ROBUST ECHO HIDING SCHEME AGAINST PITCH-SCALING ATTACKS

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

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

ABSTRACT

The pitch scaling conducts serious attacking on audio watermarked signals because it leads to a de-synchronization problem. Based on the proposed analysis-by-synthesis echo watermarking scheme, this work develops a watermark extracting process to improve the robustness performance against pitch-scaling attacks. The positions of embedded echoes are altered in the cepstrum domain due to the pitch-scaling attacks that change the frequency content of audio signals. Hence, it is unable to determine the echo position *a priori*. In this work, the cepstrum values in a reasonable range of delay offsets affected by pitch-scaling attacks are addressed from auditory point of view. The moving average function is performed on several segments to eliminate the impact of host signals on cepstrum. In this extended range for searching delay offsets, the peak value of the cepstrum is adopted to determine the embedded watermark datum. Experimental results exhibit that the proposed robust echo watermarking scheme can be superior to the conventional audio watermarking schemes under pitch-scaling attacks and other attacks.

Index Terms— Echo watermarking, analysis-by-synthesis, interlaced kernels, frequency hopping, pitch-scaling attacks

1. INTRODUCTION

Digital watermarking is an effective technique to provide the copyright protection for digital multimedia contents [1]. The watermarking technique should permanently embed the copyright information into host media signals using a perceptually transparent manner. Additionally, the embedded watermark should be correctly extracted when the watermarked signals experienced the common signal processing functions. Furthermore, the embedded watermark should not be detected or removed even for those who are familiar with the principle of the embedding techniques.

Many audio watermarking schemes have been developed to deal with high sensitivity of the human auditory system in the past decades including low-bit coding, phase coding [1], spread spectrum [2], patchwork [3] and echo hiding [1]. Among these schemes, the echo hiding is preferred because the embedded echo signals have the same statistical and perceptual characteristics as host signals.

Several modifications have been made to improve the performance of echo hiding. Oh *et al*. introduced the positive-negative echo hiding scheme [4]. Combining positive and negative echoes at nearby locations can significantly decrease the noise perceptibility. An echo hiding scheme with backward and forward kernels presented by Kim *et al*. can enhance the robustness of echo hiding [5]. To increase security, Ko *et al*. presented a time-spread echo hiding scheme that involved spreading the delay offsets of echoes using a pseudo noise sequence [6]. However, these schemes are very vulnerable to a pitch-scaling attack which can be regarded as one of the severest attacks. Accordingly, a log-scaling watermark detection is proposed to improve the robustness against pitch-scaling attacks [7]. Like others, the analysis-by-synthesis approach proposed previously by us is also vulnerable to such an attack, although it is quite robust against other attacks [8]-[11].

In this work, based on the analysis-by-synthesis approach [8]-[11], the adequate extracting process is developed to improve the robustness against pitch-scaling attacks. The positions of embedded echoes are altered when the pitching-scaling attack is applied. This action enables that the detections of echoes in specific positions yield wrong results. Therefore, the cepstrum values at a reasonable range of echoes' positions varied by pitch-scaling attacks need be investigated where the range is decided by the human auditory requirement. The peak of cepstrum values in these positions is employed to correctly determine the embedded watermark datum. Simulation results demonstrate that the proposed scheme has higher robustness than the conventional echo watermarking schemes not only for the cases of typical common attacks but also for the case of a pitch-scaling attack.

2. EFFECT OF PITCH-SCALING ON CEPSTRUM

Pitch scaling of audio signals is the process of scaling the signal's pitch without changing its playback time. For

This work is partially supported by National Science Council, Taiwan, under the contract number of NSC 96-2628-E-194-004-MY2.

signals with a frequency f, the operation of pitch scaling can be regarded as frequency modification, and the scaled frequency f' is βf where β (>0) is the pitch-scaling factor. According to the scaling property of the Fourier transform, the relationship between time-duration and bandwidth can be formulated as follows:

$$F(x[\beta n]) = \frac{1}{|\beta|} X\left(e^{j\frac{\omega}{\beta}}\right), \tag{1}$$

where F is the Fourier transform. If $\beta > 1$, the pitch becomes higher and $x[\beta n]$ is the function $x[n]$ with a time scale compressed by a factor β. Contrarily, If $\beta < 1$, the pitch becomes lower and $x[\beta n]$ is an expansion of $x[n]$.

For a simple echo hiding, an echo kernel with only one echo is represented as

$$h[n] = \delta[n] + \alpha\delta[n-d], \tag{2}$$

where $\delta[n]$ is the Kronecker delta function, and α (< 1) is the amplitude of the embedded echo. The delay offset d of the echo signal is determined by the values of the hidden watermark data $I \in \{0, 1\}$. The watermarked audio signal $y[n]$ can thus be obtained by convolving the echo kernel $h[n]$ and the host signal $x[n]$. If the watermarked audio signal is modified by pitch-scaling with pitch scaling-factor β, the watermarked signals after attack, $\hat{y}$ can be written as

$$\hat{y}[n] = y[\beta n] = h[\beta n] * x[\beta n]. \tag{3}$$

Extraction of the embedded watermark data involves the detection of the echoes' positions by using the cepstrum of the watermarked audio signals. Using Eqs. (2) and (3), the cepstrum of the watermarked signals after a pitch-scaling attack can be derived as

$$\begin{aligned} c_{\hat{y}}[n] &= F^{-1}(\ln F(h[\beta n])) + F^{-1}(\ln F(x[\beta n])) \\ &= F^{-1}\left(\ln\left(\frac{1}{|\beta|}\left(1+\alpha e^{j\frac{\omega}{\beta}d}\right)\right)\right) + F^{-1}(\ln F(x[\beta n])) \\ &= F^{-1}\left(\ln\left(\frac{1}{|\beta|}\right)\right) + \alpha\delta\left[n-\frac{d}{\beta}\right] - \frac{\alpha^2}{2}\left[n-\frac{2d}{\beta}\right] + \ldots \\ &\quad + F^{-1}(\ln F(x[\beta n])). \end{aligned} \tag{4}$$

Notably, the cepstrum value α of the kernel is located at the delay offset d/β instead of d as shown in the second term of Eq. (4). Therefore, if the watermark data is extracted by checking the cepstrum value at the delay offset d, which is the position of echo embedded in the embedding process, it would lead to an erroneous result. Because the pitch-scaling factor is unknown *a priori* in the watermark extracting process, the correct position is unable to determine in advance. In the following, a robust method is presented to solve this problem.

3. PROPOSED ECHO WATERMARKING SCHEME

The embedding process of the proposed analysis-by-synthesis echo watermarking scheme is shown in Fig. 1. The host audio signals are partitioned into several segments by means of the frequency hopping scheme before the watermark embedding process to effectively embed many watermark data [11]. Owing to the frequency hopping scheme, the security of the proposed scheme can be enhanced. Additionally, the difficulty of embedding echoes in long silent intervals of the host signals can be overcome simultaneously. Based on the analysis-by-synthesis approach, the embedded data are extracted from the watermarked audio signals during the watermark embedding process to determine whether they can be accurately recovered in the decoding process. The amplitudes of the embedded echoes are adapted according to the properties of the host audio signals to ensure that the embedded data are recovered accurately with minimal impact on the host audio signals [8]-[11]. The scheme of interlaced kernels is employed to embed the echo signals. Figure 2 depicts the impulse responses of the "Zero" and "One" kernels which are adopted to embed the binary watermark data "Zero" and "One", respectively [10]. The positions of embedded echoes in the front and rear parts of a segment are interchanged.

In the watermark extracting process, the cepstrum values used to determine the watermark data are not calculated at specific positions in which echo signals are embedded. Because the pitch-scaling attack causes the positions of embedded echoes shifting in the cepstrum domain, the cepstrum values in a reasonable range of delay offsets are calculated. This range of delay offsets can be given by a value to accommodate the degree of the pitch-scaling effect that does not seriously degrade audio quality of watermarked signals. If the pitch-scaling attack changes the watermarked audio signals by $\pm\delta$ semitones at most, then the corresponding changes for the embedded echo's position d_{a_P} are

$$\begin{cases} d_{a_P}^U = \left\lceil 2^{\frac{\delta}{12}} d_{a_P} \right\rceil \\ d_{a_P}^L = \left\lfloor 2^{\frac{-\delta}{12}} d_{a_P} \right\rfloor \end{cases}, \tag{5}$$

where $\lceil\cdot\rceil$ and $\lfloor\cdot\rfloor$ round to the nearest upper and lower integers, respectively, $d_{a_P}^U$ and $d_{a_P}^L$ are the upper and lower bounds of the delay offsets changed by the pitch-scaling attack, respectively. The range of the delay offset d_{a_P} is thus set as

$$\mathbf{D}_{\mathbf{a_P}} = [d_{a_P}^1, d_{a_P}^2, \cdots, d_{a_P}^m] = [d_{a_P}^L, d_{a_P}^L + 1, \cdots, d_{a_P}^U - 1, d_{a_P}^U]. \tag{6}$$

The corresponding changes of the pitch-scaling factor in semitone are $\mathbf{\Delta} = [\delta^1, \delta^2, \cdots, \delta^m]$, where

$$\delta^i = 12\log_2\left(\frac{d_{a_P}^i}{d_{a_P}}\right), \quad i = 1,2,\cdots,m. \tag{7}$$

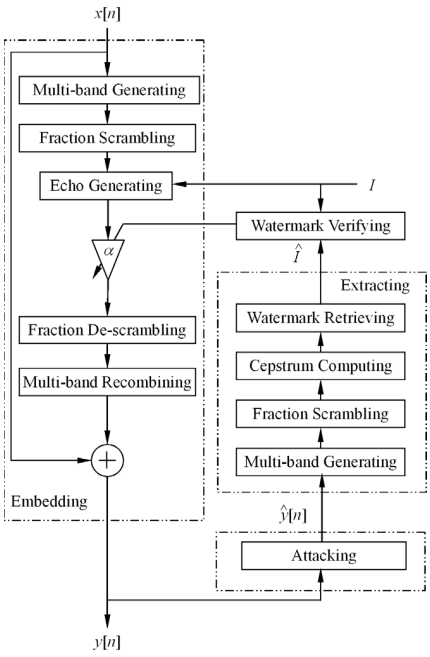

Fig. 1 Block diagram of the proposed watermark embedding process.

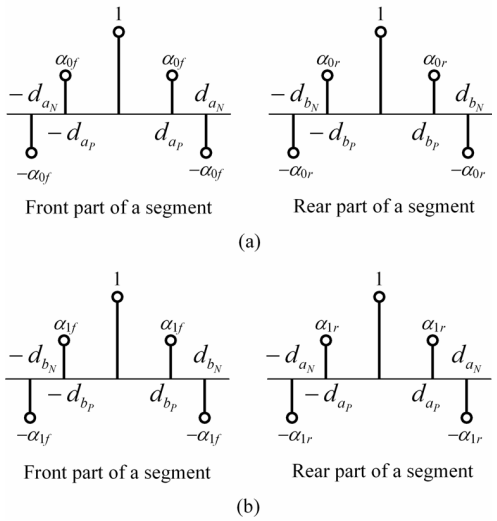

Fig. 2 Impulse responses of the interlaced kernels. (a) "Zero" kernel. (b) "One" kernel.

According to the range of the pitch-scaling factor, Δ, the element in the delay offset $\mathbf{D}_{\mathbf{a_N}}$ can be calculated as

$$d_{a_N}^{i} = \left\lfloor 2^{\frac{\delta^i}{12}} d_{a_N} \right\rfloor, \quad i = 1,2,\cdots,m. \tag{8}$$

The ranges of delay offsets $\mathbf{D}_{\mathbf{b_P}}$ and $\mathbf{D}_{\mathbf{b_N}}$ can be obtained correspondingly.

Next, the cepstrum values of watermarked audio signals in the *l*-th segment are calculated at delay offsets corresponding to the range of the pitch-scaling factor.

$$\mathbf{C}_{\hat{\mathbf{y}}}^{\mathbf{l}} = [c_{\hat{y}}^{l1}, c_{\hat{y}}^{l2}, \cdots, c_{\hat{y}}^{lm}], \tag{9}$$

To minimize the influence of host signals' cepstrum, the moving average function is conducted on the cepstrum values over several segments,

$$\tilde{c}_{\hat{y}}^{li} = \frac{1}{M} \sum_{k=l-M/2}^{k=l+M/2} c_{\hat{y}}^{ki}, \quad i = 1,2,\cdots,m, \tag{10}$$

where M is the order of moving average process. The delay offset ξ^l in which the peak value of cepstrum occurs at the *l*-th segment is determined by

$$\xi^l = \arg\max_i \left(\left| \tilde{c}_{\hat{y}}^{li} \right| \right) \tag{11}$$

The embedded watermark data in the *l*-th segment is extracted as follows:

$$\hat{I}^l = \begin{cases} 0 & if\ \tilde{c}_{\hat{y}}^{l\xi} \geq 0 \\ 1 & if\ \tilde{c}_{\hat{y}}^{l\xi} < 0 \end{cases}. \tag{12}$$

By using the abovementioned extracting process, the proposed analysis-by-synthesis approach can correctly determine the watermark data, although the pitch-scaling attack is imposed.

4. EXPERIMENTAL RESULTS

Six audio pieces are used to examine the performance of the proposed scheme. Each audio piece with 2,400,000 samples is sampled at 48 kHz and quantized by 16 bits. In the embedding process, the host audio signals are divided into many segments each of which has 4,410 samples. Each segment of the host audio signals is windowed with a Hanning window to smooth the boundaries of the segments. The delay offsets of the embedded echoes to represent the binary watermark data are set as follows: $d_{a_P} = 100$, $d_{a_N} = 103$, $d_{b_P} = 110$, and $d_{b_N} = 113$. The proposed Robust Echo Hiding scheme using Analysis-By-Synthesis (REHABS) was compared with the two conventional schemes: the Echo-Hiding scheme using the Positive and Negative echo kernels (EHPN) [4] and the Echo-Hiding scheme using the Positive, Negative, Backward, and Forward echo kernels (EHPNBF) [5]. Six types of attacks listed in Table 1 are applied on the watermarked signals to test the robustness performance of watermarking schemes. Generally, pitch-scaling attacks seriously deteriorate audio quality. Human can easily perceive the degradation on audio quality when the pitch of audio signals is altered by 3 semitones. Hence, in the proposed extracting process, the range of delay offsets is set with considering pitch scaling with ±3 semitone changes, ranging from 84 to 119 for the delay offset d_{a_P}. Additionally, the order of the moving average function is set to 30.

Table 1 Types of attacks and their conditions.

Attacks	Conditions
No Attack	
MPEG 1 Layer III	64 kbps
Quantization	8 bit
Random Noise	SNR = 20dB
Band-Pass Filtering	100Hz ~ 10 kHz
Pitch-Scaling	Increasing 2 semitones

Table 2 Robust performance of the proposed and conventional schemes with equivalent echo amplitudes smaller than or equal to 0.2.

Attacks \ Schemes	EHPN α_{PB} =0.2 α_{NB} =0.2	EHPNBF $\alpha_{PB}=\alpha_{NB}$ =0.1 $\alpha_{PF}=\alpha_{NF}$ =0.1	EHABS α_{0f}, $\alpha_{1f}\le$0.1 α_{0r}, $\alpha_{1r}\le$0.1	REHABS α_{0f}, $\alpha_{1f}\le$0.1 α_{0r}, $\alpha_{1r}\le$0.1
No Attack	90.9%	91.5%	98.7%	98.7%
MP3	87.5%	87.9%	94.3%	92.9%
Quantization	81.0%	81.1%	90.6%	86.2%
Random Noise	70.1%	70.9%	80.2%	67.8%
BP Filtering	86.7%	88.2%	95.6%	94.6%
Pitch-Scaling	59.3%	58.7%	64.0%	78.5%

Table 3 Robust performance of the proposed and conventional schemes with equivalent echo amplitudes smaller than or equal to 0.4.

Attacks \ Schemes	EHPN α_{PB} =0.4 α_{NB} =0.4	EHPNBF $\alpha_{PB}=\alpha_{NB}$ =0.2 $\alpha_{PF}=\alpha_{NF}$ =0.2	EHABS α_{0f}, $\alpha_{1f}\le$0.2 α_{0r}, $\alpha_{1r}\le$0.2	REHABS α_{0f}, $\alpha_{1f}\le$0.2 α_{0r}, $\alpha_{1r}\le$0.2
No Attack	96.7%	97.9%	99.7%	99.7%
MP3	95.6%	96.5%	98.6%	98.6%
Quantization	90.2%	91.7%	96.6%	95.6%
Random Noise	80.4%	82.0%	88.9%	85.8%
BP Filtering	95.8%	96.6%	99.1%	99.1%
Pitch-Scaling	64.0%	64.2%	66.9%	95.4%

Tables 2 and 3 list the robustness performance of the proposed and conventional echo hiding schemes with equivalent echo amplitudes smaller than and equal to 0.2 and 0.4, respectively. Simulation results reveal that the recovery accuracy rate of the proposed scheme under various attacks is much higher than those of conventional schemes. Of these attacks, the pitch-scaling attack most reduces the recovery accuracy. However, as shown in the last column in Tables 2 and 3, the proposed scheme can effectively improve the performance against such an attack. Because the proposed watermark extracting process considers a wide range of delay offsets to determine the location of the peak cepstrum for the case of the pitch-scaling attack, it is unavoidable to little lower the recovery accuracies of the other attacks as compared to our previous scheme (EHABS) only using the embedding positions for extracting. This situation is ameliorated when the equivalent echo amplitude is increased. Table 3 lists that the recovery accuracy of the proposed scheme can greatly increase for the pitch-scaling attack and very slightly decrease for other attacks. Certainly, if the range of delay offsets is lessened, the degradation of recovery accuracy by additional searching at the other attacks can be minimized.

5. CONCLUSION

This work presents a robust echo watermarking scheme. In the extracting process, the cepstrum values in a reasonable range of delay offsets are considered to overcome the pitch-scaling attack. Additionally, the moving average function is performed to alleviate the effect of host signals on the cepstrum. Under such approach, the watermarked data can be correctly extracted even that the positions of embedded echoes are altered by frequency-altering attacks. The simulation results reveal that the proposed robust watermarking scheme is superior to the conventional echo watermarking schemes in terms of robustness, especially for the case of the pitching-scaling attack.

REFERENCES

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

[2] I. J. Cox, J. Kilian, F. T. Leighton and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. on Image Processing*, vol. 6, pp. 1673-1687, 1997.

[3] M. Arnold, "Audio watermarking: feature, applications and algorithms," *Proc. of IEEE ICME*, vol. 2, pp. 1013-1016, 2000.

[4] 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 ICASSP*, pp. 1341-1344, 2001.

[5] 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.

[6] B.-S. Ko, R. Nishimura and Y. Suzuki, "Time-spread echo method for digital audio watermarking," *IEEE Trans. on Multimedia*, vol. 7, no. 2, pp. 212-221, 2005.

[7] B.-S. Ko, R. Nishimura and Y. Suzuki, "Log-scaling watermark detection in digital audio watermarking," *Proc. of IEEE ICASSP*, vol. III, pp. 81-84, May 2004.

[8] 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 ICSIP*, pp. 365-369, Aug. 2003.

[9] W. C. Wu and O. T.-C Chen, "An analysis-by-synthesis echo watermarking method," *Proc. of IEEE ICME*, vol. 3, pp. 1935-1938, June 2004.

[10] W.-C. Wu and O. T.-C. Chen, "Analysis-by-synthesis echo hiding scheme using mirrored kernels," *Proc. of IEEE ICASSP*, vol. II, pp. 325-328, May 2006.

[11] W. C. Wu and O. T.-C Chen, "Analysis-by-synthesis echo hiding scheme using frequency hopping," *Proc. of IEEE ICME*, pp. 1766-1769, July 2007.