AN ECHO-HIDING WATERMARKING TECHNIQUE BASED ON BILATERAL SYMMETRIC TIME SPREAD KERNEL

*S.A. Chou and S.F. Hsieh**[1]

Department of Communication Engineering
National Chiao Tung University Hsinchu, Taiwan 300 Republic of China
Tel: 886-3-5731974, E-mail: dbchou@gmail.com, sfhsieh@mail.nctu.edu.tw

ABSTRACT

Echo hiding is an audio watermarking method where information is embedded in the echo delay. A bilateral time spread echo kernel is proposed to improve robustness against malicious attacks as well as enhance the detection performance by increasing the peak value of the kernel cepstrum at the echo delay time. Based on the spreading property of PN sequences and expansions of logarithm and binomial, a closed-form detection gain formula is derived analytically. Computer simulation confirms the proposed method is effective and outperforms the previous unilateral method.

1. INTRODUCTION

Recently, digital copyright protection has been of much interest due to the widely illegal use of the multimedia data. There have been many audio watermarking schemes proposed to protect the intellectual property rights by taking advantages of the perceptual masking on different domain to maintain imperceptibility and ensure the security [1][2]. Among them, echo hiding [3] embeds watermark information bit in the delay time, as human hearing is not sensitive to the echoes in a short time. However, this kind of echo hiding scheme without time spreading is vulnerable to malicious attack [2][4]. A unilateral time spread pseudo-random noise (PN) kernel [4] avoids this attack by spreading the echoes using a PN sequence. By adding imperceptible pre-echoes, a bilateral PN time spread echo kernel is proposed to further improve the detection gain, whose closed-form formula is also derived theoretically and shown to outperform the previous unilateral kernel approach.

In Section 2, we introduce our system model. In Section 3, a new bilateral time spread echo hiding kernel is proposed and its effectiveness is proved analytically. Simulations support our analysis in Section 4 and conclusion will be given in Section 5.

2. SYSTEM MODEL

The echo hiding scheme includes convolutional encoding to embed the watermark in the PN kernel delay and cepstrum decoding to recover the watermark.

2.1. Convolutional Encoding

The watermarked signal $y(n)$ is the convolution of the original audio $s(n)$ with the kernel $h(n)$. The embedding process can be written as

$$y(n) = s(n) \otimes h(n), \tag{1}$$

where $\otimes$ means convolution. The bilateral symmetric time spread echo kernel in Fig. 1 is given by

$$h(n) = \delta(n) + \alpha p(n-d) + \alpha p(n+d), \tag{2}$$

where $\delta(n)$ is an unit impulse, α is the echo gain , and the echo delay d can be either d_0 or d_1, which is usually about 100-150 samples [6], depending on the watermark bit is 0 or 1. $p(n)$ is a pseudorandom sequence of a sufficient large length L,

$$p(n) = \sum_{k=0}^{L-1} p_k \delta(n-k), \text{ where } p_k \in +1, -1, \tag{3}$$

with an important spreading property

$$p(n) \otimes p(-n) \cong L \cdot \delta(n) \text{ or } P(w) \cdot P^*(w) \cong L, \tag{4}$$

where $P(w)$ is the Fourier transform of $p(n)$, and $P^*(w)$ is the conjugation of $P(w)$. To ensure the security of our echo hiding scheme, the length L has to be large enough to maintain the robustness.

[1] This work was supported by the National Science Council under Grant NSC 94-2213-E-009-063 and MediaTek.

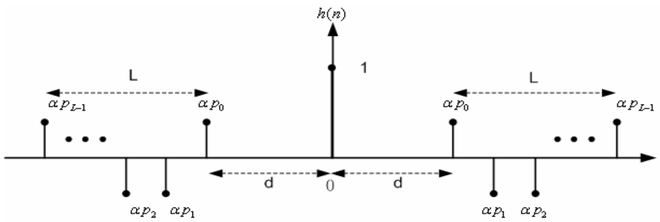


Fig. 1. Bilateral time spread echo kernel.

Besides, using (4), a series convolution of $p(n)$ i times and $p(-n)$ j times can be written as

$$p^{\otimes i}(n)\otimes p^{\otimes j}(-n)=\begin{cases} p^{\otimes(i-j)}(n)\otimes\left[L^j\cdot\delta(n)\right]=L^j\cdot p^{\otimes(i-j)}(n) & ,i>j \\ L^i\cdot\delta(n) & ,i=j \\ p^{\otimes(j-i)}(n)\otimes\left[L^i\cdot\delta(n)\right]=L^i\cdot p^{\otimes(j-i)}(-n) & ,i<j \end{cases}.$$

(5)

Our kernel in Fig. 1 is composed of three major parts. The first one is a unit single impulse at the origin, which can duplicate the original audio signal without any delay after convolution. The causal delay $\alpha p(n-d)$ contributing a chain of post-echoes $\alpha\sum_{k=0}^{L-1}p_k\cdot s(n-d-k)$ after the delay time *d* [4]. We propose to add the anti-causal delay $\alpha p(n+d)$ which is symmetrical to its causal delay counterpart. With these added pre-echoes, the detection gain can be improved significantly as will be seen later.

2.2. Cepstrum Decoding

In the decoding stage of echo hiding system in Fig. 2, cepstrum analysis is often used because the echo delay embedded in the watermarked signal can be efficiently recognized in the cepstrum domain at the delay time d_0 or d_1 [4].

First, the cepstrum of the watermarked signal $y(n)$ is given by

$$\xi_y(n)=F^{-1}\left\{\log\left\{F\left\{y(n)\right\}\right\}\right\}, \tag{6}$$

where F means the Fourier transformation and F^{-1} means the inverse Fourier transformation. From (1), the cepstrum of the watermarked signal can be written as

$$\xi_y(n)=\xi_s(n)+\xi_h(n) \tag{7}$$

where $\xi_s(n)$ and $\xi_h(n)$ denote the cepstrum of original audio signal *s(n)* and the kernel *h(n)*, respectively. The decoded cepstrum $\tilde{\xi}_y(n)$ is computed by correlation of $\xi_y(n)$ and $p(n)$, denoted by

$$\tilde{\xi}_y(n)=\xi_y(n)\otimes p(-n). \tag{8}$$

There will be a peak value for $\tilde{\xi}_y(n)$ at delay *d* as shown in Fig. 3.

Fig. 3 displays an example of $\tilde{\xi}_y(n)$ and $\xi_y(n)$. Fig. 3(a) shows $\xi_y(n)$ and we note that there is a train of sequence $g\cdot p(n-d)$ after the echo delay *d*, where *g* is a very small echo gain to be shown later. We can see that time spread echoes effectively disguise the watermarked delay *d* from hackers' attack. Fig. 3(b) and 3(c) show the decoded cepstra $\tilde{\xi}_y(n)$ with watermark bit 0 or 1, respectively. We can easily find that there is indeed a peak value located at d_0 =*100* and d_1 =*110* samples,. The procedure of deciding the watermark bit is to compare the value of $\tilde{\xi}_y(d_0)$ and $\tilde{\xi}_y(d_1)$.

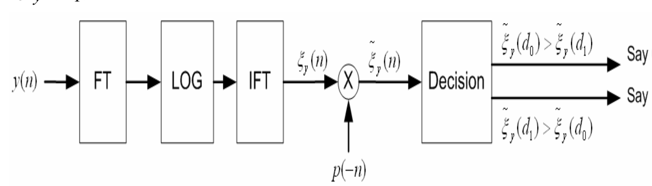
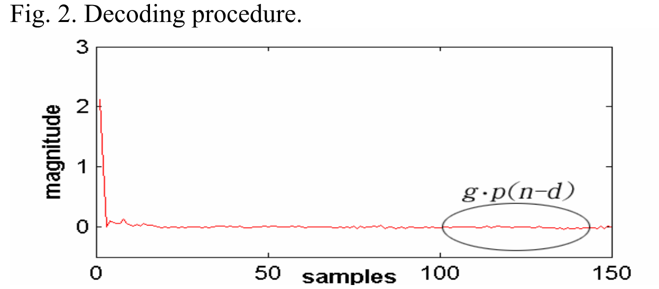
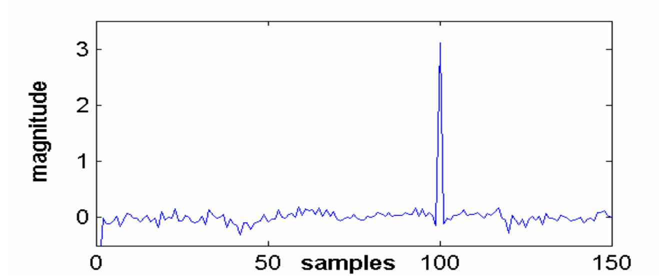


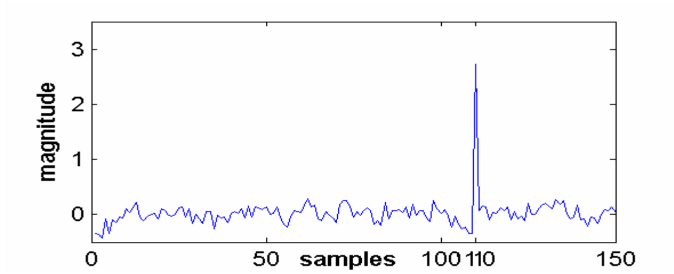
Fig. 2. Decoding procedure.



(a) watermarked cepstrum $\xi_y(n)$



(b) decoded cepstrum $\tilde{\xi}_y(n)$ with *d*= d_0 =*100*



(c) decoded cepstrum $\tilde{\xi}_y(n)$ with $d = d_1 = 110$

Fig. 3. Detection example of watermarked cesptrum $\xi_y(n)$ and decoded cepstra.

3. CEPSTRUM OF THE KERNEL

Now we will proceed to derive the cepstrum of the kernel $\xi_h(n)$ to clarify previous example in Fig. 3. In the beginning, the cepstrum of *h(n)* is

$$\xi_h(n) = F^{-1}\left\{\log\left\{F\left\{h(n)\right\}\right\}\right\}$$
$$= F^{-1}\{\log\{1+\alpha e^{-jwd}P(w)+\alpha e^{jwd}P^*(w)\}\} \quad (9)$$

Using the power series expansion of the logarithm

$$\log(1+x) = \sum_{m=1}^{\infty}\frac{1}{m}(-1)^{m+1}x^m \quad (10)$$

with $x = \alpha e^{-jwd}P(w)+\alpha e^{jwd}P^*(w)$, (9) can be written as follows

$$\xi_h(n) = \sum_{m=1}^{\infty}\xi_{h,m}(n) \quad (11)$$

where

$$\xi_{h,m}(n) = \frac{1}{m}(-1)^{m+1}\alpha^m F^{-1}\left\{(e^{-jwd}P(w)+e^{jwd}P^*(w))^m\right\} \quad (12)$$

is the *m*-th term of the kernel cepstrum spread by the power series of the logarithm. Note that the expansion in (10) is only valid when $\left|\alpha e^{-jwd}P(w)+\alpha e^{jwd}P^*(w)\right| < 1$.

By binomial expansion, (12) can be further expanded into

$$\xi_{h,m}(n) = \frac{1}{m}(-1)^{m+1}\alpha^m\sum_{l=0}^{m}C_l^m F^{-1}\left\{(e^{-jwd})^{m-2l}P^{m-l}(w)(P^*(w))^l\right\}$$
$$= \frac{1}{m}(-1)^{m+1}\alpha^m\sum_{l=0}^{m}C_l^m\delta(n-d(m-2l))\otimes p^{\otimes(m-l)}(n)\otimes p^{\otimes l}(-n). \quad (13)$$

We will focus on the coefficient at the first echo delay *d* that bears the watermark shown in Fig. 3(a), i.e.,

$$m-2l = 1\text{, or } l = \frac{m-1}{2}. \quad (14)$$

m is therefore an odd positive integer which implies that only the odd terms will contribute to this first echo delay. For simplicity, we assume that the other echo delays with $m-2l \neq 1$ can be neglected for a nearly ideal PN sequence with a very small echo gain α. According to (5) and (14), the coefficients in (13) of our concern at the first echo delay can be written as

$$\xi_{h,m}^{(d)}(n) = \left[\frac{1}{m}\alpha^m C_{\frac{m-1}{2}}^m L^{\frac{m-1}{2}}\right]p(n-d) \quad (15)$$

With *m* being an odd integer, the kernel's cepstrum of our interest at delay d becomes

$$\xi_h^{(d)}(n) = \sum_{m:odd}\xi_{h,m}^{(d)}(n) = \left[\sum_{r=0}^{\infty}\frac{1}{2r+1}\alpha^{2r+1}C_r^{2r+1}L^r\right]p(n-d)$$
$$= g\cdot p(n-d) \quad (16)$$

where *g* is a small gain in Fig. 3(a). (16) clearly shows that the kernel cepstrum includes a PN sequence at the echo delay. If we correlate $\xi_h(n)$ with $p(n)$ as shown in Fig. 2, the detection gain can be defined as the peak value at delay time *d*,

$$\beta = \left\{\xi_h^{(d)}(n)\otimes p(-n)\right\}_{@n=d}$$
$$= \sum_{r=0}^{\infty}\frac{1}{2r+1}\alpha^{2r+1}C_r^{2r+1}L^{r+1} = gL. \quad (17)$$

Notice that this closed-form detection gain formula can be increased either by increasing the echo gain α or the PN sequence length *L*. As a result, the embedded watermark can be detected more easily at the risk of perceptible audio echoes in the watermarked signal $y(n)$.

The detection gains for conventional methods can be easily obtained by setting the echo kernel as $h_{uni}(n) = \delta(n)+\alpha p(n-d)$ [4] or setting *L=1* [5]. By comparison, both $\beta_{uni} = \alpha L$ and $\beta_{L=1} = \sum_{r=0}^{\infty}\frac{1}{2r+1}\alpha^{2r+1}C_r^{2r+1}$ are smaller than the bilateral detection gain βin (17).

4. COMPUTER SIMULATIONS

For watermarking tests, we use five music clips with *44.1* KHz sampling frequency and *16*-bits quantization. The echo gain is α=0.006 and the length of the PN sequence is *L=511*. We embed bit "0" and "1" by setting the echo delay $d_0 = 100$ and $d_1 = 110$ samples, respectively. Thirty men and women were asked to distinguish between the watermarked audio and the original one. The chance of correctly identifying the watermarked audio is about *50%*, which means the watermarked audio is unrecognizable. The result is unsurprising, since the time spread echo hiding scheme can be rarely perceived [4][6].

Fig. 4 shows that the detection gain β increases with the echo gainα. We can see that the theoretically derived detection gain in (17) is very close to the simulated

Music type	Jazz		Male speech		Pops		Piano	
Method	B	U	B	U	B	U	B	U
No attack	0.99	0.88	0.88	0.85	0.92	0.79	0.92	0.74
Down sample	0.85	0.61	0.87	0.62	0.82	0.57	0.76	0.61
Mp3	0.90	0.74	0.88	0.72	0.93	0.70	0.85	0.72
Real audio	0.98	0.80	0.95	0.80	0.89	0.71	0.86	0.64
Requan-tization	0.98	0.84	0.88	0.73	0.85	0.63	0.66	0.49

Table 1. Probability of correct watermark detection of bilateral (B) and unilateral (U) time spread echo hiding schemes.

detection gain $\beta_{sim} = \left\{\xi_y^{(d)}(n) \otimes p(-n)\right\}_{@n=d}$. Besides, the bilateral time spread echo approach outperforms the unilateral case especially for a larger echo gain α. Next, we fix α=0.0156. Fig. 5 shows that as the PN sequence length L increases, the detection gain also improves and once again, the bilateral performance is better.

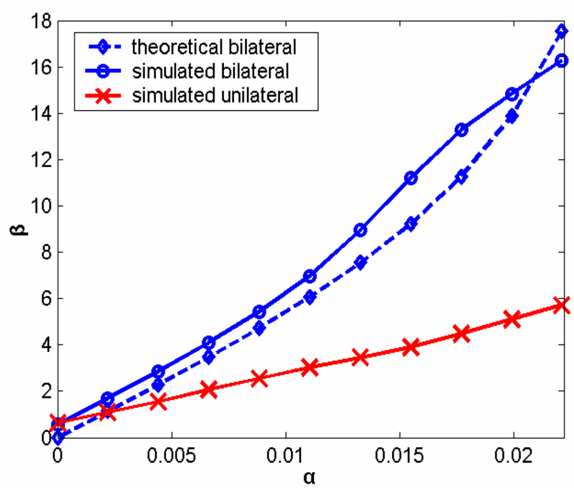


Fig. 4. Plots of detection gain β vs. echo gain α.

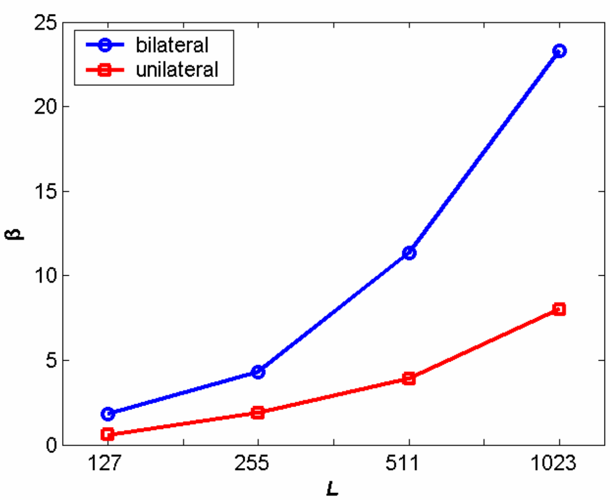


Fig. 5. Plots of detection gain β vs. PN sequence length L.

Some commonly used attacks are tested for our scheme, including downsampling, MP3 compression, real audio compression, and requantization. By fixingα=0.0156 and L=255 , the watermark bits are embedded into the music clips at a rate of 10 watermark bits per second. The probability of correct watermark detection is shown in Table 1 which confirms that the bilateral echo spread scheme is more robust to attacks due to its higher detection gain.

5. CONCLUSION

We have proposed a bilateral symmetric kernel for the time spread echo hiding scheme. Being an extension to previous methods, this bilateral PN kernel has the benefits of an improved detection gain and robustness to malicious attacks. Theoretical detection gain is derived and shown by simulation to outperform the conventional unilateral method.

6. REFERENCES

[1] C. I. Podilchuk, and E. J. Delp, “Digital watermarking algorithms and applications,” *IEEE Signal Processing Mag.*, vol 18, July 2001, pp. 33 – 46.

[2] F. A. P. Petitcolas, R.J. Anderson, and M. G.. Kuhn, “Information hiding: a survey,” in *Proc. IEEE*, vol. 87, July 1999, pp. 1062-1078.

[3] D. Gruhl, A. Lu, and W. Bender, “Echo hiding,” in *Pre-Proceedings: Information Hiding*, Cambridge, U.K., May 1996, pp. 295-316.

[4] B.-S. Ko, R. Nishimura, and Y. Suzuki, “Time-spread echo method for digital audio watermarking using pn sequences,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing,* vol. 2, 2002, pp. 2001-2004.

[5] H. J. Kim, and Y. H. Choi, “A novel echo hiding scheme with backward and forward kernels,” *IEEE Trans. Circuits Syst. Video Technol.,* vol.13, Aug. 2003, pp. 885-889.

[6] H. O. Oh, J. W. Seok, J. W. Hong, and D. H. Youn, “New echo embedding technique for robust and imperceptible audio watermarking,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol.3, May 2001, pp. 1341-1344.