Int J Speech Technol (2015) 18:443–448
DOI 10.1007/s10772-015-9287-3

Audio dual watermarking scheme for copyright protection and content authentication

Jinquan Zhang[1]

Received: 29 January 2015 / Accepted: 30 May 2015 / Published online: 11 June 2015

Abstract We propose a new multipurpose audio watermarking scheme in which two watermarks are used. For intellectual property protection, audio clip is divided into frames and robust watermark is embedded. At the same time, the feature of each frame is extracted, and it is quantized as semi-fragile watermark. Then, the frame is cut into sections and the semi-fragile watermark bits are embedded into these sections. For content authentication, the semi-fragile watermark extracted from each frame is compared with the watermark generated from the same frame to judge whether the watermarked audio is tampered, and locate the tampered position. Experimental results show that our scheme is inaudibility. The two watermark schemes are all robust to common signal processing operations such as additive noise, resampling, re-quantization and low-pass filtering, and the semi-fragile watermark scheme can achieve tampered detection and location.

Keywords Multipurpose audio watermarking · Robust watermark · Copyright protection · Semi-fragile watermark · Content authentication

1 Introduction

With the rapid development of the Internet, concerns on intellectual property protection have been raised in recent years. A way to protect the copyright of audio works is to embed robust watermark into it. On the other hand, audio contents are tampered easily by many operations, such as insertion, deletion and replacement. So, embedding (semi-) fragile watermarking into an audio work to authenticate the integrity is also received widespread concern.

Many robust audio watermarking schemes were proposed in recent years. (Ma and Han 2006) set up a mathematical relationship between the coefficient where watermark was embedded and the audible quality of the audio file with watermark. (Wu et al. 2005) embedded the synchronization codes and the hidden informative data into the approximation frequency coefficients in DWT (discrete wavelet transform) domain. By exploiting the time-frequency localization characteristics of DWT, the computational load in searching synchronization codes had been dramatically reduced. (Wang et al. 2009) embedded the watermark bits into the statistics average value of approximation frequency components in DWT. The proposed scheme was inaudible and robust against common signals processing and some de-synchronization attacks. Making use of the multi-resolution of DWT and the energy compression of discrete cosine transform (DCT), (Wang and Zhao 2006) embedded the watermark into the hybrid domain. Their algorithm was robust and the impairment to watermarked audio was inaudible. (Lei et al. 2011) proposed a robust audio watermarking scheme based on SVD–DCT (singular value decomposition). The scheme embedded the watermark into the high-frequency band of the SVD–DCT block blindly.

Some (semi-)fragile watermarking schemes were proposed in these years. In literature (Gulbis et al. 2008), authors extracted the energy of the critical bands in each segment as the feature, and embedded it into each segment by modifying the coefficients in DCT domain. (Wang and Fan 2010) computed the centroid of each audio frame,

✉ Jinquan Zhang
zhjq@cuit.edu.cn

[1] College of Information Security Engineering, Chengdu University of Information Technology, Chengdu 610225, China

quantized it as watermark. They embedded watermark in DWT-DCT domain. (Lei et al. 2010) proposed a semi-fragile audio watermarking scheme. They used binary image as fragile watermark, and embedded the watermark signal into the average value of the wavelet coefficients.

In some circumstances, copyright protection and content authentication are necessary simultaneously. Many multi-purpose watermarking techniques have been proposed to achieve these goals. (Wang and Xu 2006) presented an audio watermarking scheme which embedded robust and fragile watermark at the same time in lifting wavelet. The robust watermark and the fragile watermark were the same binary image. (Chen and Zhu 2008b) proposed a scheme which embedded semi-fragile watermark into the audio signal first. In order to extract semi-fragile accurately, zero-watermark technology was used to embed robust watermark. (Lei and Soon 2012) embedded robust watermark by modifying low frequency component and fragile watermark by modifying high frequency component in DCT domain. The fragile watermark was a chaotic sequence. (Liao et al. 2009) embedded robust watermark by quantizing the difference of the sum of the odd and the even coefficients in DWT domain and fragile watermark by modifying the first level detail coefficients. The robust watermark and the fragile watermark were both binary images. (Chen and Zhu 2008a) proposed a scheme which embedded semi-fragile watermark into the audio signal. In order to extract semi-fragile accurately, zero-watermark technology was used to embed robust watermark.

These multipurpose schemes mentioned above adopted binary image or chaotic sequence as fragile watermark. It will greatly increase the false alarm probability of tamper detection and decrease the security of watermark system, as is discussed in (Wang and Fan 2010). In our scheme, we divide an audio clip into non-overlapping frames first. Then DWT–DCT is performed on each frame. Robust watermark bits are embedded by modifying the coefficient in hybrid domain. Simultaneously, we extract the features of each frame when robust watermark is embedding, quantizing them as semi-fragile watermark and embedding them into this frame. The proposed scheme achieves right protection and content authentication for audio signal at the same time.

2 Embedding and extraction of robust and semi-fragile watermark

In our scheme, the robust watermark is a random sequence $w_R \in \{-1, 1\}$, which may be a chaotic sequence, or generated by secure hash algorithm, such as SHA-3. The features of each frame are obtained when robust watermark bits are embedded.

2.1 Robust watermark embedding process

After the original audio is divided, each frame is embedded into watermark bits as follow. The following watermark embedding rule hides several bits of the watermark in each frame. The embedding process contains the following steps.

Step 1. Let $X = (x_1, x_2, \ldots, x_{N_R})$ represents a frame audio signal with N_R samples.

Step 2. d-level DWT is applied on X, then DCT is performed on the obtained coarse signal, and we get the DCT coefficients $C = (c_1, c_2, \ldots)$. After computing the absolute value of each coefficient in C, we obtain $C' = (|c_1|, |c_2|, \ldots)$. Then sorting C' in descend order and we get $T = (|t_1|, |t_2|, \ldots)$. The first n coefficients with original sign t_1,t_2,…,t_n are selected to embed watermark. For the coefficient t_1, the watermark bit is embedded as follows (Chen and Wornell 2001)

$$t_1' = \begin{cases} round\ (t_1/S_R) \times S_R, & if\ w_R = 1 \\ floor\ (t_1/S_R) \times S_R + 0.5S_R, & if\ w_R = -1 \end{cases} \quad (1)$$

where $S_R > 0$ denotes the embedding strength.

Step 3. For the second coefficient t_2, after embedding the next watermark bit by performing on Eq. (1), we obtain t_2'. if $|t_1'| < |t_2'|$, then $|t_1'| = |t_1'| + S_R$.

Step 4. Continuing to embed watermark bit following step 3.

At last, the n coefficients are satisfied with the relationship $|t_1'| \geq |t_2'| \geq \ldots \geq |t_n'|$.

After t_n' is gotten, there may be some un-watermarked coefficients that are larger than t_n' in absolute value. We must decrease these un-watermarked coefficients in value to assure watermark can be extracted correctly. In order to improve the robustness of the scheme, it is necessary to slightly decrease these un-watermarked coefficients whose absolute value are adjacent to $|t_n'|$.

Step 5. Inverse DCT and Inverse DWT are performed on the modified coefficients. Then, the watermarked audio is obtained.

Step 6. We quantize these coefficients t_1',t_2',…,t_n' as semi-fragile watermark as described in Sect. 2.2.

The remaining watermark bits are embedded in the same way in other frames.

After all watermark bits are embedded, in order to improve the robustness against various attacks, the watermark bits are embedded repeatedly in the remaining audio frames.

2.2 Generation of semi-fragile watermark

As mentioned in Sect. 2.1 step 6, the semi-fragile watermark is generated when robust watermark is embedded.

For a certain coefficient $t_i^{'}$ in hybrid domain, we compute $m_i = \lfloor (|t_i^{'}| + 0.5Q_F)/Q_F \rfloor$, where $\lfloor \cdot \rfloor$ returns the largest integer less than the original value. Then m_i is converted into binary sequence b_i, and the lowest l bits is selected as a feature for this frame,denoted as $b_i^{'}$. If the length of b_i is shorter than l, 0 is padded in the head of b_i until the length is l. It looks like $0 \ldots 0 || b_i$. We concatenate $b_i^{'}$ and get $w = b_1^{'} || b_2^{'} || \ldots || b_n^{'}$, n is the number of hybrid domain coefficients which are chosen to embed watermark bits, as described in Sect. 2.1 step 2. That is, the length of semi-fragile watermark for each frame is nl bits.

Then, a secure, open stream cipher algorithm, such as ZUC, is chosen to encrypt w. Assume the key stream is K, the encryption rule is as follows:

$$W_F = w \oplus K \tag{2}$$

where $\oplus$ means XOR operation.

2.3 Embedding the semi-fragile watermark

The embedding process of semi-fragile watermark is described as follows.

The watermarked frame X is divided into non-overlapping sections, denoted as $X_i \mid i = 1,2,\ldots, nl$. That is to say, the length of a section is N_R/nl. Then DCT is performed on each audio section, that is, $D_i = \mathrm{DCT}(X_i) \mid i = 1, \ldots, nl$, where D_i is the DCT coefficient vector of the i-th section.

In each section, only one watermark bit is embedded. The similar rule, as robust watermark is embedded, is adopted to embed W_F by modifying a certain coefficient of D_i. Usually, the 2-th or 3-th coefficient is chosen to embed watermark bit. The embedding strength is S_F.

2.4 Extracting the robust watermark

The extracting process of the robust watermark contains the following steps.

Step 1. Assume the corresponding audio frame $X = (x_1, x_2, \ldots, x_{N_R})$.

Step 2. The same procedure is performed as in the watermark embedding process. At last, the n coefficients $t_1^*, t_2^*, \ldots, t_n^*$ are selected to detect watermark. For the coefficient t_i^*, the watermark is extracted as follows:

$$w_i^* = \begin{cases} -1, & S_R/4 \le \mathrm{mod}\ (t_i^*, S_R) \le 3S_R/4 \\ 1, & \text{otherwise} \end{cases} \tag{3}$$

Extract watermark bit repeatedly from the remaining audio. The final watermark w^* can be obtained from the extracted watermarks according to the majority rule.

The semi-fragile watermark is also generated when robust watermark is extracted. For the coefficient t_i^* in hybrid domain, compute

$$t_i^{*\prime} = \begin{cases} \lfloor (t_i^* + S_R/2)/S_R \rfloor * S_R + S_R/2, & S_R/4 \le \mathrm{mod}(t_i^*, S_R) \le 3S_R/4 \\ \lfloor (t_i^* + S_R/2)/S_R \rfloor * S_R, & \text{otherwise} \end{cases},$$

then compute $m_i^{\prime} = (|t_i^{*\prime}| + 0.5Q_F)/Q_F$. $m_i^{'}$ is converted into binary sequence. The remainder step is as similar as Sect. 2.2, and we get the $W_F^{'}$.

2.5 Extraction of the semi-fragile watermark

Since the embedding rule of semi-fragile is similar to the robust watermark, the extraction rule is also similar to it. In the extraction procedure, the embedding strength is S_F, and $W_F^{*\prime}$ is obtained.

2.6 Locating the tampered position

In order to judge whether the watermarked audio is tampered, and locate the tampered position, we will compare $W_F^{'}$, the semi-fragile watermark generated from watermarked frame, with $W_F^{*\prime}$, the extracted watermark from the watermarked audio.

To an audio signal, the amplitude of samples will change after signal process operation such as resampling, re-quantization, low-pass filtering, etc., which will influence the extraction of watermark. Define the authentication sequence as follows:

$$e(i) = W_F^{\prime}(i) \oplus W_F^{*^{'}}(i), \quad i \in [1, nl] \tag{4}$$

For a certain i, if $e(i) = 1$, it means samples in this section which watermark bit is embedded into in the watermarked audio is changed to a certain extent due to an certain signal process operation.

For each frame, Define T_P as follows:

$$T_P = \begin{cases} 1, & T_F < \sum_{i=1}^{nl} e(i) \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where T_F is a threshold. That is to say, if more than T_F sections are modified, $T_P = 1$, the algorithm judges the audio frame has been tampered. Otherwise, $T_P = 0$, it represents the audio frame has not been tampered.

3 Experiments and analysis

We test our algorithm on different audio clips including pop, light, march, piano, jazz and rock with different

lengths. The experimental results are similar for all audio files tested. We report the results with three audio clips that are the pop music clip, light music clip, and dance music clip. They are in WAV format, mono, 16 bits/sample, 15 s, and 44.1 kHz sampling frequency. The length of robust watermark is 64 bits.

In experiments, $N_R = 4096$ and $S_R = 0.2$. Db4 wavelet basis is adopted and 2-level DWT is performed. In each audio frame, the first 4 largest coefficients are chosen to embed robust watermark bits. During the generation of semi-fragile, $Q_F = 0.01$, $l = 8$. We select the 3-th coefficient in DCT domain to embed semi-fragile watermark. $S_F = 0.03$. All these parameters are chosen to achieve a good compromise between the conflicting requirements of imperceptibility and robustness.

3.1 Inaudibility tests

In order to evaluate the inaudibility of the proposed scheme comprehensively, signal-to-noise ratio (SNR) and perceptual evaluation of audio quality (PEAQ) are both used in objective evaluation tests, and Mean Opinion Score (MOS) is used in the subjective listening test.

In audio watermarking, the SNR is a difference indicator between the watermarked and the original audio. The definition of SNR is shown as follows:

$$SNR = 10\lg\frac{\sum_{i=1}^{l} x_i^2}{\sum_{i=1}^{l}(x_i - x_i^{'})^2} \tag{6}$$

where x_i and $x_i^{'}$ are the original and watermarked audio signal respectively.

As can be seen from Eq. (6), for a given audio signal, the sum of $E_C = \sum_{i=1}^{l}(x_i - x_i^{'})^2$ is the only factor influencing the SNR. Assume the sum E_{CR} and E_{CF} for the embedding of robust and semi-fragile watermark respectively. In Table 1, the effect of robust and semi-fragile watermark to SNR is given. Simultaneously, the SNR, ODG and MOS values are shown.

Table 1 The results of inaudibility tests

	E_{CR}	E_{CF}	SNR (dB)	ODG	MOS
Dance	2.13	0.35	37.8	−0.11	5.0
Pop	2.53	0.34	36.8	−0.24	5.0
Country	2.51	0.34	35.5	−0.06	5.0

3.2 Robustness tests for common signal process operations

The audio signal processing operations shown in Table 2 are performed on the watermarked audio signals. These operations include: additive noise (SNR = 65 dB/55 dB), resampling (44.1->22.05/ 11.025->44.1 kHz), re-quantization (16->8->16 bits) and low pass filtering(cutoff frequency 8 kHz). In experiments, we take into account the ODG value after the watermarked audio is treated. As can be seen in the table, the proposed scheme for robust watermark is robust to common signal process operations.

In tests, we set $T_F = 0.1$. For most audio clips, after common signal process operations, the results that the semi-fragile watermark extracted from each frame is compared with the watermark generated from the same frame are shown as Fig. 1(a). That is, as the common signal operations don't modify the content of audio signal, the scheme achieves the content authentication. Occasionally, a small number of error detected frame is shown as Fig. 1(b). It takes place when resampling (11,025 Hz) or additive noise (55 dB) is performed. We notice that the ODG value is less than −2 in this case. That is to say, the quality of audio decreases dramatically, and the noise is audible.

3.3 Tampering location test

In order to evaluate the ability to locate the tampered position against malicious operations, the attack that one part of audio is replaced by another part is performed on watermarked audio signal. Figure 2(a) shows the maliciously tampered watermarked audio signals. The watermarked audio samples from the 150,000th to the 180,000th were replaced by another part from 60,000th to 90,000th. Figure 2(b) shows the results of tamper location.

Table 2 Robustness of robust watermark tests

Signal processing	Dance		Pop		Country	
	ODG	BER	ODG	BER	ODG	BER
Additive noise (65 dB)	−0.91	0	−1.87	0	−0.41	0
Additive noise (55 dB)	−3.15	0	−3.31	0	−1.75	0
Resampling (22050 Hz)	−0.82	0	−1.14	0	−2.03	0
Resampling (11025 Hz)	−2.75	0	−2.87	0	−2.81	0
Re-quantization	−1.13	0	−2.04	0	−0.70	0
Low pass filtering (8 kHz)	−0.91	0	−0.59	0	−0.68	0

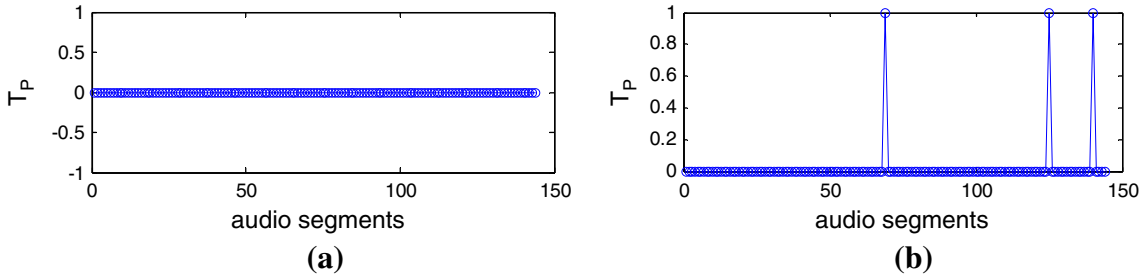

Fig. 1 Results of content authentication tests. (**a**) most cases, (**b**) rare cases

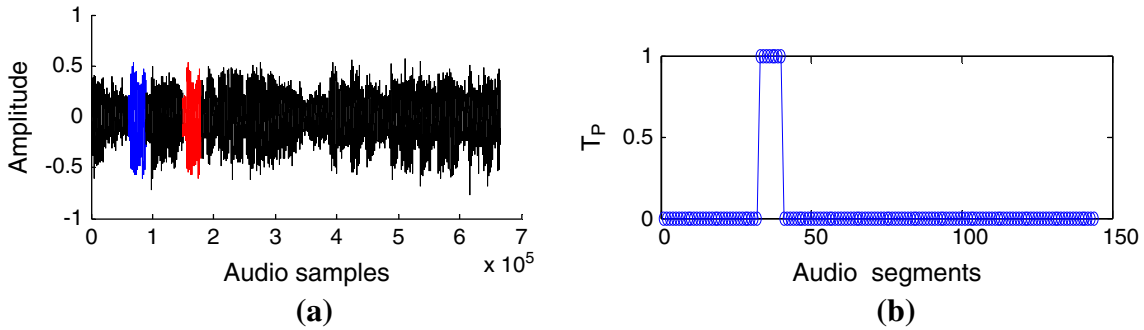

Fig. 2 Tamper Location. (**a**) Waveform performed tamper attack on a watermarked, (**b**) tamper location results of a tampered watermarked audio

According to our algorithm, the attacker may successfully tamper the watermarked audio with the probability of $1/N_R$, N_R is the length of a frame for embedding n bits robust watermark, as described in Sect. 2.1. But, even the modified watermarked audio can survive our algorithm, it won't be coherent. Usually, clear "click" may be listened.

3.4 Algorithm analysis

In our scheme, an audio clip first is partitioned into frames, and each frame is divided into sections. Robust watermark bits are embedded into the first n larger DWT–DCT coefficients of a frame, and these coefficients are quantized. As a result, the binary sequence is semi-fragile watermark bits. The semi-fragile watermark is embedded into DCT domain of each section.

When the semi-fragile watermark is embedded, the 2nd or 3rd coefficient of DCT domain in each section is modified, which will introduce noise into the audio clip which robust watermark bit has been embedded into. As previously described, the length of a frame is N_R, and the length of a section is N_R/nl. In our experiments, $N_R = 4096$, $n = 4$, and $l = 8$, if the 2-th coefficient of DCT domain in each section is modified, according to literature (Zhang and Wang 2013), it means that after DCT is performed on a frame with the length of N_R, the value of the nl-th (here is 32th) coefficient will be modified for embedding the semi-fragile watermark. As for robust watermark, in experiments, we notice that the chosen n coefficients usually distribution among (Lei and Soon 2012; Lei et al. 2010, 2011; Ma and Han 2006; Wang and Fan 2010; Wang et al. 2009; Wang and Xu 2006; Wang and Zhao 2006; Wu et al. 2005; Zhang and Wang 2013). That is, the change of samples due to embedding semi-fragile watermark has no influence on the extraction of the robust watermark.

On the other hand, when robust watermark are extracted from the watermarked audio, we compute the semi-fragile watermark. Before the DWT–DCT coefficients are quantified, we modulate these coefficients first, as described in Sect. 2.4. It deduces the false alarm rate dramatically.

4 Conclusions and future work

A multipurpose audio watermark scheme is proposed in this paper. After common signal processing operations, robust watermark can be extracted correctly, and semi-fragile one can also be detected. When the watermarked audio is modified, the scheme can detect it and accurately locate the tampered position.

There are several works to further the research introduced in this paper. One of the future works is to find a good synchronization code algorithm. This is also an open problem in the industry. With the help of good

synchronization code algorithm, our scheme can locate the position where watermarked audio is deleted or added some samples. Secondly, the embedding strength in our scheme has nothing to do with the audio content. New schemes are necessary to be presented to solve this problem. Finally, our scheme could be implemented in real scenarios to test the performance for copyright protection and tamper location.

Acknowledgments This work is supported by the Scientific Research Foundation of CUIT under Grant No. KYTZ201420.

References

Chen, B., & Wornell, G. W. (2001). Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. *IEEE Transactions on Information Theory, 47*(4), 1423–1443.

Chen, N., & Zhu, J. (2008a). Multipurpose audio watermarking algorithm. *Journal of Zhejiang University -Science A., 9*(4), 517–523.

Chen, N., & Zhu, J. (2008b). Multipurpose audio watermarking algorithm. *Journal of Zhejiang University Science A., 9*(4), 517–523.

Gulbis, M., Muller, E., & Steinebach, M. (2008). Content-based authentication watermarking with improved audio content feature extraction. *IEEE IIHMSP*, pp 620–623.

Lei, B., & Soon, I. Y. A. (2012). Multipurpose audio watermarking algorithm with synchronization and encryption. *Journal of Zhejiang University-Science C, 13*(1), 11–19.

Lei, B., Soon, I. Y., & Li, Z. (2011). Blind and robust audio watermarking scheme based on SVD–DCT. *Signal Processing, 91*(8), 1973–1984.

Lei, M., Yang, Y., Luo, S., et al. (2010). Semi-fragile audio watermarking algorithm in dwt domain. *China Communications, 7*(4), 71–75.

Liao, W., Zhang, Y., Li, D., & Zhang, M. (2009). Audio fragile-robust dual watermarking scheme based on digital wavelet transform[J]. *Journal of Zhejiang University, 43*(4), 721–726.

Ma, Y., & Han, J. (2006). Audio watermarking in DCT: Embedding strategy and algorithm. *Acta Electronica Sinica, 34*(7), 1260–1264.

Wang, H., & Fan, M. (2010). Centroid-based semi-fragile audio watermarking in hybrid domain. *Science in China Series E-Information Sciences, 53*(3), 619–633.

Wang, X., Niu, P., & Yang, H. (2009). A robust digital audio watermarking based on statistics characteristics. *Pattern Recognition, 42*(11), 3057–3064.

Wang, R., & Xu, D. (2006). Multiple audio watermarks based on lifting wavelet transform[j]. *Journal of Electronics and Information Technology, 28*(10), 1820–1826.

Wang, X., & Zhao, H. (2006). A novel synchronization invariant audio watermarking scheme based on dwt and DCT. *IEEE Transactions on Signal Processing, 54*(12), 4835–4840.

Wu, S., Huang, J., Huang, D., et al. (2005). Efficiently self-synchronized audio watermarking for assured audio data transmission. *IEEE Transactions on Broadcasting, 51*(1), 69–76.

Zhang, J., & Wang, H. (2013). Analysis on law of distortion of audio signal for embedding watermark in DCT and DWT. *Acta Electronica Sinica, 41*(6), 1193–1197.