STOCHASTIC MODEL BASED AUDIO WATERMARK AND WHITENING FILTER FOR IMPROVED DETECTION

Hongseok Kim

Access Network Lab., Korea Telecom
62-1 Hwaam-Dong, Yusung-Gu, Taejon, Korea
e-mail: innotobe@kt.co.kr

ABSTRACT

Digital watermarking is a means of copyright protection by embedding an imperceptible signal into the multimedia data. Nearly all watermarking schemes developed basically employ spread spectrum technique, and the extraction of embedded watermark depends on the correlation method. In this paper, we propose a stochastic model based public audio watermark. Based on the stochastic model and detection theory, the watermark shaping effect on detection is well explained. We also provide the theoretical whitening filter SNR gain and show that the detection performance can be significantly improved with whitening filter. As regards robust detection with whitening filter, we propose a robust detection parameter and demonstrate the reliability of the proposed estimated whitening filter to various kinds of audio degradation and distortions, e.g., MPEG compression, filtering, requantization, resampling, etc.

1. INTRODUCTION

Digital watermarking is the process of encoding hidden copyright information in digital data by making imperceptible modification to the data samples. Embedded watermark carries a certain amount of information relating to the owner, distributor or receiver of digital data, or the multimedia content itself. When the ownership of a digital work is in question, the embedded information can be extracted to completely characterize the owner. Mostly adopted scheme in watermark is spread spectrum (SS) technique [1, 2, 3, 4, 5, 6].

When the host signal, i.e., unwatermarked signal is unavailable in detection process, the host signal is like channel noise in communication. It is referred to public watermark detection. It is well known in detection theory that the matched filter, usually used for watermark detection, is optimal in the sense of Signal to Noise Ratio (SNR) in additive white Gaussian noise channel [8]. Generally the host media, i.e., image, video and audio are far from the Gaussian random process, and it leads us to the optimal detection problem using whitening matched filter.

There is a literature dealing with the whitening filter in images [5] and one heuristic approach using whitening filter concept in audio [6]. In this paper, we propose a new audio watermark based on stochastic model. Whitening filter is proposed to minimize the noise interference effect of audio signal, and the detection performance with whitening filter is provided. Experiments will verify the proposed theory, and conclusion follows it.

2. FORMULATION OF THE PROBLEM

For modeling the problem, we consider two random process, $\mathcal{S}$ for audio signal, and $\mathcal{W}$ for watermark signal in a short time duration. For the theoretic derivation, we assume the followings. (i) Audio stochastic process $\mathcal{S}$ is zero mean, wide sense stationary and ergodic Gaussian random process. (ii) The autocorrelation matrix of $\mathcal{S}$ is not singular. (iii) The inserted watermark $\mathcal{W}$ is generated from a key-dependent pseudo random sequence which can be modeled as a zero mean, independent identical distributed (i.i.d) Gaussian random process. If the key is unknown, the embedded watermark can not be statistically estimated.

3. WHITENING FILTER IN FINITE OBSERVATION INTERVAL

For a watermark signal $w(n)$ and the audio signal $s(n)$ taken from the random process $W(n)$ and $S(n)$, watermark is inserted as follows.

$$r(n) = s(n) + w(n), \quad n = 0, \cdots, N-1 \tag{1}$$

A capital represents a random process, and a small letter is for one sample among random process. Given the stochastic property of host signal, we can use the whitening filter in the finite observation interval as

$$\begin{aligned} r_h(n) &\triangleq \sum_{k=0}^{N-1} r(k)h(n,k) \\ &= w_h(n) + s_h(n), \quad n = 0, \cdots, N-1 \end{aligned}$$

where $h(n,k)$ is the filter impulse response at time n for the input at time k. Note that $h(n,k)$ is neither a causal filter nor time invariant filter since the detection is processed off-line. If N is sufficiently large, we can assume that the output of the whitening filter $\mathbf{r_h}$ for a given $\mathbf{w}$ has the white Gaussian distribution, and Likelihood Ratio Test (LRT) is

$$\frac{P(\mathbf{r_h}|\mathbf{H_1})}{P(\mathbf{r_h}|\mathbf{H_0})} = \frac{\prod_{n=0}^{n=N-1} \frac{1}{\sqrt{2\pi}\sigma_{s_n}} exp(-\frac{(r_h(n)-w_h(n))^2}{2\sigma_{s_n}{}^2})}{\prod_{n=0}^{n=N-1} \frac{1}{\sqrt{2\pi}\sigma_{s_n}} exp(-\frac{r_h(n)^2}{2\sigma_{s_n}{}^2})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta.$$

0-7803-6293-4/00/$10.00 ©2000 IEEE.

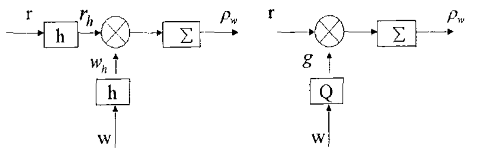

Figure 1: Whitening filter watermark scheme

Substituting $w_h(n)$, $s_h(n)$ and $r_h(n)$ into both side, and defining two new terms $Q(k,k') = \sum_n h(n,k)h(n,k')$, $g(k) = \sum_{k'} Q(k,k')w(k')$ and taking logarithm on both sides, the test statistics is obtained to [8]

$$\rho = \sum_{n=0}^{N-1} r(n)g(n) \underset{H_0}{\overset{H_1}{\gtrless}} \frac{1}{2}\sum_{n=0}^{N-1} w(n)g(n) + \sigma_{s_n}{}^2 \ln \eta \qquad (2)$$

and $\mathbf{Q} = \mathbf{R_s^{-1}}$ where $R_s(n,k) = E[S(n)S(k)]$. So the whitening filter is reduced to the right in Figure 1. Let $\mathbf{V} = [\mathbf{v_1} \cdots \mathbf{v_N}]$ be an eigenvector matrix of $\mathbf{R_s}$, $\mathbf{v_i}$ be its normalized eigenvector, and $\mathbf{\Lambda} = diag(\lambda_1, \cdots, \lambda_N)$ be its eigenvalue matrix. Using eigenvalue decomposition, $\mathbf{R_s}$ is decomposed to $\mathbf{V\Lambda V^t}$, and

$$\mathbf{g} = \mathbf{R}_s{}^{-1}\mathbf{w} = \mathbf{V\Lambda}^{-1}\mathbf{V^t w}. \qquad (3)$$

From the assumption in section 2, $\mathbf{R_s}$ is a toepliz matrix, and $E[S(n)S(n-k)]$ is calculated off-line as

$$R_S(k) \;=\; \frac{1}{M}\sum_{n=<M>} s(n)s(n-k), \quad M \geq N$$

where $n =< M >$ means the extended local region of length M around the observed audio signal $s(n)$. However, the detector does have little knowledge of $s(n)$, $r(n)$ is used to compute $\hat{\mathbf{R}}_\mathbf{s}$. Since the watermark signal power is negligible to audio signal, it makes sense. In the sequel, we will call it weak signal assumption.

4. AUDIO WATERMARK DETECTION AND ANALYSIS

4.1. Stochastic Model Based Audio Watermark

The watermark is usually shaped in some manner in accordance with the audio signal before embedded. One reason for watermark signal shaping is the watermark necessary condition for robustness to the signal distortion, i.e., compress, filtering, D/A, A/D conversion, cropping, etc. The watermark is usually embedded where the audio signal power is strong in either time or frequency to be robust to distortion [1][2]. The other reason for watermark signal shaping is the watermark necessary condition for the inaudibility to human ears. The i.i.d random process watermark is easily audible to human ears even though its power is relatively small. Instead of weakening the watermark signal power to be inaudible, it is better to use the audio masking effect. In [4], the watermark shaped by frequency masking and temporal masking looks like the audio signal in time wave form, and its power spectrum is also similar to that of audio signal. In [6], Preuss et al. shape the watermark signal in preselected suuband to make the watermark signal have similar spectrum to audio signal. From these facts, we deduce that the autocorrelation of shaped watermark will have the similar pattern to that of the audio signal. So we impose some stochastic property on the watermark random process $\mathcal{W}$ in accordance with the audio random process $\mathcal{S}$, and propose a new watermark based on stochastic model as

$$\mathbf{W} = \alpha \mathbf{V\Lambda}^\beta \mathbf{V^t U} \qquad (4)$$

where α is scaling factor to ensure the same watermark energy of $\mathbf{W}$ and $\mathbf{U}$, β is signal shaping parameter, and $\mathbf{U}$ is modeled as a zero mean i.i.d Gaussian random process, $E[U(n)^2] = \sigma_u{}^2$. From $E[\mathbf{W^t W}] = E[\mathbf{U^t U}]$, α is $\sqrt{\frac{N}{tr(\mathbf{R_s}^{2\beta})}}$ where $\mathbf{R_s}^\beta \triangleq \mathbf{V\Lambda}^\beta\mathbf{V^t}$.

The compatibility of the proposed watermark can be determined by model similarity measure μ as

$$\mu = \max_\beta E\left[\frac{\mathbf{W^t W_p}}{\|\mathbf{W}\|\,\|\mathbf{W_p}\|}\right] \qquad (5)$$

where $\mathbf{W_p}$ is watermark proposed in other algorithms. β for maximum μ can be used to correspondent signal shaping parameter for $\mathbf{W_p}$. When the watermark is embedded as [6] in whole subbands, μ is found to be over 0.9, and β is $0.3 \sim 0.4$.

4.2. Watermark Detection

To compare the detection performance easily between whitening filter and matched filter, we will use the Maximum Likelihood (ML) criterion. For watermark $\mathcal{W}$ and audio signal $\mathcal{S}$, the test statistics with whitening matched filter is obtained as in (2)

$$\rho_{wht} = (\mathbf{W}+\mathbf{S})^t\mathbf{G} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\mathbf{1}}{\mathbf{2}}\mathbf{W^t G} \qquad (6)$$

where $\mathbf{G} = \mathbf{R_s}^{-1}\mathbf{W} = \alpha\mathbf{V\Lambda}^{\beta-1}\mathbf{V^t U}$. Even thought ρ_{wht} is not theoretically Gaussian random variable, its experimental probability density function is similar to Gaussian. For the analysis, we assume that ρ_{wht} is a Gaussian random variable. The mean of ρ_{wht} is

$$\begin{aligned} E[\rho_{wht}] &= E[\mathbf{W^t G}] \\ &= E[\alpha^2\mathbf{U^t V\Lambda^{2\beta-1}V^t U}] \\ &= \alpha^2 tr(\mathbf{R_s}^{2\beta-1})\sigma_u^2. \end{aligned} \qquad (7)$$

From the weak signal assumption,

$$\begin{aligned} var[\rho_{wht}] &= E[\mathbf{G^t S S^t G}] \\ &= E[\alpha^2\mathbf{U^t V\Lambda}^{\beta-1}\mathbf{V^t R_s V\Lambda}^{\beta-1}\mathbf{V^t U}] \\ &= \alpha^2 tr(\mathbf{R_s}^{2\beta-1})\sigma_u^2 \end{aligned} \qquad (8)$$

We define $SNR \triangleq \frac{\sigma_u{}^2}{\sigma_s{}^2}$ and SNR multiplier for whitening filter as

$$\mathcal{M}_{wht}(\beta, \mathbf{R_s}) \triangleq \sigma_s{}^2 \frac{tr(\mathbf{R_s}^{2\beta-1})}{tr(\mathbf{R_s}^{2\beta})}. \qquad (9)$$

Then, from (7), (8) and (9)

$$P_e = Q\left(\frac{1}{2}\sqrt{N \cdot SNR \cdot \mathcal{M}_{wht}(\beta, \mathbf{R_s})}\right). \qquad (10)$$

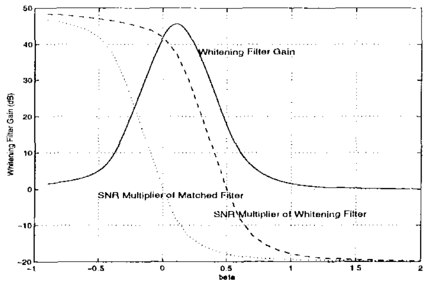

Figure 2: SNR multiplier and Whitening filter gain dots - $\mathcal{M}_{mat}(\beta, \mathbf{R_s})$, dash - $\mathcal{M}_{wht}(\beta, \mathbf{R_s})$, solid - $\mathcal{G}(\beta, \mathbf{R_s})$ for Mozart Piano concerto 23

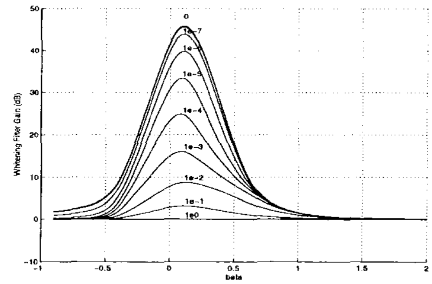

Figure 3: The effect of Robust Detection Parameter γ on Whitening Filter Gain $\mathcal{G}(\beta, \mathbf{R_s})$

Note that $\mathcal{M}_{wht}(0.5, \mathbf{R_s}) = 1$. The similar results is obtained for matched filter detection by setting $\mathbf{G} = \mathbf{W}$ and similar derivation. The results are

$$E[\rho_{mat}] = \alpha^2 tr(\mathbf{R}^{2\beta})\sigma_u^2 \quad (11)$$

$$var[\rho_{mat}] = \alpha^2 tr({\mathbf{R_s}}^{2\beta+1})\sigma_u^2 \quad (12)$$

$$\mathcal{M}_{mat}(\beta, \mathbf{R_s}) \triangleq {\sigma_s}^2 \frac{tr({\mathbf{R_s}}^{2\beta})}{tr({\mathbf{R_s}}^{2\beta+1})}, \quad (13)$$

$$P_e = Q\left(\frac{1}{2}\sqrt{N \cdot SNR \cdot \mathcal{M}_{mat}(\beta, \mathbf{R_s})}\right) \quad (14)$$

The whitening filter gain is defined as the ratio of (9) and (13),

$$\mathcal{G}(\beta, \mathbf{R_s}) = \frac{tr({\mathbf{R_s}}^{2\beta+1})tr({\mathbf{R_s}}^{2\beta-1})}{(tr({\mathbf{R_s}}^{2\beta}))^2}, \quad (15)$$

Figure 2 shows $\mathcal{G}(\beta, \mathbf{R_s})$ together with $\mathcal{M}_{mat}(\beta, \mathbf{Rs})$ and $\mathcal{M}_{wht}(\beta, \mathbf{R_s})$. As seen in Figure 2, the detection performance is heavily dependent on β.

4.3. SNR Multiplier and Whitening Filter Gain

If β is negative, the watermark signal is designed in the direction where the host signal power is relatively small. $\mathcal{M}_{wht}(\beta, \mathbf{R_s})$ is quite high, over 40 dB. Negative β is mostly preferred in the communication, but it can not be used for watermark signal design because of robustness and inaudibility condition.

If β is zero, i.e., flat shaping case, $\alpha = 1$ and $\mathbf{W} = \mathbf{U}$, i.e., the embedded watermark is i.i.d random process. For the robust watermark, the detection performance at $\beta = 0$ is the upper limit from Figure 2, and the whitening filter gain $\mathcal{G}(\beta, \mathbf{R_s})$ is maximized around $\beta = 0$. Substituting $\beta = 0$ into (15), and using $\text{tr}(\mathbf{R_S}) = \sum \lambda_n$, and $\text{tr}({\mathbf{R_S}}^{-1}) = \sum {\lambda_n}^{-1}$, this property holds with Schwarz and Holder inequality.

$$\mathcal{G}(0, \mathbf{R_s}) = \lambda_{ave} \cdot \lambda_{ave}^{-1} \geq 1 \quad (16)$$

where $\lambda_{ave} = \frac{1}{N}\sum \lambda_n$, and $\lambda_{ave}^{-1} = \frac{1}{N}\sum {\lambda_n}^{-1}$. The equality holds when the stochastic process of host signal is i.i.d. process., and all eigenvalues are same. However, audio signal has many sinusoidal components, and quasi periodic autocorrelation property in a short time duration. Moreover, there are many zero-like eigenvalues. This characteristic of audio signal guarantees that the high gain and the importance of whitening filter.

When $0 < \beta \leq 0.5$, it is the region of interest. SNR multiplier $\mathcal{M}_{wht}(\beta, \mathbf{R_s})$ abates rapidly. Most other watermarking algorithms fall in this case, $0 < \beta \leq 0.5$ since the watermark signal is shaped in accordance with the robustness and inaudibility condition. Especially, when $\beta = 0.5$, i.e., the watermark signal is shaped critically, $\mathcal{M}_{wht}(0.5, \mathbf{R_s}) = 1$. In this case, the watermark signal is designed to have the same power distribution with the host signal in each direction of eigenvectors. β is the adjusting parameter for the robustness and inaudibility, and we may find appropriate β and design the watermark to be robust and inaudible enough and to be efficient enough in detection.

When $\beta > 0.5$, i.e., over shaping case, $\mathcal{M}_{wht}(\beta, \mathbf{R_s})$ begins to be under 1. The detection performance will be bad as shown in Figure 2. This case is also useless for the watermark signal design as in negative shaping case.

4.4. Robust Detection, Tamper Proof Watermark and Whitening Filter

From Figure 2, the maximum whitening filter gain is about 45 dB. High whitening filter gain may be conducive to instability. Since the simple whitening filter is the inverse operation of the eigenvalue, the zero-like eigenvalue occurs the instability. It is referred to singular detection in detection theory [8] and undesirable for robust detection. It goes worse if the watermarked signal is distorted. To solve this problem, the robust detection parameter is proposed in whitening filter. Let all of the eigenvalues λ_i, $i = 0, \cdots, N-1$ be augmented by γ in (4.2): $\mathbf{\Lambda} \leftarrow \mathbf{\Lambda} + \gamma \mathbf{I_N}$

Figure 3 shows $\mathcal{G}(\beta, \mathbf{R_s})$ for various γ. Note the whitening filter gain $\mathcal{G}(\beta, \mathbf{R_s})$ should be reduced to attain the robust detection. Robust detection parameter γ is a compromising factor like β. As γ goes to infinity, the whitening matched filter goes to matched filter. Some heuristic approaches, e.g., imposing some lower limit on eigenvalues or ignoring eigenvalues smaller than some threshold are also possible.

As for the tamper proof watermark, zero γ and small positive β is good since the tamper proof watermark should be fragile. Extremely High gain often means instability, and the signal distortion will seriously affect the detection. If the watermarked signal is not distorted in some manner,

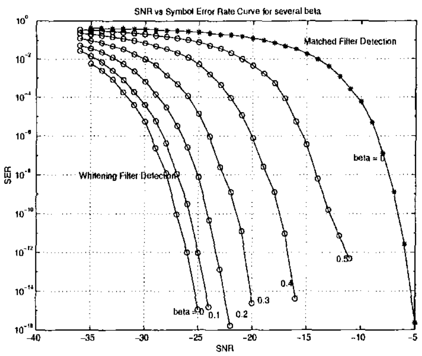

Figure 4: The effect of signal shaping parameter β *:Matched filter detection, o:whitening filter detection

the watermark can be detected with high reliance. When the audio signal is, however, distorted, watermark is not detected because of the instability of itself. For the tamper proof watermark, the detector knows that the received signal already contains a tamper proof watermark, and the miss detection of watermark means that the watermarked signal is changed.

5. EXPERIMENTAL RESULTS

For the proposed watermark with the length of $N = 512$, 11.6 ms in 44.1 KHz sampled data, various SNR and β , detection error rates under ML criterion are calculated in Figure 4. As β increases, the detection performance degrades as anticipated in Figure 2. Mozart's concerto for piano 23 is used for the experiment. The meaningful range of SNR is below at least -26 dB [3]. For the simplicity, we do not insert the watermark repeatedly in time as done in [4]. We estimate the autocorrelation of audio signal in a second. It is verified to be reasonable by experiments to assume the short time audio signal is w.s.s process.

We compare the proposed whitening filter detection with subband equalization in [6]. Preuss *et al.* analyze the power of audio in preselected subband. After inserting the watermark as in [6], we detect it with three methods: proposed whitening filter detection, subband equalization detection and matched filter detection. The result for mpeg1-layer3 compression is shown in Figure 5. This result is obtained for 512*100 detection interval, or time frame (51200/44100 = 1.161 sec), i.e., 100 times redundant watermark detection is performed, Averaged SNR of inserted watermark is -26 dB. Correlation coefficient is normalized so that 1 represents H_1 and 0 represents H_0. The test is performed on 7 time frames. 'O' indicates the averaged correlation coefficient, and 'I' bar shows the estimated standard deviation of 'O'. It is easy to confirm that whitening filter detection is most reliable. The proposed whitening filter scheme also works well under requantization, resampling, moving average filering, etc.

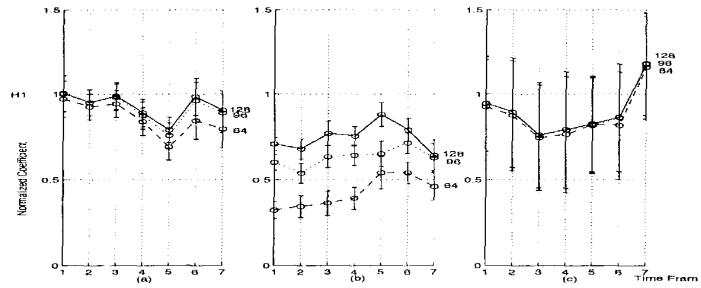

Figure 5: Watermark Detection after 128, 96, 64 Kbps MP3 compression (a) whitening filter detection (b) subband equalization detection (c) matched filter detection

6. CONCLUSION

In this paper, we proposed a stochastic model based audio watermark and showed the whitening filter for audio signal improved the watermark detection performance by decades of dB in SNR. Using the weak signal assumption, the whitening filter can be directly constructed from received signal, and makes whitening filter available. We computed the theoretical whitening filter SNR gain based on stochastic watermark and detection theory. To be robust and reliable detection, we proposed the robust detection parameter. The theoretical requirement for tamper proof watermark was also derived. The proposed whitening filter detection is verified reasonable by numerous experiments.

7. REFERENCES

[1] M.D. Swanson, M. Kobayashi, and A. H. Tewfik, "Multimedia Data-Embedding and Watermarking Technologies." *Proceedings of the IEEE, June 1998, vol. 86, no. 6, pp.1064-1087.*

[2] I.J. Cox, J. Killian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia", *IEEE trans. on Image Processing, December 1997, vol. 6, no. 12, pp. 1673-1687.*

[3] P. Bassia and I. Pitas, "Robust audio watermarking in the time domain", *EUSIPCO 1998, pp. 25-28*

[4] M. D. Swanson, B. Zhu, A. H. Tewfik, and L. Boney, "Robust audio watermarking using perceptual masking", *Signal Processing 66 (1998) pp. 337-335.*

[5] G. Depovere, T. Kalker, and J. P. Linnartz, "Improved watermark detection reliability using filtering before correlation", *Proc. IEEE Int. Conf. on Image Processing, 1998, pp. 430-434.*

[6] Preuss, *et al.*, "Embedded signaling", *US patent 5,319,735*

[7] Hyuk Choi, Hongsoek Kim, Taejeong Kim, "Robust sinusoidal watermark for images", *IEE Electronics Letters, 22nd, July 1999, Vol. 35, No.15, pp. 1238-1239*

[8] H. L. Van Tress, "Detection, Estimation, and Modulation Theory I", *Wiley, New York, 1968.*