A Robust Audio Watermarking Technique Operates in MDCT Domain based on Perceptual Measures

Maha Bellaaj
U. R. Signals and Mechatronic Systems, UR13ES49
School of Engineers of Carthage, ENICarthage, Carthage University
Tunis

Kaïs Ouni
U. R. Signals and Mechatronic Systems, UR13ES49
School of Engineers of Carthage, ENICarthage, Carthage University
Tunis

***Abstract*—the review presents a digital audio watermarking technique operating in the frequency domain with two variants. This technique uses the Modified Discrete Cosine Transform (MDCT) to move to the frequency domain. To ensure more inaudibility, we exploited the proprieties of the psychoacoustic model 1 (PMH1) of MPEG1 encoder layer I in the first variant and those of psychoacoustic model 2 (PMH2) of MPEG1 encoder Layer III in the second alternative to search the places for insertion of the watermark. In both variants of the technique, the bits of the mark will be duplicated to increase the capacity of insertion then inserted into the least significant bit (LSB). For more reliability in the detection phase, we use an error correction code (Hamming) on the mark.**

Next, to analyze the performance of the proposed technique, we perform two comparative studies. In the first, we compare the proposed digital audio watermarking technique with her two variants and those achieved by Luigi Rosa and Rolf Brigola, 'which we download the M-files of each'. The technique developed by Luigi Rosa operates in the frequency domain but using the Discrete Cosine Transform (DCT) as transformation and that proposed by Rolf Brigola uses the Fast Fourier Transform (FFT). We studied the robustness of each technique against different types of attacks such as compression / decompression MP3, stirmark audio attack and we evaluated the inaudibility by using an objective approach by calculating the SNR and the ODG notes given by PEAQ. The robustness of this technique is shown against different types of attacks. In the second, we prove the contribution of the proposed technique by comparing the payload data, imperceptibility and robustness against attack MP3 with others existing techniques in the literature.

Keywords*—*digital audio watermarking; Hamming; LSB; psychoacoustic model1, 2; MDCT; DCT; FFT; SNR; ODG

I. Introduction

Internet development and more generally the new means of communication oriented Western society into an era where digital takes a place increasingly important: Gradually digital cameras supplanting the old chemical film, DVD players replace VCRs as the Compact Discs were able to do with the disc vinyl.

In fact, digital copies are perfect now, whereas before each generation of analogical copies introduced further degradation. In addition, peer to peer networks for exchanging files can easily exchange huge volumes of multimedia data. Thus, providers of multimedia content quickly saw their sales fall significantly. They are very attentive to some new technology that will improve the digital rights management and prevent illegal redistribution of multimedia content.

However, digital technologies pose serious problems since it is easy to copy and deal with the computer digital documents. As a result, copyright is becoming increasingly unsafe and which leads to an illegal redistribution of data.

Hence the value of digital watermarking [1] as an effective solution to these problems and with the basic idea is to insert it in the digital document (image, sound, video ...) a signature so robust and imperceptible.

The watermark W is inserted into the multimedia document M to obtain watermarked document M' by application of an insert function I of the mark and using a key C. The marked media M' can undergo transformations T. The resulting document M" through an extraction function D for the detection of the brand W' (which may be identical or not to mark the initial W) or to confirm the presence or absence of the brand.

- *Robustness:* Being robust is equivalent to finding the inserted mark regardless of the changes that may infect the watermarked object. These changes are due to several types of treatments (attacks), as the passage in an analog channel resampling (print / SCANNING eg for images) compression with loss of information (such as JPEG compression [2] for images or MP3 to sounds), nonlinear distortions, additive channel noise ... However, the most marking system resist these simple transformations, but does not support the combination of them. Hence the idea of the malicious attack "StirMark" for the audio and image

- *Insertion capacity (Ratio):* This characteristic is important; it is the quantity of information that can be inserted into the original message. It is also called the payload or capacity of watermarking. Generally, in watermarking, we not fit voluminous raw data but enough useful information.

- *Imperceptibility:* In order to the mark is not easily crushed it should be imperceptible to any human: pirates, client, web visitors.... Too little perceptible mark is not robust and, worse, could be wrongly detected.

II. State of the Art

This section presents some digital audio watermarking techniques that exist in the literature.

- Boney, Tewfik and Hamdy [3] proposed a watermarking technique operates in the temporal domain. In this watermarking method, they build a brand that cans cryptographies. The psychoacoustic model calculates the masking threshold for the audio signal and filtering the brand. Then, the mark is added in the time domain to the audio signal.
- Fallahpour and Megias [4] proposed a technique using FFT transform. It involves inserting a stream of bits (secret bits) in selected FFT magnitudes while selecting the following parameters: scale factor, frequency band and frame size. Then, the FFT samples in the selected frequency band are divided into frames of size d. Then, they calculate the linear regression of FFT magnitudes for each frame. Thereafter, they calculate the average of the regression of magnitudes of the FFT samples in each frame. Finally, by means of a function f defined in [3] they insert the secret bits. The watermarked audio signal in the time domain is obtained by applying an inverse FFT. The extraction of the mark is performed using the FFT and the previously selected parameters and transmitted to the decoder in a secure manner.
- Fallahpour and Megias [29] propose another watermarking technique in the frequency domain using FFT transform. This technique consists integrate data and extract them in a bit-exact manner by modifying some values of the amplitudes of the FFT spectrum. The main idea is to divide the FFT spectrum in short frame and modify the amplitude of the selected samples using Fibonacci numbers that allow changing the frequency samples adaptively.
- S.T. Chen, H.N. Huang and C.J. Chen [5] proposed an adaptive audio watermarking method using the wavelet-based entropy (WBE). After the application of discrete wavelet transforms (DWT), they convert low-frequency coefficients into the WBE domain. Thereafter, they calculate the average values of each audio as well as derivation of some essential properties of WBE. From WBE and DWT coefficients, they calculate a characteristic curve. The basis of the integration process lies on the approximately invariant property demonstrated from the mean of each audio and the characteristic curve. In the extraction process, they use only values of the WBE.

A. *Objectives*

The main objective of this work is to develop a new watermarking system with high performance in which we seek to optimize these attributes:

- Increase the insertion capacity defined by the total number of bit to be inserted in the audio signal.
- The effect of this important capacity on the resulting signal. We seek to find a solution with which we get a tattooed signal faithful to the original.
- We estimate also reduce the complexity of the algorithms of both insertion and detection process.
- We aim to develop a technique that can resist against the maximum number of attacks and manipulations.

III. Proposed Technique

For after what we have presented above and after an extensive literature review [6, 3, 7, 8, 9, 10] led us to choose the frequency domain as an area of insertion of the mark point of view robustness and inaudibility hence the idea of using MDCT [19] to the time-frequency mapping [11, 3, 8]. On the other hand, MDCT allows a finer frequency resolution and studies the effects of edges. In addition, temporal methods based on LSB provide good results in terms of inaudibility and integration capability, but in part against, they introduce the brand damage which reduces the robustness factor. For these reasons, we applied this method (LSB method), but in the frequency domain. Parallel with this spectro-temporal processing, substitutive methods offer the possibility of using an error correcting code to reduce the error rate. For this, we have added to the proposed technique an error correcting code: Hamming code.

To reduce the audibility of the brand when inserting the bits of the brand, we have exploited the properties of the psychoacoustic model 1for variant 1 and the psychoacoustic model 2for variant 2 of the MPEG standard in the search for insertions positions. In the final step, we duplicated the bits of the brand to enhance the robustness of the proposed method.

- MDCT:

The MDCT is a lapped transform that allows doing the time-frequency mapping. It's widely used in audio processing and in particular it is used during MPEG audio compression procedure [12].

The MDCT coefficients are separated into "Low-Frequency" part and "High-Frequency" part. The major interest of this transform is that the coefficients are real and there are also robust to manipulations changing their values.

The general expression of the MDCT applied on blocks of N samples is as follows:

$$X(f) = \frac{2}{\sqrt{N}} \sum_{n=0}^{N-1} x(n) w_a(n) \cos\left(\frac{2\pi}{N}(n+n_0)\left(f+\frac{1}{2}\right)\right) \quad (1)$$

With:

- $x(n)$: Frame of N time domain sample.
- $f \epsilon \left[0, \frac{N}{2} - 1\right]$
- $n_0 = \frac{1}{2} + \frac{N}{4}$
- w_a : The analysis window with duration N.

The expression of the inverse transform (IMDCT) is:

$$y(n) = w_s(n)\frac{2}{\sqrt{N}}\sum_{f=0}^{\frac{N}{2}-1} X(f)\cos(\frac{2\pi}{N}(n+n_0)(f+\frac{1}{2})) \qquad (2)$$

With:

- $n \in [0, N-1]$
- w_s : The synthesis window with duration N.

We note that the MDCT of N sample of the signal gives only N / 2 coefficients, so it's not an invertible transform.

However, by tacking a recovery of 50% between successive blocks and ensuring some conditions on the analysis and synthesis window we can achieve a perfect reconstruction of the time signal from unmodified MDCT coefficients.

A. *Insetion of the mark: first variant of the technique*

In this first version of the proposed watermarking technique we will integrate the psychoacoustic model 1 of the MPEG standard Layer I. In the following, we will explain in detail the different modules constituting the algorithm of the insertion process:

1) Original signal: The initial input signal is an audio file of format (.wav), stereo or mono, sampled with a sampling frequency Fs = 44.1 kHz.

2) Decomposition into blocks: To increase the insertion capacity (number of bits to be inserted), we proceeded to the block decomposition of the signal. After reading and verifying the original audio, we obtain a vector containing the time samples. If the latter is stereo, we transform it into a mono audio signal. The final signal will be broken down into blocks of 1024 samples each. Thus we obtain a matrix ‹‹ block (i,k)›› of 1024 rows and NB_block columns: i=1…1024 and k=1…NB_block =length(original_audio)/1024. We can formulate this step by the following equation:

$$original\_audio = \sum_{k=1}^{NB\_block} Block(k) \qquad (3)$$

3) MDCT: This version of the technique operates in the frequency domain. To make the time-frequency mapping, we applied the MDCT on each block of 1024 samples. The coefficients of the MDCT are separated into two bands: band of "low frequency" and band of "high frequency". The major interest of this transformation is that the coefficients are real and also robust to manipulations changing their values. The output of this module is a matrix block_MDCT containing the set of blocks of 1024 frequency samples obtained by the MDCT. By applying equation (1) we get for a block k:

$$block\_MDCT(:,k) = MDCT(Block(:,k)) \qquad (4)$$

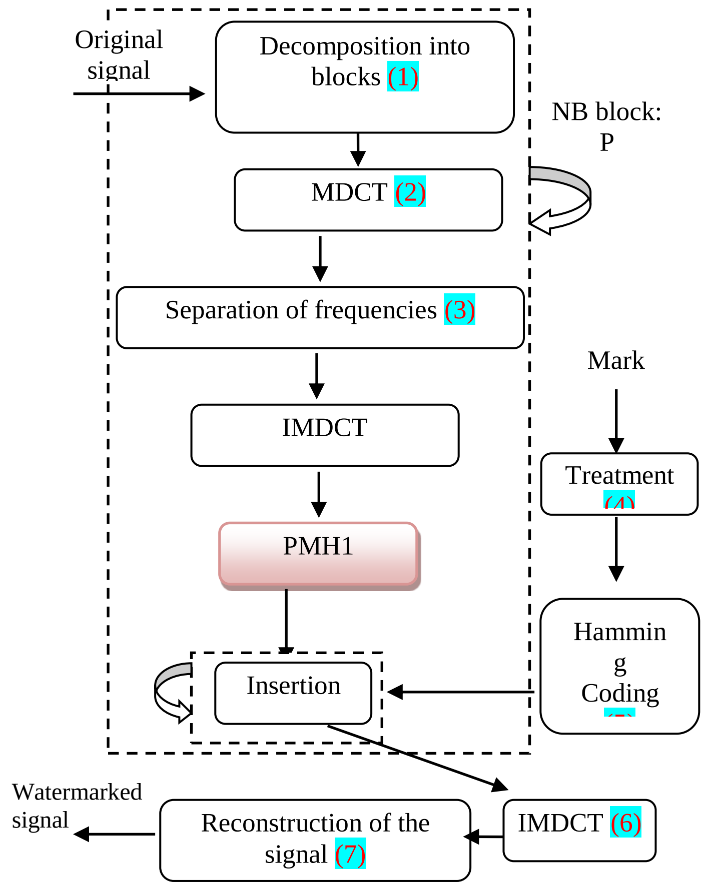

Fig. 1. Insertion schema of the mark: version 1

4) Separation of frequencies: The purpose of this module is to limit the low frequency band where we will perform the insertion. The matrix block_MDCT forms the input of this module. We take every block that is to say each column of the matrix block_MDCT, and we set the low-frequency band. The choice to integrate watermarking bits in the low frequency band due to the fact that the latter is much less sensitive to the attacks applied on the watermarked signal that the high frequency band which will be changed and even eliminated (especially compression / MP3 decompression).

We obtain at the end of this step a matrix low_freq containing all the low frequencies for each block k.

5) IMDCT: As the psychoacoustic model of the MPEG 1 standard operates on temporal samples. So it is necessary to go back to the time domain. By applying equation (2) of the IMDCT on each column of the matrix low_freq (which correspond to the low frequency band of each block k), we get the matrix low_freq_IMDCT of i = 1 … .512 rows and k = 1 … .NB_block columns:

$$low\_freq\_IMDCT(k) = IMDCT(low\_freq(k)) \quad (5)$$

6) PMH1: This module consists to look for insertion positions watermarking bits using a human psycho acoustic model. It takes as input the columns of the matrix low_freq_IMDCT formed by temporal samples. We will inject the bits of the brand in the components located under the masking curve calculated by the PMH1. And since, below this curve no modification is noticeable; this guarantees a good criterion of inaudibility of the brand. The application of the PMH1 and calculates of the masking curve will be for each block of 512 temporal samples of the low frequency band. We will have at the end of this step X insertion positions where we will substitute watermarking bits. The masking curve for a block k (one column of the matrix low_freq_IMDCT) is given by:

$$masking\_curve(k) = PMH1(low\_freq\_IMDCT(k)) \quad (6)$$

$and;$

$$X\_position(k) = components\_below\_masking\_curve(k) \quad (7)$$

7) Treatment: The process of inserting of the proposed technique can integrate any type of mark (text, image, and beep). But next in the experiments and evaluation of this version of the technique, we tried with text mark (string). Parallel to the above steps, we will then binarizing ASCII codes. The length of the mark (text) is chosen multiple of 8. We then get after binarization and shaping of the brand a binary vector of length multiple of 8. This choice will then be useful to make a Hamming (12, 8) coding to each byte of the binary vector.

8) Hamming coding: Following the insertion, detection of signature and attacks or manipulations which may act on the watermarked audio signal, the inserted bits of the mark may be changed (inversion from 0 to 1 or 1 to 0). Therefore, to improve the mark detection we proceed to the coding of the binary vector obtained by the Hamming coding after the step of binarizing and shaping described above, to ensure bit correction if it's necessary.

9) Insertion: The general principle of insertion of the brand can be described as follows: For each block k, the current bit of the message (obtained from the previous step of the coded binary vector) is substituted with the least significant bit "LSB" of the sample X_Position (i) searched in the step (6). The watermarking technique is substitutive and the insertion is performed on all blocks by inserting N times each bit of the mark (duplication of the mark bits). N is calculated based on the number of the components which are below the masking curve and the mark size.

$$N = \frac{\sum_{k=1}^{NB_{Block}} X\_position(k)}{length\_mark} \quad (8)$$

We get at the end of this step the signal block_wat "watermarked block".

10) IMDCT: After the substitutive insertion of the watermarking bits in each block, we get NB_Block watermarked in the frequency domain. In order to reconstruct the final watermarked audio signal, we need to move to the time domain.

11) For this reason, we will apply the IMDCT but this time on blocks of length 1024 samples. We obtain at the end of this step a matrix block_wat_tim containing time blocks watermarked and of dimension i = 1...1024 lines and k = 1...NB_block columns. So for a block k of size 1024, we get:

$$block\_wat\_tim(k) = IMDCT\ (block\_wat(k)) \quad (9)$$

12) Reconstruction of the signal: This module is the final step of the insertion process. Given all the time NB block watermarked block_wat_tim obtained by applying the IMDCT, we join them to form the watermarked audio signal audio_wat mono of format ".wav", sampled with the same frequency sampling the original audio signal Fs = 44.1 kHz.

B. Detection of the mark: first variant of the technique

This section details the different fundamental steps of the detection process. In the proposed technique, the detection is performed inversely to the insertion. One of the merits of this first variant of the proposed technique is that the detection is blind and does not require the original audio signal for detecting the mark. The set of parameters required when detecting and must be secured by the tattooist are:

- The number N of duplication that we can insert a bit.
- The list of positions of the components located under the masking curve sought by the human psycho acoustic model 1 in the insertion phase.

In the following, we will detail the various modules constituting the detection algorithm.

1) Decomposition into blocks: First, we decompose into blocks of 1024 samples the watermarked audio signal obtained after insertion of the brand.

2) MDCT: The time frequency mapping is performed by applying the MDCT transform equation (1).

3) Detection, Elimination of duplication, Hamming decoding: Then from the secret key, which is an array containing the positions of the components that are below the masking curve sought by the psychoacoustic model 1 in the insertion process, we can detect the bits of the brand inserted in the samples corresponding to these positions. As a result, we will then obtain a binary vector containing the bits corresponding to the coded signature. The size of the detected mark is multiple of 12. Subsequently and after the elimination of duplication, we decode the signature obtained by Hamming decoding, after which the detection and correction of one error, if it exists, it will extract the 8 data bits corrected of each 12 data bits, to find only a corrected code signature multiple of 8.

4) Formatting the mark: Finally, we proceed to format the brand to form the corresponding mark already inserted.

Figure 2 conceptualize the different steps necessary for the detection procedure of the mark.

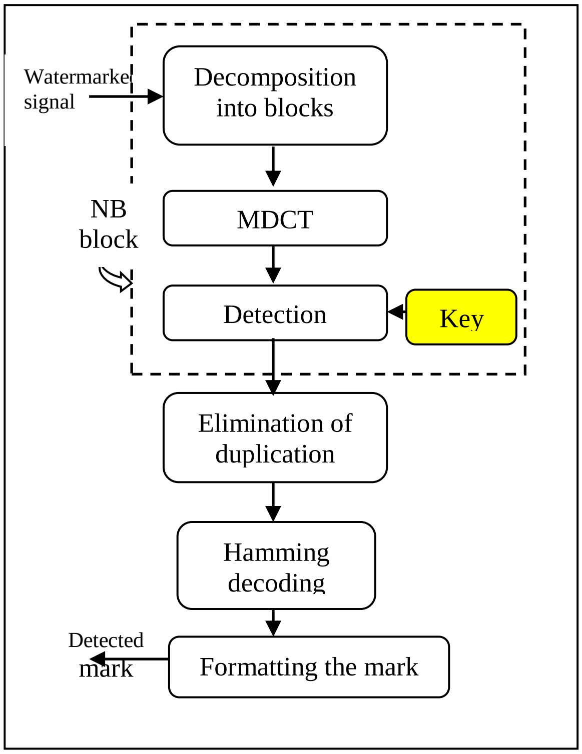

Fig. 2. Detection schema of the mark: version 1

C. Insetion of the mark: second variant of the technique

In this second variant of the proposed watermarking technique, we will integrate the psychoacoustic model 2 of MPEG 1 layer III to exploit its properties in search of the insertion positions of the mark. This version operates in the frequency domain. The time-frequency mapping is performed using the MDCT transform applied to sub-blocks of the original audio signal (.wav). The steps of (1) to (7) are the same as of the first version of the technique. After the decomposition into blocks of the original audio signal and steps (2) and (3) we will look for the insertion positions of the watermark bits.

To search for the places of insertion them less audible to the human ear, we will apply the psychoacoustic model 2 (PMH2) of the MPEG standard on the temporal samples of each sub-block of 1024 samples. This model will calculate for each block a final threshold of energy hearing thr_w . Watermarking bits will be injected into the components located under the thr_w. The number and the positions of these components differ from one block to another and from a signal to the other (dynamic). This makes the number of bits to be inserted is not the same for all the blocks and will offer a good compromise between robustness and inaudibility.

After steps (4) and (5), each bit of the coded mark is duplicated N times where N is calculated based on number of components that are below the final threshold of energy hearing and the size of the brand.

$$N = \frac{sum(components_{below_{final-threhold}})}{size_{brand}} \quad (10)$$

Next, we will make a substitutive insertion of each bit of the mark in the least significant bit (LSB) of the components searched by the PMH2. All the previous steps will be repeated NB block times (number of blocks in the audio signal) and the insertion is done on all the blocks of the audio signal. Thereafter, we apply the IMDCT on the frequency watermarked blocks of 1024 samples to obtain watermarked blocks in the time domain. The last step is to reconstruct the watermarked audio signal.

This approach offers a good compromise between robustness and inaudibility.

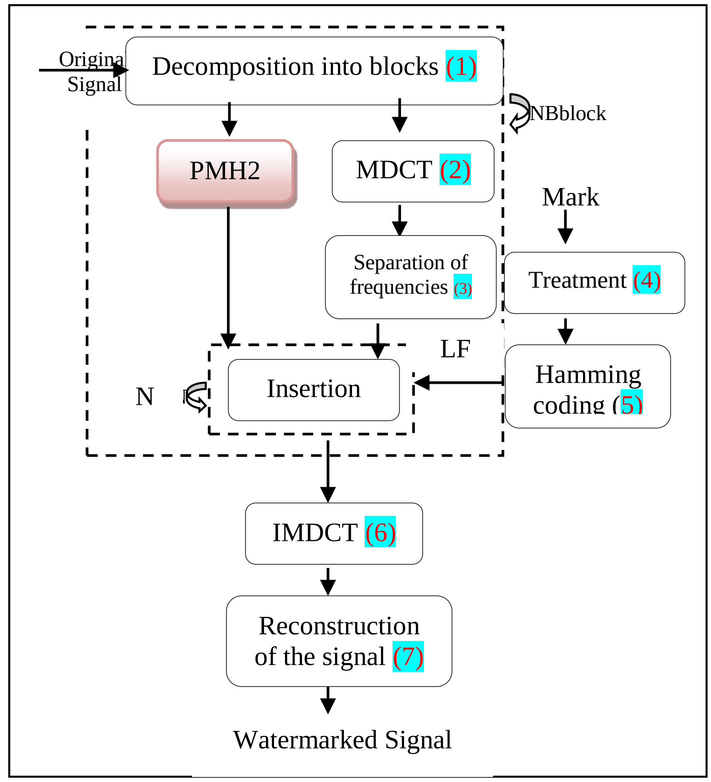

Fig. 3. Insertion schema of the mark: version 2

D. Detection of the mark: second variant of the technique

All stages forming the detection scheme for this variant are the same as for the detection in the first version (Fig. 2), except that the key will change. In this version, it is the set of positions of the components which are below the final hearing of the energy generated by the model 2 for each block. At the end of this process, we obtain the brand inserted into the original signal.

IV. Comparative Study 1: Comparison with the Techniques of Rosa and Brigola

In this section, we will present the results of the comparative study between the proposed technique with her two variants and two other techniques from the literature. First, we present the technique proposed by Luigi Rosa, then the one suggested by Rolf Brigola.

A. Technique proposed by Luigi Rosa

This technique is based on the algorithm of A. Piva whose general principle is:

- The watermarking method operates in the frequency domain using the DCT (Discrete Cosine Transform).
- L.Rosa will insert a random sequence of real numbers in the selected coefficients of the DCT.
- To ensure more inaudibility of the mark, the places of insertion are selected using a psychoacoustic model.
- The detection scheme is semi-private since it not used the original document and only gives an answer on the presence of the brand:
    - 0: could not detect the mark.
    - 1: could detect the mark.

1) Insertion scheme of the mark:

In the beginning, the author divided the original signal into blocks of N samples. Time-frequency mapping is made by the DCT transform. Then, he selects the k among the N coefficients obtained by the DCT, and he applies a psychoacoustic model on these samples to look for the places of insertion. Thereafter, he proceeds to an additive insertion of the brand.

2) Detection scheme of the mark:

The detection scheme of the mark in this technique is semi-private. The first step of detection process is to decompose the watermarked audio signal into blocks of size N. Thereafter, the author applies the DCT to move to the frequency domain. Then, he selected the watermarked coefficients to generate the vector T'. The comparison between T' and the mark itself gives an answer about the presence of the mark (it was possible to detect the brand or not).

B. Technique proposed by Rolf Brigola

1) Insertion scheme of the mark:

First, R. Brigola makes silence treatment. He jumped 31 samples from the beginning of the original signal. Later, he decomposed the signal into blocks of 8820 samples. And as the scheme of watermarking operates in the frequency domain, he applied the FFT transform to each block to do the mapping. The insertion of bits of the brand is done in the peaks of the FFT corresponding to 5 Hz, which belong to the low frequency band (under the scope of the audible spectrum). For this, he looks for each block all the peaks corresponding to 5 Hz. Thereafter, he applies the IFFT transform on the frequency watermarked blocks of 8820 samples to obtain watermarked blocks in the time domain.

The last step is to reconstruct the watermarked audio signal. This technique is performed on stereo audio. Therefore, all previous steps will be made on each channel.

2) Detection scheme of the mark

The detection scheme of the brand is the reverse of that of insertion. Except that, we need the watermarked audio signal and the inserted mark to verify the size of the detected mark. As in the insertion process, we start with a treatment of the silence for the watermarked audio. After decomposition into blocks of 8820 samples, the author applies the FFT transform to move to the frequency domain.

Thereafter, he looks for all positions where it has been inserted the mark to detect the bits of the brand previously injected. For the detection, it needs the brand already inserted to check the size of the detected mark. The final step is to format the detected bits of the mark.

C. Experimental results

In the following, we present various experimental results obtained in this study.

For this reason, we download the M-files of the previous two techniques (that of Brigola and Rosa). We tested the robustness and the inaudibility of the three techniques on an experimental corpus containing 12 recordings. The latter signals are sampled at CD quality (at a sampling frequency Fe = 44.1 kHz), duration 20s on average and different style: symphony orchestras, spoken voices (male and female), jazz, rock, singing voice… We inserted the text mark "audiowatermarking" of length 136 bits and after the hamming coding its length reaches 204 bits (after that each bit will be duplicated N times).

1) Inaudibility:

a) Objective evaluation of sound quality by the PEAQ algorithm:

For the objective evaluation we used the PEAQ algorithm [18] gives a note of ODG (Objective Difference Grade).

This algorithm compares the original signal and the watermarked signal and assigns a score between 0 and -4. If the note ODG = 0 means no degradation, if we get an ODG score that varies between -0.1 and -1 this means that the degradation is perceptible but not annoying, for an ODG score that varies between -1.1 and -2 this means that the degradation is slightly annoying, if the ODG score obtained varies between -2.1 and -3 this means that the degradation is annoying, finally, if the ODG score obtained is in the interval [-3, 1; -4] the distortion is very annoying.

The following table shows the meanings of each note.

TABLE I. SIGNIFICATION OF ODG NOTES

Signification	ODG
Imperceptible	0,0
Perceptible but not annoying	-0,1 to -1
Slightly annoying	-1,1 to -2
Annoying	-2,1 to -3
Very annoying	-3,1 to -4

The results will be displayed in the figure 4. The vertical axis represents the value of the note ODG given by PEAQ. According to the results in figure 4, we note that the proposed technique and more precisely the second variant of the proposed technique gives the most interesting results point of view inaudibility. The values given by PEAQ ranged from 0 (imperceptible) for the two variants of the proposed technique to extracts jazz, Svega, gainsbourg, feelgood ,LoopyMusic and to -0.81 and -0,75 respectively for variant 1 and variant 2 (perceptible but not annoying) to Beethoven. For the two other techniques, we note that for the two extracts "jazz, LoopyMusic" we find a note ODG equal to - 1.98, -1.90 for that of L.Rosa and -0.64, - 0.93 for that of R. Brigola.

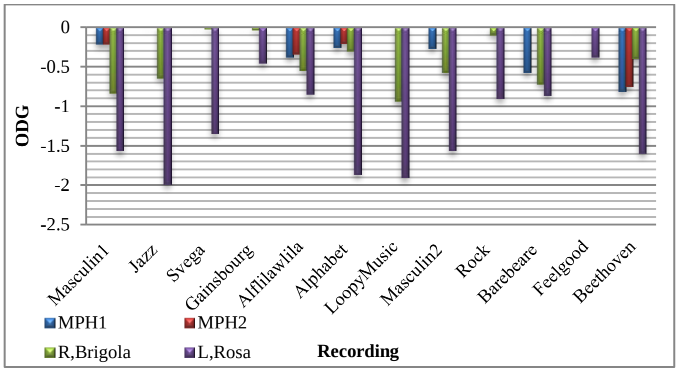

Fig. 4. Graphical representation of the absolute values of the ODG notes

b) Evaluation of sound quality by calculating the SNR

Another way to evaluate the sound quality is to compute the SNR (signal-to-noise ration). It is a measure that calculates the similarity between the original audio and the watermarked audio. The results for the three techniques are shown in the following figure. We also note that the SNR shows more the inaudibility guaranteed by the performed technique. Based on the results shown in Figure 5, we notice that the new proposed technique with its two variants provides a very important criterion of inaudibility. We also points out that the variant using the psychoacoustic model 2 gives the best results.

For example, we remark that the range of values produced by the proposed technique, variant 2, varies between 47, 89 dB et 62, 96 dB. For variant 1, the range of values varie between 46, 26 dB et 58, 55 dB. While for the technique of L.Rosa this range varies between 25,32 db and 27 db. In addition, the minimum value of the SNR obtained by the technique of R.Brigiola is 20dB while the maximum value is 35,8 dB.

Consider the example of the recording (feelgood.wav), we have reached a value of 62.96 dB with the developed technique, variant 2, 57, 34 dB with variant 1 and a value of 35, 87 dB for the one by R. Brigola and 26, 07 dB for the technique of L. Rosa, which proves again the inaudible character of the variant 2 of the developed technique.

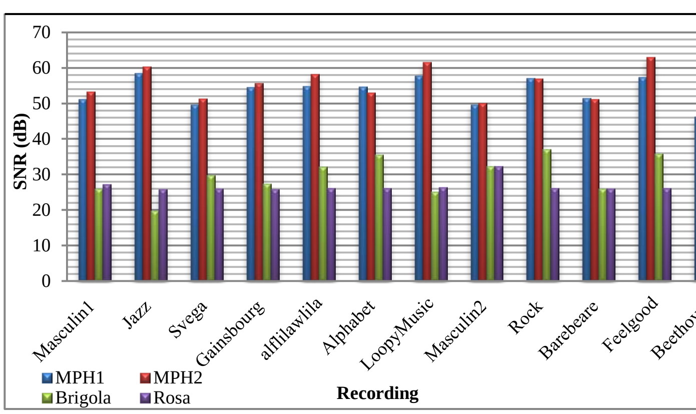

Fig. 5. Graphical representation of the values of the SNR

2) Robustness Tests:

In this section we give the results of robustness tests of the three techniques that we have made against some forms of attack. Disturbances considered in evaluating watermarking systems are realized using the tool "Stirmark Audio" and the encoder MPEG1 "lame.exe" to do the compression / decompression MP3 with three different rates 128kbps, 96kbps and 64kbps.

a) Against attacks stirmark audio

Stirmark Audio [19, 20] is a standard used by researchers in watermarking. This is an executable tool proposed by Lang [21] which tends to become a benchmark for evaluation procedure of the robustness of audio watermarking. This tool generates all versions of a degraded audio signal by various disturbances, both legal and illegal. Stirmark Audio" offers different types of attacks [22]: addition of an echo "echo", the application of a high pass "filter "rc_highpass", the application of a low pass filter "rc_lowpass" smoothing "smooth" , addbrumm, ...

The results of robustness against these attacks are shown in the figure below. On the abscissa are mentioned types of attacks "stirmark audio" used, and the on ordinate is mentioned, with each attack, the number of records that who resisted this type of attacks.

Looking at the figure, we see that the developed technique with the two variants gives the same results. It's robust against a large number of attacks and for all records. It's more robust than the two other techniques against the most attacks. For example, the number of audio robust against attacks compressor, fft_invert, rc_highpass, fft_hlpass, flippsample, fft_real_reverse, addbrumm, addsinus, … is12 recordings. For the technique of L. Rosa we find that it give poor results in terms robustness. This technique is not robust against attacks addsinus and fft_real_reverse and that for all records. We also note that the technique of R. Brigola is robust against attack "amplify" that means the increase or diminution of the amplitude of the audio signal, while the two others did not. Finally, we notice that the three techniques are not robust against attacks echo, copysample, resampling and smooth as they significantly degrade the watermarked audio signal.

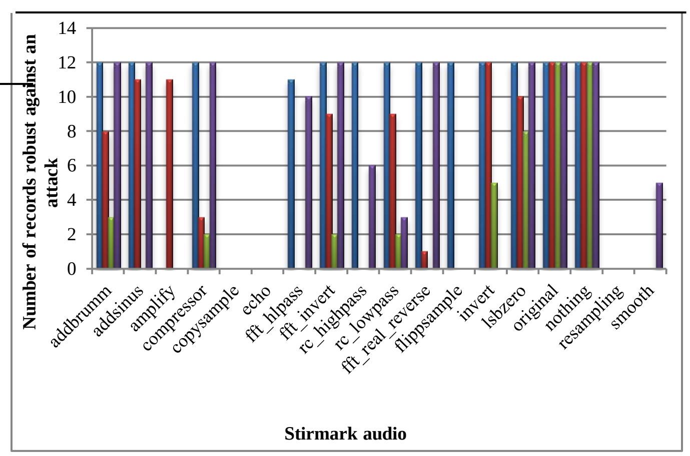

Fig. 6. Detection results after applying stirmark audio attacks

b) Against compression/decompression MP3 attack

We will perform the compression / decompression MP3 with "lame.exe" at different compression rates: 128 kbit / s, 96 kbit / s and 64 kbit / s. Test results are displayed in the following figure:

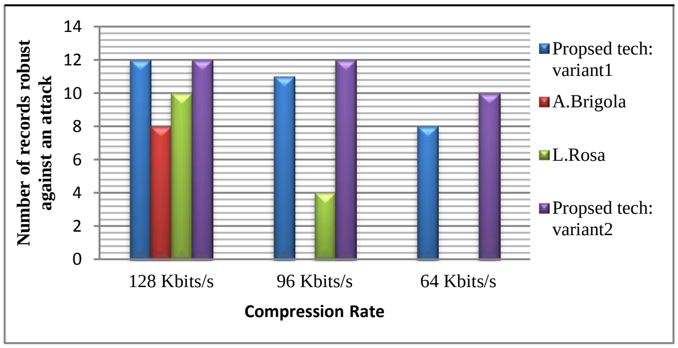

Fig. 7. Detection results after applying stirmark audio attacks

From the results shown in Figure 7 we note that:

- For all records, the variant 1 of the proposed technique is always robust against the attack of compression / decompression at a rate of 128 Kbits / s. This robustness decreases to a rate of 96 Kbits / s to become increasingly weak to 64 Kbits / s but still very interesting (8 records robust among 12).
- For all records, the variant 2 of the proposed technique is always robust against the attack at a rate of 128 Kbit / s and 96 Kbits / s. At a rate of 64 Kbits/ s, the robustness decreases slightly but still very important (10 records robust among 12). These results show the interest of this second variant in terms robustness.
- For the other two techniques, we find that this robustness is low especially for a rate of 96 and 64 Kbit / s. For example the technique of A. Brigola does not resisted against this attack for any records with the two rates of compression / decompression: 96Kbit / s and 64 Kbit / s. While that of L. Rosa has 4 records that have resisted against this type of attack with the rate 96 kbit / s and 0 record with the rate 64Kbit / s.

V. Comparative Study 2: Comparison with Others Existing Techniques

To highlight the results, we will compare in this section the proposed technique with others techniques existing in the literature. Before presenting the comparison, we will define the payload.

The data payload represents the number of bits that are embedded into the audio signal within a unit of time and is measured in the unit of bps (bits per second).

Suppose that the length of the audio signal is S seconds, and the number of bits to be inserted in the signal is K bits. So the data payload D is defined as following:

$$D = \frac{K}{S}\ (bps) \qquad (11)$$

We choose to present the results above a payload data between 72.85 bps and 185.45 bps (depending on length in seconds of tests extracts) and we can even insert more.

This performance is superior to others audio watermarking existing algorithms.

In the table below we show the contribution of the proposed technique in terms of inaudibility of the brand. We will give the average ODG values obtained for the watermarking systems.

TABLE II. Comparison of ODG Average Values

References	Payload (bps)	ODG average values
Proposed	72,85-185,45	-0,2
Khaldi & al.[23]	46,9-50,3	-0,5
Baras & Moreau [24]	83	-0,4

TABLE III. Comparison of Methods Against MP3 Attacks Sorted by Attempted Payload

Reference	Payload (bps)	Robustness to MP3 (Kbps)
Proposed	72,85-185,45	64
Khaldi & al.[23]	46,9-50,3	32
Lie & Chang [25]	43	80
Yeo & Kim [26]	10	96
Thahibana & al.[27]	8,5	96
Xiang & al. [28]	2	64

We can conclude that after the results presented in the two tables II and III that the proposed technique presents interesting performance, point of view imperceptibility of the brand (ODG = -0.2), important payload data (185.45 bps) and robustness against MP3 attack at a rate equal to 64 Kbps. Despite the fact that the robustness against MP3 attack at a compression ratio equal to 64 Kbps remains interesting, we note that the technique proposed in [23] guarantee robustness at a rate of 32 Kbps, which is still interesting and a challenge to reach.

VI. Contribution

We will mention in the following main benefits guaranteed by the two versions of the proposed technique:

- Both versions of the technique provide a very high capacity of insertion. The insertion is done in each block on the whole of the watermarked signal. In addition, each bit of watermarking message is inserted N times and at the end of the insertion process we will have a very important total inserted bits equal to N * number of bit of the message after Hamming coding.
- The number of duplication N is variable (varies from one signal to another). This dynamically calculated according to the message size and the number of insertion positions found using one of the two models used. This provides additional security for the inserted message other than the Hamming coding.
- Despite this important capability of insertion, we have not altered without applying attacks, the quality of the

audio signal and we got a watermarked file faithful to the original file.

- The detection is perfectly blind and does not require a lot of additional data (only a secret key and the number of duplication N) for extracting the injected mark. This latter provides a gain in terms of execution time of the detection process (fast algorithm).

- The use of MDCT allows a finer frequency resolution and treats the block effects that cause noticeable distortion.

VII. CONCLUSION

In this paper, we proposed a robust watermarking audio technique operating in the frequency domain. The time-frequency mapping is done by an MDCT transform applied on blocks of 1024 samples each. This transformation provides a separation between the high frequencies and low frequencies.

The inaudibility of the brand is favored by the insertion of bits of the brand in the LSB of the components looked for by the psychoacoustic model 1 of the MPEG standard in the first variant and the psychoacoustic model 2 of the MPEG standard in the second variant .

Duplication of bits of the mark across the signal for the two variants increases the robustness of the technique against a large number of attacks and manipulations, and also allows for a high embedding capacity.

This important capacity does not alter the sound quality of audio signals. Moreover, the original brand is well identified during the detection phase. This detection is improved by using Hamming coding.

In addition, we presented two detailed studies covering the performed technique and others exist in the literature. The results show the performance of the proposed technique point of view robustness, imperceptibility and capacity of insertion (scheme of the developed technique has higher payload).

As perspective, we look to enlarge the comparative study and to use other measures to test the inaudibility of the mark in the insertion phase and in order to best enhance the robustness criterion and to better show the performance of the technique carried out against the attacks we plan to calculate the bit error rate in the detection and after application of a different types of attacks.

REFERENCES

[1]. W. Bender, D. Gruhl, N. Morimoto, and A. Lu.:'Techniques for data hiding'. IBM Systems Journal, vol. 35, pp. 313–336, 1996.

[2]. M.W. Marcellin1, J Michael. Gormish, A. Bilgin, M. P. Boliek, "An Overview of JPEG-2000", Proc. of IEEE Data Compression Conference, pages 523-541, 2000.

[3]. L. Boney, A.H Tewfik, K.N. Hamdy, "Digital Watermarks for Audio Signals" , IEEE Int. Conf. on Multimedia Computing and Systems June 17-23, Hiroshima, Japan, pages 473-480, 1996.

[4]. M. Fallahpour and D Magias, "High capacity robust audio watermarking scheme based on FFT and linear regression", International Journal on Innovative Computing Information and Control:Vol. 8, Issue 4, 2012.

[5]. Shuo-Tsung Chen, Huang-Nan Huang1 and Chur-Jen Chen, "Adaptive Audio Watermarking via the Optimization Point of View on the Wavelet-Based Entropy", Digital Signal Processing, Volume 23, Issue 3, Pages 971–980, May 2013.

[6]. C. BARAS. : "Tatouage informé de signaux audio numériques". PhD Thesis, Ecole Nationale Supérieure des Télécommunications, 2005.

[7]. J. Pinel, L. Girin and C. Baras.: ' Une technique de tatouage haute capacité pour signaux musicaux auformat CD-audio'. 10ème congrès français d'acoustique. Lyon , Avril 2010, pages 12-16.

[8]. M. CHARFEDDINE, M. EL'ARBI and C. BEN AMAR. :' A blind audio watermarking scheme based on Neural Network and Psychoacoustic Model with Error correcting code in Wavelet Domain '. ISCCSP. Malta, 2008, pages 12-14.

[9]. D. labarre. :' Exploitation de la cyclostationnarité pour le tatouage des signaux audio (watermarking)'. Diplôme d'Etudes Approfondies (DEA), Ecole doctorale des sciences physiques et de l'ingénieur de l'université Bordeaux I, 2001.

[10]. N. Cvejic and T.Seppanen.:'Robust Audio Watermarking in Wavelet Domain Using Frequency Hopping and Patchwork Method'. Proceedings of the 3rd International Symposition on Image and Signal Processing and analysis, 2003.

[11]. J.Pinel, L.Girtin, C.Baras and M. Parvais. :' A hight-capacity watermarking technique for audio signals based on MDCT –domain quantization'. Proceedings of 20 th International Congress on Acoustic, ICA . Sydney, Australia, 2010, pages 23-27.

[12]. J.S. Jacaba . :'Audio Compression Using Modified Discrete Cosine Transform: The Mp3 Coding Standard'. A research paper submitted to the Department of Mathematics College of Science The University of the Philippines Diliman, Quezon City, 2001.

[13]. Norme internationale, ISO/CEI 11172-3, Technologies de l'information codage de l'image animée et du son associé pour les supports de stockage numérique jusqu'à environ 1,5 Mbit/s, partie 3 : Audio.

[14]. J. Prévost. :'Standards de compression audio-vidéo '. (CEA, GIP RENATER, Aristote), Mars 2007.

[15]. R.W.:'Hamming, error detecting and error correcting codes'. Bell Syst. Tech. , 1950, J. 26 (2), pages 147–160.

[16]. M. Bellaaj, K. Ouni. Comparative analysis of audio watermarking technique in MDCT domain with other references in spectral domain. In: 9th international multi-conference on systems, Signals and Devices, Chemnitz, 2012.

[17]. R. Brigola. A simple example on audio watermarking in the frquency domain. April 2010.

[18]. Union Internationale des Télécommunications (UIT): Recommandation B.S. 1387: Méthode de mesure objective de la qualité du son perçu, 2001.

[19]. F. Petitcolas, R. Anderson and M. Kuhn. :'Attacks on copyright marking systems'. Proceedings of the Second International Workshop on Information Hiding, 1998, volume 1525, pages 219-239.

[20]. Stirmark.http://www.petitcolas.net/fabien/watermarking/stirmark, accessed, june 2010.

[21]. http://www.iti.cs.uni-magdeburg.de/~alang/ , accessed june 2010.

[22]. A. Lang, J. Dittmann, R. Spring and C.Vielhauer. :' Audio Watermark Attacks: From Single to Profile Attacks'. 2005.

[23]. K. Khaldi and A.Boudraa, 'Audio Watermarking Via EMD'. Ieee Transactions on Audio, Speech, And Language Processing, Vol. 21, No. 3, pages 675-680, March 2013.

[24]. C. Baras and N. Moreau. 'Modulation CDMA informée dans un système de tatouage audio'.COmpression REpresentation des Signaux Audiovisuels (CORESA 2005), Nov 2005, Rennes, France. 2005.

[25]. W.-N. Lie and L.-C. Chang, "Robust and high-quality time-domain audio watermarking based on low frequency amplitude modification," IEEE Trans. Multimedia, vol. 8, no. 1, pp. 46–59, Feb. 2006.

[26]. I.-K. Yeo and H. J. Kim, "Modified patchwork algorithm: A novel audio watermarking scheme," IEEE Trans. Speech Audio Process., vol. 11, no. 4, pp. 381–386, Jul. 2003.

[27]. R. Tachibana, S. Shimizu, S. Kobayashi, and T. Nakamura, "An audio watermarking method using a two-dimensional pseudo-random array," Signal Process., vol. 82, no. 10, pp. 1455–1469, 200.

[28]. S. Xiang, H. J. Kim, and J. Huang, “Audio watermarking robust against time-scale modification and MP3 compression,” Signal Process., vol. 88, no. 10, pp. 2372–2387, 2008.

[29]. Fallahpour, M., Megías, D. ,”Audio watermarking based on Fibonacci numbers “. IEEE Transactions on Audio Speech and Language Processing. Num. 8, Pag. 1273-1282. ISSN.1558-7916. DOI:10.1109/TASLP.2015.2430818, 2015.