A Digital Audio Watermark Embedding Algorithm

Xianghong Tang[1], Yamei Niu[1], Hengli Yue[1], Zhongke Yin[2]

[1] School of Communication Engineering, Hangzhou Dianzi University, Hangzhou, Zhejiang, 310018, China

tangxh@hziee.edu.cn, niuyamei1120@sina.com, yuehengli283@sohu.com

[2] School of Computer & Communications Engineering, Southwest Jiaotong University, Chengdu 610031, China

zkyin@home.swjtu.edu.cn

Abstract

This paper proposed an audio digital watermarking algorithm based on the wavelet transform (WT) and the complex cepstrum transform (CCT) by combining with human auditory model and using the masking effect of human ears. This algorithm is realized to embed a binary image watermark into the audio signal and improved the imperceptibility of watermarks. Experimental results show that this algorithm has a better robustness against common signal processing such as noise, filtering, resampling and lossy compression.

Keyword: watermarking, wavelet transform, cepstrum.

I. Introduction

It’s a problem to be resolved by entire the world to prevent information of digital products rights, pirated and tampered optionally. The digital watermarking techniques are considered an effective solution to the problems of copyrights. It embedded the secret information about copyright owners using the masking effect of human visual system (HVS) and human audible system (HAS) in order to prove the copyright ascription of the information. From the view of the recent application, watermarking can be divided into two classes: image watermarking and audio watermarking. There are many researches on image watermarking but few on audio watermarking.

According to the implementation process of audio watermarking algorithms, they can be divided into the time domain methods and transform domain methods. From the view of the performance of watermarks against attacks, the performance of the transform domain methods are commonly considered better than that of the time domain methods [1].

The complex cepstral analysis is a homomorphic mapping and is the most effective extraction method in audio identification. Since the advantage of cepstral transform (CCT) is that the coefficients of the complex cepstral transform are much decorrelation. S.K.Lee and Y.S.Ho[2] discussed the audio watermarks embedding algorithm with the cepstral transform, a digital watermark was inserted into the cepstral components of the audio signal using a technique analogous to spread spectrum communications, hiding a narrow band signal in a wideband channel. In this method, a pseudo-random sequence was used to watermark the audio signal. The watermark is then weighted in the cepstrum domain according to the distribution of cepstral coefficients and the frequency masking characteristics of HAS. C.T.Hsith and P.Y.Tsou[3] discussed the detection of audio blind watermarking in cepstrum domain using the concepts of energy characteristic radix in frequency domain.

The wavelet transform is a new tool of signal processing in recent years. The principal advantage of wavelet transform is that it can decompose signals into different frequency components and analyzes signals in the time domain and frequency domain simultaneously. So the wavelet transform is used widely in many fields of signal processing[4], also in the watermarking techniques field, specially in the digital image watermarking techniques[5].

In this paper, we discuss the audio watermark embedding based on the human auditory characters using the wavelet and cepstrum transform, and propose a novel watermark embedding algorithm to embed binary image watermarks into original audio signal. The experimental results show that this algorithm has a better robustness against common signal processing such as noise, filtering, resampling and lossy compression.

Section 2 of this paper introduces discrete wavelet transform (DWT), complex cepstral transform and the novel watermark algorithm. The experiment results and analysis are given in Section 3, and our conclusions are given in Section 4.

II. The Watermarking Algorithm Based on WT and CCT

Usually the basic principal of digital audio watermark embedding is that the imperceptible components of audio signals are replaced by the watermarks. This substitution is constrained by the inaudibility and robustness [6]. The inaudibility, some times it is called perceptual transparency, is that humans cannot hear the differences between the watermarked audio signal and original audio signal. The main requirement for watermarking is perceptual transparency. The embedding process should not introduce any perceptible artifacts, that is, the watermark should not affect the quality of the original audio signal. However, for robustness, the watermark energy should be maximized under the constraint of keeping perceptual artifacts as low as possible. Thus, there must be a trade-off between perceptual transparency and robustness. This problem can be solved by applying HAS in the watermark embedding process [7]. The robustness is that after all kinds of the common signal processing operations on the watermarked audio, such as the transmission, filtering, resampling and lossy compression, the watermarks are damaged little and still can be detected or retrieved on the basis of the certain correct probability.

From the theory of human psychoacoustic auditory model, we know that the sense of the human ears on audio signal spectrum is conditional. The weaker signal that will not be perceptible by the ears near a stronger signal. This effect is the masking effect. Here, the stronger sound is called masking sound and the weaker one is called masked sound. The maximum sound pressure level of imperceptible masked sound is called masking threshold [8]. The human ears cannot hear the sounds under the masking threshold. Using the auditory masking effects to embed the watermarks can

realize their inaudibility better. The calculating formula of the audio signal masking threshold is as follows [9]:

$$M(f) = 3.64(f/100)^{-0.8} - 6.5e^{-0.6(f/1000-3.3)^2} + 10^{-3}(f/1000)^4\, dB(\text{SPL}). \quad (1)$$

Where M represents the masking thresholds, f represents the frequency of the audio signal.

A. *Discrete Wavelet Transform*

Suppose an original audio signal is $f(n)$ and the filters corresponding to the scale function and wavelet function are the $\overline{H}$ and $\overline{G}$ respectively [4]. The discrete wavelet transform (DWT) of audio signal $f(n)$ is showed as Fig.1. CA is the approximate components sub-band which mainly represent the low frequency components of the audio signal, and CD is the detail components sub-band which mainly represent the high frequency components of the audio signal. If continued to decompose the approximate components CA with J levels, we can obtain the wavelet decomposing components in different decomposing levels. This decomposing process is called the wavelet multiresolution analysis with J levels to the audio signal $f(n)$. The detail components and the approximate component obtained are respectively CD_i and CA_J, where the CD_i ($i = 1, 2, \cdots, J$) represents the i-th level detail component. This multi-resolution decomposing process is approximate to the human auditory system [4].

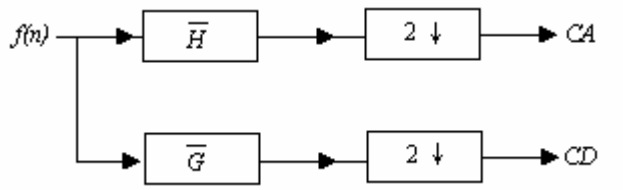

Fig. 1. The diagram of the discrete wavelet transform

Since the hearing of human ears is not much sensible to the minute change of the wavelet high frequency components and the coefficients of the high frequency component are smaller, so we can embed the watermarks into the high frequency components sub-band to realize the inaudibility of watermarks effectively. But the high frequency components are easily to be destroyed or removed by some kind of common signal processing, so that the robustness of the watermarks can't be ensured. While the low frequency components are the main components of the signal, because in the high frequency sub-band, the coefficients of the high frequency components are bigger and they carry most of the energy of the audio signal. So selecting the low frequency components sub-band to embed the watermarks can obtain an excellent robustness.

Although the wavelet approximate components are affected little by the outer environment and their stabilities are good, selecting this area to embed the watermarks can enhance the robustness of the watermarks, but at the same time they carry the main energy of the audio signal and are the main components of the signal. So the change to the approximate components can easily destroy the quality of the original audio signal. In other words, the watermarks are not perceptually transparent. In order to assure the perceptual transparency and the robustness of the embedded watermarks, we can divide the coefficients of the approximate components CA_J into two classes: the important coefficients and the common coefficients. The important coefficient is that its magnitude is bigger than the given threshold, otherwise it is the common coefficient. The important coefficients are the watermarks embedding objects, but the watermarks are not embedded into the important coefficients directly.

B. Complex Cepstral Transform

The complex cepstrum analysis is a homomorphic mapping and is the most effective characteristic extraction technique in audio identification. The principal advantage of cepstral transform (CCT) is that the coefficients of the complex cepstral transform are highly decorrelated after the audio signal is transformed by the complex cepstrum. Besides the high-order complex cepstral are numerically quite small and have a wide range of variances when going from low to higher-order complex cepstral coefficients. When the complex cepstral coefficients are attacked by the common signal processing, the variances obtained from the attacked cepstral coefficients are much smaller than that of the original sample in time domain, so the complex cepstral coefficients are fit to embed the watermark [10].

In order to improve the robustness and inaudibility of the embedded watermarks, the watermarks are not embedded into the important coefficients directly but into the complex cepstral components of the important coefficients transformed by the complex cepstrum. Let the $c(n)$ represent the complex cepstral of the coefficients $z(n)$. Then the definitions of the complex cepstral transform (CCT) and its inverse transform are as follows:

$$c(n) = F_1^{-1}\{\ln F_1[z(n)]\}, \tag{2}$$

$$z(n) = F_1^{-1}\{\exp F_1[c(n)]\}. \tag{3}$$

Where $F_1[\cdot]$ and $F_1^{-1}[\cdot]$ represent the Fourier transform and the inverse Fourier transform respectively.

C. Watermark Embedding and Extracting Algorithm

According to the properties of the wavelet transform and complex cepstral transform, based on the marking effect of human audible system, our watermark embedding scheme is depicted in Figure 2. The steps of watermark embedding process is as follows:

(1) The original audio signal is processed firstly by the DWT with J levels (generally the J = 3 or 4). Then the important coefficients $z(n)$ are selected from the approximate components sub-band CA_J with the threshold for the watermarks embedding objects. In order to compromise the robustness and inaudibility of the watermark, the threshold is calculated to be half of the maximum absolute coefficient in CA_J. The detailed components CD_i $(i = 1, 2, 3)$ was acted as the secrete key to the watermarking detection.

(2) Making use of the masking effect to jig the coefficients that fitting to embed watermark from the important coefficients $z(n)$. These coefficients are denoted as $z_1(n)$.

(3) Getting the cepstral coefficients $c(n)$ with the CCT to the $z_1(n)$.

(4) Before embedding watermarks, the original watermarks are pre-processed by the confusion technology in order to improve the security of the watermarks. The original watermarks in this paper was not a traditional psycho-random sequence, it was a binary image, thus the watermarks were needed to be processed by inducing their dimensions so that the two-dimension binary image was transformed into the one-dimension sequence [11]. Suppose $I = \{i(l, j), 0 \le l \le M_1, 0 \le j \le M_2\}$ is the two-dimension binary image, $i(l, j) \in \{0, 1\}$, thus the one-dimension sequence $W = \{w(k) = i(l, j)$, $0 \le l \le M_1$,

$0 \le j \le M_2$, $k = lM_1 + j\}$. In order to remove the correlation between the neighbouring pixels and improve the robustness, the elements of the sequence W are randomly permuted by a liners feedback shift register(LFSR), $W_p = Permute(W) = \{w_p(k) = w(k')$, $0 \le k, k' \le M_1 M_2\}$.

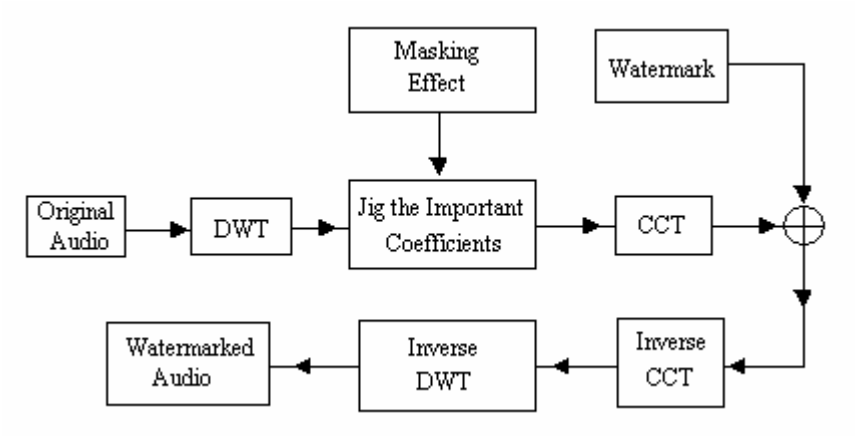

Fig. 2. The flow diagram of watermarks embedding

(5) Embedding watermarks through the formula:

$$c'(n) = c(n)(1 + \alpha w_p) . \quad (4)$$

Where w_p was the sequence of watermarks, and the parameter α is the factor of the watermark embedding strength. The parameter α should be adjusted fitly to ensure the robustness and inaudibility of the watermarks embedding.

(6) Making use of the inverse CCT and inverse DWT to obtain the watermarked audio signal.

The watermarks recovery process is the inverse process to the embedding process. The process of watermarks recovery needs the original audio signal.

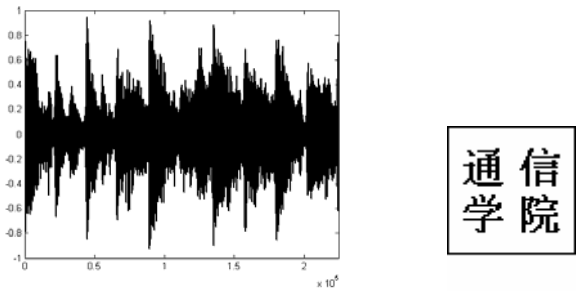

Fig.3 (a) The original audio signal and watermarks

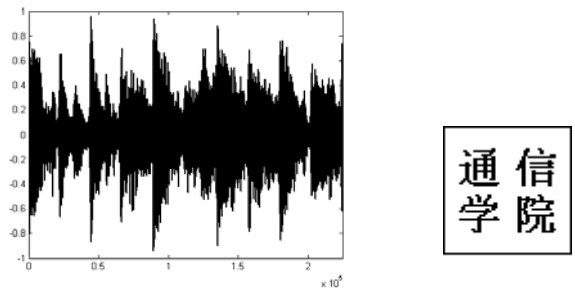

Fig.3 (b) The watermarked audio signal and the recovery watermarks

III. Experiment Results and Analysis

In order to evaluate the imperceptibility and robustness of the audio proposed embedding algorithm, we have done a lot of the simulation experiments on the computer, some results are given as follows. Figure 3(a) gave the original audio and watermarks, the original audio is a segment of the rock music whose length is 5.93 seconds, single-channel, the sample rate $f_s = 44.1KHz$ and the resolution is 16 bits. The original watermark is a binary image whose size was 64×64. The wavelet is the Db-4, and the level of the wavelet decomposition $J = 3$.

In order to remove the effects of the subject factor, we adopted the SNR and the normalized correlation (NC) to measure the performance of the embedding algorithm in this paper. The normalized correlation (NC) is defined as follow:

$$NC = \left| (\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} w(i,j)w'(i,j) / \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} w(i,j)w(i,j)) \right| . \quad (5)$$

Where $w(i, j)$ is the watermark of the original binary image, $w'(i, j)$ is the watermark of the recovery binary image, $N_1 \times N_2$ is the size of the binary image watermark.

The experiment results show that the magnitude of the parameter α over range of 0.01~0.45 (Figure 5). In order to compromise the inaudibility and robustness of the watermark, $\alpha = 0.02$ for the following experiment results.

The Figure 3(b) shows the watermarked audio signal and the recovery watermark whose SNR=39.0716*dB*.The embedded watermark had a good inaudibility.

In order to detect the robustness of watermarks, the watermarked audio signal was done with several digital signal processes such as the lowpass filter, additive white Gauss noise (AWGN), denoising, resampling and lossy compression. The lowpass filter is 3th-order Butterworth lowpass filter whose cutoff frequency is 9kHz, the *j*-times resampling means that insert *j* zeros into neighboring two points, and denoising included adding AWGN, then filtered by the lowpass filter [12], and MP3 lossy compression was adopted in this paper.

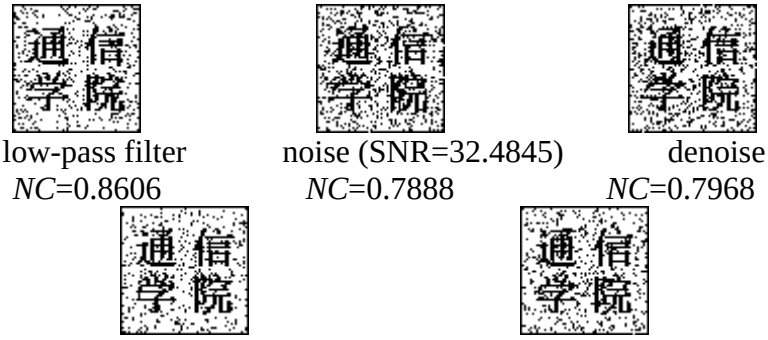

Fig. 4. Recovery watermark which suffered by different attacks

The recovered watermarks are showed in Figure 4 respectively when the watermarked audio signal was done by the above five signal processing operations. Figure 5 shows how the watermark embedding intensity α affected on the watermarked audio signal, we can see that the interval of magnitude of the parameter α is about 0.01~0.45. The change of the normalized correlation (NC)

with different common signal processing is given respectively in Figure 6, Figure 7 and Figure 8. We can see from the result of the experiment that the watermark embedding algorithm based on the human auditory model and the complex cepstral technology had a better robustness and inaudibility. The normalized correlation (NC) between the recovery watermarks and original watermarks is higher. And the recovery watermarks had a better visual identification.

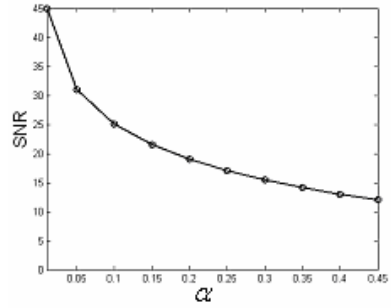

Fig.5. The change of SNR with α

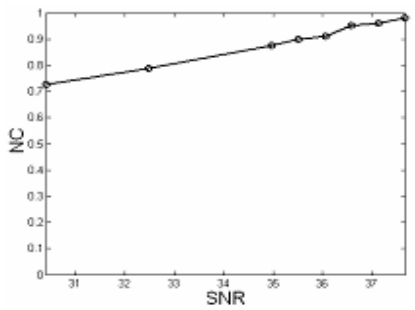

Fig. 6. The robust to AWGN

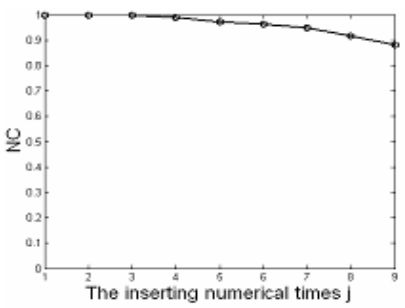

Fig.7. The robust to up-sampling

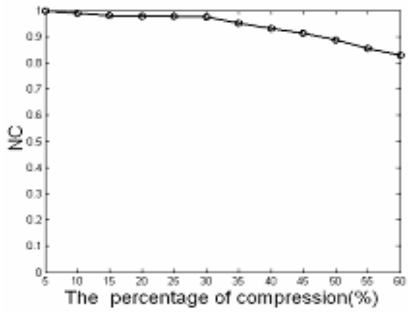

Fig.8. The robust to MP3

IV. Conclusion

Based on the human auditory model (HAS) and the masking effect of human ears, making use of the principal advantages of the wavelet transform (WT) and the complex cepstrum transform (CCT), this paper proposed an audio watermarking schema. The experiment results show that the watermarked audio signal was basically same as the original signal in time domain, the embedded watermark had a good inaudibility in hearing and has a good robustness against the common signal processing tasks such as filtering, adding noise, denoising, resampling especially in lossy compression by using this algorithm to embed watermark. This algorithm does not compromise the robustness and inaudibility of the watermark effectively.

References

[1] I. J. Cox, J. Kilian, and T. Shamoon, *Secure Spread Spectrum Watermarking for Image, Audio and Video*. IEEE trans. on Image Processing, 1997, vol.6, pp.1673-1687.

[2] Sang-Kwang Lee, Yo-Sung Ho, *Digital Audio Watermarking in the Cepstrum Domain*. IEEE trans. on Consumer Electronics, 2000, vol.46 (3), pp.744-750.

[3] C. T. Hsieh, P. Y. Tsou, *Blind cepstrum Domain Audio Watermarking based on Time Energy Features*. 2002 14th International Conference on Digital signal Processing Proceeding, Greece: Santorini, 2002, pp.705-708.

[4] Xiang-hong Tang, Zhen-hua He, Shu-qin Xie, *Analysis and application of wavelet*. Sichuan publishing company of science and technology, Chengdu, China, 1999, pp.78-155.

[5] M. Barni, F. Bartolini, and A. Piva, *Improved Wavelet-Based Watermarking Through Pixel-Wise Masking*. IEEE trans. on Image Processing, 2001, vol.10(5), pp.783-791.
[6] M. D. Swanson, M. Kobayashi, and A. H. Tewfik, *Multimedia Data-Embedding and Watermarking Technologies*. Processing of the IEEE, 1998, vol.86 (6), pp.1064-1086.
[7] M. Swanson, B. Zhu, A. Tewfik, and L. Boney, *Robust audio watermarking using perceptual masking*. Signal Processing, 1998, vol.66(3), pp.337-355.
[8] Zhen-hua, Lu, *Application of the auditory features in digital audio compression coding*. Audio Engineering, 1998(5), vol.11, pp. 6-12.
[9] Li Zhu, Cong-liang Guo, *Application of the psycho-acoustic model in digital audio*. Audio Engineering, 2002(8), vol.15, pp.11-14.
[10] Min-rui Zhang, Ke-chu Yi, *Audio and image watermarking algorithms in cepstrum domain*. Journal of Xidian University, 2003, Vol.30(6), pp.730-738.
[11] Yue Sun, Hong Sun, and Tian-ren Yao, *Digital audio watermarking algorithm based on quantization in wavelet domain*. Journal of Huazhong University of Science and Technology, 2002, Vol.30(5), pp.12-15 .
[12] Hong-yi Zhao, Chang-nian Zhang, *Digital signal processing and realization in MATLAB.* Publishing company of chemical industry, Beijing, 2001, pp.129-131.

Xiang-hong Tang was born in Sichuan, China, in 1962. He received the Ph.D. degree in signal and information processing from University of Electronic Science and Technology of China. He is currently a professor with the School of Communication Engineering of Hangzhou Dianzi University. His research interests include image and video processing, spread spectrum communication, digital watermarking, and wavelet transform theory.

Yamei Niu was born in Shandong, China, in 1978. She was received the M.S. degree in information technique from Hangzhou Dianzi University in 2004. Her research interests are in the areas of signal and audio processing, image and video coding and data compression

Heng-li Yue was born in Henan, China, in 1977. She was received the M.S. degree in information technique from Jianghan Petroleum University in 2000. She is currently a docent with the Department of telecommunication engineering of Hangzhou Dianzi University. Her research interests are in the areas of signal and image processing, and include image and video coding, data compression.

Zhong-ke Yin was born in 1969, in Henan, China. He graduated from University of Electronic Science and Technology of Electronics of China in 1997 with a PhD degree. He now works as a professor at the School of Information Science and Technology of Southwest Jiaotong University. His interest covers signal and information processing, image processing and transmitting. He is senior member of Chinese Electronic Association.