AN AUDIO WATERMARKING SCHEME BASED ON AN EMBEDDING STRATEGY WITH MAXIMIZED ROBUSTNESS TO PERTURBATIONS

*C. Baras, N. Moreau**

GET - Télécom Paris
TSI Department
46 rue Barrault, 75013 Paris, FRANCE

P. Dymarski

Technical University of Warsaw
15/19 Nowowiejska, 00-665 Warsaw, POLAND

ABSTRACT

A particular application of audio watermarking systems consists in using the audio signal as a transmission channel for binary information. In this context, reliability criteria with respect to transmission rate, such as Bit Error Rate (BER), is a major issue that defines system performance. Current research has already proven the advantage of using informed embedding rather than blind embedding. In this paper, we present a new informed watermarking system based on an embedding strategy that maximizes robustness to perturbations such as compression and analog transmission. The proposed system performance is then compared to the equivalent blind embedding system performance. Experimental results permit to assess the achieved efficiency. For transmission rates, up to 500 bits/s, BERs are divided by 10 when channel perturbation is compression and divided by 4 in the case of analog transmission.

1. INTRODUCTION

Digital representation of audio signals has made data access easier and fostered illegal data exchange possibilities. Copyright protection has therefore become a major issue. For several years, watermarking techniques have been developed as a possible solution to face this problem. Scientific studies pointed out new watermarking application fields (broadcast monitoring or transactional watermarking for example). Our main concern is the particular case where audio signal can be viewed as a transmission channel, which can embed binary information. The watermarking process is then designed as a communication system with additive noise. The useful information (the watermarking) is hidden by noise (the "host" audio signal). In this context, the watermark has to be robust to classical distortions (which will be referred to as channel perturbations) such as compression or digital-to-analog and analog-to-digital conversion. The embedding strategy aims at transmitting as much information as possible with the best reliability. Thus, transmission rate (in bits per second) and Binary Error Rate (BER) define system performance.

When choosing the embedding process, it is necessary to conciliate perceptual distortion and watermarking detection constraints. These constraints can be represented by two regions of the audio signal space [1]. The embedder role consists in choosing a watermark which lies in the intersection of these two regions. The way of choosing the watermark depends on the required embedding strategy. The most promising, pointed out by Cox in [2], exploits the similarity between an additive watermarking system and a communication system with side information. Using the *a priori* knowledge of the host signal for defining an embedding strategy improves data hiding capacity. Indeed, Costa [3] proved that the capacity of such a communication system is only dependent on both channel perturbation and watermarking power values but not on the host signal. Moreover, he proposed an embedding scheme based on a structured codebook. Nevertheless, Costa's embedder implementation is limited by the codebook size. Current researches aim at approaching this model by proposing structured codebooks, designed with quantization processes or methods of fast codebook search. Most work use theoretical channel capacity as a design criterion as suggested by Costa, not taking into account that channel perturbation power may not be known during the embedding process. We propose a watermarking system using side-information and maximizing robustness to channel perturbation without making any hypothesis on perturbation power. The system is designed as a closed loop scheme introducing a local copy of the detection process at the embedder to take into account the knowledge of the audio signal. Its performance will be compared to the equivalent blind embedding scheme.

The outline of the paper is the following. In section 2, audio watermarking principles and the reference blind embedding system (BSE) are described. In section 3, the closed loop watermarking scheme and the embedding strategy maximizing robustness to perturbation which will be referred to as ESMR are presented. Experimental results are given in section 4 which allows us to analyze the impact of our embedding strategy on system performance.

2. AUDIO WATERMARKING SYSTEM PRINCIPLES

A reference audio watermarking system processing digital signals was developed in [4]. It was designed as a communication system, as shown in figure 1.

Source encoding process maps the hidden message into a sequence of L symbols $\{k_l\}_{l=1..L}$. Each symbol is chosen among the set $\{1, ..., M\}$ and codes for $N_{bs} = \log_2(M)$ binary digits. The embedding process requires an embedding codebook $\mathcal{S}$ containing M waveforms with length N : $\mathcal{S} = \{\underline{s}_m\}_{m=1..M}$. The modulation interface maps each symbol k_l into the k_l-th codebook waveform so that the modulated signal on each symbol interval $[(l-1)N...lN-1]$ is : $\underline{v} = \underline{s}_{k_l}$. To satisfy the inaudibility constraint the power spectral density of the embedded signal should be lower than a masking threshold given by a PsychoAcoustic Model (PAM). This threshold is to be taken into account when designing the filter $H(f)$ with impulse response $h(n)$. Fil-

*Thanks to ARTUS RNRT project for funding (http://www.telecom.gouv.fr/rnrt/projets/res_01_37.htm).

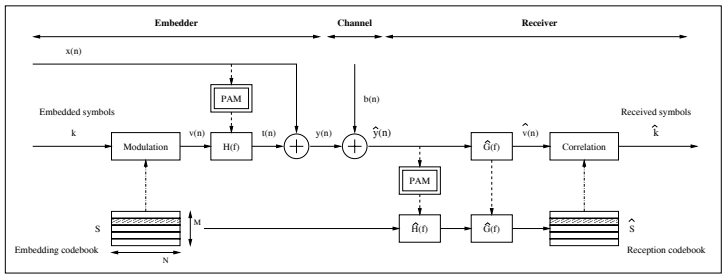

Fig. 1. Reference audio watermarking scheme.

tering $\underline{v}$ by $H(f)$ allows us to increase the embedded signal power and still respect the threshold constraint. $H(f)$ is designed so that the spectral shape of its response to a white noise with unit power matches with the masking threshold. As a consequence, embedding codebook waveforms are chosen to be white, gaussian and with unit power. The perceptual filter $H(f)$ yields the watermarking signal $t(n)$. The watermarked audio signal $\underline{y}$ is finally obtained by adding the watermarking signal $\underline{t}$ and the audio signal $\underline{x}$.

In the receiver scheme it is supposed that there is no symbol interference. Thus the received signal $\hat{y}$ is processed in each symbol time slot. It is first filtered by the whitening filter $G(f)$ obtained by performing the linear prediction of $x(n)$. Let $g(n)$ be the impulse response of this filter. Indeed in the case of a channel free from perturbation, $\hat{v}(n) = g(n) * x(n) + g(n) * h(n) * v(n)$ is the received signal in the well-known AWGN configuration, since $g(n) * x(n)$ is supposed to be white and gaussian. The watermarked signal at the receiver stage has only M possibles forms that define the reception codebook : $\hat{\mathcal{S}} = \{\underline{\hat{s}}_m\}_{m=1..M}$. Each of them is obtained by filtering the embedding codebook waveforms by $H(f)$ and $G(f)$. Then the optimum receiver is a correlation demodulator that selects the reception codebook waveforms whose correlation with the filtered signal is the highest. In order to minimize BERs [5] embedding waveforms are chosen to be biorthogonal. As $H(f)$ and $G(f)$ are not available at the receiver stage, they can be approximated by two filters $\hat{H}(f)$ and $\hat{G}(f)$. $\hat{H}(f)$ is designed based on the masking threshold of $\hat{y}(n)$ and $\hat{G}(f)$ whitens $\hat{y}(n)$. Finally, the reception dictionary $\hat{\mathcal{S}}$ is the filtered version of the embedding one by $\hat{H}(f)\hat{G}(f)$.

In our application channel we introduce two distortions types represented by an additive noise $\underline{b}$:

- a MPEG compression that yields strong distortion in the high frequency region of the audio signal, which led us to use band-limited waveforms [5] at the embedder,
- analog transmission that yields desynchronization of the information sequence, which led us to add a synchronisation mechanism to the previous watermarking system. This mechanism consists in embedding regular synchronisation patterns known at the receiver. The reception process starts with searching the synchronisation patterns in the received signal then detects information that follows each pattern.

3. CLOSED LOOP WATERMARK EMBEDDING

3.1. Using the local receiver

To take into account the knowledge of the host signal during the embedding process, a closed-loop watermarking scheme is designed, by adding to our reference scheme a local copy of the reception process at the embedder. The obtained scheme is presented in figure 2.

Supposing that $H(f)$ and $G(f)$ are good approximations of $\hat{H}(f)$ and $\hat{G}(f)$, the local copy of the receiver allows us to work out the whitened audio signal $\underline{\tilde{r}}$, the received watermarked signal $\underline{\tilde{w}}$ and the reception codebook $\tilde{\mathcal{S}} = \{\underline{\tilde{s}}_m\}_{m=1..M}$ for each analysis window. Thus, to transmit symbol k with no error, the following inequality must be satisfied at the input of the correlator :

$$\forall m \in \{1..M\}, m \neq k, (\underline{\tilde{w}} + \underline{\tilde{r}} + \underline{\tilde{b}})^t \underline{\tilde{s}}_k > (\underline{\tilde{w}} + \underline{\tilde{r}} + \underline{\tilde{b}})^t \underline{\tilde{s}}_m, \quad (1)$$

where $\underline{\tilde{b}}$ is the channel noise filtered by $G(f)$.

Let us now find a watermarking signal that satisfies the following constraints : respect the inaudibility constraint and be solution of the previous errorless transmission inequality. The inaudibility constraint is related to the design of $H(f)$, that imposes the acceptable distortion region given by :

$$\sigma_v^2 = \frac{\underline{v}^t \underline{v}}{N} \leq 1. \quad (2)$$

Errorless transmission depends on channel distortions which are supposed to be unknown during the embedding process. Consequently, we introduce a parameter σ_b^2 for each analysis window, that characterizes the system robustness to perturbations, similarly to Miller's embedding strategy, namely Maximizing Robustness [1]. Yet, errorless transmission conditions are parameterized by σ_b^2. It puts the detection region for a robust transmission into the following inequality :

$$\forall m \in \{1..M\}, m \neq k, (\underline{\tilde{w}} + \underline{\tilde{r}})^t (\underline{\tilde{s}}_k - \underline{\tilde{s}}_m) > \sigma_b^2. \quad (3)$$

The probability of error is controlled by σ_b^2 since transmission of k is free from error and robust to noise $\underline{b}$ when $\forall m \neq k, -\underline{b}^t(\underline{s}_k - \underline{s}_m) < \sigma_b^2$. Therefore, our goal is to find a watermark that lies in the intersection of these two regions with a maximum robustness parameter. We propose two approaches to solve this problem : first, to choose an adapted embedding codebook and second, to find a watermark adapted to a given codebook.

3.2. Choice of the codebook

As in Costa's model we structure the embedding codebook as a set of M sub-codebooks $\{\mathcal{S}_m = \{\underline{s}_m^p\}_{p=1..N_{vs}}\}_{m=1..2^{N_{bs}}}$ of size N_{vs}. To modulate symbol k the most appropriated waveform in $\mathcal{S}_k$ must be chosen. Equation (3) shows that the higher the correlation between $\underline{\tilde{r}}$ and $\underline{\tilde{w}}$ is, the better the detection. Therefore, k is modulated by the waveform $\underline{s}_k^{opt}$ whose correlation with $\underline{\tilde{r}}$ is the

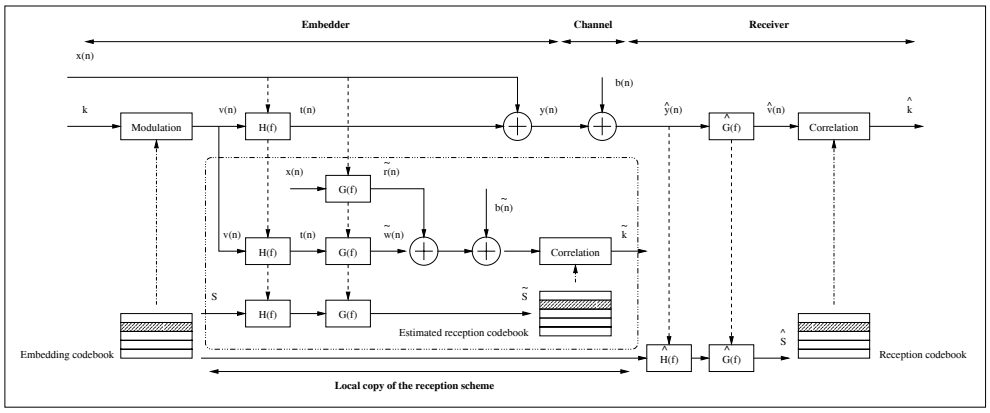

Fig. 2. Closed-loop watermarking scheme.

greatest. Robust transmission inequality defining detection region becomes :

$$\forall m \in \{1..2^{N_{bs}}\}, m \neq k, \forall p \in \{1..N_{vs}\}, \\ (\underline{\tilde{w}} + \underline{\tilde{r}})^t (\underline{\tilde{s}}_k^{opt} - \underline{\tilde{s}}_m^p) > \sigma_b^2. \quad (4)$$

3.3. Choice of watermarking signal

Considering the previous codebook, we aim at finding an inaudible watermarking solution of (4) with a maximum robustness parameter. Since intersection of acceptable distortion and detection regions can be empty, we first need to find under which conditions there is a solution $\underline{v}$ for an unspecified robustness parameter. Then we can find the maximum admissible robustness.

3.3.1. Existence of the solution

The existence of a solution can be seen as an optimisation problem. It consists in finding $\underline{v}^{opt}$ giving minimum audible distortion under the constraint that it satisfies the inequality (4). This optimisation problem is solved by introducing $(2^{N_{bs}} - 1) \times N_{vs}$ Lagrangian multipliers $\{\lambda_{m,p}\}$ so that :

$$\begin{cases} \underline{v}^{opt} = \arg\min_{\underline{v}} J(\underline{v}, \lambda_{m,p}) \\ J(\underline{v}, \lambda_{m,p}) = \frac{\underline{v}^t \underline{v}}{N} + \\ \quad \sum_{\substack{m=1 \\ m \neq k}}^{2^{N_{bs}}} \sum_{p=1}^{N_{vs}} \lambda_{m,p} (\sigma_b^2 - (\underline{\tilde{w}} + \underline{\tilde{r}})^t (\underline{\tilde{s}}_k^{opt} - \underline{\tilde{s}}_m^p)). \end{cases} \quad (5)$$

As the received signal is expanded over the $M = 2^{N_{bs}} N_{vs}$ reception codebook waveforms, $\tilde{w}(n)$ can be chosen in the signal space defined by the reception codebook $\tilde{\mathcal{S}}$. Thus, due to filtering linearity, $\underline{v}^{opt}$ is searched as a linear combination of embedding codebook waveforms : $\underline{v}^{opt} = \sum_{m=1}^{2^{N_{bs}}} \sum_{p=1}^{N_{vs}} \alpha_{m,p}^{opt} \underline{s}_m^p$.

Coefficients $\{\alpha_{m,p}^{opt}\}$ are obtained using Uzawa's method described in [6]. If the power of $\underline{v}^{opt}$ does not respect the acceptable distortion constraint, then intersection of acceptable distortion region and detection region is empty. Consequently, no solution of equation (3) is acceptable and the robustness parameter must be decreased to enlarge the detection region. On the contrary, if $\underline{v}^{opt}$ power satisfies the acceptable distortion constraint, embedding $\underline{v}^{opt}$ yields an inaudible and robust transmission for the given robustness parameter σ_b^2.

3.3.2. Maximizing robustness to perturbations

Maximizing robustness to perturbation consists in maximizing robustness parameter σ_b^2 so that the previous optimisation problem still has an acceptable solution $\underline{v}^{opt}$. The acceptable solution is related to $\underline{v}^{opt}$ power and we prove that $\underline{v}^{opt}$ power is an increasing function of σ_b^2. Therefore finding $\underline{v}^{opt}$ with maximum acceptable power (that is 1) is equivalent to maximizing σ_b^2. Finally we use an iterative algorithm gradually increasing σ_b^2 up to its maximum value and controlling $\underline{v}^{opt}$ power. With certain signal segments, maximum value of σ_b^2 may be negative. It means that no signal to be watermarked permits a transmission free from error, given the chosen codebook.

4. EXPERIMENTAL RESULTS

4.1. Test plan

System performance is evaluated by measuring the average BER for the binary transmission rate $R = \frac{N_{bs} F_e}{N}$. These average BERs are obtained by watermarking L binary digits on a set of 5 different digital audio signals (monophonic or polyphonic instrumental pieces, classical or light music), sampled at 44.1 kHz and which duration $\frac{L}{5R}$. Since results accuracy for a 70% reliability range is $\Delta = \sqrt{\frac{BER(1-BER)}{L}}$, we have chosen to transmit $L = 25000$ binary digits to achieve compromise between accuracy (lower than $\Delta = 3.10^{-3}$) and processing time.

System performance is presented for the following channels : a channel with no perturbation, a single compression process, a single analog transmission and a channel with both compression and analog conversion. Compression process is performed by an MPEG 1 Layer 3 digital encoder, operating at 96 kbits/s. Analog transmission is performed through an analog line connecting two PCs.

The used codebooks have the following properties. The size of the codebook for the Blind Embedding Strategy (BES) is 4 with $N_{bs} = 2$. The size for our Informed Embedding Scheme (IES) is 16 with $N_{bs} = 2$ and $N_{vs} = 4$. In both cases codebook waveforms are chosen to be biorthogonal and having a cut off frequency of 11kHz.

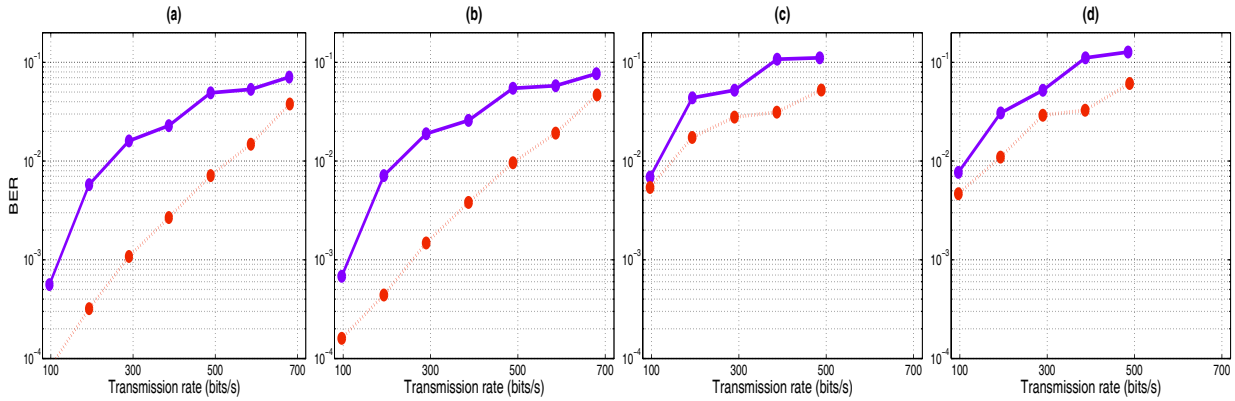

Fig. 3. BERs versus transmission rate with respect to embedding strategy for different channel configurations. Embedding strategy are the blind one (solid line) and the informed one (dotted line). Channel perturbations are the following : (a) without perturbation, (b) MPEG compression, (c) analog transmission, (d) MPEG compression and analog transmission.

4.2. Results

Perceptual quality of watermarked audio signal was first evaluated by the authors through informal listening tests, which are based on the recommendation ITU-R BS.1116. We concluded that watermarks were almost inaudible.

Subsequently system performance was measured. BERs versus transmission rate plots are presented in figure 3 for different channel perturbation and different embedding strategies. These figures shows the impact of IES on system performance. In the case of a channel free from perturbation presented in figure 3 (a), BERs with IES are divided by 10 compared to BERs with BES, for a transmission rate lower than 500 bits/s. A similar improvement is obtained in the case of an MPEG compression, which is presented in figure 3 (b). In both cases, the impact of IES is less significant beyond 600 bits/s. Indeed, with such transmission rate, symbol interference (mostly due to filtering) becomes more and more important : the hypothesis that there is no symbol interference at the receiver stage is any more valid. Improvement in the case of analog transmission with and without MPEG compression (respectively presented in figure 3 (d) and (c)) are also shown. With both embedding strategies, performance is roughly related to the synchronisation mechanism. Thus acceptable transmission rates are restricted by synchronisation unlocking, which appears starting from 600 bits/s. Up to 500 bits/s BERs with IES are divided on average by 4 compared to BERs with BES. In a nutshell, when a transmission reliability of 10^{-2} is required, IES permits a transmission robust to MPEG compression at 500 bits/s, whereas the acceptable transmission rate with BES is 250 bits/s. If desynchronisation of information sequence is introduced, the acceptable transmission rate with IES is decreased to 200 bits/s, whereas it is almost 130 bits/s for BES.

5. CONCLUSION

In this paper a new informed embedding scheme for audio watermarking has been presented. The embedder exploits a local copy of the receiver scheme to take into account the *a priori* knowledge of the audio signal. Our embedding strategy enables to choose an adapted watermark so that transmission robustness to additive channel perturbations is maximized. Impacts of this process on system performance, evaluated by Bit Error Rate (BER), have been described for channel perturbations of two types : an MPEG compression and a desynchronisation operation. This informed embedding scheme enables to significantly improve transmission reliability compared to a blind embedding scheme. Up to 500 bits/s, BERs are divided by 10 in the case of MPEG compression and by 4 in the case of a desynchronisation operation. Therefore a robust transmission through an audio channel with a 10^{-2} reliability can be achieved at 500 bits/s with a MPEG perturbation and at 200 bits/s with a desynchronisation perturbation.

Error correction codes and specific modulations should be introduced to improve the encoding procedure and prevent system from bursty errors. In particular, we are planning to evaluate the contribution of convolution codes, cyclic Hamming codes and treillis coded modulation to system performance. Testing has already started and preliminary results are very encouraging.

6. REFERENCES

[1] M. Miller, I. Cox, and J. Bloom, “Informed embedding : exploiting image and detector information during watermark insertion,” in *Proc. of Int. Conf. on Image Processing*, (Vancouver, Canada), IEEE, September 2000.

[2] I. Cox, M. Miller, and A. McKellips, “Watermarking as communications with side information,” *Proceedings of the IEEE*, vol. 87, pp. 1127–1141, July 1999.

[3] M. Costa, “Writing on dirty paper,” *IEEE Transactions on Information Theory*, vol. 29, pp. 439–441, May 1983.

[4] C. Baras, P. Dymarski, and N. Moreau, “Système de tatouage audio en boucle fermée,” *GRETSI*, 2003.

[5] J. Proakis, *Digital communications*. McGraw-Hill, 2001, 4th edition.

[6] D. Luenberger, *Linear and Nonlinear Programming*. Addison-Wesley, 1989, 2nd edition.