Journal of Southeast University (English Edition) Vol. 23, No. 2, pp. 206 – 210 June 2007 ISSN 1003—7985

Blind multiple watermarking algorithm for audio

Chen Ning Zhu Jie

(Department of Electronic Engineering, Shanghai Jiaotong University, Shanghai 200240, China)

Abstract: To make digital watermarking accomplish several goals, a new method for simultaneously embedding multiple watermarks into the same audio signal is proposed. First, the original audio signal is segmented into frames of appointed lengths and all the element watermarks are encoded to achieve a mixed watermark. Then, the binary bits in the mixed watermark are embedded into the audio frames with the echo hiding technique. The watermark extraction can be performed without an original audio signal. Furthermore, in order to enhance the extraction accuracy and the robustness of the proposed algorithm against common signal manipulations, the autocorrelation of the power cepstrum is utilized to estimate the echo delays in the watermarked audio frames to extract the mixed watermark and the corresponding decoding method is applied to achieve the element watermarks. Computer simulation results indicate that the proposed scheme has great robustness against common signal manipulations of Mp3 compressing, re-sampling, re-quantizing, low-pass filtering and white noise addition.

Key words: multiple watermarking; echo hiding; power cepstrum; blind detection

With the fast development of digital signal processing techniques and Internet, a number of new audio applications, such as voice-over-internet-protocol (VoIP), have emerged and, therefore, how to achieve the copyright protection and the integrity verification of digital audio signals has become a hot potato. The multiple watermarking algorithm, which can embed multiple watermarks into the same cover signal simultaneously, is considered as a good choice for the above problem. And a number of multiple watermarking algorithms in various technologies[1-4] have been reported recently. However, all these algorithms are designed for images. In this paper, a new multiple watermarking scheme for audio, which is based on an echo hiding technique[5] and a power cepstrum analysis[6], is illustrated. Computer simulation results verify the effectiveness of the proposed algorithm in terms of extraction accuracy and robustness against common signal manipulations.

1 Echo Data Hiding

The embedding process of the watermarking system based on echo hiding techniques can be represented as a system with two possible system functions (see Fig. 1). The delay between the original signal and the echo is dependent on the system function chosen. Specifically, if binary 1 is to be embedded, the original signal is echoed with delay d_1 and if binary 0 is to be embedded, it is echoed with delay d_0 (shown in Fig. 2, where a_0 and a_1 denote the amplitudes of the echoes). And when the delay d_0 (or d_1) is small enough, say 2 ms, the human auditory system cannot perceive a clear echo[7]. In order to embed more than one bit, the original signal is segmented into smaller frames. Each individual frame can then be embedded with the desired bit by considering each as an independent signal. The final encoded signal is the recombination of all independently echoed signal frames.

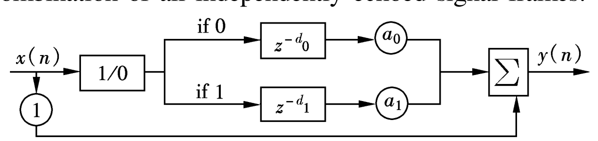

Fig. 1 Block diagram of echo hiding

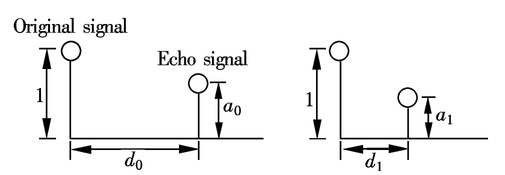

Fig. 2 Echo kernels

The extraction of the embedded information involves the detection of the echo delay. Usually, the complex cepstrum is employed in this stage because a waveform that contains an echo of delay d_1 (d_0) will exhibit its peak in the location of d_1 (d_0) in its complex cepstrum. But in most cases, the spike corresponding to the echo delay in the complex cepstrum can be corrupted by a complex cepstrum of the original signal or noise, so autocorrelation is applied to the complex cepstrum to estimate the echo delay[8-12]. However, simulation results

Received 2006-12-19.

Biographies: Chen Ning (1979—), female, graduate; Zhu Jie (corresponding author), male, doctor, professor, zhujie@sjtu.edu.cn.

show that the detection ratio of this decoding scheme is relatively low and its robustness against common signal manipulations is not satisfactory, the essences of which are the inherent phase unwrapping and aliasing problems during the calculation of the complex cepstrum.

2 Echo Detection Based on Power Cepstrum Analysis

The power cepstrum analysis[13] can be represented by the block diagram shown in Fig. 3, in which the aliasing problem can be avoided by applying a smoothing window function (e. g., hamming window) immediately before or after the logarithm operation and adding zeros at the end of the frame.

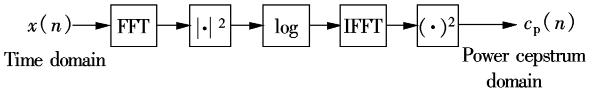

Fig. 3 Block diagram of power cepstrum analysis

And it is considered efficient in recognizing wavelet arrival time and amplitude[6]. To explain the principle of echo detection based on the power cepstrum, an example of a single additive echo is given and the total signal can be denoted as

$$x(t) = s(t) + as(t - \tau) \tag{1}$$

The power spectrum of $x(t)$ is

$$\phi_x(\omega) = \phi_s(\omega)[1 + a^2 + 2a\cos\omega\tau] \tag{2}$$

Thus, the spectral density of a signal with an echo has the form of an envelope that modulates a periodic function of frequency. The envelope is the spectrum of the original signal and the periodic function of frequency is the spectrum contribution of the echo. By taking the logarithm of Eq. (2), this product is converted to the sum of two components:

$$\log\phi_x(\omega) = \log\phi_s(\omega) + \log(1 + a^2 + 2a\cos\omega\tau) \tag{3}$$

Furthermore, if $|a| \ll 1$, a^2 can be neglected in Eq. (3) and remembering that

$$\log(1 + x) = \sum_{m=1}^{+\infty} \frac{1}{m}(-1)^{m+1}x^m \quad -1 < x \leqslant 1 \tag{4}$$

$\log\phi_x(\omega)$ can be well-approximated as

$$\log\phi_x(\omega) \cong \log\phi_s(\omega) + 2a\cos\omega\tau \tag{5}$$

Then, $\log\phi_x(\omega)$ is viewed as a waveform that has an additive periodic component whose fundamental frequency is the echo delay τ[14]. In conventional analysis of time waveforms, such periodic components show up as sharp peaks in the corresponding Fourier spectrum. Therefore, the spectrum of the log spectrum would likewise show a peak when the original time waveform contains an echo. This is the reason why the power cepstrum can be used for estimating the echo delay. Usually, the following equation is used to calculate the power cepstrum of signal $x(n)$.

$$c_p(n) = (\mathrm{IFFT}(\log|\mathrm{FFT}(x(n))|^2))^2 \tag{6}$$

It can be seen that phase information is lost in the power cepstrum; therefore, it does not have phase unwrapping problem, which is sensitive to noise. In this paper, the autocorrelation of the power cepstrum is utilized to estimate the echo delay. The performance comparison results between the proposed echo detection algorithm and the routine one (see section 1) are shown in Tab. 1 (where BAR represents the bit accuracy rate), which verifies that the proposed algorithm has greater robustness against common signal manipulations than the old one.

Tab. 1 Performance comparison results

Operation		BAR	
		New method	Old method
Mp3 compression	14.7:1	92	43
	8.8:1	100	40
	5.5:1	100	36
Re-sampling/kHz	44.1-22.05-44.1	99	38
	44.1-11.025-44.1	99	39
Re-quantizing/(bit·sample^{-1})	16-8-16	100	39
Low-pass-filtering/kHz	11.025	99	40
	22.050	100	40
Noise-addition/dB	30	90	36
	40	99	30

3 Multiple Watermarking Algorithm

3.1 Embedding scheme

Suppose that the multiple watermarks are composed of three element watermarks ($\boldsymbol{w}_1 = 1\ 0\ 1$, $\boldsymbol{w}_2 = 0\ 1$ and $\boldsymbol{w}_3 = 1$) of binary sequences. And there are three echo delays (d_1, d_0 and λ) that are utilized to represent the binary 1, binary 0 and NULL in watermarks, respectively. Then the embedding procedure of multiple watermarks is as follows:

① The original audio signal $s(n)$ is segmented into non-overlapping frames $\boldsymbol{F}_i$ ($i = 1, 2, \ldots, L$) of appointed length. And if L_w denotes the length of the longest watermark among the element watermarks, then $L \geqslant L_w$. Those watermarks whose lengths are smaller than L_w are added by NULL at the end.

② The first binary bit in each element watermark (namely, $\boldsymbol{w}_1(1) = 1$, $\boldsymbol{w}_2(1) = 0$ and $\boldsymbol{w}_3(1) = 1$) are embedded into the first frame $\boldsymbol{F}_1$ by echoing it with delays d_1, $d_1 + d_0$ and $d_1 + d_0 + d_1$ respectively to get the watermarked frame $\boldsymbol{F}_1'$ (see Fig. 4). And the second bit of each element watermark (namely, $\boldsymbol{w}_1(2) = 0$, $\boldsymbol{w}_2(2) = 1$ and $\boldsymbol{w}_3(2) = \mathrm{NULL}$) are embedded into the second frame $\boldsymbol{F}_2$ by echoing it with delays d_0, d_0+d_1 and $d_0+d_1+\lambda$ respectively to get $\boldsymbol{F}_2'$. All the following watermarked frames are obtained by the same method.

③ All the watermarked frames $\boldsymbol{F}_i'(i=1,2,\ldots,L_w)$ are connected end to end to get the watermarked audio signal $s'(n)$.

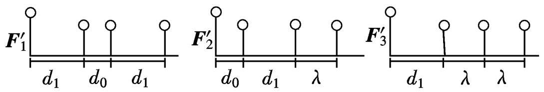

Fig. 4 Embedding method

3.2 Extraction scheme

At the receiver, the echo detection method proposed in section 2 is utilized to estimate the echo delays in the watermarked audio frames to extract the watermarks (without the original audio signal). The specific procedure is as follows:

① The watermarked audio signal $s''(n)$ (after certain manipulations) is segmented into non-overlapping frames $\boldsymbol{F}_i''(i=1,2,\ldots,L)$ of appointed lengths.

② The autocorrelation of the power cepstrum ACPC$_i(i=1,2,\ldots,L)$ of each frame $\boldsymbol{F}_i''$ is calculated.

③ In each ACPC$_i$, the three highest peaks P_1, P_2 and P_3 are detected. Then, the intervals between P_1 and original, P_2 and P_1, P_3 and P_2 in ACPC$_i$ correspond to the i-th bit in the extracted element watermarks $\boldsymbol{w}'_1, \boldsymbol{w}'_2$ and $\boldsymbol{w}'_3$. For example, if the interval of P_2 and P_1 in ACPC$_2$ is d_1, then the second bit of $\boldsymbol{w}'_2$ is 1. By this method, all the element watermarks can be simultaneously extracted without the original audio signal.

4 Simulation Results and Analysis

A piece of classical music of 11-second length and sampled at 44.1 kHz with 16 bit/sample is chosen as a frame of original audio signal. The echo delays used to represent the binary 1, binary 0 and NULL in watermarks are $d_1=45$ units, $d_0=89$ units, and $\lambda=67$ units, respectively. The echo decay rate is $\alpha=0.4$. Then the original audio and the watermarked one are shown in Fig. 5. It is difficult for the human hearing system to distinguish between them.

Fig. 6 shows the power cepstrum (PC) of the watermarked audio frame. The normalized amplitude of peaks in the PC as a function of echo decay rate is shown in Fig. 7. It is obvious that the amplitudes of three peaks increase with the enhancement of echo decay rate. Namely, the robustness of the proposed algorithm is dependent on the echo decay rates.

In order to test the robustness of the proposed algorithm, various common signal manipulations were

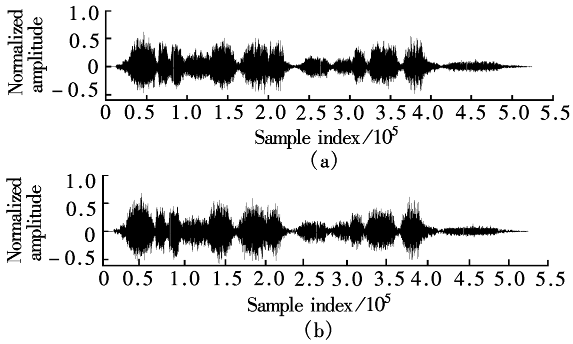

Fig. 5 Audio. (a) Original audio; (b) Watermarked audio

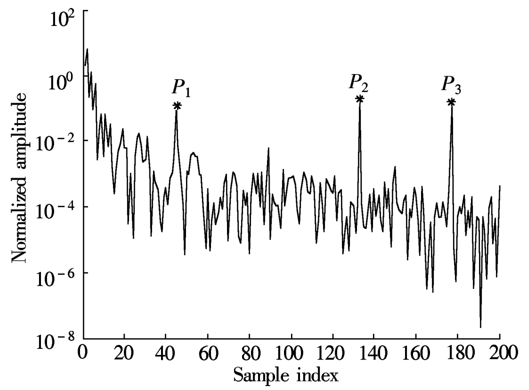

Fig. 6 PC of watermarked audio signal

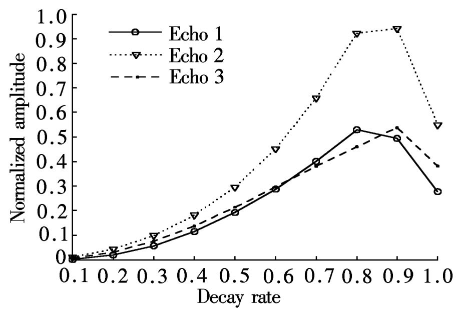

Fig. 7 Amplitudes of peaks as a function of echo decay rate

applied to the watermarked audio signal and then the watermark was extracted again.

1) Mp3 compression

The power cepstrum of the watermarked audio signal that goes through Mp3 compression at 48 kbit/s (namely, compression rate is 14.7 : 1) is shown in Fig. 8. And the changes of the amplitudes of three peaks with the increase in compression rate are shown in Fig. 9. All these indicate that the proposed algorithm has great robustness against Mp3 compression manipulation.

2) Re-sampling

A watermarked audio signal which is sampled at 44.1 kHz is sub-sampled down to 22.5 kHz and then up-sampled back to 44.1 kHz. Then, the power cepstrum of the re-sampled watermarked signal is shown in Fig. 10.

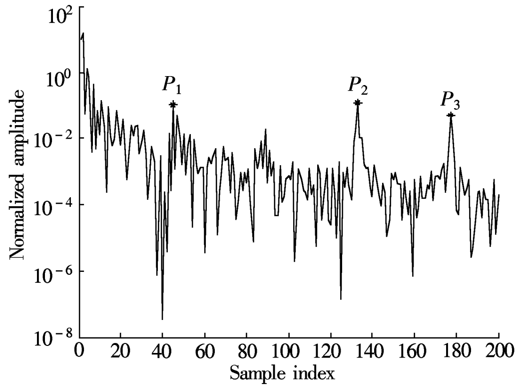

Fig. 8 PC of watermarked audio signal after Mp3 compression manipulation

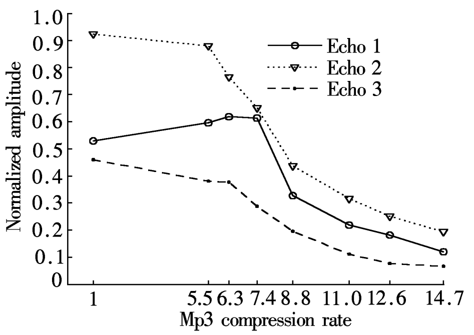

Fig. 9 Amplitudes of peaks as a function of Mp3 compression rate

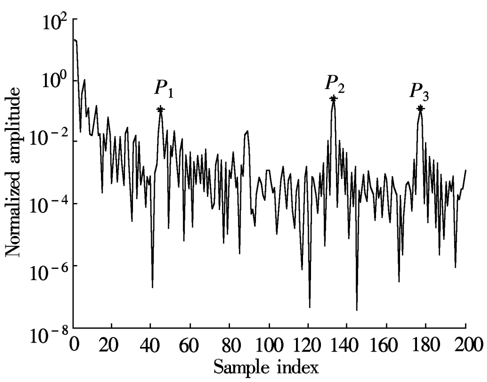

Fig. 10 PC of watermarked audio signal after re-sampling manipulation

3) Re-quantization

The watermarked audio signal originally quantized at 16 bit/sample is re-quantized down to 8 bit/sample and back to 16 bit/sample. As shown in Fig. 11, there appear three obvious peaks corresponding to the echo delays in the power cepstrum of the re-quantized watermarked audio signal.

4) Low-pass filtering

A 6-tap butter-worth low-pass filter with the cut-off frequency of 11.025 kHz was applied to the watermarked audio signal. The power cepstrum of the low-pass filtered watermarked audio signal is shown in Fig. 12.

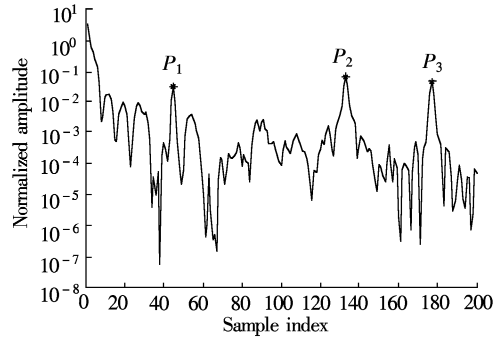

Fig. 11 PC of watermarked audio signal after re-quantizing manipulation

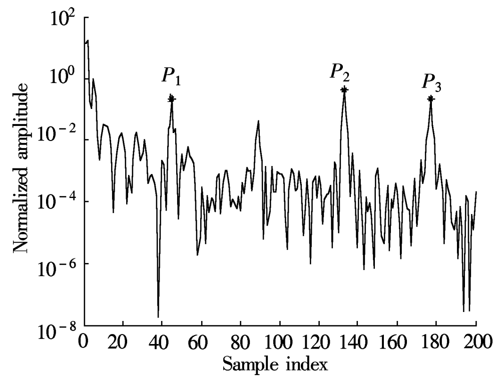

Fig. 12 PC of watermarked audio signal after low-pass filtering manipulation

5) Noise addition

White noise with a constant level of 20 dB is added to the watermarked audio signal. The power cepstrum of the watermarked audio signal that has been mixed with white noise is shown in Fig. 13. Simulation results verify that the proposed algorithm had great robustness against white noise addition manipulation and that when the noise was gradually enhanced, the amplitude of the peaks reduced gradually and the perceptual quality deteriorated at the same time.

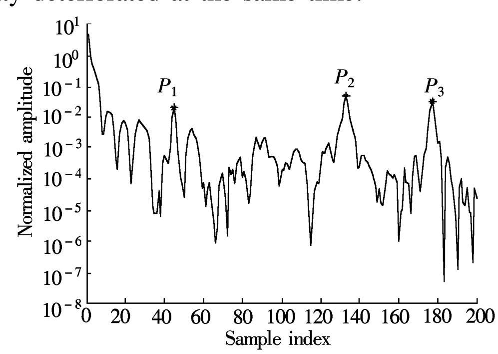

Fig. 13 PC of watermarked audio signal after white noise adding manipulation

5 Conclusion

A new multiple watermarking algorithm for digital audio signals is proposed. All the element watermarks are encoded first to generate a mixed watermark and then it is embedded into the original audio frames with echo hiding technique. At the receiver, the autocorrelation of the power cepstrum is utilized to estimate the echo delays in the watermarked audio frames to extract the mixed watermark and the corresponding decoding method is applied to obtain all the element watermarks. All these are performed without the original audio. Computer simulation results indicate that the proposed algorithm has great robustness against the common signal manipulations of Mp3 compression, the re-sampling, re-quantizing, low-pass filtering and white noise addition. Our future work will focus on combining a psychoacoustic model with the proposed algorithm to enhance the perceptual quality of the watermarked audio signal and the content of the proposed algorithm.

References

[1] Wong P H W, Au O C, Yeung Y M. A novel blind multiple watermarking technique for images [J]. *IEEE Trans on Circuits and Systems for Video Technology*, 2003, **13**(8): 813 - 830.

[2] Ji Zhen, Xiao Weiwei, Wang Jianhua, et al. A multiple watermarking arithmetic for digital image based on chaotic sequences [J]. *Chinese Journal of Computers*, 2003, **26**(11): 1555 - 1561. (in Chinese)

[3] Giakoumaki A, Pavlopoulos S, Koutsouris D. A multiple watermarking arithmetic applied to medical image management [C]//*Proc of the 26th Annual Intl Conf on IEEE-EMBS*. San Francisco, 2004: 1 - 5.

[4] Wang Rangding, Xu Dawen, Li Qian. Multiple audio watermarks based on lifting wavelet transform [C]//*Proc of the 4th Intl Conf on Machine Learning and Cybernetics*. Guangzhou, 2005: 18 - 21.

[5] Bender W, Gruhl D, Morimoto N, et al. Techniques for data hiding [J]. *IBM System Journal*, 1996, **35**(3/4): 313 - 336.

[6] Kemerait R C. Signal detection and extraction by cepstrum techniques [D]. Florida: University of Florida Gainesville, 1971.

[7] Hyen O Oh, Jong Won Seok, Jin Woo Hong, et al. New echo embedding technique for robust and imperceptible audio watermarking [C]//*Proc of IEEE Intl Conf on Acoustics, Speech and Signal Processing*. Salt Lake City, USA, 2001, **3**: 1341 - 1344.

[8] Gruhl Daniel, Bender Walter, Lu Anthony. Echo hiding [C]//*Proc of the First Intl Workshop on Information Hiding*. London, 1996, **1174**: 293 - 315.

[9] Kim Hyoung Joong, Choi Yong Hee. A novel echo-hiding scheme with backward and forward kernels [J]. *IEEE Trans on Circuits and Systems for Video Technology*, 2003, **13**(8): 885 - 889.

[10] Ko B, Nishimura R, Suzuki Y. Time-spread echo method for digital audio watermarking [J]. *IEEE Trans on Multimedia*, 2005(2): 212 - 221.

[11] Chou S A, Hsieh S F. An echo-hiding watermarking technique based on bilateral symmetric time spread kernel [C]//*Proc of ICASSP*. Toulouse, 2006, **3**: 1100 - 1103.

[12] Wu Wen-Chih, Chen O T-C. Analysis-by-synthesis echo hiding scheme using mirrored kernels [C]//*Proc of ICASSP*. Toulouse, 2006, **2**: 325 - 328.

[13] Bogert B P, Healy M J R, Tukey J W. The quefrency analysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe cracking [C]//*Proc of the Symposium on Time Series Analysis*. New York, 1963: 209 - 243.

[14] Oppenheim A V, Schafer R W. From frequency to quefrency: a history of the cepstrum [J]. *IEEE Signal Processing*, 2004, **21**(5): 95 - 106.

一种可盲检测的多功能音频水印算法

陈　宁　　朱　杰

(上海交通大学电子工程系,上海 200240)

摘要:为了使数字水印实现多种功能,提出了一种能在同一音频信号中同时嵌入多个水印的新方法. 首先,载体音频信号被分隔成指定长度的帧,并且所有成员水印被编码成一个混合水印. 然后,利用回声隐藏技术将混合水印中的二进制比特嵌入到音频信号帧中. 水印提取不需要原始音频信号. 为了提高水印算法的提取正确率和抵御常规信号处理的鲁棒性,功率倒谱的自相关被用于检测携带水印的音频信号帧中的回声延时以提取出混合水印,并且相应的解码算法被用于分离成员水印. 计算机仿真结果表明所提出的水印算法对常规的信号处理,如 Mp3 压缩、重采样、重量化、低通滤波以及白噪声具有很强的鲁棒性.

关键词:多功能水印;回声隐藏;功率倒谱;盲检测

中图分类号:TN912. 3