Audio asymmetric watermarking technique

Teddy Furon, Nicolas Moreau, Pierre Duhamel

AUDIO PUBLIC KEY WATERMARKING TECHNIQUE

T. Furon,

THOMSON multimedia, UIIS Lab
1, av Belle Fontaine
35511 Cesson Sévigné
furont@thmulti.com

N. Moreau, and P. Duhamel

ENST Paris, Laboratoire TSI
46, rue Barrault
75634 Paris Cedex 13
{moreau,duhamel}@sig.enst.fr

ABSTRACT

This paper presents the application of the promising public key watermarking method[1] to the audio domain. Its detection process does not need the original content nor the secret key used in the embedding process. It is the translation, in the watermarking domain, of a public key pair cryptosystem [1]. We start to build the detector with some basic assumptions. This leads to a hypothesis test based on probability likelihood. But real audio signals do not satisfy the assumption of a Gaussian probability density function. Moreover, the use of an advanced human perception model to hide the watermark makes the detection issue a tough problem. Our works result in a new detection process offering a good test's power for a low probability of false alarm.

1. INTRODUCTION

To build a copy protection system for consumer electronic devices, we are looking for a technique, which could embed in an original content a signal commonly called watermark. Compliant devices such as players or recorders are able to detect the presence of this watermark. In this particular case, its presence means that the content is protected and thus it is illegal to copy it. This embedded watermark must not be perceptible. Watermarking schemes have been developed since several years for audio, video or still image.

To assess the security, we made a threat analysis of these techniques. They have achieved good results in non-perceptibility and robustness, but all of them are symmetric schemes. Symmetric means that the detection process needs to know the parameters used by the embedding process. The knowledge of these parameters allow pirates to forge illegal content by modifying or removing watermark. This set of parameters is called the secrete key and must be stored in a safe and secure place. This is not possible in consumer electronics. Tamper proof device is too expensive.

This is the reason why asymmetric watermarking schemes inspired from the cryptography domain have been recently studied ([2],[3] and [1]). They should be as robust as symmetric techniques with a detector needing a set of parameters called the public key different from the embedding's secret key. It is neither possible to deduce the private key nor possible to remove the watermark knowing the public key.

The method we proposed in [1] was experimentally approved for still images where the perceptual model was just a modulation by a local gain. We suspect it to be usable in many domains, for instance, in the audio domain. This paper first clarifies what assumptions on signals are mandatory. Secondly, the detection process as described is defeated because the perceptual model in audio is not based on modulation but on filtering. An improvement is proposed and tested at the end of the paper.

[1] French patent application number 99-07139 filled on the first of June 1999

2. DETECTION PROCESS

2.1. Notation

We briefly describe the method presented in [1]. Let the sequence $\{x_n\}$ of real numbers represent the original content (sequence of base-band samples or coefficients given by a common transformation (DCT, MDCT, wavelet, DFT...)). The watermark signal $\{w_n\}$ is created by filtering a white noise $\{v_n\}$. The secret key is used to generate this white noise and to choose a filter h which frequency response module matches with a given spectrum's template $|H(f)|^2$. Then, this watermark sequence $\{w_n\}$ will be added to the sequence $\{x_n\}$ with a suitable attenuation μ in order to remain non-perceptible. The detection process verifies if the sequence $\{r_n\}$, representing the received content, has a power spectrum density shaped like the template $|H(f)|^2$. This template constitutes the public key. Of course, knowing the public key, it is not possible to guess the private key: the phase of the filter and the sequence $\{v_n\}$ are missing. With several assumptions (h is SISO but neither a maximum nor minimum phase filter, $\{v_n\}$ Gaussian white noise), it is proved that the pirate can not estimate the watermark signal in order to remove it from the protected content.

2.2. Simple assumptions.

Let use a simplified embedding process defined by (1):

$$y_n = x_n + \mu.w_n \text{ with } w_n = (h \otimes v)_n \quad \forall n \in [0..N-1] \quad (1)$$

where $\{v_n\}$ is a central Gaussian random process of variance unity, $\{h_n\}$ is the impulse response of the filter h, which is normalized so that $\int_{-0.5}^{0.5} |H(f)|^2 .df = 1$.

The hypothesis here are :

- $\{x_n\}$ is a Gaussian white noise of variance σ_x^2.
- μ is a constant embedding power, bound to a relative watermark to original content power ratio: $\gamma^2 = \frac{\mu^2}{\sigma_x^2}$.

2.2.1. Simple hypothesis test for Gaussian random processes

The detection process verifies two hypothesis on the received content $\{r_n\}$:

- $\mathcal{H}_0$: $\{r_n\}$ represents a non watermarked content, so it is a Gaussian white noise.
- $\mathcal{H}_1$: $\{r_n\}$ represents a watermarked content, so it is a Gaussian colored noise like the $\{y_n\}$ sequence.

The detection will make its decision comparing the likelihood ratio to a given threshold T_1. With the Bayesian rule, it becomes:

$$R_L = \frac{p(\mathcal{H}_1|\{r_n\})}{p(\mathcal{H}_0|\{r_n\})} \geq T_1 \rightarrow \frac{p(\{r_n\}|\mathcal{H}_1)}{p(\{r_n\}|\mathcal{H}_0)} \geq T_1.\frac{p(\mathcal{H}_0)}{p(\mathcal{H}_1)} \quad (2)$$

According to the assumptions, $(r_0, ..., r_{N-1})\prime$ is a N dimension Gaussian vector $\underline{r}$, so that:

$$p(\{r_n\}|\mathcal{H}_i) = \frac{1}{(2\pi)^{\frac{N}{2}} \det(B_{g_i})^{\frac{1}{2}}} \exp(-\frac{1}{2}(\underline{r}' B_{g_i}^{-1} \underline{r}))$$

where B_{g_i} is the covariance matrix associated to the spectral density function g_i. As log is a strictly increasing function, the test is finally expressed in (3):

$$\begin{aligned} \text{Test}(\underline{r}) &= (\underline{r}'(B_{g_1}^{-1} - B_{g_0}^{-1})\underline{r}) - \log(\left|\frac{\det(B_{g_0})}{\det(B_{g_1})}\right|) \\ &\leq -2.\log(T_1.\frac{p(\mathcal{H}_0)}{p(\mathcal{H}_1)}) = T_2 \end{aligned} \quad (3)$$

Indeed, the detection compares the difference of the log likelihood $L_N(\underline{r}|\mathcal{H}_1) - L_N(\underline{r}|\mathcal{H}_0)$ to a given threshold, knowing that for Gaussian stationary process $\underline{r}$:

$$L_N(\underline{r}) = -\frac{1}{2}(N\log(2\pi) + \log(\det(B_g)) + \underline{r}' B_g^{-1} \underline{r}) \quad (4)$$

2.2.2. Efficiency of the test

The critical region $\mathcal{D}_1$ is the region of $\mathcal{R}^N$ where the received content $\underline{r}$ is detected as a watermarked content because $\text{Test}(\underline{r}) < T_2$ whereas the region $\mathcal{D}_0$ is defined by $\{\underline{r} \in \mathcal{R}^N \ |\text{Test}(\underline{r}) > T_2\}$. Probability of false alarm is noted $P_{fa} = P(\underline{r} \in \mathcal{D}_1 \mid \mathcal{H}_0)$. This is the probability that a received content is declared protected whereas it was not watermarked. The power of the test is noted $P_{de} = P(\underline{r} \in \mathcal{D}_1 \mid \mathcal{H}_1)$. Simulations will lead to the graph of $P_{de} = P_{de}(P_{fa})$ for some parameters. This will help to set these parameters in order to maximize the power of the test P_{de} for a given probability of false alarm P_{fa}.

2.3. Less simple assumptions

If the sequence $\{x_n\}$ is a colored noise, then we use a pair of interleavers in order to break its statistical structure. If we assume that the de-interleaver in the detection process acts like a random permutation π on the samples, then $\{x_{\pi(n)}\}$ becomes a white noise. We have to interleave with the inverse permutation π^{-1} the sequence $\{w_n\}$ before adding it to the original content. Finally, after the random permutation at the detection process, we retrieve the sequence $\{w_n\}$ plus some white noise. Equation (1) becomes:

$$y_{\pi(n)} = x_{\pi(n)} + \mu.w_n \text{ with } w_n = (h \otimes v)_n \quad \forall n \in [0..N-1]$$

If the received sequence $\{r_n\}$ is not Gaussian distributed, the log likelihood expression is different from (4).The test (3) holds no more. This is the case if $\{x_n\}$ is not Gaussian distributed or if the embedding power is not constant: in order to embed the watermark in a non-perceptible way, a human perception model could issues a sequence $\{\mu_n\}$ changing in time.

2.4. Principal part of likelihood function

Even for simple cases, the expressions of B_g^{-1} and $\det(B_g)$ in (4) are really cumbersome. Moreover, this test is only valid for Gaussian distributed sequences. We looked for a more robust test.

2.4.1. Definition

Many works have be done in order to render likelihood function more practical. Since maximum likelihood estimators or hypothesis tests may be considered as optimal only when $N \rightarrow \infty$, Whittle suggested in [4] replace the log likelihood $L_N(\underline{r})$ by its principal part $\widetilde{L}_N(\underline{r})$, which satisfies (5):

$$\frac{(L_N(\underline{r}) - \widetilde{L}_N(\underline{r}))}{\sqrt{N}} \overset{N\rightarrow\infty}{\rightarrow} 0 \quad \text{(converg. in probability)} \quad (5)$$

$\widetilde{L}_N(\underline{r})$ can be chosen to be simpler than $L_N(\underline{r})$. At the same time, estimators maximizing $\widetilde{L}_N(\underline{r})$ and test hypothesis on $\widetilde{L}_N(\underline{r})$ are proved to be asymptotically equivalent to those related to $L_N(\underline{r})$.

In [5], one finds the following theorem:

Theorem 1 *Let the spectral density g and the covariance function β of a stationary random process $\{x_n\}$ satisfy the following conditions: $\exists m > 0 \mid g(f) \geq m$ and $\sum_{k=1}^{\infty} k\,|\beta(k)|^2 < \infty$. Then relation (5) is valid where*

$$\begin{aligned} \widetilde{L}_N(\underline{r}) &= -\frac{N}{2}\{\log(2\pi) + \int_{-\frac{1}{2}}^{\frac{1}{2}} \log(g(f)) + \frac{I_N(f)}{g(f)}.df\} \\ I_N(f) &= \frac{1}{N}\left|\sum_{k=0}^{N-1} r_k e^{2\pi j f k}\right|^2 \end{aligned} \quad (6)$$

$I_N(f)$ is the periodogram of the sequence $\{r_n\}$.

If in (6), the integral forms are replaced by their Riemann sums, this expression will give the same asymptotic results when N is large. This expression leads to another

interpretation of Whittle's approximation. For this, [6] recall us that under sufficiently mild conditions the random variables $I_N(\frac{k}{N})$ are asymptotically mutually independent and identically distributed as a central χ_2 with 2 degrees of freedom. Instead of the probability density function of $\underline{r}$, let us consider the probability density function of the vector $\underline{I} = (I(\frac{1}{N}), \ldots, I(\frac{\lfloor\frac{(N-1)}{2}\rfloor}{N}))'$. Then, asymptotically $N \to \infty$,

$$L_N(\underline{I}) = \sum_{k=0}^{\lfloor\frac{(N-1)}{2}\rfloor} (\log(g(\frac{k}{N})) + \frac{I_N(\frac{k}{N})}{g(\frac{k}{N})})$$

Note that these considerations do not employ the Gaussian assumption since the asymptotic properties of the periodogram values $I_N(\frac{k}{N})$ mentioned above are valid under much broader conditions on $\underline{r}$ (cf. [6]).

2.4.2. Application to the detection process

The test previously defined in (3) can be replaced by:

$$N. \sum_{k=0}^{\lfloor\frac{(N-1)}{2}\rfloor} \log(\frac{g_0(\frac{k}{N})}{g_1(\frac{k}{N})}) + I_N(\frac{k}{N})(\frac{1}{g_0(\frac{k}{N})} - \frac{1}{g_1(\frac{k}{N})}) \geq T_2 \quad (7)$$

This test is equivalent to the previous one only when $N \to \infty$ but does not require the assumption of a Gaussian random process $\{r_n\}$.

In the case of perceptual model based on modulation, the embedding power changes from sample to sample. We need to express the power spectrum density $g_1(f)$ and see if the estimation of $\{\mu_n\}$ can helps the detection. Knowing that the central random process $\{w_n\}$ is independent from the other processes, the power spectrum density under hypothesis $\mathcal{H}_1$ becomes:

$$g_1(f) = \sigma_x^2 + g_M(f) \otimes |H(f)|^2$$

where $g_M(f)$ is the Fourier transform of the auto-correlation function of the sequence $\{\mu_n\}$.

Example: If the original content $\{x_n\}$ is not naturally a white noise, we shall use interleavers in the embedding process and in the detection process. This pseudo-random permutation might also whiten the sequence $\{\mu_n\}$ issued from the human perceptual model:
$g_M(f) = \sigma_\mu^2 + E[\mu_n]^2.\delta_0(f)$ ($\delta_0(f)$ is the Dirac "function"). Finally, we have:

$$\begin{aligned} g_1(f) &= \sigma_x^2 + \sigma_\mu^2 + E[\mu_n]^2.|H(f)|^2 \\ &= \sigma_r^2 + E[\mu_n]^2.(|H(f)|^2 - 1) \end{aligned}$$

The conclusion is that, for basic perceptual models - based on gain modulation, we do not need the exact values of the sequence $\{\mu_n\}$, but only its covariance function β_M, and in the case of the example, only its mean $E[\mu_n]$.

3. IMPLEMENTATION

The issue now is to prove that the method is valid for audio content, despite the filtering required by the advanced perceptual model.

3.1. Advanced human perception model

An advanced human perception model represents the masking phenomenon in the form of a discrete filter h_{PM}. The watermark sequence $\{w_{\pi^{-1}(n)}\}$ must be filtered by h_{PM} before being added to the original sequence $\{x_n\}$ in order to be non-perceptible. This changes the embedding equation in:

$$y_{m,n} = x_{m,n} + (h_{m,PM} \otimes w_{\pi^{-1}})_n \text{ with } w_n = (h \otimes v)_n$$

The use of an advanced human perception model for audio contents was first described in [7]. $x_{m,n}$ represents the nth sample of the mth window of length N_{PM} from the sampled signal (the algorithm is repeated for each window m). A perceptual analysis, which is indeed exactly the same as the one used in compression schemes like MPEG2 layer I, II or III, gives the global masking function Ψ_m. The filter $h_{m,PM}$ is built such that its frequency response respects the following constraint:

$$|H_{m,PM}(f)|^2 \leq \Psi_m(f) \ \forall f \in [0, \frac{1}{2}]$$

From $\rho_m(\tau)$, which is the inverse Fourier transform of $\Psi_m(f)$, the classical Levinson's algorithm gives one AR filter of order L, $H_{m,PM}(z) = \frac{\mu_m}{1+\sum_{i=1}^{L} a_{m,i} z^{-i}}$ matching the constraint.

3.2. On the detector side

In the detector, the result of the de-interleaving of the received sequence $\{r_n\}$ can be expressed as:

$$\begin{aligned} r_{\pi(n)} &= x_{\pi(n)} + \varepsilon_{\pi(n)} + \mu_m.(h \otimes v)_n \ \forall n \in [0..N-1] \\ \varepsilon_n &= -\sum_{i=1}^{L} a_{m,i}.(h \otimes v)_{\pi^{-1}(n)-i} \text{ and } m = \lfloor\frac{n}{N_{PM}}\rfloor \end{aligned} \quad (8)$$

This could remind the example of subsection (2.4.2): the embedding power μ_m varies because the content lasts more than one window of N_{PM} samples (feeding the advanced human perception model) and the action of the interleaver is spread on several consecutive windows. The detector estimates the mean of μ_m doing exactly the same processing than the embedding's one. As the watermark energy is very low compared to the energy of the original, the perceptual model should give almost the same filter $h_{m,PM}$ window by window.

Yet, the sequence $\{\varepsilon_n\}$, statistically independent from $\{x_n\}$, is an extra noise for the detector. Moreover, the energy of a watermark sample $w_{\pi^{-1}(n)}$ has been spread up to L consecutive samples in $\{y_n\}$, where L is the order of the filters $H_{m,PM}$. We may compare Equation (8) to a well-known problem in digital communications: the inter-symbol interferences stemming from the filtering due to channel effect [8]. This gives us the idea of an equalizer. A zero-forcing pre-filter located before the de-interleaver, highly boosts indeed test's performances.

4. SIMULATION

4.1. Gaussian random sequences

The embedding process is described by Equation (1). $\{x_n\}$ and $\{v_n\}$ are Gaussian white noise sequences of length $N =$

1024, 4096 or 16384. The embedding power γ is set to -15dB, -20dB or -26dB. For each experiment, we compare results from the hypothesis test based on likelihood (3) to the one based on its principal part (7). Each test is repeated a large number $S(N) \propto \sqrt{N}$ of times to estimate the probability of false alarm or the power of the test. Each experiment is repeated three times to see how accurate is the result. Simulations ends to the graph of $P_{de} = P_{de}(P_{fa})$. The two kinds of test perform equally for $N \propto 1000$, which is large enough in the case of Gaussian distributed sequences.

4.2. Non Gaussian distributed signals

$\{x_n\}$ is now the sampling of a musical signal at $F_e = 32$kHz. We choose real signals in order to be close to the assumption of wide stationarity (no fast changes in power, no drums...). It lasts about 8 seconds which corresponds to 2^{18} samples. The interleaver is a random permutation of the same size.

4.2.1. Robustness of the periodogram test

The previous experiment is repeated with real signals, the embedding formula is still (1). Decisions is made with the hypothesis test based on periodogram (7). The results are almost as good as in the Gaussian case. This confirm the robustness of this test at least when $N \propto 1000$.

4.2.2. Pre-filtering

The watermark is now added with the advanced human perception model. The application of the perceptual model to the signal received by the detector gives an estimation of the embedding power (cf. 2.4.2). The experiment leads to very disappointing results: even for attenuation of -15dB which is the limit of audible watermark, it gives extremely poor test's power. This lack of efficiency is due to the use of the perceptual model based on convolution and not on modulation defeats the previous detector (cf. 3.2). As the detector can estimate the filters $H_{m,PM}(z)$ with the help of the perceptual model, we are able to filter the received sequence, window by window, with $H^{-1}_{m,PM}(z)$ before the de-interleaver and the hypothesis test. This will minimize the term ε_n in (8) and moreover, group back to one sample the energy spread on L consecutive samples. We compare this strategy with the zero-forcing method used in an equalizer in a digital communication scheme [8]. The detection results are much better and even similar to performances obtained in (4.2.1). Due to the lack of space, we only show in Fig. (1) the result of this last experiment.

5. CONCLUSION

The paper justifies the use of the periodogram based hypothesis test in this public key watermarking technique. The remaining assumption is only that signal are wide stationary. Despite the convolution imposed by advanced perceptual model, a simple zero-forcing filter boosts detection performances. Simulations prove the feasibility of this technique. Our work focuses now on the robustness to malicious attacks [3].

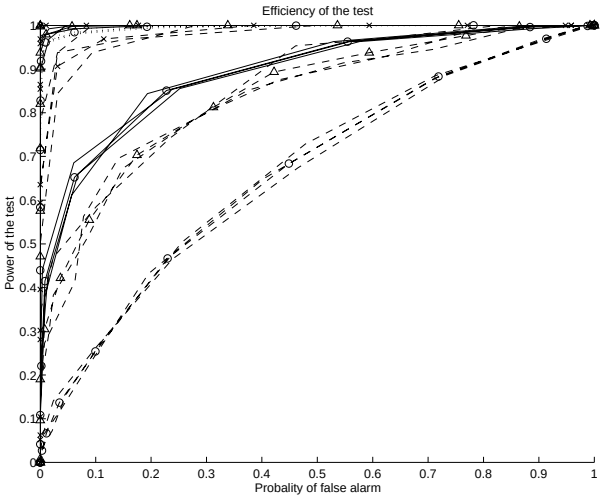

Figure 1: Hypothesis test on pre-filtered sequences . $\circ N = 1024$; $\triangle N = 4096$; $\times N = 16384$. dashed line $\gamma = -26$dB; solid line $\gamma = -20$dB; dotted line $\gamma = -15$dB

6. ACKNOWLEDGMENTS

We thank Leandro de Campos Teixeira Gomes (Université Paris V), Mathieu Carré (ENST Paris / CCETT) and Marcos Perreau-Guimaraes (Université Paris V) for providing the matlab code of the advanced human perceptual model.

7. REFERENCES

[1] T. Furon and P. Duhamel "An Asymmetric Public Detection Watermarking Technique" in *Proc. of the 3rd Int. Work. on Information Hiding*, Dresden, Sept 1999.

[2] R.G. van Schyndel, A.Z. Tirkel, and I.D. Svalbe "Key Independent Watermark Detection" in *ICMCS'99*, Florence, Italy, 1999.

[3] J. Eggers and B. Girod "Robustness of Public Key Watermarking Schemes", V^3D^2 Watermarking Workshop, Erlangen, Germany, Oct 1999.

[4] P. Whittle, *A Study in the Analysis of Stationary Time Series*, Almquist and Wiksell, 1954.

[5] K. Dzhaparidze, *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*, Springer Series in Statistics, Springer-Verlag, 1986.

[6] R.A. Olshen, "Asymptotic properties of the periodogram of a discrete stationary process", in *J. Appl. Probab.,v 4*, 1967.

[7] M.D. Swanson, B. Zhu, A.H. Tewfik and L. Boney, "Robust Audio watermarking using perceptual masking", in *Signal Processing, v 66 no 3*, May, 1998.

[8] J. G. Proakis, *Digital Communications*, Electrical Engineering Series, McGraw-Hill International Editions, third edition, 1995.