DCT BASED BLIND AUDIO WATERMARKING SCHEME

Charfeddine Maha, Elarbi Maher, Koubaa Mohamed and Ben Amar Chokri
REGIM, Research Group on Intelligent Machines, University of Sfax
National School of Engineers (ENIS), BP 1173, Sfax, 3038, Tunisia

Keywords: Data Security, Discrete Cosine Transforms, Neural Network, Hamming Codes.

Abstract: During these recent years, different digital audio media are available due to internet and audio processing techniques. This is beneficial to our daily life but it brings the problem of ownership protection. As a solution, audio watermarking technique has been quickly developed. It consists of embedding information into digital audio data. This paper proposes an audio watermarking scheme operating in frequency domain using Discrete Cosine transforms (DCT). We take the advantage of hiding the mark in the frequency domain to guarantee robustness. To increase robustness of our schemes, the watermark is refined by the Hamming error correcting code. We study the band used to hide the watermark bits according essentially to the effects of the MP3 compression on the watermarked audio signal. To assure watermark embedding and extraction, we use neural network architecture. Experimental results show that the watermark is imperceptible and the algorithm is robust to many attacks like MP3 compression with several compression rates and various Stirmark attacks. Furthermore, the watermark can be blind extracted without needing the original audio.

1 INTRODUCTION

Today, music distribution over the Internet is an increasingly important technology. Most of the music content is compressed in order to save disk space and speed up transmission over band limited channels. However, controlling the use of the distributed works or tracking illicit copies still presents major problems.

Within this context, the digital watermarking (Bender, Gruhl, Morimoto, and Lu, 1996), art of hiding information into the audio signal, can provide a useful mechanism to track such illicit copies or to attach property right information to the material.

Watermarking covers any type of digital documents that include text, sound, image and video (Cox, I., Miller, M., and Bloom, 2002).

Inaudibility (perceptual transparency), robustness and capacity of the hiding algorithm are three contradictory requirements for any audio watermarking scheme, called “the magic triangle” (Barnett, 1999).

Research in audio watermarking is not as mature, compared to research in image and video watermarking (Arnold, M., Wolthusen, S. and Schmucker, 2003). This is due to the fact that, the human auditory system is much more sensitive than the human visual system, and that inaudibility is much more difficult to achieve than invisibility for images (Al-Haj and Mohammad, 2010). Even so, many audio watermarking techniques have been proposed in literature in recent years. Most of these algorithms attempt to satisfy watermarking requirements by exploiting the imperfections of the human auditory system (Katzenbeisser, S., and Petitcloas, 2000).

In this paper, we present a blind audio watermarking scheme operating in the frequency domain (Martin and Burg, 2007). The original audio is divided into 512 block samples. Some blocks are chosen randomly and then transformed to the DCT domain to embed watermark. Watermark embedding and extraction are based on memorizing the relationships between a transformed central sample and its 8 neighbours using Neural Network NN architecture (Haykin, 1995). The band used to hide the watermark bits is studied according to the effects of the MP3 compression on the watermarked audio signal and to the inaudibility watermarking propriety. The extraction is the inverse process and is based on blind technique (Hsieh and Tsou, 2002). In fact, we are capable of recovering the watermark data without requiring access to the original audio.

The rest of this paper is organized as follows: section 2 describes the proposed audio watermarking

Maha C., Maher E., Mohamed K. and Amar Chokri B. (2010).
DCT BASED BLIND AUDIO WATERMARKING SCHEME.
In *Proceedings of the International Conference on Signal Processing and Multimedia Applications*, pages 139-144
DOI: 10.5220/0002922101390144

scheme. Section 3 shows the obtained results. Finally, section 4 concludes the paper.

2 THE PROPOSED AUDIO WATERMARKING SCHEME

Similar to static image, the realizable audio watermarking has two approaches: time domain and transform domain. In fact, some schemes embed the watermark directly in the time domain (Laftsidis, Tefas, Nikolaidis and Pitas, 2003), and others operate in the transform domain usually using Discrete Fourier Transform DFT (Malik, Khokhar, Rashid and A, 2003), Discrete Cosine Transform DCT (Wang, Ma, Cong, Yin, 2005) or Discrete Wavelet Transform DWT (Wu, S.Q., Huang, J.W., Huang, D.R., Shi and Y.Q., 2004).

According to the audio watermarking related work survey, we contribute by proposing an algorithm which not only assures blind detection without resorting to the original digital audio signals, but also exploits the DCT domain to assure robustness property thanks to the advantages of this domain. Besides, it uses the NN for both insertion and detection process to ameliorate robustness contrary to many schemes that exploit the NN only for the detection process (Wang et al., 2005). The algorithm adds also an Error Correcting Code to refine the embedded watermark during detection. Details of the proposed audio watermarking scheme are described in the following subparts.

2.1 Watermark Embedding

The main steps of the embedding procedure developed are depicted in figure 1.

First, an original audio signal is divided into non-overlapping blocks of 512 samples. This number is chosen because a signal of 512 samples is constantly stable. Next, a DCT transform is applied to those blocs. At the same time, a watermark is decomposed into sets of length 8. Each set is encoded with a Hamming error correcting code (12, 8) (W and Hamming, 1950). After that, we generate a pseudo-random index sequence Z(x) where x ∈ (1,…..,pq), is the encoded watermarking size. Simultaneously, a NN is trained to be used later in the watermark insertion and extraction process. Given the pq selected blocks, we locate the band of middle frequencies of each block and we proceed then to the watermark insertion using NN. Finally we construct the watermarked audio signal after performing the IDCT transform.

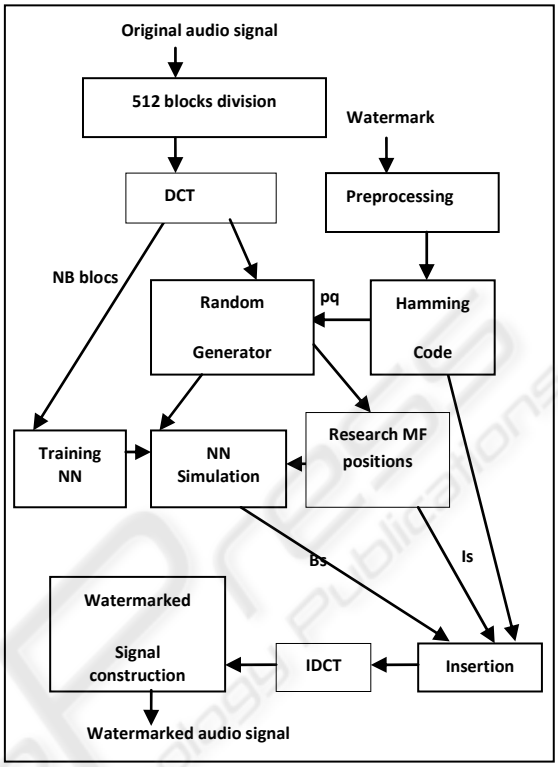

Figure 1: DCT embedding process.

2.1.1 Locating Middle Frequencies and Finding the Position of Insertion

For each block, we locate the band of middle frequencies MF to insert the watermark bit contrary to various algorithms that choose both low and MF (Wang et al., 2005) (Zhiping, Lihua, 2007). In fact, we choose to hide the watermark bits in the MF while the high frequencies will be modified and even deleted by signal manipulations such as MP3 compression, therefore we will lose our bits if we hide them in the high frequencies. Moreover, if we insert the bits in the low frequencies, we will get a distinguishable watermark and so an affected audio signal. We have studied for several blocks the PSD power spectral density before and after MP3 compression.

The figure 2 shows that the MP3 compression with 128Kbps rate modifies the signal sampled at 44.1 KHz starting with a frequency of 11 Khz which corresponds to the index band [100:258] from a block of 512 samples. For each block and after obtaining the band of MF, we locate the closed sample value to the average value of the actual band and we deduct its index position.

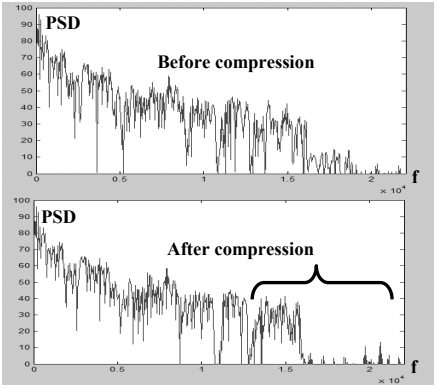

Figure 2: The power spectral density before and after MP3 compression (128Kbps).

2.1.2 Watermark Insertion Strategy

We use the Hamming error correcting code (12,8) to enhance the robustness. The Hamming code can surmount the corruption of a watermark, and can help it endure serious attacks.

We establish the relationship between frequency samples around a frequency central sample by using the back-propagation neural networks (BPNN) model. The BPNN architecture contains three layers: the input layer with eight neurons, the hidden layer with nine neurons, and the output layer with a single neuron. For a selected sample I(x), the network is trained with its 8 neighbours as input vector and the value of the sample as output. Since the training process for the NN is completed, a set of synaptic weights, characterizing the behaviour of the trained neural network, can be obtained and used in the NN simulation process. Each watermark bit Wi is embedded in a selected block by modifying the frequency sample at the position of insertion. The insertion is based on a comparison between Bs and Is values where Bs is the NN output and Is the value of the original central sample at the position of insertion. We use an adaptive watermarking strength st, equal to abs (Bs-Is) to accentuate the difference between Bs and Is. Then we perform a substitute insertion depending on the result of this difference and the watermark bit values (1 or 0).

After performing the IDCT transform, we get the audio watermarked signal.

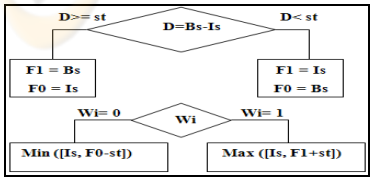

Figure 3: The Insertion strategy.

2.2 Watermark Retrieval

The detection process showed in figure 4 is the inverse embedding one and does not need the original audio signal. We require in this process all NN synaptic weights, random generator results and pq inaudible positions. Those parameters constitute the secret key of our technique. Through the watermark-extraction process, the NN is employed to estimate the central sample.

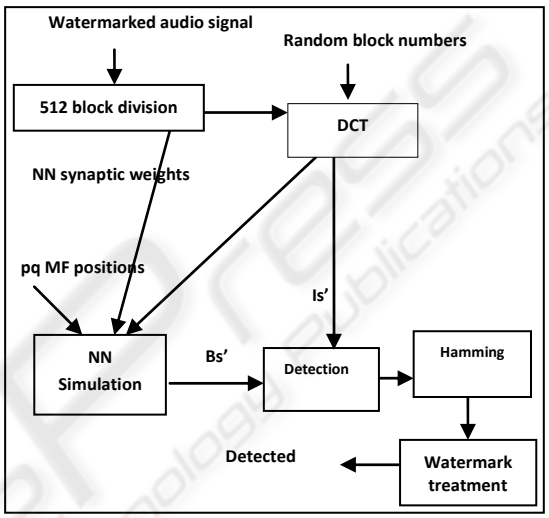

Figure 4: DCT retrieval process.

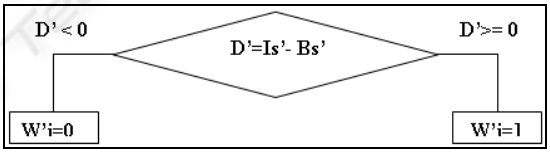

Figure 5: The detection strategy.

The length of the mark detected is compulsory a multiple of 12 because of the Hamming coding.

We extract the watermark bits and we correct them using Hamming decoding. We get ultimately a corrected watermark multiple of 8.

3 EXPERIMENTAL RESULTS

In order to test the inaudibility and the robustness of the proposed algorithm, we perform several experiments. Various audio "wav" files with 44.1 KHz sampling rate and 16 bits per sample are used in our tests and some files are presented in table 1. The chosen watermark can be a text "author_n" or a binary image of size 32×32 showed in figure 6.

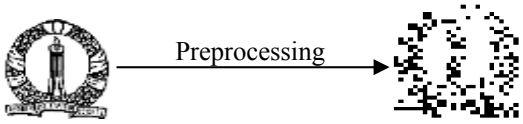

Figure 6: Image watermark.

The normalized cross-correlation (Tung and Ling, 1997) NC is calculated to evaluate similarity between the extracted and the inserted watermarks.

$$NC = \frac{\sum_{i,j=1}^{n} bin(i,j) * bin'(i,j)}{\sqrt{\sum_{i,j=1}^{n} bin'(i,j)^2 * \sum_{i,j=1}^{n} bin(i,j)^2}} \quad (1)$$

The more NC is close to 1, the more binary detected watermark "bin'" is similar to the binary inserted watermark "bin". Our technique guarantees free error detection (NC=1) in the case of an ideal exchange (without signal processing manipulations). In order to regard these eventual operations and to evaluate the robustness performance of the scheme, we performed several tests in which the watermarked audio is subjected to frequently encountered degradations. To estimate the audio quality after watermark embedding, we use signal-to-noise ration (SNR) as an objective measure, and listening tests as a subjective measure.

3.1 Robustness Results

The manipulations used to evaluate our scheme include MP3 compression with three compression rates 128kbps, 96kbps and 64kbp and a variety of stirmark attacks (Lang, Dittmann, Spring and Vielhauer, 2005). Robustness has been assessed using the NC correlation measure between the embedded watermark and the identified watermark consecutively from the watermarked and manipulated signals. Figure 7, figure 8, figure 9 and figure 10 summarize the watermark detection results for these degradations. In this paper, we present the robustness results for the audio files "svega.wav", "Tunisia_hymn.wav" and "coran.wav".

Table 1: Audio signals and theirs descriptions.

Signal	Name	Description
1	Svega.wav	Female song voice
2	Coran.wav	Male spoken voice
3	Jonass.wav	Piano with basic music
4	Romeo.wav	Violin
5	Tunisia_hymn.wav	Rhythmic music
6	Pant_rose.wav	Cartoon song

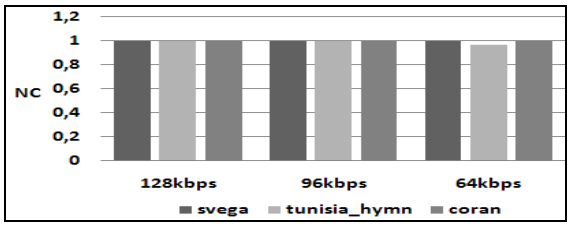

Figure 7: NC values after MP3 compression (watermark is a text).

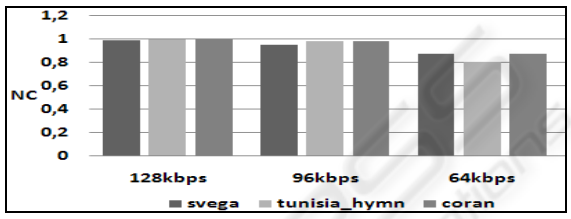

Figure 8: NC values after MP3 compression (watermark is an image).

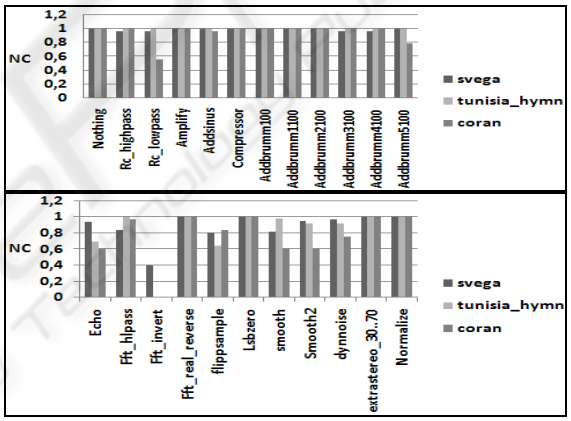

Figure 9: NC values after Stirmark attacks (watermark is a text).

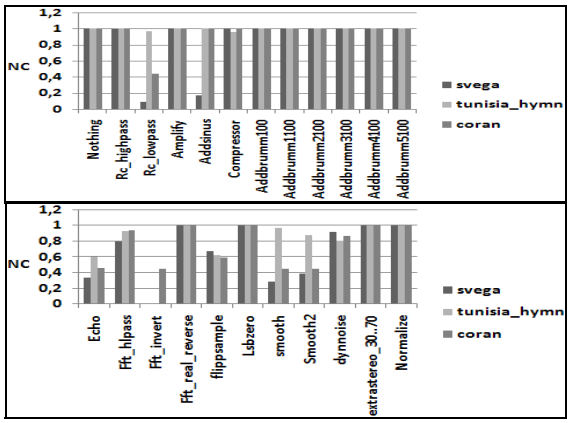

Figure 10: NC values after Stirmark attacks (watermark is an image).

We note that the technique is robust to MP3 compression in the rates 128, 96 and 64 kbps when the mark is a text and in the rates 128, 96 kbps when the mark is an image. For Stirmark attacks, we observe that the technique is robust to several signal manipulations. For example, for the rc_highpass, amplify, compressor, addbrummn with different values, Fft_real_reverse, lsbzero, extrastero and normalize attacks, we have good robustness results whatever the mark is a text or an image. On the contrary, we can conclude that the method cannot resist the echo and fft_invert attacks. It is necessary to note that adding for example echo, affects greatly the watermarked signal which isn't interesting for the watermarker to have a watermarked signal very different to the original one.

In general, if we compare the robustness results against the attacks proposed in this paper versus the results against the same attacks in the paper (Al-Haj and Mohammad, 2010) based on DWT and SVM, we can remark that the performance of the proposed algorithm is better than the (Al-Haj and Mohammad, 2010), the (Cox, I., Kilian, J., Leighton, T. and Shamoon. 1997), and the (Ozer, H., Sankur, B. and Memon. 2005) algorithms, especially in the MP3 compression with different rates and addbrumn with different values.

3.2 Inaudibility Results

To measure imperceptibility, we use the signal-to-noise ration (SNR) as an objective measure, and listening tests that involved 10 persons after watermark embedding as a subjective measure. The listeners are presented with the original and the watermarked audio and are asked to report any differences between the two signals. No listener has noticed remarkable difference between the two signals. The SNR formula is:

$$SNR = 10 \log_{10} \left(\frac{\sum_i Y_i^2}{\sum (Y_i - y_i)^2} \right) \quad (2)$$

Where Y and y are original audio signals and watermarked audio signals respectively.

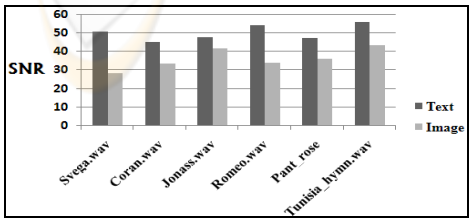

Figure 11: SNR values for DCT technique.

According to the International Federation of the Phonographic Industry (IFPI), the SNR of the watermarked audio signal should be greater than 20 dB (Sriyingyong and Attakitmongcol, 2006). It is remarkable that our technique possesses high SNR values and then presents good inaudibility results. In addition, if we observe the inaudibility results of the (Al-Haj and Mohammad, 2010) scheme (the mean of the SNR values presented in this paper is 25.24 dB) and of some traditional techniques (Sehirli, M., Gurgen, F and Ikizoglu, 2004) (each SNR value is computed as the mean of the SNR values of different audio signals each marked by different watermarks), showed in the table 2, we can deduct that our proposed DCT technique presents the highest SNR values (the mean of the SNR values presented in figure 11 is 43.10 dB) except for the LSB method with a 67.91 SNR but poor robustness results.

Table 2: SNR values of traditional techniques.

Reference	Algorithm	SNR
(Uludag, U., Arslan and L, 2001)	DC-level Shifting	21.24
(Bender et al.,1996)	Echo	21.47
(Bender et al.,1996)	Phase	12.20
(Bender et al.,1996)	LSB	67.91
(Cox et al. ,1997)	Spread Spectrum	28.59
(Swanson, M., Zhu, B., Tewfic, A. and Boney, 1998)	Frequency Masking	12.87
(Al-Haj and Mohammad, 2010)	DWT SVM	25.24
Our DCT technique	DCT NN	43.10

4 CONCLUSIONS

In this paper, we proposed an imperceptible and robust audio blind watermarking technique based on DCT transform. Operating in this domain makes the watermark more resistant to wide range of attacks. This robustness is improved by using Hamming code because it overcomes the corruption of the watermark. It's instructive to point out that the proposed algorithm satisfies the desired features of optimal audio watermarking, which have been set by the IFPI. IFPI states that the watermark should not degrade perception of audio, the algorithm should offer more than 20 dB SNR, the watermark should be able to resist the most common audio processing operations and attacks, and the watermark should prevent unauthorized removal unless the quality of audio becomes very poor. Referring to the figures above, it's easy to conclude that the performance of

the proposed algorithm fulfils the desired IFPI required performance.

The proposed scheme can be improved by studying others error correcting codes like the convolution-codes, to enhance the robustness of the watermark.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the financial support of this work by grants from General Direction of Scientific Research (DGRST), Tunisia, under the ARUB program.

REFERENCES

Al-Haj and Mohammad, (2010). Digital Audio Water-marking Based on the Discrete Wavelets Transform and Singular Value Decomposition. *European Journal of Scientific Research* Vol.39 pp.6-21

Arnold, M., Wolthusen, S. and Schmucker, (2003). Techniques and applications of digital watermarking and content protection. *Artech House. ZwickerE, Fastl H. Psychoacoustics:Facts and models. Springer-Verlag.*

Barnett, (1999). Digital watermarking: applications, techniques and challenges, *Electronics and Communication Engineering Journal*, pp. 173–183.

Bender, Gruhl, Morimoto and Lu, (1996). Techniques for data hiding, *IBM Systems Journal*, vol35, pp.313–336.

Cox, I., Kilian, J., Leighton, T. and Shamoon, (1997). Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing;* 6(12): 1673-1687.

Cox, I., Miller, M. and Bloom, (2002). Digital Watermarking, *USA: Academic Press.*

Haykin, (1995). Neural Networks: a comprehensive foundation, *Macmillan College Publishing Company, New York, NY, USA*.

Hsieh and Tsou, (2002). Blind cepstrum domain audio watermarking based on time energy features. *13th International Conference on Digital signal Processing Proceeding*, Greece: Santorini, pp. 705-708.

Katzenbeisser, S.and Petitcloas, (2000). Information Hiding Techniques for Steganography and Digital Watermarking. *USA: Artech House.*

Laftsidis, Tefas, Nikolaidis and Pitas, (2003). Robust multibit audio watermarking in the temporal domain. *ISCAS (2)*, pp. 933-937.

Lang, Dittmann, Spring and Vielhauer, (2005). Audio Watermark Attacks: From Single to Profile Attacks.

Malik, Khokhar, Rashid and A, (2003). Robust audio watermarking using frequency selective spread spectrum theory, *(ICASSP '03)*, pp.385-388.

Martin and Burg, (2007). Digital Audio Representation, Comparison of DCT and DFT. *Supplement to Chapter 4 of The Science of Digital Media* .

Ozer, H., Sankur, B. and Memon, (2005). An SVD-based audio watermarking technique. *Proceedings of the Multimedia and Security Workshop, NY*, 51-56.

Swanson, M., Zhu, B., Tewfic, A. and Boney, (1998). Robust audio watermarking using perceptual masking. *Signal Processing*; 66(3): 337-355.

Sehirli, M., Gurgen, F. and Ikizoglu. (2004). Performance evaluation of digital audio watermarking techniques designed in time, frequency and cepstrum domains. *Proceedings of the International Conference on Advances in Information Systems*, 430–440.

Sriyingyong and Attakitmongcol, (2006). Wavelet-Based Audio Watermarking Using Adaptive Tabu Search, *Wireless Pervasive Computing, 2006 1st International Symposium on 16-18 an. 2006 Page(s):1 – 5.*

Tung and Ling, (1997). Digital watermarking for video, *Proceedings 1997 13th International Conference on Digital signal Processing*, Vol: 1, pp. 217 -220, 2-3.

Uludag,U., Arslan and L., (2001).Audio watermarking using DC-level shifting. *Project report, Bogazici University http://busim.ee.boun.edu.tr/~speech/publications/audio_watermarking/uu_la_audio_wm2001.pdf*

W and Hamming, (1950). Error detecting and error correcting codes, *Bell Syst. Tech. J.26*(2), pp.137–160.

Wang, Ma, Cong, Yin, (2005). An Audio Watermarking Scheme with Neural Network. Advances in Neural Networks–ISNN 2005 Springer.

Wu, S. Q., Huang, J. W., Huang, D. R., Shi and Y, (2004) Selfsynchronized audio watermark in dwt domain, Proceedings of the 2004 International Symposium on Circuits and Systems (ISCAS '04), pp.712-715.

Zhiping, Lihua, (2007). A Novel Algorithm for Robust Audio Watermarking Based on Quantification DCT Domain, Third International Conference on International Information Hiding and Multimedia Signal Processing, vol.1, pp.441-444.