

# Digital Blind Audio Watermarking for eLearning Media

**Youngseock Lee**

Dept. of Digital Broadcasting Engineering, Chungwoon University  
San 29, Namjang-Ri, Hongsung-Kun, Hongsung-Eup, Chungnam 350-701, Korea  
yslee@chungwoon.ac.kr

**Jongweon Kim**

Dept. of Copyright Protection, Sangmyung University  
7, Hongji-dong, Jongno-gu, Seoul 110-743, Korea  
jwkim@smu.ac.kr

**Abstract**— Digital audio watermarking techniques are widely applied to protect the ownership of digital audio media. This study presents an audio watermarking scheme for audio e-learning contents. The proposed audio watermarking method can significantly reduce the bit error rates, so enhance the robustness against common attacks. The experimental results under common attacks demonstrate that our watermarking scheme can be applied to the ownership-proof and the protection of audio eLearning contents.

**Keywords**— audio e-learning contents, common attacks, Digital audio watermarking

## I. INTRODUCTION

Recently the rapid changes from a product-based to a knowledge-based society has resulted in an increased demand for knowledge workers who are capable of higher order thinking and reasoning to solve intricate problems in the work place. This requires organizations to educate and train anyone, anytime, and from anywhere [1].

For this task, asynchronous eLearning, defined as instructional content or learning experience delivered or enabled by electronic technologies including the Internet, intranets, and extranets, breaks the limitations of time and space and also creates many benefits, including reduced cost, regulatory

compliance, meeting business needs,. The impact of eLearning is real and it has received fairly extensive attention from practitioners and information system researchers. eLearning is reported to be a means of solving learning and performance problems and has become an increasingly critical issue.

As digital multimedia for eLearning (video, audio and images) become available for re-transmission, re-production, and publishing over the Internet, a real need for protection against unauthorized copy and distribution is increased. These concerns motivate researchers to find ways to forbid copyright violation.

The most of eLearning contents can be categorized into audio and video (including images) contents. Typically the audio contents play a major role and the video contents, a secondary role to eLearning subjects, i.e., the absence or distortion by malicious attacks of audio contents in eLearning system by the audio and video contents can cause eLearning subjects to misunderstand about eLearning contents.

Therefore at the viewpoint of information system researchers, the ownership-proof and the protection of audio eLearning contents must be heavily emphasized. The most promising solution for this challenging problem seems to lie in information hiding techniques. Information hiding is the process of embedding a message into digital media.

The embedded message should be imperceptible; in addition to that the fidelity of digital media must be maintained.

To protect eLearning audio contents, watermarking is very popular method to hide information about themselves in the audio contents.

In general, an effectiveness of audio watermarking scheme must satisfy the following requirements: (i) Imperceptibility: The quality of the audio should be retained after adding the watermark. Imperceptibility can be evaluated using both objective and subjective measures. According to the recommendation of IFPI (International Federation of the Phonographic Industry), a watermarked audio signal should maintain more than 20 dB SNR. (ii) Security: Watermarked signals should not reveal any clues about the watermarks in them. Also, the security of the watermarking procedure must depend on secret keys, but not on the secrecy of the watermarking algorithm. (iii) Robustness: Ability to extract a watermark from a watermarked audio signal after various signal processing attacks. (iv) Payload: The amount of data that can be embedded into the host audio signal without losing imperceptibility. For audio signals, data payload refers to the number of watermark data bits that may be reliably embedded within a host signal per unit of time, usually measured using bits per second (bps). There should be more than 20 bps data payload [2].

In next section, we shall show the related works in terms of audio watermarking. Then a new blind audio watermarking method for audio eLearning contents will be proposed to improve the (BERs) and robustness.

## II. RELATED WORKS

Most of the existing audio watermarking methods can be categorized into the spread-spectrum watermarking, phase coding watermarking and echo-hiding watermarking. The spread-spectrum

watermarking methods are generally robust to attacks but the watermarks embedded are audible, i.e., spread-spectrum watermarking adds audible noise to the audio signal. In addition to that, computation cost for synchronization is extremely high at the decoding process.

In contrast, the phase coding watermarking introduces less audible noise to the audio signal but is not robust against most of the common attacks. Most phase coding watermarking methods are non-blind, i.e., the original audio signal is needed at the decoder to extract watermarks from the watermarked signal.

Compared with the spread-spectrum watermarking and the phase coding watermarking, echo hiding watermarking has more balanced performance both in imperceptibility and robustness against attacks. Also, the watermarks can be easily embedded into the audio signal [3].

In echo-hiding watermarking, echo kernel plays a vital role and determines, to a large extent, the performance of watermarking. In single-echo watermarking, one large echo is used to hide single watermark, which results in low imperceptibility. In order to overcome this problem, [4] and [5] use multiple small echoes with different delays to hide a watermark. However, the detection rate of this approach is low. In [6], Oh et al. develop a different echo hiding scheme, which consists of positive and negative echo kernels, to enhance imperceptibility.

Moreover, since slightly smaller echo kernels can be used in this scheme, marginal improvement of imperceptibility can be achieved. For the watermarking schemes in [4]-[8], security is a major problem since none of them uses secret key. Pirates with basic knowledge in audio watermarking can easily identify the embedded watermarks and thus remove (or disable) them from the watermarked signal.

The time-spread echo method in [9] significantly enhances security by using a

pseudonoise (PN) sequence as a secret key. That is, watermarks embedded in the audio signal cannot be extracted without knowing the PN sequence.

In this paper, we will propose the new blind echo hiding watermarking model for audio e-learning contents to improve the robustness and imperceptibility under common attacks.

### III. PROPOSED AUDIO WATERMARKING MODEL

In echo hiding systems [9], each original audio signal segment is convolved with a kernel signal to make watermarked signals. The kernel for time spread echo hiding system is as

$$h(n) = \delta(n) + \alpha \cdot p(n-d)$$

$p(n)$  is a pseudo-random noise whose amplitude is  $\pm 1$ ,  $\delta(n)$  is Dirac delta function,  $\alpha$  is a small value as echo coefficients and  $d$  is a delay that is selected between two values corresponding to one or zero bit embedding. By this kernel the watermarked signal is a copy of echo of original signal.

If we use  $x(n)$

$$y(n) = x(n) + \alpha \sum_{i=0}^N [p(i)x(n-d-i)] \quad (2)$$

$N$  is the  $p(n)$  length. Here if we define the complex cepstrum transform as below,

$$c_y(n) = F^{-1}(-1) \left( \log |F(y[n])| \right) \quad (3)$$

We can use this transform for discovering the watermarking at the receiver [8]:

$$c_y(n) = c_x(n) + \alpha \cdot p(n-d) \quad (4)$$

if  $\alpha \ll \frac{1}{\text{Max}(|p(n)|)}$

$p(n)$  is the Fourier transform of  $p(n)$ . After generating right  $p(n)$  by authorized receiver,

the final step is to take a cross correlation between  $c_y(n)$  and  $p(n)$

$$\begin{aligned} cc(n) &= c_y(n) \otimes p(n) \\ &= n_y(n) + \alpha \cdot p(n) \cdot p(n-d) \\ n_y(n) &= c_x(n) \otimes p(n). \end{aligned} \quad (5)$$

Here  $\otimes$  is a symbol for cross correlation and  $cc(n)$  presumably has a peak at  $d$ , so the receiver decides the embedding bit based on the delay that is corresponding to this peak. Here two drawbacks to solve are the first term of (4) that may make the detection process erroneously and another problem is comprehensible that watermark is not embedded into the whole of signal. To solve these two problems, in [10], they proposed a novel time spread echo hiding model. The proposed model is the modified form of encoder kernel of (2),

$$h'(n) = \delta(n) + \alpha \cdot b \cdot p(n) \quad (6)$$

$$y(n) = x(n) + \alpha \cdot b \sum_{i=0}^N [p(i)x(n-i)] \quad (7)$$

Unlike the conventional time spread echo hiding, the  $p(n)$  sequence is embedded to the entire original audio signal from first bit to the end.  $b \in \pm 1$  is the bit to be embedded and decoded in the decoder,  $N$  is the audio signal length. After applying the real cepstrum transform in (8) to the watermarked signal, we will have (9),

$$\text{Recep}y(n) = F^{-1}(\log_e |F(y[n])|) \quad (8)$$

$$\begin{aligned} \text{Recep}y(n) &= \text{Recep}x(n) \\ &+ \frac{\alpha}{2} P(n) + \frac{\alpha}{2} P(-n) \end{aligned} \quad (9)$$

The normalized correlation amount of (10) is used for detecting the watermarking bits. The correlation amount in the last equation of (10) have two terms, left term that is a noise section due to the original signal effect in the detector and is considered

as the source of error in detection process and the right term that the watermark bit  $b$  is in it.

$$C = \frac{1}{N} \sum_{n=1}^N R_{ceps}x(n) + \frac{1}{2} \alpha \cdot b \quad (10)$$

This system is considerably different from that of time spread echo hiding. In the encoder the watermark is spread into the whole of the signal and the watermark bit is a sign bit, not a special delay. The system decoder is relied on a correlation amount, instead of a peak at the decoder. The left term of the decoding equation (10) is the main source of error in the detection process, even in the no-attacks environments. In [10], they proposed the echo hiding model to remove main source error in left term of (10) as

$$y(n) = x(n) + (\alpha \cdot b - \gamma) \sum_{i=0}^N [p(i)x(n-i)] \quad (11)$$

$$\gamma = \frac{1}{N} \sum_{n=1}^N R_{ceps}x(n) \cdot p(n).$$

From (11), their normalized correlation amount is

$$C = \frac{1}{2} \alpha \cdot b. \quad (12)$$

But their normalized correlation amount of (11) is wrong in our calculation. In the echo hiding model in (7), the error source term of the normalized correlation amount  $C$  is not eliminated and  $C$  is a function of  $\gamma$ . The echo hiding model in (7) can be rewritten by convolution operator  $*$  (

$$y(n) = x(n) + (\alpha b - \gamma)[p(n) * x(n)], \quad (13)$$

and its cepstrum is

$$R_{ceps}y(n) = R_{ceps}x(n) + (\alpha b - \gamma) \times \left[ \frac{1}{2} p(n) + R_{ceps}x(n) \right]. \quad (14)$$

And new correlation amount  $C'$  is following as

$$C' = -\gamma^2 - (\alpha b + 0.5)\gamma + 0.5 \alpha b. \quad (15)$$

New correlation amount  $C'$  is the quadratic form of  $\gamma$

$\gamma = 0.5$  is a trivial solution by (11), i.e.,  $\gamma = 0.5$

$$\gamma = -\frac{\alpha \cdot b + 0.5}{2}. \quad (16)$$

#### IV. EXPERIMENTAL RESULTS

In his section, the performances of the proposed echo hiding watermarking method by new decoder in (15) are evaluated and compared with the performances of [9]. Five random selected audio clips in eLearning contents were used, which consist of music (vocal and instruments) and speeches. All audio clips were sampled at 44.1 kHz with 16bit quantization and the length of segment was 44100 samples for each clips.

We first assumed that the watermarked audio signal was not under any attacks. Table I shows that the bit error rates (BERs) of our scheme and method [9] versus echo coefficient parameter  $\alpha$

TABLE I  
BER VERSUS  $\alpha$

$\alpha$	BER (%)	
	Proposed	Method in [9]
0.001	12.5	19.6
0.002	10.4	16.5
0.003	9.5	12.7
0.004	8.6	11.9
0.005	7.7	10.6
0.006	6.5	9.3
0.007	5.3	8.2
0.008	3.1	7.2

As expected, BERs are anti-proportional to value of  $\alpha$ . However, our scheme has lower BER than the one in [9], regardless of the value of  $\alpha$ . Then we tested the robustness of these techniques against some common attacks in Table II.

**TABLE II**  
SOME COMMON ATTACKS

Name of attacks	Description
Re-sampling	Watermarked signals were re-sampled at 16 kHz
Re-quantization	Watermarked signals were quantized with 8 bits
Time-scaling	15% (speed up)
MP3 attack	Compressing the watermarked signal by Mpeg-3 layer1 (lossy compression)

**TABLE III**  
ROBUSTNESS AGAINST COMMON ATTACKS

Attacks	BER (%)	
	Proposed	The method in [9]
Re-sampling	45.2	55.5
Re-quantization	57.4	62.9
Time-scaling	14.6	18.5
MP3 attack	42.3	49.2

It can be seen from Table III that the proposed watermarking scheme achieves higher BERs under all attacks than the time-spread echo method in [9].

## V. CONCLUSIONS

In this paper, we propose a new echo hiding audio watermarking scheme based on the modified correlation amount. It is shown that the proposed watermarking method can significantly reduce the BERs, thus enhances the robustness of our watermarking scheme. The experimental results demonstrate that our watermarking scheme can be applied to the ownership- proof and the protection of audio eLearning contents.

## ACKNOWLEDGMENT

This research is supported by the Ministry of Culture, Sports and Tourism (MCST) and KOrea Culture Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program 2009.

## REFERENCES

- [1] S. Downes. (2005) Univ. of Colombo homepage. [Online]. Available: [http://www.cmb.ac.lk/news-letter/ext\\_pages/Vlc/E-learning%202.pdf](http://www.cmb.ac.lk/news-letter/ext_pages/Vlc/E-learning%202.pdf).
- [2] V. Bhat K, I. Sengupta, and A. Das, "An adaptive audio watermarking based on singular value decomposition in the wavelet domain," *Digital Signal Processing*, vol. 20, pp. 1547-1558, 2010.
- [3] I. Natgunanathan, Y. Xiang, "A novel pseudonoise sequence for time-spread echo based audio watermarking," in *Proc. GLOBECOM '2009*, 2009.
- [4] C. Xu, J. Wu, Q. Sun, and K. Xin, "Applications of digital watermarking technologies in audio signals," *J. Audio Eng. Soc.*, vol. 47, no. 10, Oct. 1999.
- [5] H. Oh, J. W. Seok, J. W. Hong, and Y. Suzuki, "New echo embedding technique for robust and imperceptible audio watermarking," *IEEE Trans. Multimedia*, vol. 7, no. 2, pp. 212-221, Apr. 2005.
- [6] H. O. Oh, H. W. Kim, J. W. Seok, J. W. Hong, and D. H. Yoon, "Transparent and robust audio watermarking with new echo embedding technique," in *Proc. IEEE Int. Conf. Multimedia Expo., 2001*, pp. 433-436.
- [7] Y. Yardymcy, A. E. Cetin, and R. Ansari, "Data hiding in speech using phase coding," in *Proc. Eurospeech 97*, vol. 3, 1997, pp. 1679-1682.
- [8] H. J. Kim, Y. H. Choi, "A novel echo-hiding scheme with backward and forward kernels," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 13, no. 8, pp. 885-889, Aug 2003.
- [9] B. S. Ko, R. Nishmura, and Y. Suzuki, "Time-spread echo method for digital audio watermarking," *IEEE Trans. Multimedia*, vol. 7, no. 2, pp. 212-221, Apr 2005.
- [10] Y. Erfani, S. Siahpoush, "Robust audio watermarking using improved TS echo hiding," *Digital Signal Processing*, vol. 19, pp. 809-814, 2009.