A NEW WIENER FILTERING BASED DETECTION SCHEME FOR TIME DOMAIN PERCEPTUAL AUDIO WATERMARKING

S. Larbi, M. Jaïdane

Ecole Nationale d'Ingénieurs de Tunis
Signals and Systems Laboratory
BP 37 Le Belvédère, 1002 Tunis, Tunisia
Sonia.Larbi@enit.rnu.tn, nej.jaidane@planet.tn

N. Moreau

GET - Télécom Paris
Signal and Image Processing Department
43 Rue Barrault, 75013 Paris, France
moreau@tsi.enst.fr

ABSTRACT

This paper presents a new detection method for a spread spectrum and perceptual watermarking system, viewed as a hidden data transmission system, where the generic detection operation is achieved by a Wiener deconvolution filter. In this paper, we point out the insufficiencies of the generic reception scheme concerning truncation errors and ill-conditioned deconvolution operation, and we propose a cascade realization of the reception filter, which significantly improves the detection performances (to a multiplicative factor of 20) at higher bit rates, even in presence of an MPEG compression.

1. INTRODUCTION

As the aimed application of watermarking in the present work is not the content protection but the hidden data transmission in an *audio channel*, we may view the watermarking scheme of fig.1 as a very particular communication system [1, 2]. Indeed, the original highly colored and powered unwatermarked audio x_n represents the additive noise, and the modulated and spectral shaped signature $\{a_k\}$ the transmitted information. The receiver is based on a blind detection, since the signal x_n is not available at reception.

The spectral shaping filter H, which purpose is to ensure the watermark inaudibility, is computed out of the masking threshold of x_n at the emitter, and out of the observed data $\hat{y}_n$ at the receiver, since both thresholds are very close because both signals hear like. Thus, we have a good approximation of the filter H on the reception side.
In this paper, we propose a new detection scheme based on this -at reception- a priori available information. Indeed, even if both detection schemes are theoretically equivalent, the proposed one reduces significantly the numerical insufficiencies of the generic receiver by improving the bit error rate (BER) to ca a factor of 20.

We present in sections 2 and 3 the considered watermarking scheme and we point out the insufficiencies of the generic receiver in terms of truncation errors and ill-conditioning. The proposed scheme is then presented in section 4 and simulation results with and without an MPEG codec in the channel are discussed in section 5.

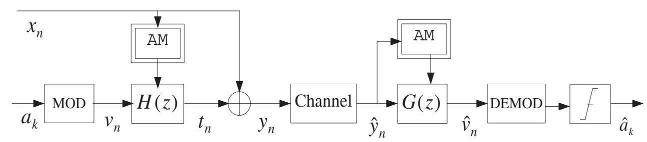

FIG. 1 –. *The generic watermarking scheme*

2. THE GENERIC WATERMARKING SCHEME

The considered watermarking system is depicted on fig.1. The emitter produces the spectral shaped watermark signal t_n and adds it to the original audio x_n in the time domain. The imperceptibly watermarked audio is then given by $y_n = x_n + t_n$. The receiver is composed of a Wiener filter G and a demodulator followed by a decision operation. It estimates the transmitted signature $\{\hat{a}_k\}$ from the observed data $\hat{y}_n$. The channel will be assumed to be ideal, the effect of the coder is introduced in section 5 as a robustness test to compression . So in the following, we will assume $\hat{y}_n = x_n + t_n$.

2.1. The emitter

The signature is a sequence of symbols a_k chosen in a finite alphabet $\mathcal{A} = \{a_0, \cdots, a_{K-1}\}$. The modulator is made out of a dictionary containing K spread spectrum and orthogonal vectors, where each vector corresponds to a given symbol a_k. The spread spectrum signal v_n is then constructed out of the signature by picking out the corresponding dictionary vectors. To ensure inaudibility, the watermark signal t_n is obtained by spectral shaping of v_n through the filter H, which squared frequency response module matches with the frequency masking threshold $M_x(f)$ of x_n, over processing windows of N samples. $M_x(f)$ is computed by an auditory model[1](AM) and is updated each N samples window. Hence, note that in the following, all filters, power spectral densities (PSD) and covariance functions and matrices are updated each processing window. The filter H is chosen pure recursive, of order P and causal with infinite impulse response (IR) h_n. The transmitted watermarked signal y_n is then given by :

$$y_n = v_n * h_n + x_n \tag{1}$$

2.2. The receiver

The detection process of the transmitted signature is achieved by a non causal and infinite IR filter G with output $\hat{v}_n$ described by :

$$\hat{v}_n = \sum_{i=-\infty}^{\infty} g_i \hat{y}_{n-i} \tag{2}$$

where g is the IR of G, optimized to minimize the mean square error MSE=E$[(v_n - \hat{v}_n)^2]$. Substituting (1) and assuming x_n and v_n uncorrelated, we prove easily [3] that the Wiener-Hopf equations to solve for g are of the form:

$$\sum_{i=-\infty}^{+\infty} g_i r_{\hat{y}}(k-i) = \sum_{l=0}^{+\infty} h_l r_v(k+l),\ k \in \mathbb{Z} \tag{3}$$

where $r_{\hat{y}}$ and r_v are the covariance functions of $\hat{y}_n$ and v_n.

To solve (3), we first assume the masking thresholds M_x and $M_{\hat{y}}$ to be very close, which is nearly true. Thus, the IR h_n can be computed by an AM at the receiver out of $\hat{y}_n$.

Secondly, to estimate $\hat{v}_n$, the infinite IRs g and h have to be truncated to finite lengths: $\underline{g} = (g_{-M_{nc}} \cdots g_{M_c})^t$ and $\underline{h} = (h_0 \cdots h_Q)^t$. The solution to (3) expressed in matrix form is:

$$R_{\hat{y}} \underline{g} = R_v \underline{h} \ \Rightarrow\ \underline{g} = R_{\hat{y}}^{-1} R_v \underline{h} \tag{4}$$

where $R_{\hat{y}}$ and R_v are the covariance matrices of $\hat{y}_n$ and v_n.

Note that at the receiver, $R_{\hat{y}}$ is estimated from only N $\hat{y}_n-$samples, because of the blockwise processing. R_v may be better estimated from a long white sequence of modulated symbols, since the emitter dictionary is available on the reception side.

The choice of the truncation orders M_c,M_{nc} is limited by the constraint $M_c + M_{nc} + 1 \leq N$ on the one hand, and by the poor accuracy of the estimated values of $r_{\hat{y}}$ for higher M_c and M_{nc} values on the other hand.

3. INSUFFICIENCIES OF THE GENERIC SCHEME

3.1. IR h_n truncation effect

Since the frequency response of the filter H is synthesized from M_x, and because M_x is nearly a -20 dB translated

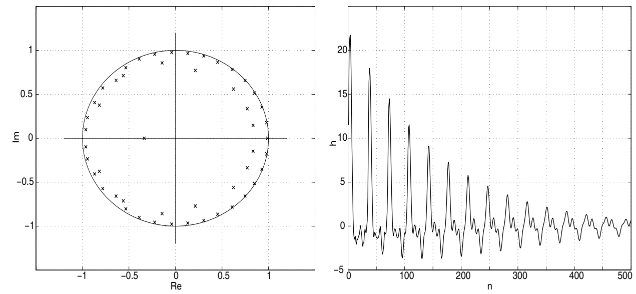

FIG. 2 –. *Example of poles ($P = 50$) and corresponding long IR h_n for a violin excerpt.*

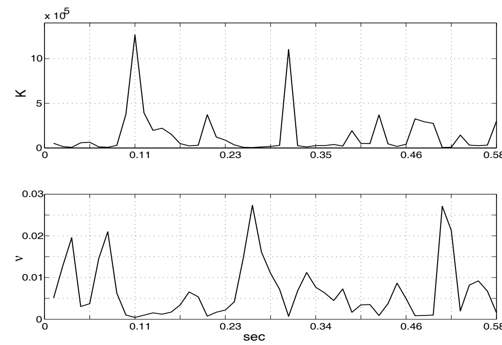

FIG. 3 –. *Condition number $K_{\hat{y}}$ and corresponding spectrum flatness $\nu_{\hat{y}}$ over time for a piano excerpt.*

version of $S_x(f)$, the PSD of x_n, $H(f)$ has the same harmonic character as $S_x(f)$. This implies that $H(z)$ has poles in proximity of the unit circle, resulting in very long IR h. Fig.2 illustrates this fact with a long IR example.
The choice of the truncation order Q is very sensitive, because it depends on the signal properties of the window being processed: harmonic character leads to long h and noisy character results in shorter IR. It is the well known over/undermodeling problem. Thus, setting the same Q for all windows leads to significant modeling errors on the solution of the system (4).

3.2. Bad conditioning of the Wiener-Hopf system

The accuracy of the solution of (4) can be measured by the condition number of the matrix $R_{\hat{y}}$, which is defined for a Toeplitz matrix by [4]:

$$K_{\hat{y}} = \max(\lambda_i)/\min(\lambda_i) \tag{5}$$

where the λ_i are the eigenvalues of $R_{\hat{y}}$. Indeed, a too large[2] condition number indicates an ill-conditioned system, since $K_{\hat{y}}$ estimates worst-case loss of precision. Furthermore, covariance matrices of correlated signals have large condition

1. The used AM is a modified version of MPEG model No 2.

2. Too large means roughly that $\log_{10}(K) \gtrsim$ precision of matrix entries.

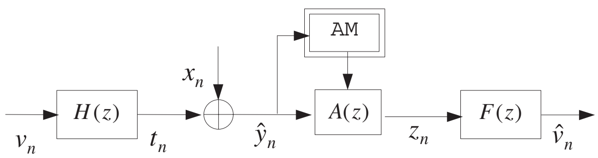

FIG. 4 –. *The proposed cascade realization of the detection scheme.*

numbers [5, 6]. In this context, we introduce the spectrum flatness measure ν [7]:

$$\nu = \frac{\left(\prod_{i=1}^{M} \lambda_i\right)^{\frac{1}{M}}}{\frac{1}{M}\sum_{i=1}^{M} \lambda_i} \tag{6}$$

where $M = M_c + M_{nc} + 1$ is the dimension of $R_{\hat{y}}$. Note that $\nu \to 1$ for a white signal and $\nu \to 0$ for a sine wave. Fig.3 illustrates $\nu_{\hat{y}}$ and $K_{\hat{y}}$ over time for a piano excerpt. It is obvious that over windows where $\nu_{\hat{y}} \to 0$, $K_{\hat{y}}$ becomes too large, indicating that the system (4) is ill-conditioned.

4. A NEW WIENER FILTERING BASED DETECTION SCHEME

We propose in the following a cascade realization of the Wiener filter given by (4). Both detection schemes are theoretically equivalent, except that the proposed one makes the most of the at detection available a priori knowledge of the spectral shaping filter $H(z)$. Indeed, the new scheme reduces the insufficiencies of the generic scheme and improves considerably its performances.

4.1. Cascade realization of $G(z)$

Since the spectral shaping filter is a causal and pure recursive filter with transfer function:

$$H(z) = \frac{1}{1 - \sum_{i=1}^{P} \alpha_i z^{-i}} = \frac{1}{A(z)}, \;\; \alpha_i \in \mathbb{R} \tag{7}$$

we may easily prove that the frequency response of the optimal Wiener filter is [3]:

$$\begin{aligned} G(z) &= A(z)\left[\tfrac{S_v(z)}{S_v(z)+A(z)A(1/z^*)S_x(z)}\right] \\ &= A(z)F(z) \end{aligned} \tag{8}$$

where S_v and S_x are the PSDs of v_n and x_n. Obviously, $G(z)$ can be viewed as the cascade of the inverse filter $A(z)$ followed by a non causal Wiener smoothing filter $F(z)$ reducing the filtered noise $b_n = x_n * \tilde{h}_n$, where $\tilde{h}$ is the IR of $A(z)$. The new detection scheme is depicted on fig.4.

One main advantage of the cascade filtering is that the IR $\tilde{h}$ of $A(z)$ is finite, thus we avoid the error introduced by the truncation of the IR h when solving the system (4).

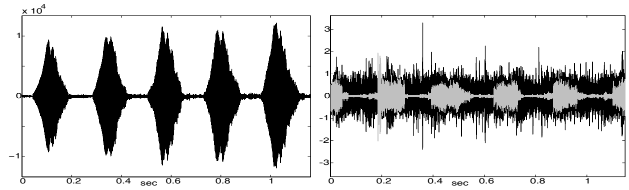

FIG. 5 –. *Nightingale song excerpt (left) and corresponding equalized outputs $\hat{v}_n$ (grey) and $\hat{v}_n^{'}$ (black).*

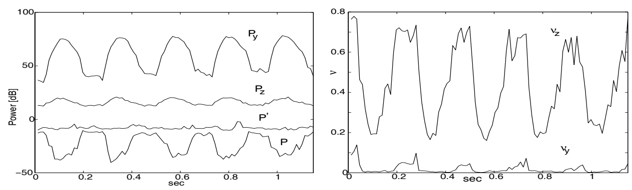

FIG. 6 –. *Comparison of instantaneous powers P_v and $P_{v'}$ (left) and spectrum flatness measures of $\hat{y}_n$ and z_n.*

Denoting z_n the input of $F(z)$, one may express the Wiener-Hopf equations with the equalized output $\hat{v}_n^{'}$ and the infinite IR f of $F(z)$ as follows:

$$\hat{v}_n^{'} = \sum_{i=-\infty}^{\infty} f_i z_{n-i} \tag{9}$$

where the transition signal z_n is given by:

$$z_n = \hat{y}_n * \tilde{h}_n \approx v_n + x_n * \tilde{h}_n \tag{10}$$

Solving for f by minimizing the MSE=E$[(v_n - \hat{v}_n^{'})^2]$, and because v_n and b_n are not correlated, we find the optimal non causal and infinite IR f:

$$r_v(k) = \sum_{i=-\infty}^{\infty} f_i r_z(k-i) \quad \forall\, k \tag{11}$$

or, in matrix form, after truncating f to $\underline{f} = (f_{-L_{nc}}, \cdots, f_{L_c})$:

$$\underline{f} = R_z^{-1}\, \underline{r}_v, \tag{12}$$

where R_z is the covariance matrix of z_n.

4.2. Properties of the transition signal z_n

The cascade decomposition (8) presents two main advantages. On the one hand, the IR $\tilde{h}$ is finite, thus we avoid truncation errors. On the other hand, the transition signal z_n has convenient properties: it is less correlated than $\hat{y}_n$ and it has nearly constant instantaneous power.

Fig.5 shows a nightingale song excerpt, having important power variations, and the corresponding outputs $\hat{v}_n$ and $\hat{v}_n^{'}$.

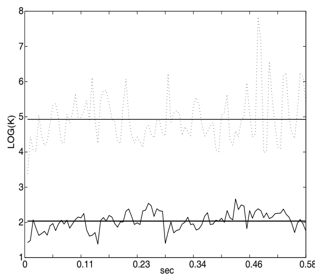

FIG. 7 –. *Comparison of condition numbers $K_{\hat{y}}$ (dotted) and K_z.*

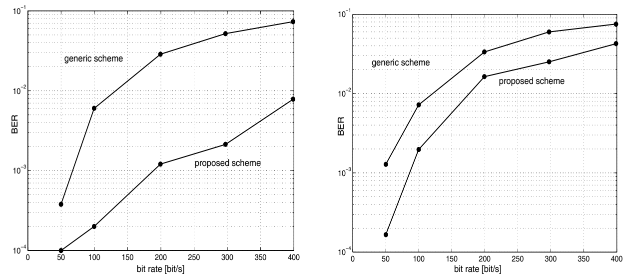

FIG. 8 –. *BERs enhancement: without (left) and with codec.*

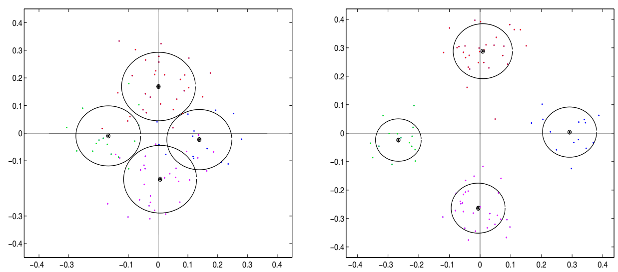

FIG. 9 –. *Example of detection constellations at 300 bits/s and without compression for the generic (left) and the proposed scheme.*

Fig.6 shows the instantaneous power values P_v, $P_{v'}$, P_y and P_z of the signals $\hat{v}_n$, $\hat{v}_n^{'}$, $\hat{y}_n$ and z_n. The fact that $P_{v'}$ has lower variations than P_v can be easily explained. Since $H(z)$ produces the watermark signal t_n from a white sequence v_n of power σ_v^2, its inverse $A(z)$ will translate the contribution of t_n in y_n to a white signal of constant power σ_v^2.

Further, the squared module of $A(f)$ is very close to that of the whitening filter $B(f)$ of the signal x_n to within a translation factor of ca -20 dB, so that the filtered noise $b_n = x_n * \tilde{h}_n$ is also less correlated. This is shown on fig.6 by the spectrum flatness measure ν_z, which is much higher than ν_y, pointing out that z_n is much less correlated than $\hat{y}_n$.

Hence, due to its convenient properties, the transition signal z_n has a better conditioned covariance matrix R_z, and the system (12) is solved with higher precision than (4). Fig.7 shows the decimal logarithm of $K_{\hat{y}}$ and K_z over time for a piano excerpt. Indeed, on this example, K_z is lower than $K_{\hat{y}}$ to within a multiplicative factor of ca 10^3.

5. SIMULATION RESULTS

The performances in terms of BER over bit rate for both detection schemes are shown on fig.8. The plots represent the BER average of 20 Monte Carlo simulations with 6 different audio excerpts. The signature was modulated at 2 bits/symb with random dictionaries of 4 orthonormal and spread spectrum vectors, resulting in 500 bits long sequences. The modulated signal v_n is white and of unit variance. The filter orders and IRs lengths were set to: $P = M_c = M_{nc} = L_c = L_{nc} = 50$ and $Q = 256$.

The proposed scheme enhances the BERs in the case without codec to within a multiplicative factor of ca 20. The detection performances in case of an MPEG codec[3] are also improved of ca a factor 2.

We depict on fig.9 an example of detection constellations for both detection methods without compression, where the barycenter and the standard deviation of the detected data are outlined by a thick point and a circle respectively.

The results of fig.8 and 9 point out that the proposed detection scheme is far more efficient than the generic one, even in presence of an MPEG codec in the channel.

3. The used coder is MPEG 1 Layer I at bit rate 96 kbits/s.

6. CONCLUSION

This paper presented an analysis of the Wiener deconvolution based detection in a perceptual watermarking scheme viewed as a hidden data transmission system, where two problems were pointed out: IR over/undermodeling and ill-conditioned deconvolution system.

The proposed cascade realization of the reception filter overcomes these insufficiencies and improves considerably the data transmission reliability in the *audio channel* at lower as at higher bit rates.

7. REFERENCES

[1] L. de C.T. Gomes, M. Mboup, M. Bonnet, and N. Moreau. Cyclostationarity-based audio watermarking with private and public hidden data. *109th AES Convention, Los Angeles*, September 2000.

[2] S. Larbi, M. Jaïdane, and N. Moreau. Audiowatermark detection for all-pass pirate attack:hybrid blind equalization/wiener deconvolution approach. In *EUSIPCO,France*, 2002.

[3] M. H. Hayes. *Statistical Digital Signal Processing and Modeling*. John Wiley, 1996.

[4] G.H. Golub and C.F. Van Loan. *Matrix computations*. Johns Hopkins University Press, 1983 (Second Edition 1989).

[5] F. Milinazzo, C. Zala, and I. Barrodale. On the rate of growth of condition numbers for convolution matrices. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-35(4):471–475, 1987.

[6] A. G. Ybarra and S.T. Alexander. Effect of ill-conditioned data on least squares adaptive filters. In *ICASSP, USA*, 1998.

[7] N. Jayant and P. Noll. *Digital coding of waveforms*. Prentice Hall, 1984.