DIGITAL AUDIO WATERMARKING USING PSYCHOACOUSTIC MODEL AND CDMA MODULATION

Wahid Barkouti[1], Lotfi Salhi[1] and Adnan Chérif[1]

[1] Department of Physics, Faculty of science, Manar University, Tunis, Tunisia.

Wahid.Barkouti@fst.rnu.tn
Lotfi.salhi@laposte.net
Adnane.cher@fst.rnu.tn

ABSTRACT

DIGITAL WATERMARKING IS USED TO INSERT INFORMATION (A SIGNATURE) IN A COMPUTER DOCUMENT. THE ADDITION OF THE SIGNATURE MUST BE IMPERCEPTIBLE AND UNDETECTABLE BY ANY SYSTEM IGNORING ITS MODE OF INSERTION. IN PARTICULAR, IT MUST BE COMPLETELY INVISIBLE TO THE HUMAN EYE. THIS METHOD IS DIFFERENT FROM CRYPTOGRAPHY, WHICH HIDES A MESSAGE BY MAKING IT UNREADABLE. DIGITAL WATERMARKING NOW INCLUDES OTHER DATA WITHIN MUSIC SIGNALS. AUDIO WATERMARKING CONSISTS IN EMBEDDING INAUDIBLE INFORMATION IN AN AUDIO SIGNAL. THE WATERMARKING SYSTEM MUST GUARANTEE ONE TRANSMISSION OF THE INAUDIBLE, RELIABLE AND ROBUST INFORMATION FACE A SET OF DISRUPTIONS. IN THIS GOAL, WE PROPOSE A NEW STRATEGY OF INSERTION ADAPTED TO A WATERMARKING SYSTEM. THIS STRATEGY PERMITS TO CONSTRUCT AN INAUDIBLE WATERMARKING AND OF MAXIMAL HARDINESS TO THE ADDITION OF A NOISE.

KEYWORDS

Audio watermarking, CDMA modulation, intellectual property & psychoacoustic model.

1. INTRODUCTION

Significant changes in the global society are caused by rapid development of the Internet and the digital information revolution, ranging from the influence on the world economy to the way people nowadays communicate. Broadband communication networks and multimedia data available in a digital format (images, audio, video) opened many challenges and opportunities for innovation [1].

History of digital watermarking is associated since its inception to the protection of copyright [2]. Digital watermarking has been discovered for protecting digital media from tampering and enforcing digital copyright protection [3].

During these last years, the numeric watermarking signal spilled a lot facing the strong industrial demand concerning the protection of the intellectual property. The watermarking reveals the possibility to use the audio signal as a support of communication. In order to tackle the problem of unauthorized copy and distribution of digital audio data, watermarking systems are employed to insert copyright data into the audio stream [4].

Otherwise, the increasing developments of the exchanges of audio files under numeric format as preserving its resonant quality gave birth to the action of the pirates. For this fact, the

watermarking has been proposed as a solution to the problems of piracy.Audio watermarking was proposed as a potential solution; this technique consists in inserting in the audio signal a mark into the signature, inaudible and undetectable. Since it is extended to new domains of application, the watermarking system is seen as a chain of communication: the audio signal (signal of music for a quality CD) is susceptible to carry binary, inaudible and indelible information. Digital watermarking is the process of embedding or inserting a digital signal or pattern in the original data, which can be later used to identify the author's work, to authenticate the content and to trace illegal copies of the work. The inserted binary information is characterized by a debit and a reliability of detection, which one wishes the most elevated possible. A watermarking system must make objectives of the satisfy with performances, defined in terms of debit and binary success rate (BER) and of hardiness to the classic distortions, such of the operations of compression - reconstruction or conversions analogical/numeric [5]. Such a chain of communication also benefits to the emitter of the knowledge of the signal support, in which is drowned the watermark information [6]. To reach optimal performances, the emitter must be adapting to the signal support rather than the receptor to annul its effects. Therefore we focused our attention on the strategy of insertion [7], aiming to choose an adapted watermarking, reconciling distortion perceptually and reliability of detection.

This article presents in the first part a watermarking system, which is the basis of our work. The different watermarking techniques are presented in the second part, finally results of implementation of watermarking system are explained in the third part.

2. WATERMARKING SYSTEM

The watermarking system can be defined as a bruited chain of communication. Indeed, the watermarking is the signal to transmit and the audio signal is considered as a noise [8]. From the watermarking system represented by the figure1, three stages of the communication chain can be dedicated: the emitter that is the seat of the insertion of information, the channel and the receptor which are the seat of the detection.
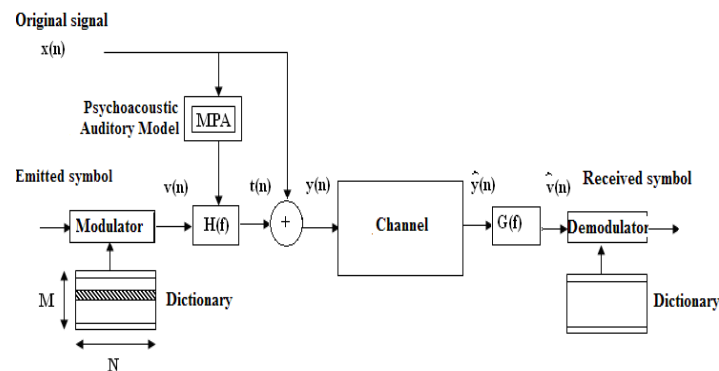

Figure 1. Watermarking System [4]

The signal modulated $v(n)$ is got by concatenation of a continuation of Sm vectors, the continuation of the indications m is being characteristic of the transmitted symbols (binary Nbs-upletses). Every vector Sm is chosen among a dictionary of S. broadcast. The modulated signal is

put spectrally in shape by a $H(f)$ filter, this filtering has for objective to adapt the spectral density of power of v to a doorstep of masking. The watermark signal is finally got by temporal summons between the audio signal $x(n)$ and the signal put in $t(n)$ shape [4].The audio signal faces two shapes of disruption; the first is an operation of compression - reconstruction, achieved by a MPEG coder. The second is a desynchronized operation achieved by the transmission via an analogical line between two PCs.

The received signal $\hat{y} = y + b$ is submitted to a filtering of Wiener that whitens the audio signal and estimates the signal module $\hat{v}$. Finally the detection of information is achieved by calculation of the cross-correlation between the signal and the vectors of a receipt dictionary, identical to the dictionary of broadcast.

2.1. Emitter

Information to give out is supposed to be a sequence of L binary M - Uplet noted under vectorial shape. The step of modulation aims to construct the signal modulated $v(n)$ carrier of the binary information while using a technique of CDMA modulation or MS. It requires a dictionary of s broadcast, containing M white vectors. The transmission debit R in communication is defined as being the number of bits of information transmitted by unit of time T_b [9]. The time bit T_b is bound to the frequency of sampling of the audio signal by the following relation:

$$T_b = N_b T_e \tag{1}$$

Where there is the number of the signal samples $x(n)$ during which one transmits a bit of information. So, the transmission debit is defined by:

$$R = \frac{1}{T_b} = \frac{f_e}{N_b} \tag{2}$$

The modulated signal $v(n)$ is put then in shape by an autoregressive filter of answer in $H(f)$ frequency. This operation aims to increase the power of the modulated signal by a report of one of the audio signal $x(n)$ while respecting the constraint of inaudibility imposed to the watermarking. This constraint can be characterized by a doorstep of $s_m(f)$ masking, homogeneous to a spectral density of power (DSP), descended of a psychoacoustic analysis of the audio signal. This doorstep establishes the limit frequencies to satisfy by the DSP of a signal that one would wish to add to the audio signal without misleading auditory deterioration. The $H(f)$ filter of formatting is conceived. Therefore, the DSP of the watermarking signal $t(n)$, resulting of the filtering of $v(n)$ by $H(f)$, coincides with the doorstep of masking. Its implementation requires a modulated signal $v(n)$ of power unit.

The audio watermark signal $y(n)$ is got then by simple addition between the audio signal $x(n)$ and the watermarking signal $t(n)$.

$$y(n) = x(n) + t(n) \tag{3}$$

2.2. Channel

The channel is the seat of the disruptions brought to the audio watermark signal $y(n)$ .It drives to the damaged audio watermark signal $y(n)$. In the particular case of a MPEG compression or operation of filtering, does it drag the suppression of the pass and the high frequencies of $y(n)$.

2.3. Receptor

The recuperation of the watermarking signal from the audio watermark signal exploits a filtering of Wiener. This filter achieves the equating of the channel and the scalding of the noise therefore, i.e. of the music signal [10].

The inversion of the channel takes place in recalculates H filters from the watermark signal. Thus, it exploits the psychoacoustic model and must be determined on identical windows of analysis to those of $H(f)$ therefore of size N_{HG} . While making the hypothesis that $y(n)$ has not been disrupted too much by the operation of compression-decompressioni.e. that $\hat{y}(n) \cong y(n)$, one has:

$$\hat{y}(n) = x(n) + t(n) \tag{4}$$

$$\hat{y}(n) = x(n) + \sum_{i \in H} h(i)v(n-i) \tag{5}$$

The audio signal $x(n)$ not being known at the time of the receipt phase, one estimates the $G(f)$ filters and $H(f)$ respectively by $\hat{G}(f)$, filter whitening of the signal $\hat{y}(n)$ and $\hat{H}(f)$ constructs from the doorstep of masking associated to $\hat{y}(n)$.

3. TECHNIQUES OF WATERMARKING

The techniques of watermarking, knew a very important development these last years in order to protect the intellectual property of the documents. Two techniques of modulations have been considered: the technical CDMA (Code Division Multiple Access) and the modulation of symbols (MS) to display the specter.

CDMA modulation allows several users to share the same transmission channel, in a simultaneous way. The technique of CDMA modulation will be the most adapted technique to the watermarking system. It permits to improve the performances of the system in different debits. This technique is clean to the numerical communications. The second modulation for the display of the specter is the modulation of symbols. It consists in associating an m - uplets to a unique element of the dictionary. This dictionary is constituted then of M=2m forms of waves chosen the some to the other to optimize the detection by the calculation of the interrelationship. To satisfy the setting in œuvre of the formatting filtering, these shapes of wave will be of power unit.

The call of the psychoacoustic is necessary in a watermarking system to establish the conditions of inaudibility of the transmission and the choice of the pace from what is constructed the modulated signal. The psychoacoustic becomes attached to modulate the human auditory system and the perception of the sound. This property consists in modifying the DSP of the code originally white in order to make it correspond to the pace of the psychoacoustic model chosen to

guarantee the inaudibility. The objective of this idea is to present the inherent properties of masking to the human ear which will be used for the realization of the formatting. The psychoacoustic auditory model is an algorithm to imitate the Human Auditory System (HAS). HAS is sensitive to a very wide dynamic range of amplitude and frequency. It is also very sensitive to additive random noise. Moreover it is difficult to reach any type of audio signal through the properties of the HAS.

4. RESULTS

The numeric conception of a watermarking system in real time under MATLAB permits to have a better analysis and development for the studied watermarking chain. Indeed, the numeric implementation under Matlab of a watermarking interface leads to the following stages:

- Calculation of the masking doorstep.
- The modulation of the signal.
- The calculations of the formatting filter coefficients (a_i and b_i).
- The last stage that consists to the formatting to have a watermark signal.

We are more especially interested by two psychoacoustic models: the first is described by Leandro D.C.T Gomes and the second by Ricardo. A. GARCIA.

4.1 Psychoacoustic Model

4.1.1. Model of Garcia

The psychoacoustic model of Garcia is a version of the model n°2 of MPEG. Garcia proposed to calculate the doorstep of masking on the critical strip scale to respect to best the physiology of the auditory perception. The DSP of the audio signal is transposed in the scale of Bark by integration of the power in every coins strip. The unique display function is applied for the modeling of the effects of masking of the membrane basilar. A transfer downwards of the found doorstep must be made with a α_G parameter. The optimal choice of this transfer parameter will give us the most effective model. For fixed to -45 dB, we get the doorstep of masking presented in the figure 3.
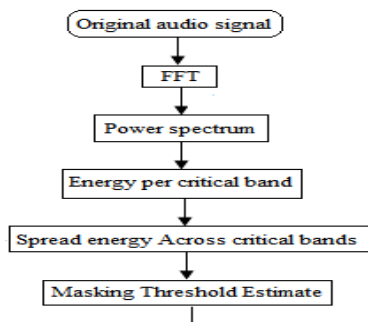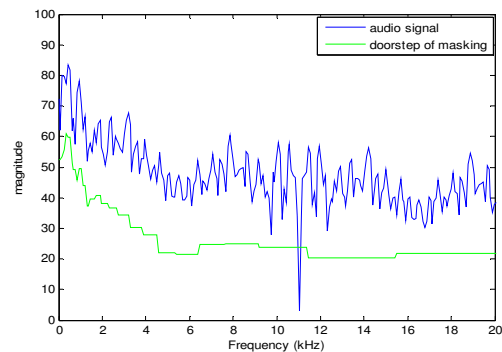

Figure 2. Garcia algorithm

Figure 3.Doorstep of masking

4.1.2. Model of Leandro

The model of Leandro is a version simplified of the model n°1 of MPEG whereas the DSP of the audio signal is decomposed in 4 coins strips. Indeed, in each of these coins strips, the DSP undergoes a dynamic compression, an operation of convolution by unique function of display (without distinction of tonality of the components) then a smoothing. The doorstep got there after translated to get a report of power between the doorstep of masking and the acceptable audio signal (RMS). This RMS influences the conditions of inaudibility.

This transfer (in the spectral domain) is programmed. It takes a factor of gain in the temporal domain. Thus, this doorstep offers the advantage of a very weak complexity of calculation (by report notably in the MPEG model).

```
┌─────────────────────────┐
│  Original audio signal  │
└─────────────────────────┘
             │
             ▼
┌─────────────────────────┐
│    Estimate of the DSP  │
└─────────────────────────┘
             │
             ▼
┌─────────────────────────┐
│    Dynamics reduction   │
└─────────────────────────┘
             │
             ▼
┌───────────────────────────────┐
│ Increase in the high frequencies │
└───────────────────────────────┘
             │
             ▼
┌───────────────────────────────┐
│ Smoothing by low pass filtering │
└───────────────────────────────┘
             │
             ▼
┌─────────────────────────┐
│    Masking threshold    │
└─────────────────────────┘
```
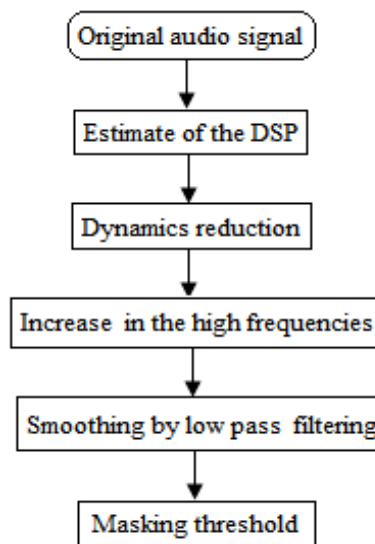
Figure 2.Leandro algorithm

The curve of the masking doorstep is displaced downwards of a parameter fixed a priori to -10 dB. This parameter can be adjusted according to the power of watermarking desired. An optimal choice of this parameter will guarantee us a model more effective psychoacoustic.
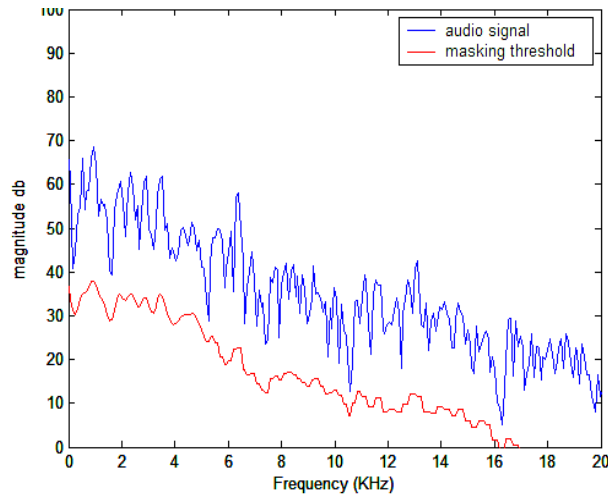
Figure 5. Doorstep of masking calculated by the Leandro model

The different doorstep observations presented in the figure 3 Show that for the Leandro model we get a curve that follows the variations of the DSP of the signal, but its dynamics and its power are less important. For the Garcia model, we obtained a continuous curve. It is the description on which the basilar membrane varies from a critical band to another.
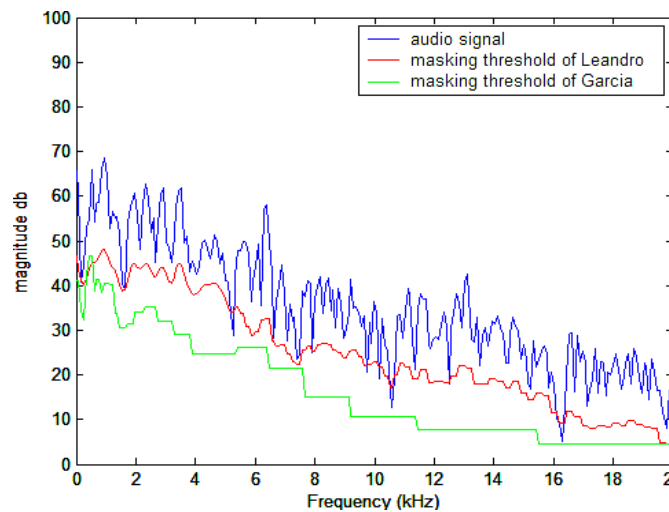

Figure 6. Doorstep of masking calculated by the two models

4.2 CDMA Modulation

We studied two techniques of modulations that are used in the watermarking chain to know the CDMA and the MS. The survey of the cost in terms of time of calculation shows that the CDMA modulation is the best. The techniques of CDMA modulation are the most adapted techniques to

the watermarking system. They permit to improve the performances of the system for transmission debits.

We deduct that the technique of simultaneous modulation CDMA presented in the table1 is more reliable at the time of the transmission of information on the watermarking chain.
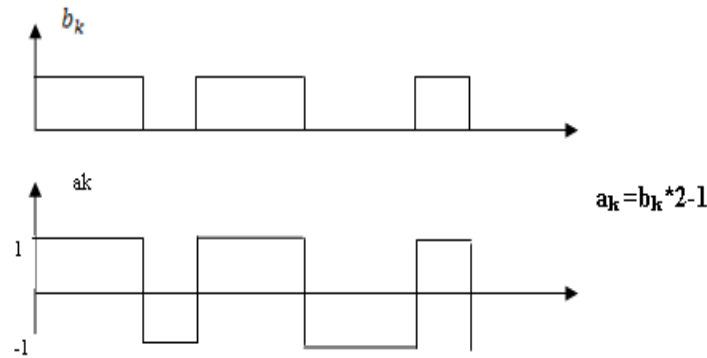

Figure 7. Coding technique

In the case of this modulation, the modulated signal $v(n)$ is a linear combination of vectors of the dictionary D whose emission coefficients are ± 1. The sequence of bits representing the binary message b_k to insert in the audio signal $x(n)$ is introduced into a coding module to form another binary sequence of coded symbols a_k where a_k is the amplitude Physics (± 1).

The encoding technique chosen uses a single code to insert a binary message. The positive version of the code is used to convert the bit 1 and the negative version is used to represent the bit 0. Figure 5 illustrates this type of coding.
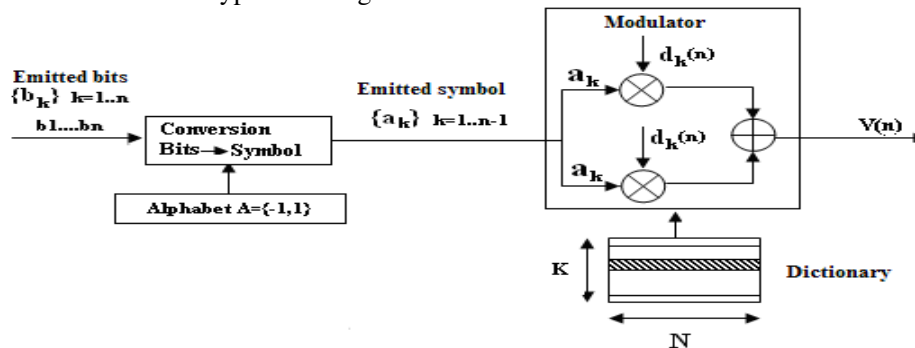

Figure 8. Principle of CDMA modulation

In this study, we consider a dictionary of M=m orthogonal waveforms, i.e. white vector length N and an m-tuple to be transmitted $(b_i ..., b_m)$.Each element b_k m-tuple is uniquely associated with a waveform dictionary $d_k(n)$. This waveform is then weighted by amplitude associated with physical a_k and b_k bits concerned; the information transmitted during the symbol time is the sum of all contributions of bits of the m-tuple, namely:

$$v(n) = \sum_{k=1}^{m} a_k d_k (n) \qquad\qquad (6)$$

The insertion of information representing the modulated signal CDMA sent in series on an original audio signal $x(n)$, we gave a signal watermarking $t(n)$ to represent in the following figure:
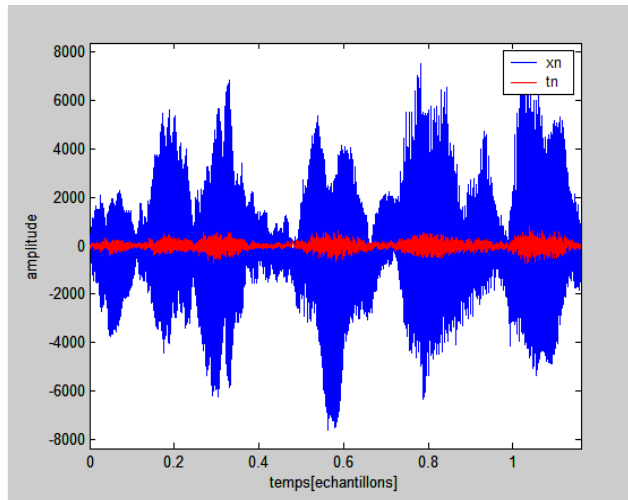

Figure 11. Evaluation of signals $x(n)$ and $t(n)$ in real times

While introducing the modulated signal CDMA messenger simultaneously and the signal of masking in a filter of formatting, we get the watermarking signal $t(n)$ visualized in the figure 12.
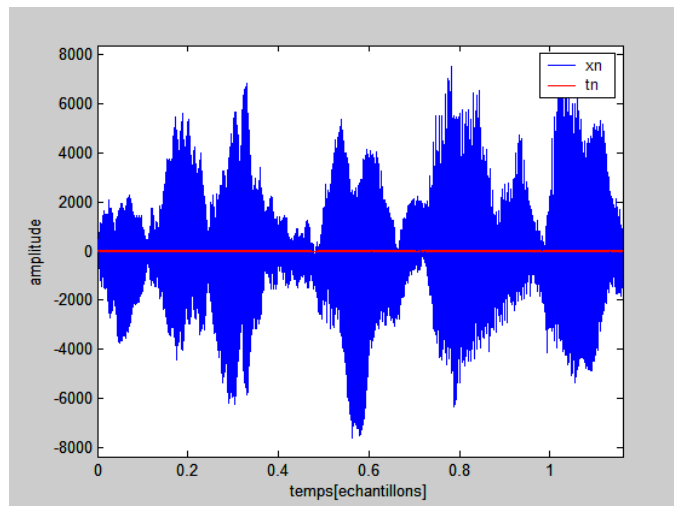

Figure 12.Evaluation of the signals $x(n)$ and $t(n)$ in real time

The sum of the watermarking signal $t(n)$ and the original signal $x(n)$ give the watermark signal figure13.
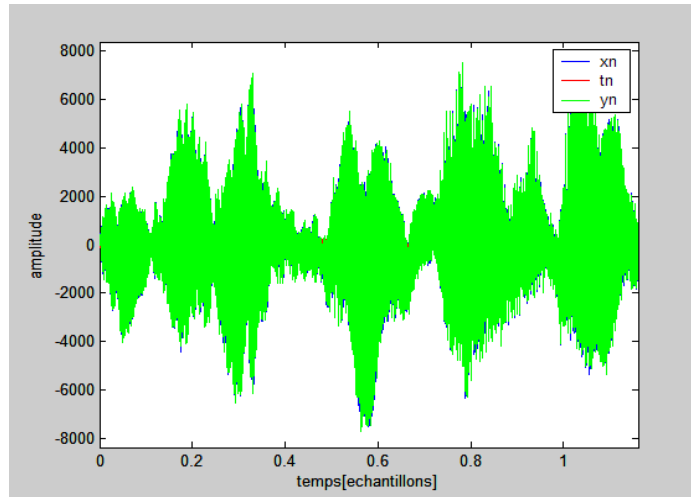

Figure 13. Temporal evaluation of the signals $x(n)$ and $t(n)$ and $y(n)$ in real time

While examining all these results, we can conclude that the watermarking system implemented following the different sections of modulation is more robust and reliable.

4.3. Detection

The phase of demodulation and decision is based on a correlation detector: to determine the similarity between information received and watermark forms taken by the tattoo at the reception. For Wiener filtering, these forms are the elements of the emission dictionary while for the whitening filter, we need to calculate a dictionary of receipt by filtering emission dictionary with filters in the chain like $H(f)$ and $G(f)$. Then, the receiver calculates the correlation between the signals received on windows of size N_S symbol times, and each element of the dictionary. After that, we choose among the M correlation values obtained the greatest. The transformation involving a symbol to an element of the dictionary is bijective so it is possible to determine the symbol associated with the chosen dictionary. We rebuilt the sequence of symbols received. The probability of transmission error chain is thus given as the ratio of the number of errors per symbol in the sequence of symbols emitted and the received sequence, and the number of symbols transmitted.

5. CONCLUSION

In order to improve the watermarking chain we have implemented two psychoacoustic models existing in the literature to know those of Leandro and Garcia and two techniques of modulation to study its performances to know the CDMA and the MS. The results of simulations using the models of Garcia and Leandro showed that the simultaneous modulation CDMA is the most reliable at the time of the transmission of information.

Our work puts in relief four constraints for the conception of the watermarking system that also define the notion of performance of the system: the transparency of insertion of the watermarking,

the transmission reliability, the transmission debit and the time of calculation to assure the implementation in real time.

REFERENCES

.

[1] Cvejic, Nedeljko, (2004) "Algorithms for audio watermarking and steganography", University of Oulu. ISBN 951-42-7383-4.

[2] Teddy Furon (2002) "Application du tatouage numérique à la protection de copie ", Doctoral thesis.

[3] Vimal Bibhu (2009) "On Line Secret Watermark Generation for Audio files", INTERNATIONAL JOURNAL OF COMPUTER AND ELECTRICAL ENGINEERING VOL.1, No.5.

[4] Senbin Yang, Wei Tan, Yanpu Chen, Wenjun Ma, (2010) "Quantization-Based Digital Audio Watermarking in Discrete Fourier Transform Domain", JOURNAL OF MULTIMEDIA, VOL.5, NO. 2.

[5] Cleo.Baras, P.D.Marski et N. Moreau. "Système de tatouage en boucle fermée". GRETSI_2003.

[6] I. J. Cox, M. L. Miller and A. McKellips. (1999) "Watermarking as communications with side information". Proceedings of the IEEE, 1999.

[7] M. L. Miller, I. Cox and J.A.Bloom, (2000) "Informed embedding exploiting image and detected information during watermark insertion", In IEEE International Conference on Image Processing.

[8] R. A. GARCIA, (1999) "Digital watermarking of audio signals using a psychoacoustic auditory model and spread spectrum Theory». The 107th Convention of Audio Engineering Society (AES), New York, USA, September 1999.

[9] Sonia Djaziri Larbi, (2005) "Structure d'égalisation en tatouage audio numérique", Doctoral thesis.

[10] Brossier (J-M), (1997) "Signal et Communication numérique Egalisation et synchronisation". Collection traitement du signal.

Authors

W.Barkouti is a researcher in signal processing laboratory at the University of Sciences of Tunis - Tunisia (FST). Degree in electronics and he received a M.Sc. degree in Automatic and Signal Processing (ATS) from The National Engineering School of Tunis (ENIT). Currently, he is a PhD student under the supervision of Prof. A. Cherif. His research interests include digital audio watermarking, psychoacoustic, and security of information.

L. Salhi is a researcher member of the Signal Processing Laboratory at the University of Tunis - Sciences Faculty of Tunis (FST). He received a M.Sc. degree in Automatic and Signal Processing (ATS) from The National Engineering School of Tunis (ENIT). Currently, he works a teacher of physics sciences and he prepares his doctorate thesis focused on the Analysis, Identification, and Classification of Pathological Voices using automatic speech processing.

A.Cherif received his engineering diploma from the Engineering Faculty of Tunis and his Ph.D. in electrical engineering and electronics from The National Engineering School of Tunis (ENIT). Actually he is a professor at the Science Faculty of Tunis, responsible for the Signal Processing Laboratory. He participated in several research and cooperation projects, and he is the author of international communications and publications.