AN AUDIO SPREAD-SPECTRUM DATA HIDING SYSTEM WITH AN INFORMED EMBEDDING STRATEGY ADAPTED TO A WIENER FILTERING BASED RECEIVER

C. Baras, N. Moreau

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

ABSTRACT

A particular application of audio data hiding systems and watermarking systems consists of using the audio signal as a transmission channel for binary information. The system should ensure a reliable and robust transmission for various channel perturbations but also propose a low computational cost for real-time applications. In this paper we present a hybrid spread-spectrum data hiding system, which combines two reference systems taken from the State-Of-The-Art: the one based on a real-time receiver and the other one based on an informed embedding strategy with maximized robustness to additive perturbations. Experimental results permit to assess the efficiency of the system in terms of: (1) transmission reliability, which is significantly improved compared to reference systems, and (2) computational costs, which allows for the feasible real-time reception process of broadcast applications with off-line embedding.

1. INTRODUCTION

Audio data hiding research was developed jointly in response to the growing use of audio signals under digital format. Data hiding is a generic term which groups processes used to embed some information into an audio signal without any perceptual degradation. Embedded information brings an added value to the audio signal, which can be interesting for many applications [1] which hacker attacks are excluded from: it can be related to content description for indexing, labelling for monitoring, advertising for broadcasting, etc. It could be used as signature for copyright protection, but it would require to take pirate attacks into account, which is out of the context of this paper.

Spread-Spectrum (SS) data hiding systems are designed as a communication channel. A binary message, the watermark information, is embedded in a noise, which is the audio signal, according to an embedding process which conciliates perceptual distorsion and information detection constraints. The data hiding system should be robust to classical distorsions (further referred to as channel perturbations) applied to audio signals. These distortions described in [2] are: filtering, format change, noise addition, dynamic change and time stretching. The system design should finally offer a low computation cost for real-time application purposes. Therefore three criteria are used to evaluate the system performance: (1) perceptual distorsion measure, (2) Bit Error Rate (BER) with respect to transmission rates and channel perturbations and (3) processing time evaluation.

State-of-the-Art in the data hiding field can be resumed according to two major research directions:

- the choice of an adapted embedding strategy: several informed strategies (even for additive SS systems [3, 4]) have already been proposed to exploit the *a* priori knowledge of the audio signal during the embedding process and have proved their efficiency to improve system performance.
- the choice of an efficient receiver scheme: in particular equalization techniques seem to be promising [5].

In this paper, we propose to take benefits from each previous directions by designing a hybrid data hiding system. The proposed system relies on the Wiener filtering based detection scheme presented in [5] and on an informed embedding strategy inspired by [3]. Its receiver allows for a real-time reception process of the transmitted information and its embedder maximizes the system robustness to additive channel perturbations.

The outline of the paper is the following. In section 2, audio data hiding principles and reference systems, which our hybrid system is based on, are described. In section 3, the design of our hybrid system is presented. Experimental results are given in section 4 to evaluate systems performances and analyse the efficiency of our system compared to reference systems.

2. REFERENCE DATA HIDING SYSTEMS

We focus on two efficient systems designed for audio data hiding. The first one, proposed in [3], is an informed embedding system. Its embedding strategy aims at maximizing system robustness to additive channel perturbations. It will be denoted by IS. The second one, proposed in [5], targets the evaluation of equalization techniques on system performance by designing an efficient receiver based on a Wiener filter. It will be denoted by WS.

2.1. Audio data hiding principles

Both systems were designed from the same generic communication system shown in figure 1. Source encoding process maps the hidden message into a sequence of L symbols $\{k_l\}_{l=1..L}$, chosen among the set $\{1, ..., M\}$. The modulation interface uses an

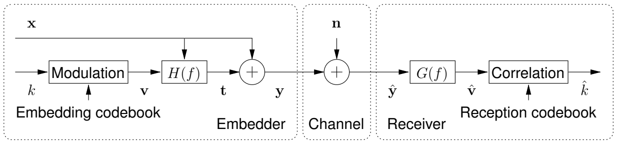

Fig. 1. Generic data hiding scheme.

0-7803-9332-5/05/$20.00 ©2005 IEEE

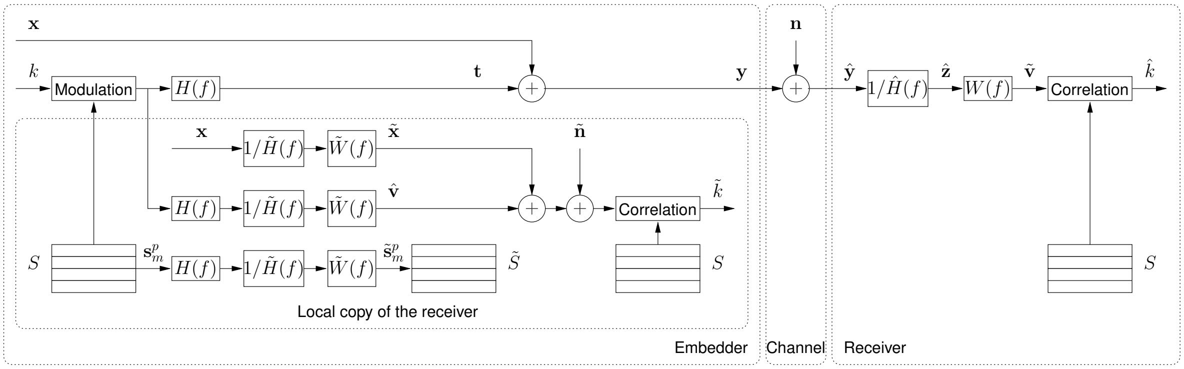

Fig. 2. Hybrid data hiding scheme.

embedding codebook $\mathcal{S} = \{\mathbf{s}_k\}_{k=1..M}$ containing M SS and biorthogonal waveforms with length N. Each symbol k_l is mapped into the k_l-th waveform of $\mathcal{S}$ so that the modulated signal on the l-th symbol interval $[(l-1)N...lN-1]$ is : $\mathbf{v} = \mathbf{s}_{k_l}$. To satisfy the inaudibility constraint, the watermarked signal $\mathbf{t}$ is constructed by filtering $\mathbf{v}$ with a psychoacoustic shaping filter $H(f)$, computed by a psychoacoustic study of the audio signal $\mathbf{x}$. The watermarked audio signal $\mathbf{y}$ is finally obtained by adding $\mathbf{t}$ to $\mathbf{x}$. The signal $\hat{\mathbf{y}}$ resulting from channel perturbations applied on $\mathbf{y}$ is first processed by a reception filter $G(f)$, yielding the received signal $\hat{\mathbf{v}}$. The detector is based on a reception codebook $\hat{\mathcal{S}}$ containing M waveforms, associated in a bijective way to the M embedding codebook waveforms. The decision interface selects on each symbol interval the reception codebook waveform whose correlation with the received signal $\hat{\mathbf{v}}$ is the highest. It supposes that the inter-symbol interference has been compensated.

2.2. Systems characteristics and performance

Systems IS and WS differ by the receiver scheme and the design of the modulated signal $\mathbf{v}$.

The reception filter of IS is the whitening filter of $\hat{\mathbf{y}}$. Consequently, the reception codebook $\hat{\mathcal{S}}$ must be evaluated on each symbol interval, since it contains the filtered versions of the embedding codebook waveforms by $H(f)$ and $G(f)$. It yields a high computational cost of IS's reception process, as it will be shown in section 4.2. WS uses an equalization receiver, made up with a zero-forcing filter followed by a Wiener filter that estimates $\mathbf{v}$. Here, reception and embedding codebooks are the same, which allows for a real-time reception process. Therefore, WS's receiver leads to a better computational cost than IS's receiver.

Let us consider the l-th symbol interval where the information k_l has to be embedded. WS's modulated signal is $\mathbf{v}_{WS} = \mathbf{s}_{k_l}$ whereas IS's one is a linear combination of the M embedding codebook waveforms: $\mathbf{v}_{IS} = \sum_{m=1}^{M} \alpha_m \mathbf{s}_m$. $\mathbf{v}_{IS}$ is specifically chosen to improve the probability of transmitting k_l with no error even when channel is distorted with an additive noise by using a local copy of the receiver scheme at the embedder. This last informed strategy results in lower BERs than those of WS as it will be proved in section 4.2.

Consequently we aim at designing a data hiding system that benefits from the advantages of both systems: an equalizer as reception scheme due to its low computational cost and an informed embedding strategy due to its high transmission reliability.

3. HYBRID DATA HIDING SYSTEM

Figure 2 illustrates the proposed hybrid data hiding system, further denoted by IWS.

3.1. Receiver scheme

The receiver scheme is identical to the WS's one and is defined by the following steps. $\hat{\mathbf{y}}$ is filtered by a zero-crossing filter $F_{zc}(f)$ to obtain $\hat{\mathbf{z}}$. $F_{zc}(f)$ is designed to compensate the perceptual shaping filter $H(f)$ computed from the audio signal $\mathbf{x}$, that is $F_{zc} = 1/H(f)$. Since $\mathbf{x}$ is not available during the receiver stage, F_{zc} uses an estimation of $H(f)$ computed from $\hat{\mathbf{y}}$. A non causal Wiener filter $W(f)$ is then designed to minimize the mean square error $MSE = E[||\mathbf{v} - \hat{\mathbf{v}}||^2]$. Its coefficients $\mathbf{w}$ are given by: $\mathbf{w} = R_{\hat{z}}^{-1} r_v$, where $R_{\hat{z}}$ is the covariance matrix of $\hat{\mathbf{z}}$ and r_v is the covariance function of $\mathbf{v}$. Finally, the detector is a correlation demodulator using the embedding codebook as reception codebook.

3.2. Embedder scheme

The embedder scheme is obtained by adding to the embedder presented in figure 1 a local copy of the previous receiver. Choosing the adapted embedding strategy deals with choosing for each symbol k_l, embedded during the l-th symbol interval, the adapted modulated signal $\mathbf{v}$ that conciliates perceptual inaudibility and detection constraints.

3.2.1. Inaudibitity and detection constraints

The inaudibility constraint is ensured by the perceptual shaping filter $H(f)$. Its design only imposes to choose a modulated signal $\mathbf{v}$ that satisfies the following inequality :

$$\sigma_v^2 = \frac{\mathbf{v}^t \mathbf{v}}{N} \leq 1. \tag{1}$$

Conditions of robust detection are given by the local copy of the receiver scheme. For the moment, let us suppose that $1/\tilde{H}(f)$ and $\tilde{W}(f)$ are designed to be good approximations of the reception filters. It allows us to estimate the signals playing part during the detection process. These signals are: the filtered audio signal $\tilde{\mathbf{x}}$, the filtered watermarking signal $\tilde{\mathbf{v}}$ and the filtered versions of each embedding waveform, represented as the codebook $\tilde{\mathcal{S}}$. Then, to transmit the symbol k_l with no error, the correlation vector between the received signal and the codebook waveforms has to reach its maximum value for the waveform $\mathbf{s}_{k_l}$ associated with k_l. This is equivalent to satifying the following $M-1$ inequalities at the input of the correlator:

$$\forall m \neq k_l, (\tilde{\mathbf{x}}+\tilde{\mathbf{v}}+\tilde{\mathbf{n}})^t \mathbf{s}_{k_l} > (\tilde{\mathbf{x}}+\tilde{\mathbf{v}}+\tilde{\mathbf{n}})^t \mathbf{s}_m, \tag{2}$$

where $\tilde{\mathbf{n}}$ is some additive channel perturbation. Since $\tilde{\mathbf{n}}$ is unknown during the embedding process, we introduce a parameter σ_n^2, as suggested in [4], that characterizes the system robustness to perturbations. Robust detection constraints (2) become:

$$\forall m \neq k_l, (\tilde{\mathbf{x}}+\tilde{\mathbf{v}})^t (\mathbf{s}_{k_l} - \mathbf{s}_m) \geq \sigma_n^2. \tag{3}$$

In this context, maximizing system robustness amounts to finding $\tilde{\mathbf{v}}$ satisfying (3) with a maximum robustness parameter σ_n^2 which is finally described by the following equation:

$$\tilde{\mathbf{v}} = \arg\max_{\tilde{\mathbf{u}}} J_1(\tilde{\mathbf{u}}), J_1(\tilde{\mathbf{u}}) = \min_{m \neq k_l} (\tilde{\mathbf{x}}+\tilde{\mathbf{u}})^t (\mathbf{s}_{k_l} - \mathbf{s}_m) \tag{4}$$

Thus, the adapted embedded strategy consists in choosing $\mathbf{v}$ (that yields $\tilde{\mathbf{v}}$ after filtering), built from the embedding codebook, that satisfies inaudibility (1) and robust detection (4) constraints.

3.2.2. Choice of the codebook and a waveform to be detected

We structure the embedding codebook $\mathcal{S}$ as a set of M sub-codebooks $\{S_m = \{\mathbf{s}_m^p\}_{p=1..P}\}_{m=1..M}$ each containing P biorthogonal waveforms. Each waveform of the sub-codebook S_{k_l} can be embedded to transmit the symbol k_l. Now, only one waveform of S_{k_l}, which is denoted $s_{k_l}^{opt}$, is the waveform which is the most likely to be detected. Indeed only one waveform maximizes the correlation with the received signal over the waveforms of S_{k_l}. Moreover, equation (2) shows that the higher the correlation between $\tilde{\mathbf{x}}$ and $s_{k_l}^{opt}$, the easier the detection. Thus, $s_{k_l}^{opt}$ is:

$$s_{k_l}^{opt} = \arg \max_{s_{k_l}^p \in \mathcal{S}_{k_l}} J_2(\mathbf{s}_{k_l}^p), J_2(s_{k_l}^p) = \tilde{\mathbf{x}}^t \mathbf{s}_{k_l}^p, \tag{5}$$

and the adapted received modulated signal $\tilde{\mathbf{v}}$ yielding a robust detection becomes:

$$\tilde{\mathbf{v}} = \arg\max_{\tilde{\mathbf{u}}} J_3(\tilde{\mathbf{u}}), J_3(\tilde{\mathbf{u}}) = \min_{m \neq k_l, p} (\tilde{\mathbf{x}}+\tilde{\mathbf{u}})^t (\mathbf{s}_{k_l}^{opt} - \mathbf{s}_m^p) \tag{6}$$

3.2.3. Choice of the watermarking signal

Considering the previous codebook and the waveform $s_{k_l}^{opt}$ which is the most likely to be detected, we intend to find $\mathbf{v}$ (and $\tilde{\mathbf{v}}$) that satisfies inaudibility (1) and detection (6) constraints. Since $\mathbf{v}$ is constructed from the embedding codebook, $\mathbf{v}$ can be expressed as a linear combination of the PM embedding codebook waveforms, rewritten under a vector representation:

$$\mathbf{v} = \sum_{m=1..M, p=1..P} \alpha_m^p \mathbf{s}_m^p = \mathcal{S}\alpha.$$

Due to filtering linearity, $\tilde{\mathbf{v}}$ becomes $\tilde{\mathcal{S}}\alpha$. The choice of $\mathbf{v}$ can now be stated as the evaluation of the coefficients α. This is related to the following optimisation problem under constraints:

$$\begin{cases} \alpha = \arg\max_{\lambda} J_4(\lambda) \\ J_4(\lambda) = \min_{m=1..M, m\neq k, p=1..P} (\tilde{\mathbf{x}} + \tilde{\mathcal{S}}\lambda)^t (\mathbf{s}_{k_l}^{opt} - \mathbf{s}_m^p) \\ \lambda^t \frac{\mathcal{S}^t \mathcal{S}}{N} \lambda \leq 1 \end{cases} \tag{7}$$

The coefficients α are obtained using a sub-optimal iterative algorithm with a step parameter ρ, inspired from [4], that proceeds as follows:

```
1.  $\alpha$ is null.
2.  Compute $\mathbf{v} = \mathcal{S}\alpha$ and $\tilde{\mathbf{v}} = \tilde{\mathcal{S}}\alpha$.
3.  If $\sigma_v^2 < 1$, find $\mathbf{s}_m^p$ with $m \neq k_l$ which minimizes $(\tilde{\mathbf{x}}+\tilde{\mathcal{S}}\alpha)^t(\mathbf{s}_{k_l}^{opt} - \mathbf{s}_m^p)$ and modify $\alpha$ as follows: $\alpha_{k_l}^{opt} = \alpha_{k_l}^{opt} + \rho$, $\alpha_m^p = \alpha_m^p - \rho$. Then go back to step 2.
4.  If $\sigma_v^2 = 1$, terminate.
```

3.2.4. Filters design

$\tilde{H}(f)$ and $\tilde{W}(f)$ are computed using the watermarked signal $\mathbf{y}_{WS}$ obtained by embedding the sequence of symbols into the audio signal with WS.

3.3. Synchronisation mechanism

A synchronisation mechanism is introduced to prevent the system from desynchronizing due to time stretching. These perturbations imply that the sampling rate at the receiver F_r is different from the audio sampling rate F_e. F_e is known from the receiver. F_r is estimated off-line using a learning sequence. Our mechanism consists of periodically embedding a synchronisation pattern into the audio signal, which splits the binary message into sub-sequences. At the receiver, each pattern is located using a sliding correlation technique with a stretched version of the synchronisation pattern with respect to the difference between F_r and F_e. Then, the following sub-sequence is detected using a stretched version of the reception codebook and the estimation of F_r is updated, by evaluating the time interval between two synchronisation patterns.

4. EXPERIMENTAL RESULTS

4.1. Test plan

System performances are evaluated through three criteria: the perceptual quality, BERs with respect to transmission rate R for various channel perturbations and computation cost. We have used a set of 20 audio signals, sampled at $F_e = 44.1$ kHz and watermarked with L binary digits to process L/R seconds of signal. We have transmited $L = 100000$ binary digits to achieve a compromise between accuracy of BER (lower than 10^{-3}) and processing time. Thus, measures of BER lower than 10^{-3} are not significant.

A various range of perturbations has been considered. We use the automated evaluation tool (and its default parameters) proposed in [2] from which we select only perturbations adapted to broadcasting application field. These are filtering operations, dynamics compression, loudness changes, echo adding and resampling. We also consider MPEG compression, performed by an MPEG 1 Layer 3 digital encoder, white noise adding with RSB=30 dB and an operation that simulates time stretching.

Systems are implemented using Matlab©version 6.1 for the embedder process and the C language for the reception process. The machine used for simulations has the following characteristics: Pentium 4, 1.80 GHz, 512MB RAM.

Performances of our hybrid scheme IWS are compared with those of the two reference schemes IS and WS. The used codebook contains $M = 2$ biorthogonal waveforms for WS and MP waveforms with $M = 2$ and $P = 4$ for IS and IWS, forming P

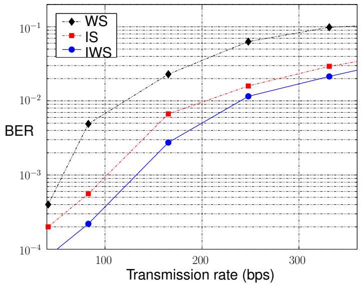

Fig. 3. BER vs transmission rate for a perturbation-free channel.

biorthogonal sub-codebooks. In both cases, codebook waveforms have a cut-off frequency of 6 kHz.

4.2. Results

Listening tests were performed to evaluate the perceptual distorsion introduced by the watermark. These tests are inspired from the UIT-R BS 1116 recommandation and were performed on a set of five people. Test results confirm that the watermark is "perceptible but not irritating" as defined on the perceptual grade.

BERs were measured for different binary transmission rate $R = \frac{\log_2(M)F_e}{N}$ and perturbations. Figure 3 presents the BERs of the three systems when the channel is free from perturbation. It confirms the efficiency of our hybrid system. Indeed, BERs with IWS are divided by more than 1.5 compared with IS's BERs and by more than 5 compared with IWS's BERs. Table 1 presents BERs obtained with IWS at $R = 165$ bits/s for perturbations yielding non desynchronisation. It proves that IWS is quite robust, since the most distorted perturbation is the MPEG compression at 64 kbits/s, for which the obtained BERs are still lower than the WS's BERs. Figure 4 shows the impact of time stretching on BERs for transmission rates lower than 165 bits/s and various time-scaling ratio $(1 - F_r/F_e)$. The use of the synchronisation mechanism leads to an increase in BER, even when the time-scaling ratio is null. Nevertheless a transmission with BERs lower than 5.10^{-3} can still be obtained when $R < 50$ bits/s or when $R = 80$ bits/s and the time-scaling ratio is low.

Finally, the computational cost has been measured as the ratio between the simulation time (in seconds) and the duration of the processed signals (in seconds) when $R = 165$ bits/s. Test results show that real-time is achieved at the receiver for IWS since the computational cost is 0.3 (whereas the IS receiver's one is 4.6). At the moment, the embedding stage can not ensure real-time processing since its computational cost is superior to the IS receiver's one. But its computational cost could be significantly reduced by optimizing the embedder implementation, which has not yet been done.

5. SUMMARY AND CONCLUSIONS

In this paper, an informed embedding scheme for SS data hiding system has been presented. It takes benefits from two State-Of-The-Art reference systems by exploiting the receiver of the one for its low computational cost and the embedding strategy of the other for its high transmission reliability. Its receiver relies on an equalization Wiener filter and its embedder uses a local copy of the receiver to choose an adapted watermark that maximizes system robustness to additive channel perturbations. System performances have been evaluated by listening tests, BERs measures for a range of classical perturbations and computational cost. The proposed scheme enables to significantly improve system performances compared to the two reference systems: a robust transmission with almost 10^{-3} reliability can be achieved at 50 bits/s even with a desynchronization perturbation. Real-time reception can be reached, thus making the application to broadcasting feasible as long as embedding is processed off-line. Yet we should turn our attention on the embedder, aiming to reducing its computational cost and eventually introducing specific modulations such as trellis coded modulation to further improve transmission reliability.

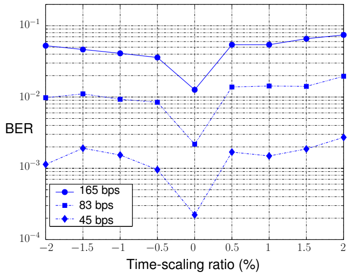

Fig. 4. BER vs time-scaling ratio in the case of time stretching.

Perturbation	BER	Perturbation	BER
any	$2.6\,10^{-3}$	high-pass filter	$5.7\,10^{-3}$
MPEG 96kbit/s	$3.3\,10^{-3}$	low-pass filter	$2.4\,10^{-3}$
MPEG 64kbit/s	$10.6\,10^{-3}$	resampling	$2.7\,10^{-3}$
white noise	$5.2\,10^{-3}$	echo	$7.6\,10^{-3}$
loudness	$5.2\,10^{-3}$	compressor	$5.8\,10^{-3}$

Table 1. BERs of IWS for Stirmark's perturbations.

6. REFERENCES

[1] L. Gomes, P. Cano, E. Gomez, M. Bonnet, E. Battle, "Audio Watermarking and Fingerprinting: For Which Applications?," in *Journal of New Music Research*, vol. 32, pp. 65-81, 2003.

[2] M. Steinebach, F. Petitcolas, F. Raynal, J. Dittmann, C. Fontaine, S. Seibel, N. Fates, L.C. Ferri, "StirMark benchmark: audio watermarking attacks", in *Proc. Int. Conf. on Information Technology: Coding and Computing*, pp. 49-54, 2001.

[3] C. Baras, N. Moreau, P. Dymarski, "An audio watermarking scheme based on an embedding strategy with maximized robustness to perturbations," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, vol. 17, pp. 357-360, 2004.

[4] M. Miller, G. Doerr, I. Cox, "Applying Informed Coding and Embedding to Design a Robust, High capacity Watermark," in *IEEE Trans. on Image Procesing*, vol. 13, pp. 792-807, 2004.

[5] S. Larbi, M. Jaidane, N. Moreau, "A new Wiener filtering based detection scheme for time domain perceptual audio watermarking," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, vol. 5, pp. 949-952, 2004.