A TEMPORAL MASKING TECHNIQUE AND ITS PERFORMACE ANALYSIS FOR AUDIO WATERMARKING

Shuang-An Chou, Shih-Fu Hsieh[1] *and Ko-Chiang Li*

DEPARTMENT OF COMMUNICATION ENGINEERING
National Chiao Tung University Hsinchu, Taiwan 300 Republic of China, Tel: 886-3-5731974
E-mail: michael.cm92g@nctu.edu.tw, sfhsieh@mail.nctu.edu.tw, calvin0050.cm94g@nctu.edu.tw

ABSTRACT

Echo hiding embeds watermark information bit in the delay time. For imperceptibility, we propose a temporal masking technique to match the psycho acoustic model. Based on the spreading property of PN sequences and expansions of logarithm and binomial, two asymptotic detection gains are derived depending on the echo gain is small or large. This performance analysis of the detection gain explains the improvement with an increment of a small echo gain and the degradation when a large echo gain is raised even higher. Computer simulation confirms our analysis.

1. INTRODUCTION

The recent growth of networked multimedia systems has increased the need for the protection of digital media. There have been many audio watermarking schemes proposed to preserve the intellectual property rights by taking advantages of the perceptual masking on different domains to keep imperceptibility and assure the security [1]-[2].

The echo hiding scheme embeds imperceptible watermark information bit in the delay time (echo), as human hearing is not susceptible to the echoes in a short time [3].To avoid attack by hackers, a time spread unilateral pseudo-random noise (PN) kernel [4] can be used. By adding imperceptible pre-echoes, a bilateral PN time spread echo kernel further improves the detection gain [5]. For imperceptibility, we propose to suppress the pre-echoes and shape the kernel to match the psycho acoustic model [6]. The detection gain, whose closed-form formula will be derived analytically.

In Section 2, a new bilateral time spread echo hiding kernel is proposed. In Section 3, after presenting the encoding and decoding process, two areas of echo gain for detection gain is found from the performance analysis. Simulation supports our analysis in Section 4. Conclusion is given in Section 5.

2. A TEMPORAL MASKING KERNEL

We know that the masking effect of the psychoacoustic model [6] has two phenomena. First, it is decaying with time. As the echo goes further away from the original audio, the strength of the mask becomes lower with time, just like an exponential decay. Second, the pre-masking decline faster than the post-masking.

In order to avoid a perceptible echo, we try to match the proposed kernel to the psychoacoustic masking model with three steps. First, we shape unilateral time spread pseudo-random noise (PN) kernel $h(n)$ by multiplying a shaping function *v(n)* which is decaying with time and delayed by d. We expand the pre-masking part with symmetry and suppress it by multiplying the suppression factor μ, where $0 \le \mu \le 1$. Finally, we add the unit impulse $\delta(n)$ into the kernel. The proposed kernel is given in Fig 1.

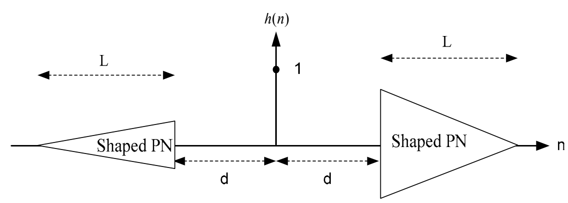

Fig. 1 Suppression shaped bilateral PN kernel

The kernel can be formulated as

$$h(n) = \delta(n) + v(n-d) \cdot \alpha\, p(n-d) + v(-n-d) \cdot \alpha\mu\, p(-n-d) \quad (1)$$

where 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. We can see that the bilateral symmetric time spread echo kernel(BTSEH) [5] is a special case with with $\mu = 1$ and $v(n) = 1$, implying $\varepsilon_v^2 = L$ without shaping. Similarly, the unilateral time spreading echo hiding kernel (UTSEH) [4] is the case with $\mu = 0$ and $v(n) = 1$.

The PN sequence of length L can be written as

[1] This work was supported by the NSC and MediaTek research center at Chiao Tung University.

$$p(n)=\sum_{k=0}^{L-1} p_k \delta(n-k) \quad \text{where }, p_k \in +1,-1 \tag{2}$$

with an important spreading property

$$p(n)\otimes p(-n) \cong L\cdot\delta(n) \text{ or } P(w)\cdot P^*(w)\cong L \tag{3}$$

v(n) is a decreasing window function which makes echoes more imperceptible. The energy of *v(n)* is defined as

$$\varepsilon_v^2 = \sum_{n=d+k}^{d+k+L-1} v^2(n) \tag{4}$$

From (2)-(4), we have

$$\begin{aligned} &v(n)p(n)\otimes v(-n)p(-n)\cong \varepsilon_v^2\cdot\delta(n) \\ &\text{or } V(w)\otimes P(w)\cdot V(w)^*\otimes P^*(w)\cong\varepsilon_v^2 \end{aligned} \tag{5}$$

where $P(w), V(w)$ is the Fourier transform of $p(n), v(n)$, and $P^*(w), V(w)^*$ is the conjugation of $P(w), V(w)$. $\otimes$ is the linear convolution operation.

Besides, using (5), a series convolution of $v(n)p(n)$ i times and $v(-n)p(-n)$ j times can be written as

$$v(n)^{\otimes i} p^{\otimes i}(n)\otimes v^{\otimes j}(-n)p^{\otimes j}(-n)\cong \begin{cases} v(n)^{\otimes(i-j)} p^{\otimes(i-j)}(n)\otimes\left[\varepsilon_v^{2j}\cdot\delta(n)\right], i>j \\ \varepsilon_v^{2i}\cdot\delta(n) \qquad , i=j \\ v(n)^{\otimes(j-i)} p^{\otimes(j-i)}(n)\otimes\left[\varepsilon_v^{2i}\cdot\delta(n)\right], i<j \end{cases} \tag{6}$$

These equations will be used later to find detection gain performance.

3. PERFORMANCE ANALYSIS

The encoding process is convolution of the audio s(n) with the kernel, and the watermarked signal will be

$$y(n)=s(n)\otimes h(n) \tag{7}$$

The decoding procedure in the first step is that we should transform the watermarked audio signal to cepstrum domain,

$$\xi_y(n)=F^{-1}\left\{\log\left\{F\left\{y(n)\right\}\right\}\right\}, \tag{8}$$

where F means the Fourier transformation and F^{-1} means the inverse Fourier transformation. From (7), the cepstrum of the watermarked signal can be written as

$$\xi_y(n)=\xi_s(n)+\xi_h(n) \tag{9}$$

where $\xi_s(n)$ and $\xi_h(n)$ denote the cepstrum of original audio signal s(n) and the kernel h(n), respectively.

3.1. Small echo gain

The next step will try to recognize where the peak value is. We call it the detection gain, which is the most important factor in $\xi_h(n)$ [5].

$$\begin{aligned} \xi_h(n) &= IFT\log\left\{FT\left[[\delta(n)+v(n-d)\cdot\alpha p(n-d) + v(-n-d)\cdot\alpha\mu p(-n-d)]\right]\right\} \\ &= IFT\left\{\log\left[1+\alpha e^{-jwd}V(w)\otimes P(w)+\alpha\mu e^{jwd}V^*(w)\otimes P^*(w)\right]\right\} \end{aligned}$$

(10)

With the logarithm expansion,

$$\log(1+x)=\sum_{m=1}^{\infty}\frac{1}{m}(-1)^{m+1}x^m \tag{11}$$

where

$$x=\alpha e^{-jwd}V(w)\otimes P(w)+\alpha\mu e^{jwd}V^*(w)\otimes P^*(w) \tag{12}$$

Note that the power series expansion in (11) will converge if

$$\left|\alpha e^{-jwd}V(w)\otimes P(w)+\alpha\mu e^{jwd}V^*(w)\otimes P^*(w)\right|<1 \tag{13}$$

It means the echo gain must be very small.

Applying the logarithm expansion, (9) can be written as

$$\xi_h(n)=\sum_{m=1}^{\infty}\xi_{h,m}(n) \tag{14}$$

where

$$\xi_{h,m}(n)=\frac{1}{m}(-1)^{m+1}\alpha^m\cdot IFT\begin{Bmatrix} e^{-jwd}(V(w)\otimes P(w)+ \\ \mu e^{jwd}V^*(w)\otimes P^*(w))^m \end{Bmatrix} \tag{15}$$

By binomial expansion

$$(a+b)^m=\sum_{l=0}^{m}C_l^m a^{m-l}b^l \tag{16}$$

and setting $a=e^{-jwd}V(w)\otimes P(w)$, $b=\mu e^{jwd}V^*(w)\otimes P^*(w)$, (15) can be further expanded into

$$\begin{aligned} \xi_{h,m}(n) &= \frac{1}{m}(-1)^{m+1}\alpha^m\sum_{l=0}^{m}C_l^m IFT\begin{Bmatrix}[e^{-jwd}V(w)\otimes P(w)]^{m-l}\cdot \\ [\mu e^{jwd}V^*(w)\otimes P^*(w)]]^l\end{Bmatrix} \\ &= \frac{1}{m}(-1)^{m+1}\alpha^m\sum_{l=0}^{m}C_l^m\{\delta(n-(m-l)d)\otimes \\ &\quad [v^{\otimes(m-l)}(n)p^{\otimes(m-l)}(n)]\otimes\delta(n+ld)\otimes[v^{\otimes l}(-n)\mu^l p^{\otimes l}(-n)]\} \end{aligned}$$

(17)

We will focus on the coefficient at the first echo delay d [5], i.e.

$$m-2l=1, \text{ or } l=\frac{m-1}{2} \tag{18}$$

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 $m-2l\neq1$ can be neglected for a nearly ideal PN sequence with a very small echo gain α. According to (6) and (18), the coefficients in (17) of our concern at the first echo delay can be written as

$$\xi^{(d)}_{h_{S\_BTS},m}(n)\cong\frac{1}{m}(-1)^{m+1}\alpha^m C^m_{\frac{m-1}{2}}\mu^{\frac{m-1}{2}}\varepsilon_v^{m-1}[v(n-d)p(n-d)] \tag{19}$$

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

$$\xi_h^{(d)}(n) = \sum_{m:odd} \xi_{h,m}^{(d)}(n)$$
$$= \left[\sum_{r=0}^{\infty} \frac{1}{2r+1} \alpha^{2r+1} C_r^{2r+1} \mu^r \varepsilon_v^{2r}\right] v(n-d)p(n-d)$$
$$= g \cdot v(n-d)p(n-d) \tag{20}$$

where , $g = \left[\sum_{r=0}^{\infty} \frac{1}{2r+1} \alpha^{2r+1} C_r^{2r+1} \mu^r \varepsilon_v^{2r}\right]$ (21)

The detection gain can be defined as the peak value at delay time d

$$\beta = \left\{\xi_h^{(d)}(n) \otimes v(-n)p(-n)\right\}_{@n=d}$$
$$= \sum_{r=0}^{\infty} \frac{1}{2r+1} \alpha^{2r+1} C_r^{2r+1} L \mu^r \varepsilon_v^{2r}$$
$$= g \cdot L \tag{22}$$

The detection gains for bilateral methods can be easily obtained by setting the energy as $\mu = 1$ and $\varepsilon_v^{\ 2} = L$, and $\beta = \sum_{r=0}^{\infty} \frac{1}{2r+1} \alpha^{2r+1} C_r^{2r+1} L^{r+1}$ are the same as before [5].Similarly, the unilateral time spreading echo hiding kernel is $\mu = 0$ and $\varepsilon_v^{\ 2} = L$,and $\beta = \alpha L$ [4].

3.2. Large echo gain

On the view of detection gain, if we use bigger echo gain α , we will have a better decoding gain at the echo delay. However, this thought is wrong!

As long as (13) is not satisfied, which means

$$\left|\alpha e^{-jwd} V(w) \otimes P(w) + \alpha \mu e^{jwd} V^*(w) \otimes P^*(w)\right| > 1 \tag{23}$$

We need to seek another asymptotical formula other than the one used in (11). That means the echo gain must be sufficiently high. Now we rewrite (10) as

$$\xi_h(n) = IFT\left\{\begin{array}{l}\frac{1}{2}\log\{1 + \alpha e^{-jwd} V(w) \otimes P(w) + \\ \alpha \mu e^{jwd} V^*(w) \otimes P^*(w)\}^2\}\end{array}\right\}$$
$$= IFT\left\{\begin{array}{l}\frac{1}{2}\log\{1 + 2(\alpha e^{-jwd} V(w) \otimes P(w) \\ +\alpha \mu e^{jwd} V^*(w) \otimes P^*(w)) + \alpha^2 \varepsilon_v^{\ 2}(1+\mu)^2\}\end{array}\right\} \tag{24}$$

Bringing $\alpha^2 \varepsilon_v^{\ 2}(1+\mu)^2$ out of the bracket

$$\xi_h(n) = \tilde{\xi}_{h,DC} + \tilde{\xi}_h(n) \tag{25}$$

where $\tilde{\xi}_{h,DC} = \frac{1}{2} F^{-1}\left\{\log\left\{\alpha^2 \varepsilon_v^{\ 2}(1+\mu)^2\right\}\right\}$ is constant, and

$$\xi_h(n) = \frac{1}{2} F^{-1}\left\{\log\left\{\begin{array}{l}1 + \frac{1}{((1+\mu)\alpha\varepsilon_v)^2} + \frac{2}{((1+\mu)\alpha\varepsilon_v)^2}(\alpha e^{-jwd} V(w) \\ \otimes P(w) + \alpha \mu e^{jwd} V^*(w) \otimes P^*(w))\end{array}\right\}\right\} \tag{26}$$

Using the logarithm expansion and binomial expansion in (11), (12), and (14), we can express (26) as a general form. Besides, by substituting $(\alpha e^{-jwd} p(w) + \alpha \mu e^{jwd} p^*(w))^l$ with $((1+\mu)\alpha\varepsilon_v)^l$, (26) can be further simplified as

$$\xi_{h,m}(n) = \frac{1}{2m}(-1)^{m+1} \sum_{l=0}^{m} C_l^m \left(\frac{1}{((1+\mu)\alpha\varepsilon_v)^2}\right)^{m-l} \left[\frac{2}{((1+\mu)\alpha\varepsilon_v)^2}\right.$$
$$\left.(\alpha e^{-jwd} V(w) \otimes P(w) + \alpha e^{jwd} \mu V^*(w) \otimes P^*(w))\right]^l \tag{27}$$

According to the spreading property in (6), we can further set (27) as

$$\xi_h(n) = \sum_{m=1}^{\infty} \xi_{h,m}(n)$$
$$= \alpha \varepsilon_v^{\ 2} \left\{\begin{array}{l}\sum_{m \in odd} \frac{1}{m} \sum_{r=0}^{\frac{m-1}{2}} C_{2r+1}^m 2^{2r} \left(\frac{1}{((1+\mu)\alpha\varepsilon_v)^2}\right)^{m-r} \\ -\sum_{m \in even} \frac{1}{m} \sum_{r=0}^{\frac{m}{2}-1} C_{2r+1}^m 2^{2r} \left(\frac{1}{((1+\mu)\alpha\varepsilon_v)^2}\right)^{m-r}\end{array}\right\} \tag{28}$$

We can observe that the peak value β decreases as echo gain α increases, if α is sufficiently large. Now we can see the important and surprising fact that an enlarged echo gain α can increase or decrease the detection gain, depending on the echo gain α itself is small enough or too large.

We can say with fair certainty that there is an area of the echoes to achieve the maximum detection peak value to make the correct detection probability largest. If the echo gain is in higher area, the increased echo gain not only pulls down the detection performance but also makes the echoes more perceptible.

4. SIMULATION

We want to simulate the formula we derived in (20) and (28) with $\mu = 1$ (BTSEH) and $\mu = 0.4$ without shaping ($\varepsilon_v^2 = L$). We choose an audio frame size which is 0.2 sec long to embed the same kernel with delay=100 samples and record the detection gain at this delay. The simulation results of detection gain at echo delay are shown in Fig. 2.and 3.

First, in Fig. 2 the theoretical dash line are derived from (20). The theoretical dash line with $\mu = 1$ (BTSEH) is very close to the simulated result. But it doesn't happen on $\mu = 0.4$.

The discrepancy may result from the effect of the wrapping phase. Previous property of the cepstrum neglects the effect of the phase. The phase of $Y_u(w)$, $S(w)$, and $H_u(w)$ are all within the region -π to π . So the phase of the watermarked audio spectrum may be different from the phase summation of original audio spectrum and the kernel spectrum. However, if we unwrap the phase, the magnitude of the audio cepstrum will become too large to cover the magnitude of the kernel. Therefore, the appearance of the PN pattern after the echo delay is no longer predominant so that the advantage of PN spreading may fail. However, the effect of phase with $\mu \simeq 1$ is of no importance since the spectrum of this kernel is zero phase. Our simulation result will confirm this.

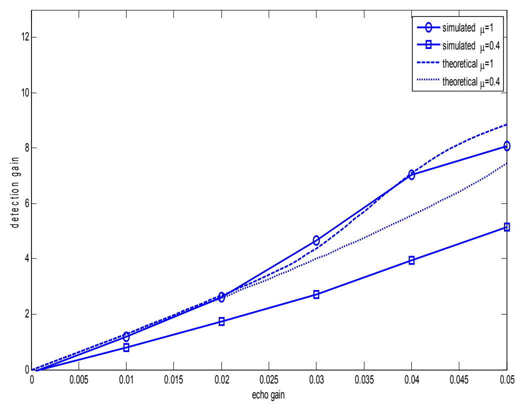

Fig.2 Simulated and theoretical detection gains versus α within the lower area

Second, in Fig 3 the theoretical dash lines are derived from (28). We can see that theoretical dash lines are similar to the simulated result.

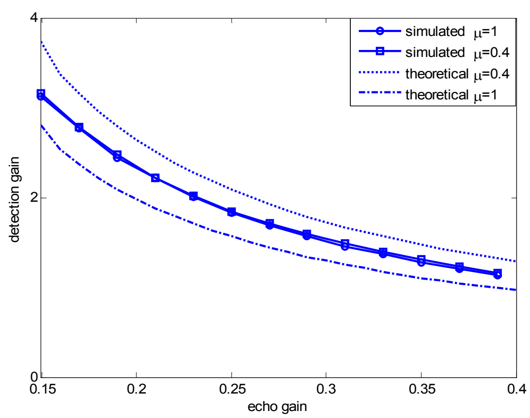

Fig.3 Simulated and theoretical detection gains versus α within the higher area

Next, we fix α=0.0156 in Fig. 4 within the lower area. We see the longer PN contributes a larger detection gain. But simulation curve is a little bit different from theoretic curve at $\mu = 0.4$ due to phase effect as mentioned before.

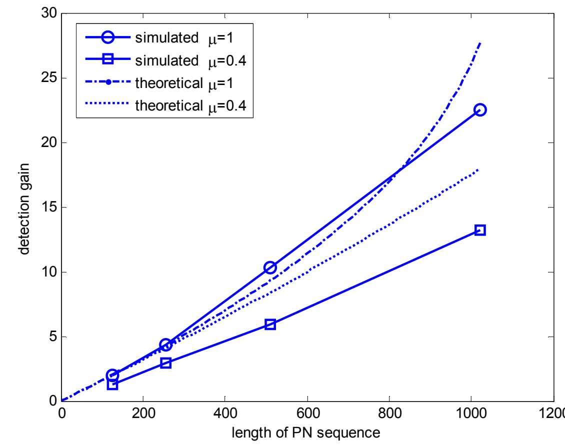

Fig.4 Simulated and theoretical detection gains versus L

5. CONCLUSION

In this paper, we have proposed a temporal masking technique for the time spread echo hiding scheme. It improves the robustness and imperceptibility of the watermarking. Theoretical asymptotical detection gains are derived and shown by simulation to match the conventional method. We also explain why the detection gain improves only when the echo gain is small enough.

6. REFERENCES

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

[2] 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.

[3] 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.

[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] S.A. Chou and S.F. Hsieh, "An Echo-Hiding Watermarking Technique Based on Bilateral Symmetric Time Spread Kernel." in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing* vol.3, 2006, pp.1100-1103.

[6] T. Painter, and A. Spanias, "Perceptual coding of digital audio," in *Proc. IEEE*, vol. 88, 2000, pp. 451-515.