	0.9514	0.7693	0.5801	0.4174	0.2953

VI. Conclusion

In this paper, an efficient and robust zero-watermarking scheme for audio is proposed. Compared with the traditional audio watermarking algorithms, the proposed scheme introduces no audio quality degradation and can solve the contradiction between robustness and imperceptibility perfectly. The multi-resolution characteristic of DWT ,the energy compression characteristic of DCT and the steady sign of certain DWT-DCT coefficients with maximal absolute value are combined to extract important features which are used to generate the secret key as used for watermark recovery. Simulation results show that the proposed scheme has strong imperceptibility, robustness and security.

Acknowledgment

This research is supported by The SRF for ROCS SEM of P.R. China; The National Science Foundation of P.R. China under Grant No. 60802058.

References

[1] F. S. Wei, F. Xue, and M. Y. Li, "A blind audio watermarking scheme using peak point extraction", IEEE International Symposium on Circuit and System, Kobe, Japan, pp.4409-4412, May 2005.

[2] 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.

[3] Y. G. Xiong, and R. D. Wang, "An audio zero-watermark algorithm combined DCT with Zernike moments", International Conference on Cyberworlds, Hangzhou, China, pp.11-15, Sept. 2008.

[4] R. D. Wang, and W. J. Hu, "Robust audio zero-watermark based on lwt and chaotic modulation", The 6th International Workshop on Digital Watermarking(IWDW 2007) , Guangzhou, China, pp.373-381, Dec. 2007.

[5] X. Zhong, X. H. Tang, and H. L. Yue," Zero -watermark scheme based on audio's statistical character",International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications, Hangzhou, China, pp.1227-1230, August 2007.

[6] N. Chen, and J. Zhu,"Robust speech watermarking algorithm", Electronics Letters, vol.43, no.24, pp.1393-1395, Nov. 2007.

[7] N. Chen, and J. Zhu,"A robust zero-watermarking algorithm for audio", EURASIP Journal on Advances in Signal Processing, vol.2008, doi:10.1155/2008/453580.

[8] X. Y. Wang, and H. Zhao, "A novel synchronization invariant audio watermarking scheme based on DWT and DCT", IEEE Trans. on signal processing, vol.54, no.12, pp.4835-4840, Dec. 2006.

[9] X. Y. Wang, W. Qi, and P. Niu, "A new adaptive digital audio watermarking based on support vector regression", IEEE Transactions on Audio, Speech, and Language Processing, vol.15, no.8, pp.2270-2277, Nov. 2007.

[10] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, " Secure spread spectrum watermarking for multimedia", IEEE Trans on Image processing, vol.6, no.12, pp.1673-1687, Dec. 1997.

[11] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception", Proceedings of the IEEE, vol.81, no.10, pp.1385-1422, Oct. 1993.