ANALYSIS-BY-SYNTHESIS ECHO HIDING SCHEME USING MIRRORED KERNELS

Wen-chih Wu[α,β], *Oscal T.–C. Chen*[α]

Dept. of Electrical Engineering, National Chung Cheng University, Chia-Yi, 621 Taiwan[α]
Dept. of Electrical Engineering, Wu-Feng Institute of Technology, Chia-Yi, 621 Taiwan[β]

ABSTRACT

Watermarking using the echo hiding technique can have fairly good performance in perceptual quality. However, the tradeoff between robustness and imperceptibility still is a challenge work. In this study, the analysis-by-synthesis echo hiding scheme based on the mirrored kernels is proposed to determine the appropriate amplitudes of the embedded echoes with considering the characteristics of host signals. The analysis-by-synthesis technique is to perform attacks on the watermarked signals during the watermark embedding process to suggest the amplitude parameters. Additionally, the attacks are modeled by the Gaussian white noise. To avoid the embedded watermarks depending on the host signals and reduce the impact of various attacks, the positions of mirrored kernels for embedding "0" and "1" are inter-changed alternatively. The experimental results demonstrate that the proposed analysis-by-synthesis echo hiding scheme is superior to the conventional schemes in terms of robustness and perceptual quality.

1. INTRODUCTION

The easy manipulation and transmission of digital media has brought a strong demand on copyright protection techniques [1]-[2]. Cryptographic systems provide a way to avoid the access of digital contents. However, as soon as the secret key is decrypted, it is unable to prevent digital contents reproduced. This drawback can be improved via digital watermarking schemes. A digital watermarking technique embeds the copyright data into the host media permanently. Additionally, the embedded data are hidden in the sense that it is perceptually and statistically undetectable. Since the human auditory system is more sensitive to the visual system, the development of a high-performance audio watermarking is indeed a challenge task.

In the literature, several audio watermarking schemes have been developed such as phase coding, spread spectrum, patchwork, echo hiding, and so on [1]-[4]. In these techniques, the echo hiding can be widely used due to good perceptual quality. Since the embedded echo signals have the same statistical and perceptual characteristics as the host signals, the echo hiding seldom generates undesirable noises. The echo hiding can be treated as the blind watermarking technique and thus can be implemented easily. The robustness of the echo hiding can be enhanced by increasing amplitudes of echo signals, but audio quality would be degraded. In the conventional echo hiding schemes, the compromise between robustness and imperceptibility need be well addressed. Accordingly, several modifications on the echo kernels have been developed to improve the performance of robustness and imperceptibility. For example, Oh *et al*. introduced the dual-kernel echo hiding scheme [3], and Kim *et al*. presented an echo hiding scheme with backward and forward kernels [4].

Among these schemes, the effort is focused on improving the performance of the echo hiding by modifying the kernels. However, they have not taken the characteristics of host audio signals into account. Certainly, there is no consideration in connection with the attacks on the watermarked audio signals. In this work, we present an echo hiding scheme based on the analysis-by-synthesis approach [5], [6]. The amplitude of the embedded echo signal is adapted during the embedding process by considering not only the properties of the host signals but also the situations when the watermarked audio signals have suffered various attacks. Furthermore, various attacks are characterized by using the Gaussian white noise model that is employed in the proposed analysis-by-synthesis approach. Finally, the positions of mirrored kernels for embedding "0" and "1" are inter-changed in each audio segment alternatively, such that there is a lot of room to select the amplitudes of the embedded echoes, and thereby these amplitudes can be effectively minimized.

2. ECHO HIDING

The echo hiding proposed by Bender *et al*. is a popular scheme for audio watermarking [1]. The watermarked audio signal $y[n]$ can be obtained by convoluting the host signal $x[n]$ and the echo kernel $h[n]$. For simplicity, the kernel with a single echo is written as follows:

$$h[n] = \delta[n] + \alpha\delta[n-d], \tag{1}$$

This work is partially supported by National Science Council, Taiwan, under the contract number of NSC 93-2622-E-194-009-CC3.

where $\delta[n]$ is the Kronecker delta function and the amplitude of the embedded echo is denoted as the decay rate α (< 1). The delay offset d of the echo signal is determined in accordance with the values of the hidden data $I \in \{0, 1\}$.

Extraction of the embedded watermark data involves the detection of the echoes' positions, that can be realized by using the cepstrum of the watermarked audio signals. The cepstrum of the watermarked audio signals is defined as follows:

$$\begin{aligned} c_y[n] &= F^{-1}(\ln F(y[n])) \\ &= F^{-1}(\ln H(e^{j\omega})) + F^{-1}(\ln X(e^{j\omega})) = c_h[n] + c_x[n] \end{aligned} , \quad (2)$$

where F and F^{-1} represent the Fourier transform and the inverse Fourier transform, respectively. The cepstrum of the watermarked audio signals comprises two parts: one is the cepstrum of the kernel $c_h[n]$ and the other is the cepstrum of the host signals $c_x[n]$. Therefore, the performance of the echo hiding scheme would be affected by the host audio signals.

Using the power series expansion, the cepstrum of the kernel function expressed in Eq. (1) can be rewritten as follows:

$$c_h[n] = \alpha\delta[n-d] - \frac{\alpha^2}{2}\delta[n-2d] + \frac{\alpha^3}{3}\delta[n-3d] - \cdots . \quad (3)$$

The cepstrum of the kernel function has a value α at the delay offset d, which corresponds to the decay rate. The echo hiding scheme proposed by Bender *et al.* uses the fixed-amplitude echoes at two different delay offsets to act as the identification tags of watermark data. The extracted watermark is then determined by checking the cepstrum values at these two delay offsets.

Oh *et al.* and Kim *et al.* have introduced the dual-kernel echo-hiding scheme, and echo-hiding scheme with backward and forward kernels, respectively. In the proposed scheme, both kernel functions are adopted to implement the echo signals and they can be represented as follows:

$$\begin{aligned} h[n] = \delta[n] + \alpha\delta[n-d_p] + \alpha\delta[n+d_p] \\ - \alpha\delta[n-d_n] - \alpha\delta[n+d_n] \end{aligned} , \quad (4)$$

where d_p and d_n denote the delay offsets of the positive and negative echoes, respectively. The watermarked audio signal can then be written as

$$y[n] = \begin{cases} x[n] + \alpha x[n-d_{ap}] - \alpha x[n-d_{an}] + \alpha x[n+d_{ap}] - \alpha x[n+d_{an}], & I = 0 \\ x[n] + \alpha x[n-d_{bp}] - \alpha x[n-d_{bn}] + \alpha x[n+d_{bp}] - \alpha x[n+d_{bn}], & I = 1 \end{cases} . \quad (5)$$

Depending on the type of hiding data, the embedded echoes are placed at the time delay of a or b.

In the extracting process, the cepstrum of the watermarked signals, $\hat{y}$, after attacks is given by

$$c_{\hat{y}}[n] = \begin{cases} c_x[n] + \frac{\alpha}{1-\alpha^2}\delta[n-d_{ap}] - \frac{\alpha}{1-\alpha^2}\delta[n-d_{an}] + \cdots + c_A[n], & I = 0 \\ c_x[n] + \frac{\alpha}{1-\alpha^2}\delta[n-d_{bp}] - \frac{\alpha}{1-\alpha^2}\delta[n-d_{bn}] + \cdots + c_A[n], & I = 1 \end{cases} , \quad (6)$$

where $c_A[n]$ represents the deviation that is resulted from the various attacks exerted on the watermarked audio signals. The difference of cepstrum coefficients at these two delay offsets, $c_{\hat{y}}(a,b)$, is

$$\begin{aligned} c_{\hat{y}}(a,b) &= c_{\hat{y}}[n-d_a] - c_{\hat{y}}[n-d_b] \\ &= \left(c_{\hat{y}}[n-d_{ap}] - c_{\hat{y}}[n-d_{an}]\right) - \left(c_{\hat{y}}[n-d_{bp}] - c_{\hat{y}}[n-d_{bn}]\right) \\ &= \begin{cases} \frac{2\alpha}{1-\alpha^2} + c_x(a,b) + c_A(a,b), & I = 0 \\ -\frac{2\alpha}{1-\alpha^2} I + c_x(a,b) + c_A(a,b), & I = 1 \end{cases} \end{aligned} . \quad (7)$$

The extracted information data, $\hat{I}$, can then be determined as follows:

$$\hat{I} = \begin{cases} 0 & c_{\hat{y}}(a,b) \geq 0 \\ 1 & c_{\hat{y}}(a,b) < 0 \end{cases} . \quad (8)$$

Therefore, in order to extract the embedded data correctly, the following condition should be satisfied

$$\begin{aligned} \frac{2\alpha}{1-\alpha^2} + c_x(a,b) + c_A(a,b) \geq 0, \quad if\ I = 0 \\ -\frac{2\alpha}{1-\alpha^2} + c_x(a,b) + c_A(a,b) < 0, \quad if\ I = 1 \end{aligned} . \quad (9)$$

Usually, the decay rate, α, is fixed without contemplating the effects of the host audio signals and various attacks. To achieve the high recovery accuracy, the echo's amplitude should be adapted dynamically.

3. PROPOSED ECHO WATERMARKING SCHEME

The analysis-by-synthesis approach is adopted in the proposed echo hiding scheme. During the watermark embedding process, we explore whether these embedded data can be accurately recovered from the watermarked audio signals, even though there may be some malicious attacks exerted on the watermarked audio signals. According to the characteristics of the host audio signals, the amplitudes of the embedded echoes are adapted appropriately so as to ensure that the embedded data are recovered accurately after various common attacks and have minimum impacts on the host audio signals. The embedding process of the proposed echo hiding scheme using the analysis-by-synthesis approach is shown in Fig. 1.

In general, the values of high-order cepstrum coefficients are smaller than those of low-order cepstrum coefficients. Furthermore, the cepstrum coefficients at two delay offsets of the host signals may not be equal. If the delay offset of the embedded echo is fixed for each type of watermark datum, it will lead to a problem that an audio segment is likely for embedding one type of watermark datum but improper for another type of watermark datum. In this work, we develop mirrored kernels to tackle this problem. For the mirrored kernels, each segment of the host audio signals is separated into two parts. In the front part of a segment, the echo is placed at the delay offsets of d_a and d_b for watermark data "0" and "1," respectively. Contrarily, the echo positions are interchanged in the rear part of this

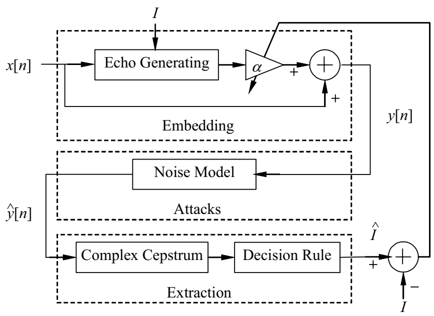

Fig. 1 Block diagram of the embedding process.

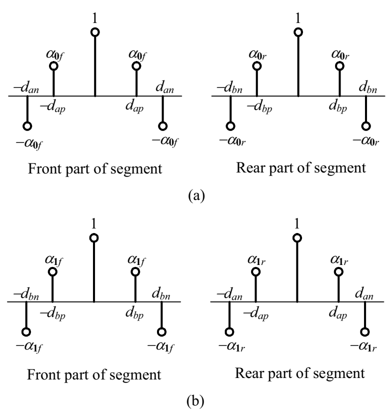

Fig. 2 Impulse responses of mirrored kernels. (a) "Zero" kernel. (b) "One" kernel.

segment. The impulse responses of the "zero" and "one" kernels are depicted in Fig. 2.

In the analysis-by-synthesis approach, we consider whether these embedded data can be accurately recovered from the watermarked audio signals during the watermark embedding process. Since the watermarked audio signals may undergo various malicious attacks before the extracting process, these attack conditions should be considered in the watermark embedding process. Because the characteristics of the common signal processing and the malicious attacks are so diverse, the real influences on each sample of the audio signals are unknown *a priori*. For simplicity, these attacks may be regarded as adding a random noise into the watermarked audio signals. Hence, the Gaussian white noise with zero mean and σ variance, $N(0, \sigma)$, is used to characterize the attacks.

Next, the difference of cepstrum coefficients at these two delay offsets of a segment, $c_{\hat{y}}(a,b)$, can be derived at

$$\begin{aligned} c_{\hat{y}}(a,b) &= c_{\hat{y}_f}(a,b) + c_{\hat{y}_r}(b,a) \\ &= \begin{cases} \left(\dfrac{2\alpha_{0f}}{1-\alpha_{0f}^2} + \dfrac{2\alpha_{0r}}{1-\alpha_{0r}^2}\right) + \left(c_{x_f}(a,b) - c_{x_r}(a,b)\right) + \left(c_{N_f}(a,b) - c_{N_r}(a,b)\right), & I = 0 \\ -\left(\dfrac{2\alpha_{1f}}{1-\alpha_{1f}^2} + \dfrac{2\alpha_{1r}}{1-\alpha_{1r}^2}\right) + \left(c_{x_f}(a,b) - c_{x_r}(a,b)\right) + \left(c_{N_f}(a,b) - c_{N_r}(a,b)\right), & I = 1 \end{cases} \end{aligned} \quad (10)$$

Accordingly, to extract the embedded data correctly, the following condition should be satisfied

$$\begin{aligned} &\left(\frac{2\alpha_{0f}}{1-\alpha_{0f}^2} + \frac{2\alpha_{0r}}{1-\alpha_{0r}^2}\right) + \left(c_{x_f}(a,b) - c_{x_r}(a,b)\right) + \left(c_{N_f}(a,b) - c_{N_r}(a,b)\right) \geq 0 \quad \text{if } I = 0 \\ &-\left(\frac{2\alpha_{1f}}{1-\alpha_{1f}^2} + \frac{2\alpha_{1r}}{1-\alpha_{1r}^2}\right) + \left(c_{x_f}(a,b) - c_{x_r}(a,b)\right) + \left(c_{N_f}(a,b) - c_{N_r}(a,b)\right) < 0 \quad \text{if } I = 1 \end{aligned} \quad (11)$$

The best condition to extract the hidden datum correctly with the smallest amplitude of the embedded echo is

$$\left(c_{x_f}(a,b) - c_{x_r}(a,b)\right) + \left(c_{N_f}(a,b) - c_{N_r}(a,b)\right) = 0 \cdot \quad (12)$$

It is possible to satisfy this condition when the host audio signals have the similar properties in the front and rear parts of a segment. Additionally, the malicious attacks on the front and rear parts of a segment may be regarded as approximately equal. Even though, this case in Eq. (12) is not achievable. The proposed analysis-by-synthesis approach would modify the values of decay rates to ensure that the condition in Eq. (11) is satisfied. However, without using the mirrored kernels, the best condition to extract the hidden datum correctly is derived at:

$$c_x(a,b) + c_A(a,b) = 0 \cdot \quad (13)$$

Generally, this condition is seldom to hold. Therefore, the mirrored kernels can yield the benefit on the robustness of the echo hiding by using a small decay rate.

4. EXPERIMENTAL RESULTS

The five audio pieces are utilized as the host signals to test the performances of the proposed scheme. Each host signal has 4,096,000 samples. The audio signals are sampled at 44.1 kHz where each sample is quantized by 16 bits. In the embedding process, host signals are divided into segments, each of which has 4,096 samples. The Hanning window is employed to smooth the boundaries of the segments. The time offsets of positive and negative echoes at the position *a* are placed at the 100th and 103th sampling positions, respectively. The time offsets of positive and negative echoes at the position *b* are placed at the 110th and 113th sampling positions, respectively.

The Echo Hiding scheme with Dual-kernel (EHD) developed by Oh *et al*. [3] and Echo-Hiding scheme with Backward and Forward kernels (EHBF) proposed by Kim *et al.* [4] are two conventional approaches for comparison in this work. Since the embedded echoes are different in these schemes, the equivalent decay rate for each scheme is employed to do fair comparison. The equivalent decay rate for the case of EHBF is calculated as follows:

$$\alpha_e = \alpha_b + \alpha_f, \quad (14)$$

where the α_b and α_f are the amplitudes of the backward and forward echo kernels, respectively. As to the proposed scheme, the following equation is used to represent the equivalent decay rate in each audio segment.

Table 1 Comparisons of the proposed and conventional schemes

Performance \ Schemes		EHD $\alpha = 0.4$	EHBF $\alpha_b = \alpha_f = 0.2$	Proposed $\alpha_{0f}, \alpha_{1f}, \alpha_{0r}, \alpha_{1r} \leq 0.2$
Robustness	Closed loop	96.3%	96.8%	98.9%
	MP3	92.6%	93.8%	97.2%
	Quantization	88.7%	91.1%	92.7%
	Noise	82.5%	83.9%	91.8%
	Re-sampling	92.1%	93.4%	96.9%
	BP filtering	94.6%	95.2%	98.4%
Audio quality	SNR(dB)	17.4	20.4	24.2
	Mean of α_e	0.4	0.4	0.124

Table 2 Effects of the proposed scheme using the mirrored kernels or not

Performance \ Schemes		Without mirrored kernels	With mirrored kernels
Robustness	Closed loop	94.1%	98.9%
	MP3	90.7%	97.2%
	Quantization	89.1%	92.7%
	Noise	86.1%	91.8%
	Re-sampling	87.5%	96.9%
	BP filtering	91.8%	98.4%
Audio quality	SNR(dB)	24.8	24.2
	Mean of α_e	0.117	0.124

$$\alpha_e = (\alpha_{of} + \alpha_{0r} + \alpha_{1f} + \alpha_{1r})/2, \tag{15}$$

In addition to the SNR of watermarked audio signals, the amplitude of the embedded echo is also used to evaluate the audio quality. Additionally, the robustness of the proposed and conventional schemes is compared at the following cases.

1) Closed-loop: The watermark data are extracted from the watermarked audio signals without any attacks.

2) MP3 attack: Encoding/decoding was performed on the watermarked audio signals by using the MPEG-1 audio layer III with a bit rate of 64kbps.

3) Quantization attack: After quantization, each sample of the watermarked audio signals is quantized to 8 bits.

4) Adding random noises: After adding random noises, the watermarked audio signals would have a signal-to-noise ratio of 20 dB.

5) Re-sampling attack: The watermarked audio signals are down sampled by a factor of 2, and then up-sampling.

6) Band-pass filtering attack: A 10th order Butterworth band-pass filter which has 100 Hz and 10 kHz cutoff frequencies is applied to the watermarked audio signals.

According to the above-mentioned cases, the performance of the proposed scheme is verified. Here, the values of decay rates are limited to ensure that the equivalent amplitudes of the embedded echoes are less than those of the conventional schemes at each segment. Such setting in the amplitudes of the embedded echoes considers mainly for fair comparison in the audio quality. Table 1 lists the performance of the proposed and the conventional echo hiding schemes. Simulation results reveal the high recovery accuracy rate of the proposed scheme under various attacks, which is much better than those of the conventional schemes. Since the maximum amplitudes of embedded echoes used for the proposed scheme are limited to less than those of the conventional schemes for each segment, the audio quality of the proposed scheme is better than that of the conventional schemes. The values of the SNR and the mean values of the equivalent decay rates validate this phenomenon. In other words, the proposed scheme can improve the robustness with even embedded echoes' amplitudes smaller than those of the conventional schemes.

Next, the impact of the mirrored kernels is discussed. The simulation results of the proposed scheme with and without the mirrored kernels are listed in Table 2. The robustness is highly improved for the echo hiding scheme using mirrored kernels. Additionally, the averaged decay rates of these two approaches are similar. Hence, the mirrored kernels can effectively reduce the influences of the host audio signals and the malicious attacks on the robustness performance.

5. CONCLUSION

In this work, the decay rates of echoes are adapted by using an analysis-by-synthesis approach to take the advantage of the properties of host signals. Additionally, a model based on the Gaussian white noise is established to simulate the conditions for various malicious attacks. Moreover, the mirrored kernels are proposed to eliminate the influences of the host signals and various attacks on the watermarked data. The simulation results reveal that the proposed scheme is superior to the conventional schemes in terms of the perceptual quality of the watermarked audio signals and the robustness for various attacks.

REFERENCES

[1] W. Bender, D. Gruhl, N. Morimoto and A. Lu, "Techniques for data hiding," *IBM System Journal*, vol. 35, nos. 3 & 4, pp. 313-336, 1996.

[2] S. Craver, N. Memon and M. M. Yeung, "Resolving rightful ownerships with invisible watermarking techniques: Limitations, attacks, and implications," *IEEE J. Selected Areas in Communications*, vol. 16, no. 4, pp. 573-586, 1998.

[3] H. O. Oh, J. W. Seok, J. W. Hong and D. H. Youn, "New echo embedding technique for robust and imperceptible audio watermarking," *Proc. of IEEE Int. conf. Acoustic, speech, and Signal Processing*, vol. 3, pp. 1341-1344, 2001.

[4] H. J. Kim and Y. H. Choi, "A novel echo-hiding scheme with backward and forward kernels," *IEEE Trans. on Circuit and System for Video Technology*, vol. 13, No. 8, pp. 885-889, Aug. 2003.

[5] W. C. Wu, O. T.-C. Chen and Y. H. Wang, "An echo watermarking method using an analysis-by-synthesis approach," *Proc. of the 5th IASTED Int. Conf. on Signal and Image Processing*, pp. 365-369, Aug. 2003.

[6] W. C. Wu and O. T.-C. Chen, "An Analysis-by-Synthesis Echo Watermarking Method," *Proc. of IEEE Int. Conf. on Multimedia and Expo*, June 2004.