Chapter 16
Digital Audio Watermarking Technique Exploiting the Properties of the Psychoacoustic Model 2 of the MPEG Standard

Maha Bellaaj and Kais Ouni

Abstract In this paper, we propose a watermarking technique for digital audio data operates in the frequency domain. Time–frequency mapping is often done using the Modified Discrete Cosine Transform (MDCT) [1]. It is based on the design of the psychoacoustic model 2 (MPH2) of the MPEG standard [2] layer 3 but specific to audio watermarking. To ensure more inaudibility, the insertion of the mark bits will be in the least significant bit (LSB). In this technique, we duplicate the bits of the mark in order to have a maximum capacity of insertion and robustness of growing against different types of attacks. In order to increase the detection rates, we used Hamming [3] code as error correction code. We studied the robustness of this technique against compression/decompression MP3 attack and we evaluated the inaudibility by calculating the Signal-to-Noise Ratio (SNR) and the objective difference grade (ODG) notes given by PEAQ. To highlight our results, we compared the proposed technique with three other existing techniques.

Keywords Watermarking · Digital audio · MDCT · Psychoacoustic model 2 (MPH2) · Inaudibility · Robustness · SNR · ODG · PEAQ

16.1 Introduction

Through the development of Internet and the emergence of new communications media digital takes a place increasingly important, which poses serious problems since it is easy to copy and deal with these computer documents. As a result,

M. Bellaaj · K. Ouni (✉)
U. R. Signals and Mechatronic Systems, Higher School of Technology and Computer Science, Carthage University, Tunis, Tunisia
e-mail: kais.ouni@esti.rnu.tn

M. Bellaaj
e-mail: maha_bellaaj@yahoo.fr

the copyrights become increasingly unprotected and we also suffer from illegal redistribution of data. As effective solution to these problems comes the digital watermarking [4], whose basic idea is to insert into the digital document (image, sound, video…) a signature in a way robust and imperceptible. Since the 1990s, the articles continue to multiply in order to find a watermarking technique which satisfies the following characteristics: robustness, large insertion capacity, and imperceptibility of the mark [5].

In this paper we propose a watermarking technique for digital audio based on a spectral approach of insertion of the mark combined with modeling of psychoacoustic phenomena to improve the robustness of the technique.

This paper is organized as follows: In Sect. 16.2, we detail the process of insertion and detection for the proposed technique. Section 16.3 presents the experimental results.

In Sect. 16.4, we compare the results obtained by the proposed technique with three other existing techniques. In the last section, we give a conclusion and perspective to this work.

16.2 Presentation of the Proposed Watermarking Technique

A detailed bibliographic study on digital watermarking [6–8] showed that the frequency domain space is a good point of view robustness and inaudibility hence the idea of using Modified Discrete Cosine Transform (MDCT) to move from time domain to the frequency domain [9, 10]. In addition, MDCT allows a finer frequency resolution.

16.2.1 Insertion Schema of the Mark

In the first step, the original audio signal (.wav) will be divided into block of 1024 samples. Thereafter, we will apply the MDCT to move to the frequency domain. This transformation will break the frame into low frequency (LF) and high frequency (HF). To separate these two frequency bands, we will use a frequency separation module. At the end of this step, we get all the low frequencies where we will insert bits of the brand. The choice of the LF band is due to the fact that the latter is much less sensitive against the attacks than the HF band (especially against MP3 compression). In parallel and to search for the places of insertion them less audible to the human ear, we will apply the psychoacoustic model 2 (MPH2) of the MPEG standard on the temporal samples of each sub-block of 1024 samples. Insertion places are located under the final threshold of energy hearing generated by this model for each block. This approach provides a good compromise between robustness and inaudibility.

Fig. 16.1 Insertion schema
of the mark
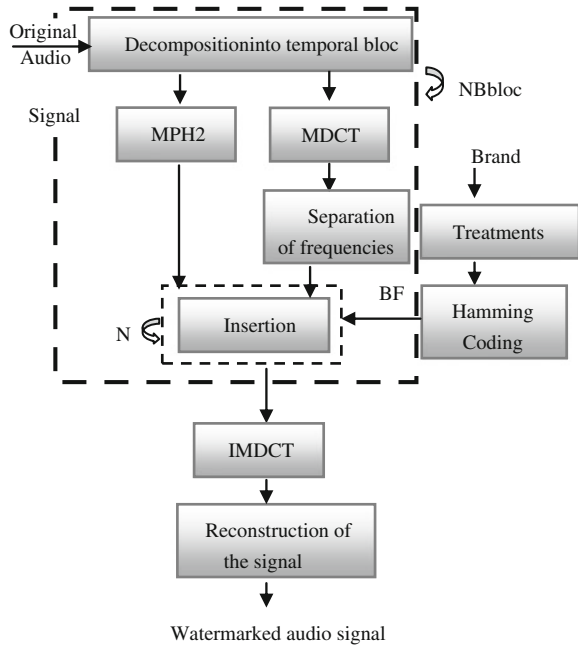
After the application of several treatments (binarization of the brand, decomposition into portions of 8 bits each) and Hamming coding (12, 8) to ensure the correction bits if necessary, since the bits of the signature can undergo changes during the insertion and detection, each bit is duplicated N times where N is calculated based on number of components that are below the final threshold of energy hearing and the size of the brand. 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 MPH2. 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.

Figure 16.1 will give the general scheme and the different steps necessary for the insertion of the brand.

16.2.2 Detection Schema of the Mark

According to the Fig. 16.2, we note that the detection scheme of the brand is the inverse of the insertion. It is a blind detection that does not require the original audio signal or the presence of the mark originally inserted. Only the secret key

Fig. 16.2 Detection schema
of the mark
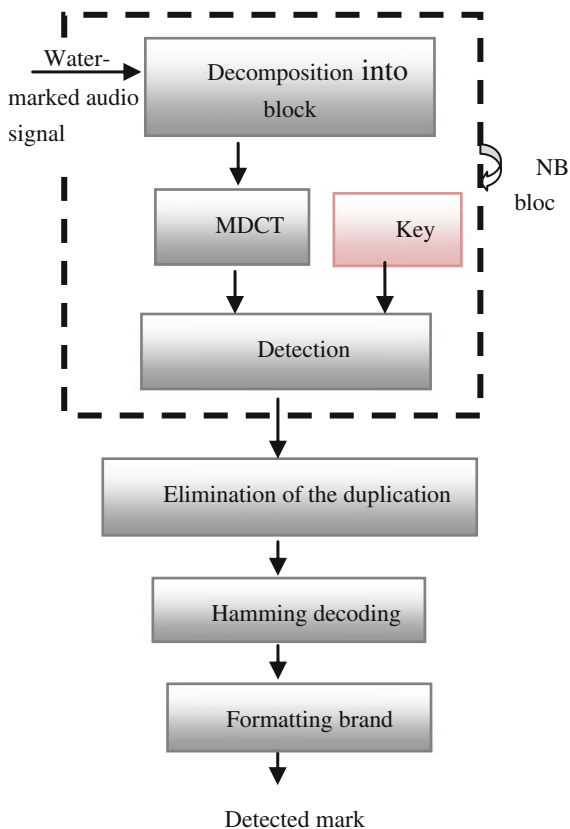
(all the positions of the less sensitive components sought by the MPH2 in the
insertion phase and the number of duplication N) is required. The output of the
detection process is the final mark decoded and formatting.

16.3 Test Result

This section will present the different experimental results obtained by this technique. These results were focused on an experimental corpus composed of 12
audio signals. These signals are sampled at CD quality (at a sampling frequency
Fe = 44.1 kHz), duration 20 s on average and different style: symphony orchestras, spoken voices (male and female), jazz, rock, singing voice...

16.3.1 Inaudibility

16.3.1.1 Spectrogram

For testing the watermarking system presented above, 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). From the tests, we were able to detect correctly and without error the mark which is identical to the original brand.

The Figs. 16.3 and 16.4 shows the spectrograms of the original audio signal and the watermarked audio signal. We will use an extract to the comparison:

Jazz.wav: extract of jazz

- Interpretation:

If we compare the spectrogram of the watermarked signal with the spectrogram of to the originals signals (by comparing Fig. 16.3 with Fig. 16.4), we notice that they are very similar.

Also, while listening to the original signal and the watermarked signal we do not perceive a difference. Despite the large number of bits already inserted, we do not perceive the existence of the signature in the watermarked signal which remains faithful to the original signal.
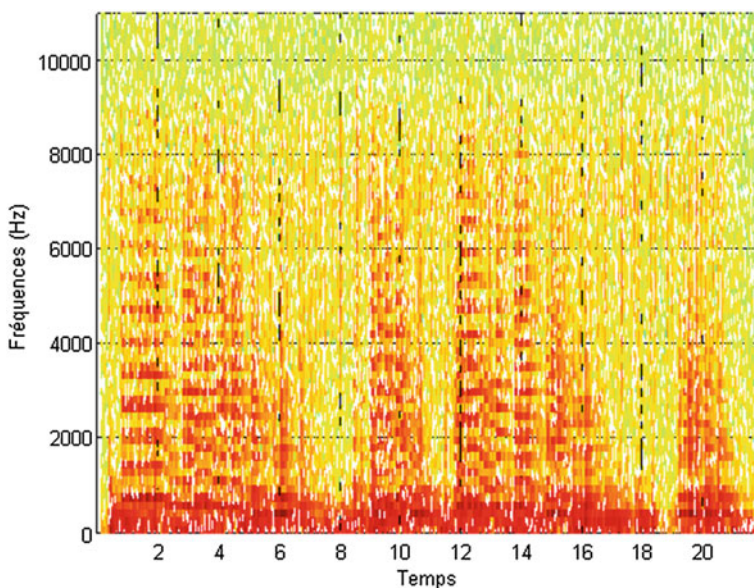

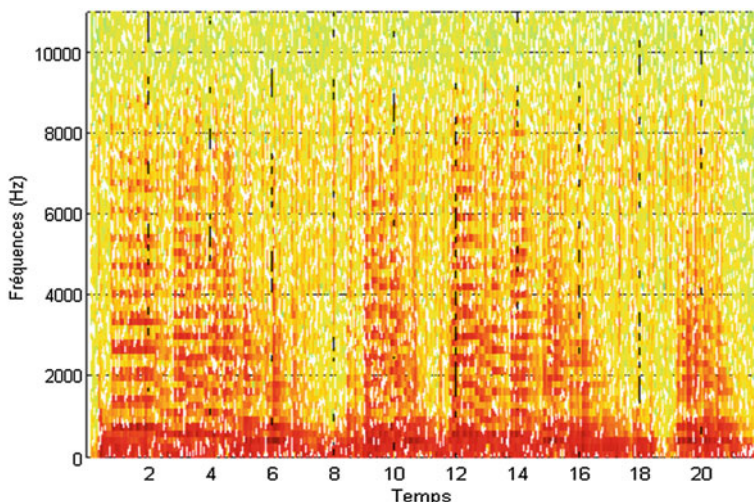**Fig. 16.3** Spectrogram of the original audio (jazz.wav)

Fig. 16.4 Spectrogram of the watermarked audio (jazz_tatoue.wav)

16.3.1.2 Evaluation of the Sound Quality by PEAQ

The PEAQ algorithm [11] allows for an objective evaluation of sound quality. It generates as output a note of objective difference grade (ODG). This algorithm compares the original signal and the watermarked signal and assigns a score between 0 and −4. The Table 16.1 presents the meaning of each note.

We note from Fig. 16.5 that the notes of ODG vary between 0 (Imperceptible) and −0.35 (Perceptible but not annoying). These values are very interesting and show that our watermarking system degrades very little the sound quality of extracts and proves that the proposed technique provides a good criterion for inaudibility of the brand during the insertion process.

16.3.1.3 Evaluation of the Sound Quality by Calculating the SNR

Another way to demonstrate the inaudibility of the mark is to calculate the Signal-to-Noise Ratio (SNR). It is a measure that calculates the similarity between the original audio and the watermarked audio.

The results for this technique are shown in the Fig. 16.6.

From the results displayed in Fig. 16.6 we can see that the values of SNR show more the inaudibility provided by our technique. These values vary between 74.1546 and 82.7722 db, they are very interesting and confirm the results previously obtained by PEAQ.

Table 16.1 Signification notes of ODG

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 of this evaluation are shown in Fig. 16.5

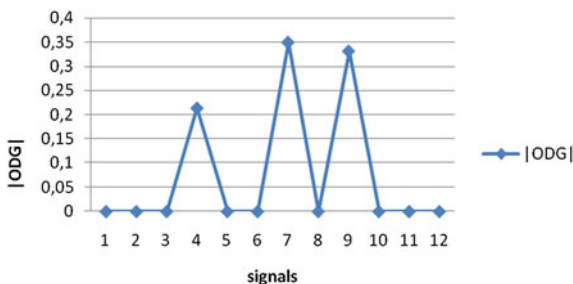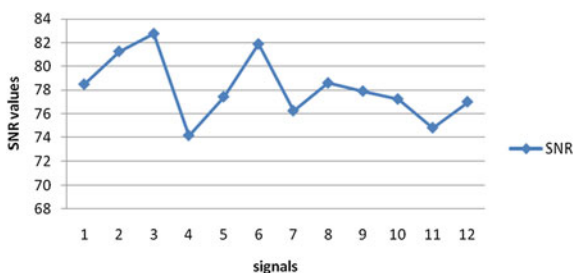**Fig. 16.5** Graphical representation of the absolute values of ODG notes

**Fig. 16.6** Graphical representation of the SNR values

16.3.2 Robustness

16.3.2.1 Robustness Against Compression/Decompression MP3

The compression/decompression MP3 is performed by "lame.exe" at three different rates: 128, 96, and 64 Kbit/s. Test results are displayed in the Fig. 16.7.

From the results displayed in Fig. 16.7, we note that the technique is always robust against the attack of compression/decompression MP3 for the two compression rate 128 and 96 Kbit/s. The strength decreases for a rate of 64 Kbit/s but still very interesting (9 records/12 records are robust against attack).

Fig. 16.7 Robustness of the technique against the attack of compression/ decompression MP3
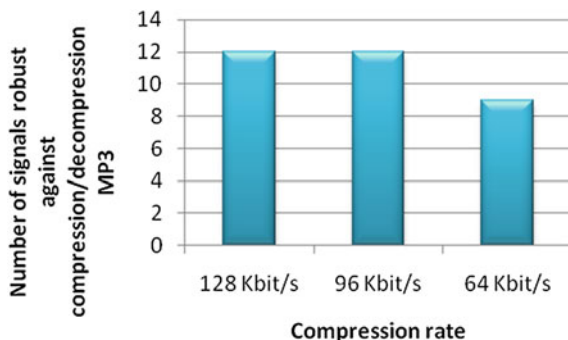
16.4 Comparison to the Existing Watermark Techniques

To highlight our results, we will compare in this section the detailed above technique with three other techniques developed in [12].

The experimental corpus used above is the same as that used in [12].

We will present in the Table 16.2 the range of values of ODG given by PEAQ and the range of values of SNR for each technique.

Table 16.3 will illustrate the number of signals robust against compression/ decompression attack for each technique.

The presented results show that the proposed technique using the MPH2 of the MPEG standard gives better results in terms of the inaudibility and robustness than the technique using the psychoacoustic model 1 of the MPEG standard, the technique proposed by R. Brigola and the technique proposed by L. Rosa.

Table 16.2 Range of values of ODG and SNR for each technique

	ODG	SNR
Proposed watermarking technique + MPH2	0 to −0.35	74.1546 –82.7722 db
Watermarking technique + MPH1	0 to −0.819	23.1483–37.7138 db
Technique of R. Brigola	0 to −0.923	19.7856–35.8742 db
Technique of L. Rosa	−0.375 to −1.98	25.6629–32.2940 db

Table 16.3 Number of signals robust against compression/decompression MP3

	128 Kbit	96 Kbit	64 Kbit
Proposed watermarking technique + MPH2	12/12	12/12	9/12
Watermarking technique + MPH1	12/12	11/12	8/12
Technique of R. Brigola	8/12	0/12	0/12
Technique of L. Rosa	10/12	4/12	0/12

16.5 Conclusion and Perspectives

In this paper, we proposed a blind watermarking technique for audio (.wav) and which operates in the frequency domain. The time–frequency mapping is done by MDCT transformation applied to blocks of 1024 samples each. The inaudibility of the mark is favored by inserting bits in the LSB of components of the LF band which is under the final threshold of energy hearing calculated by the MPH2 of MPEG standard. The duplication of bits of the mark throughout the signal increases the robustness of the technique against attacks and allows having a high capacity of insertion. This important capability of insertion does not affect the sound quality of audio signals. In addition, the original brand is well identified in the detection phase. This detection is improved by using of Hamming coding. As perspective, we aim to test our technique against other types of attacks such as stirmark audio attacks.

References

1. Mu-Huo C, Yu-Hsin H (2003) Fast IMDCT and MDCT algorithms—a matrix approach. IEEE Trans Signal Process 51:221–229
2. 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
3. Hamming RW (1950) Error detecting and error correcting codes. Bell Syst Tech J 26(2):147–160
4. Bender W, Gruhl D, Morimoto N, Lu A (1996) Techniques for data hiding. IBM Syst J 35:313–336
5. Barnett R (1999) Digital watermarking: applications, techniques and challenges. Electron Commun Eng J 11(3):173–183
6. Baras C (2005) Tatouage informé de signaux audio numériques. Doctoral thesis, High National School of Telecommunications
7. Boney L, Tewfik AH, Hamdy KN (1996) Digital watermarks for audio signals. In: IEEE international conference on multimedia computing and systems, Hiroshima, Japan pp 473–480 June 17–23, 1996
8. Pinel J, Girin L, Baras C (2010) Une technique de tatouage haute capacité pour signaux musicaux au format CD-audio. In: 10 ème congrès français d'acoustique, 12–16, Lyon
9. Cvejic N, Seppanen, T (2003) Robust audio watermarking in wavelet domain using frequency hopping and patchwork method. In: Proceedings of the 3rd international symposium on image and signal processing and analysis, Rome, Italy
10. Charfeddine M, El Arbi M, Ben Amar C (2008) A blind audio watermarking scheme based on neural network and psychoacoustic model with error correcting code in wavelet domain. In: ISCCSP, Malta, pp 12–14
11. Union Internationale des Télécommunications (UIT): Recommandation B.S. 1387: Méthode de mesure objective de la qualité du son perçu (2001)
12. Bellaaj M, Ouni K (2012) 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