



HAL
open science

Iterative Informed Audio Data Hiding Scheme Using Optimal Filter

A. Loboguerrero, Patrick Bas, J. Lienard

► **To cite this version:**

A. Loboguerrero, Patrick Bas, J. Lienard. Iterative Informed Audio Data Hiding Scheme Using Optimal Filter. 2003 International Conference on Communication Technology ICCT2003, Sep 2003, Bejin, China. pp.electronic version. hal-00166579

HAL Id: hal-00166579

<https://hal.science/hal-00166579>

Submitted on 7 Aug 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Iterative Informed Audio Data Hiding Scheme Using Optimal Filter

Alejandro LoboGuerrero, Patrick Bas, Joël Liénard

Laboratoire d'Images et Signaux
961 rue de la Houille Blanche, Domaine Universitaire BP 46
38042 Saint Martin d'Hères Cedex France
email: alobogue@lis.inpg.fr, patrick.bas@lis.inpg.fr, joel.lienard@lis.inpg.fr

Abstract- Audio watermarking is a method that allows the insertion of an imperceptible mark on an audio data set. Although the watermarking is often used to guarantee copyrights, it can also be used to increase the information transmitted in a communication context. In typical watermarking applications, the audio data set represents the host that supports the embedded information and it is considered as “noise”. However, we can use the fact that this signal is not random but completely known at the transmitter to reduce imperceptibility and improve robustness. Based on this idea, we derive a model from a classical data transmission technique and we take advantage of the knowledge of the host information to design an informed embedding scheme. This model has been then improved using an iterative controlling of the transmitted power and optimal filter. This informed embedding scheme allows us to provide a robust data-hiding system for several treatments, and it's compared to non informed and matched filter embedding schemes.

Keywords- watermarking (audio) ; spread spectrum ; datahiding (audio.)

I. INTRODUCTION

Watermarking is usually presented as a copyright protection technique. The basic idea consists on adding a mark (also called a signature) into the data to be protected[1]. This mark must be unnoticeable: it's not allowed to include any audible distortion while treating audio signals. Audio watermarking must also be robust to usual signal processing as well as to the malevolent attacks. However, this mark must be detectable by the authorized people.

In addition to its use for copyrights protection, the watermarking techniques can be employed in data transmission technologies for audio data hiding purposes [2,3]. Several applications are possible in this context. For example, if the priority information to be transmitted is the speech voice, then it may be useful to transmit video analysis information such as the lips motion parameters, the speaker's face or any other information identifying the speaking face. Such an application is addressed in the project ARTUS¹ which aims to embed animation parameters into audio and video contents.

In this paper, we want to present an enhanced audio watermarking scheme that is applied to digital communications .

Well defined constraints are required in this framework:

- The transmission rate of the inserted message must be as high as possible.
- The inserted message must be imperceptible to the listeners (No audible distortion is permitted).
- In addition to channel's noise, detection is perturbed by the audio signal used as support of the information.
- The embedded information must be preserved after audio compression. MPEG layer 3 (MP3) will be often used [4].

Contrary to copyright protection based watermarking techniques, the mark embedded message does not represent significant data and consequently is not subject to intentional attacks.

II. IMPLEMENTATION

A. General Scheme

The proposed watermarking method relies on a CDMA-like watermarking scheme [7]. This technique allows simultaneous multi-user communications and can be also used for low SNR (Signal to noise ratio) mono-users applications. It's based on the Spread Spectrum modulation technique that allows a homogeneous power distribution in all available bandwidth.

1) Embedding

The modulation processes uses just one single code to carry a binary message in this implementation. The values +1 and -1 represent this binary message. Positive version of the code corresponds to bit 1 and, negative version will be used to carry bit 0. (Fig. 1).

¹ This project was funded by national RNRT Artus Project

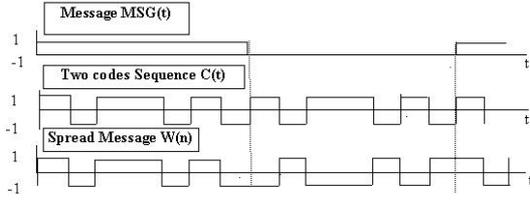


Figure. 1 Single Code Modulation

The meaning of this product can be expressed as the calculation of

$$W(t) = MSG(t) \times C(t) \quad (1)$$

In this equation MSG(t) represents the message that will be sent, C(t) the single code and W(t) the signal obtained from modulation process. This signal W(t) is then added to the speech signal to provide $S_w(t)$.

$$S_w(t) = S(t) + W(t) \quad (2)$$

In order to diminish the audibility of the message, the inserted codes have to fulfill several power and spectral distribution requirements. These conditions will be studied further in this paper.

2) Optimal Detection

At the reception stage, the SNR has been maximized. This is done in complement of the classical spread spectrum data hiding scheme in which the noise is considered to be white and gaussian. In data hiding applications, the audio signal plays the role of a noise. In fact, this “noise” (music, speech) has a very dynamical frequency behavior and it can be treated in an optimized way by the use of the Optimal Filter [8].

This kind of filtering maximizes the SNR for signals immersed in a non-white noise. The equation characterizing this filter is:

$$Fop(f) = \frac{C^*(f)e^{2\pi fT}}{S_s(f)} \quad (3)$$

This filter can be described in two parts. The first one has been noted as $C^*(f)e^{2\pi fT}$ and it is the matched filter to the sent signal C(t). The argument of the exponential represents a temporal delay and T is the signal duration. The second one, noted as $S_s(f)$ refers to a whitening process of the audio signal. In fact, it denotes the Spectral Power Density SPD of the signal to be whitened.

The detection is then computed by convolution of the received signal (sent signal + channel noise) with the Optimal Filter. This convolution is then sampled every T seconds in a synchronized way. The decision will be taken by comparing its value with a threshold fixed beforehand. The result of this operation will be called Y(t).

In this system the value of the threshold is zero (binary message). The decision law is described by the two classical following equations:

$$\text{If } Fop(f) \otimes S_w(f) \Big|_{t=T} > 0 \Rightarrow \text{Detected Bit} = 1 \quad (4)$$

$$\text{If } Fop(f) \otimes S_w(f) \Big|_{t=T} < 0 \Rightarrow \text{Detected Bit} = 0$$

B. Additional Improvements

Once the watermarking method has been defined, several modifications have been implemented in order to enhance the system behavior. The main idea of this implementation lies on the use of the value of each sound realization to attain the desired conditions. The resulting scheme is referred as informed embedding scheme [5] because the host signal is used to construct the emedded mark. This scheme has been improved take into account the modifications included by this implementation (informed embedding). This amelioration has been done in a iterative way yielding to an the iterative informed data hiding scheme.

This process can be decomposed in four stages:

1. Controlling the Codes Power.
2. Iterative Adjusting of the Power Control.
3. Spectral weighting of the code by the use of a human like psychoacoustics model.
4. A synchronization stage will be also presented

1) Informed Embedding by Codes Power Control

The proposed strategy allows us to minimize the perception of the added message. This is done by finding the minimal necessary power to obtain free error detection accordingly with the equations presented in (4). This control power is possible because of the knowledge of the

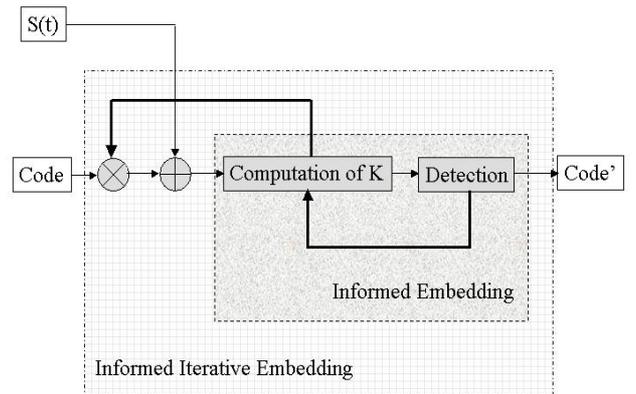


Figure 2. Informed Embedding and Iterative Embedding

audio signal in the transmission stage. This knowledge allows us to do a deterministic approach to the support signal and calculate the desired values. The use of this amount of power combined with the utilization of a psychoacoustics model guarantees not only free error detection but also a minimal audible distortion. To find this quantity of power, it is necessary to determinate the minimal coefficient value for which each code must be multiplied. Based on the equations (4) (Detection System) the Calculation yields the following result:

$$K \geq - \frac{\int_{\tau} S(\tau) Fop(t-\tau) \partial \tau}{\int_{\tau} W_m(\tau) Fop(t-\tau) \partial \tau} \Bigg|_{t=T} \quad (5)$$

In this equation C_m denote the code weighted by the human psychoacoustics model, F_{op} is the same as in (3).

In the Fig. 2, the region labeled "informed embedding" represents the calculation of K is done from a local replica of the detection system.

The modifications yield by this implementation affect the final design of the scheme, particularly the message detection step and the implementation of the psychoacoustics model. Before explaining those points, it is important to clarify the generated modifications.

- *Modification of $S_s(t)$.* In the equation (4) the Spectral Power Density of $S(t)$ and is substituted by $S_w(t)+N(t)$. Although there is a little difference between them, it can generate a high number of errors. The calculation of the coefficient K (5) was thus modified in order to decrease the effect produced by this approximation. This is explained in the paragraph 2).
- *Modification of $W_m(t)$.* In reception, It is necessary to determine the appearance of the transmitted code. This code has been spectral-weighted by the psychoacoustics model (stage 3)). To that end, we have assumed that the difference obtained while computing the psycoacoustics model from $S(t)$ and that one computed from $S_w(t)$ is negligible. This assumption has been corroborated by the final tests.

2) Power Control by Iterative System

We have modified the calculation of K regarding the difference between the desired and available Spectral Power Density. The idea is to recalculate the necessary power from the modified Spectral Power Density called S_{s+w} obtained from $S_s+W(f)$. This development takes us to the implementation of the iterative system represented in the Fig. 2 by the region labeled "Informed Iterative Embedding". The block "computation of K " shown in this figure is the same as that one described in (5). While calculating every single iteration, the system affects the computation until free error detection has been attained.

The resulting equations are:

$$K_{i+1} \geq - \frac{\int_{\tau} S(\tau) Fop_i(t-\tau) \partial \tau}{\int_{\tau} K_i W(\tau) Fop_i(t-\tau) \partial \tau} \Bigg|_{t=T} \quad (6)$$

$$Fop_i(t) = TF^{-1} \left[\frac{C^*(f) e^{2\pi j T}}{S_{s+K_i W}(f)} \right] \quad (7)$$

With $K_0=1$.

3) Psychoacoustics Model

The psychoacoustics stage allocates the spectral representation of the code in the frequencies where the human audio system is less sensitive. This is done by a frequency modulation of the code using a model derived from the compression standard MPEG-1 [6]

4) Synchronization

The synchronization stage is guaranteed by the addition of a synchronization code that is detected using classical sliding correlation techniques.

C. Definitive Implementation

Once we have totally developed each one of the stages involved in this development we can do a presentation of the final scheme (Fig. 3).

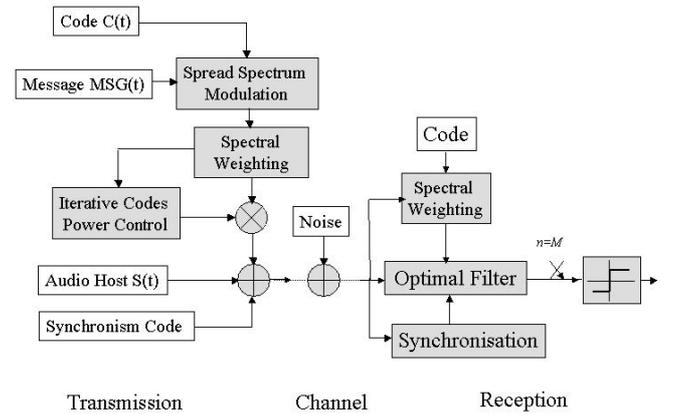


Figure 3. Definitive Implementation

III. RESULTS

To compare the proposed scheme to classical embedding systems, three computations of the Signal to Message Ratio have been done.

The first one corresponds to the proposed scheme (Fig. No. 3).

In the second test, the matched filter has replaced the optimal filter. The obtained power ratio compared to the result obtained in the first test allows us to compare the performances between the optimal and matched filter.

The third test is done to compare the behavior of our informed power control system vs. a classic implementation (open loop with constant power). In order to achieve this, we have used the constant power that is necessary to obtain the same number of errors and we have measured the Signal to Message Ratio. The comparison between the obtained power ratio and the first test shows the desired result.

Ten signals have been processed and the obtained results have been evaluated for a bit rate of 86 bit/s. The used sounds have been sampled to 44.1 kHz for 16 bits-bytes and resultant files have 116 seconds duration every one. This corresponds to a 10.000 information bits sent.

The Chart No. 1 synthesizes these results in dB's.

Although this implementation guarantees free error detection in the case of an ideal channel without additive noise, an Additive Gaussian Noise Channel (AGNC) with a SNR of 40 dB's has been simulated. In order to consider this eventual presence of noise, the computation of K has been adjusted. The BER obtained is 0.082. In these conditions the added message stays inaudible.

Proposed Scheme	Matched Filter Scheme	Constant Power
Optimal Filtering	Matched Filtering	Optimal Filtering
Iterative Power Control	Iterative Power Control	No Power Control
63,7	61,8	43,2

Chart No. 1 Obtained Signal to Message Ratios (dB's)

IV. CONCLUSIONS AND PERSPECTIVES

The data-hiding scheme based on the deterministic knowledge has been presented. The fact that the optimal filtering detection is more performant than the classical correlation detector has been corroborated by measuring the necessary power for free error detection. By the other hand, this approach outlines the fact that the audio host can be treated in a different way from a noise. An iterative data hiding system has been developed based on this idea and it has been shown that acceptable bit rates can be reached without perceptible audio distortion using this technique. This idea can be used to improve the performances of the system synchronisation.

V. REFERENCES

- [1] V. Basia and I. Pitas. Robust Watermarking in time domain" Proc EUSIPCO, volume 1, Rhodes, 1998.
- [2] A. LoboGuerrero, P. Bas, J. Lienard, "Implémentation d'un système de tatouage pour la transmission de données" JEP 2002, pp. 93-96. Nancy, France.
- [3] L. Boney, A. H. Tewfik and K. N. Hamdy. Digital Watermarks for Audio. In Eusipco, pages 473-480, Trieste, 1996.
- [4] The Fraunhofer Institute Internet HomePage <http://www.iis.fhg.de/amm/techinf/layer3/index.html>
- [5] M.L. Miller, I.J. Cox, J.A. Bloom, "Informed Embedding: Exploiting Image and Detector Information During Watermark Insertion," IEEE International Conference on Image Processing, vol. 3, pp. 1-4, 2000.
- [6] T. Painter, A. Spanias, "A Review of Algorithms for Perceptual Coding of Digital Audio Signals" in Proc. DSP-97, 1997.
- [7] J.G. Proakis, *Digital Communications*, McGraw-Hill, New York, 1995.
- [8] J. Max, *Methodes et Techniques de Traitement du Signal et Applications aux Mesures Physiques*, Vol II, Masson, Paris, 1986.