International Journal on
"Technical and Physical Problems of Engineering"
(IJTPE)
Published by International Organization of IOTPE

ISSN 2077-3528
IJTPE Journal
www.iotpe.com
ijtpe@iotpe.com

September 2013 | Issue 16 | Volume 5 | Number 3 | Pages 18-26

BLIND ROBUST AUDIO WATERMARKING BASED ON REMAINING NUMBERS IN DISCRETE COSINE TRANSFORM

M. Mosleh [1] N. Hosseinpour [2]

1. Department of Computer Engineering, Dezfoul Branch, Islamic Azad University, Dezfoul, Iran, mosleh@iaud.ac.ir
2. Department of Information Technology, Dezfoul University of Medical Science, Dezfoul, Iran, hoseinpour@dums.ac.ir

Abstract- The rapid growth and increasing of information technology and computer network have caused that data transition in digital form is considered more and more. The main disturbance of owners and producers of digital products is to protect distribution and unauthorized copying. An effective solution for this problem is digital watermarking techniques. The purpose of audio digital watermarking is to insert a series of hidden information into audio file so that it is not heard and robust against signal processing attacks. This paper proposes a new audio watermarking method that is able to embed watermark sequence into host signal based on remaining numbers. The extraction process is carried out blindly without any knowledge about original audio signal. For comprehensive evaluation of the proposed method, several experiments including subjective and objective tests (MOS, SNR and SNR_{seg} measures), robustness and performance (BER, NC, TP and FP measures) against Stirmark attacks as well as capacity have been performed. Experimental results show that the proposed method is able to improve trade-off among three measures transparency, robustness, and capacity.

Keywords: Audio Watermarking, Discrete Cosine Transform (DCT), Embedding and Extraction Algorithms.

I. INTRODUCTION

Rapid growth and increasing communication has caused the digital information transmission to be considered, strongly [1, 2]. In order to protect the security of such information, different encryption and watermarking methods have been developed [3]. The encryption methods are able to convert information to code by using special algorithms, and in the opposite side, the receiver device retrieve the information by knowing decoding key. Although, the encryption techniques are used to secure the transmission of information, but there is no grantee for security after decryption.

For solving such problem, the watermarking techniques were presented to provide security of information after decoding. In other words, the watermarking is a technique which can embed special data in original signal to realize security of information against unauthorized copying, false claim of owner and exclusiveness. In recent years, most presented watermarking techniques have been focused on digital images and video clips, but audio watermarking has become recently an important research that is of interest for many scientists. There are two main reasons that audio watermarking is more complicated than image and video watermarking: First, the Human Auditory System (HAS) has greater sensitivity than the Human Visual System (HVS), because the human ear is capable to detect amplitude and frequency changes of signal, precisely. Second, the size and duration of audio signals are very shorter than image files and video clips, and therefore, the amount of information embedded in the audio signals is very large in comparison with image files and video clips, and consequently this information reduce the audio signal quality.

An audio watermarking method should be proper from transparency, robustness and capacity viewpoints [4]: *Transparency*: This means that the difference between the watermark and original signal is not understandable. *Robustness*: This shows resistance of watermark signal against common signal processing and malicious attacks. *Payload*: This is the number of bits that can be embedded into audio signal per time unit. It is mentioned that there is no known method that can satisfy all of above, well. Also, presented methods depending on their applications can only make trade-off among these measures. In general, the watermarking methods can be divided into two categories, regarding the extraction process: non-blind and blind methods [5]. In the non-blind watermarking, the extractor unit is able to extract watermark sequence by using original signal. Such systems have high-performances. In the opposite, although, the blind watermarking methods do not need to original signal for extracting the watermark sequence, but their complexity is high.

This paper proposes a new blind audio watermarking method that is capable to embed watermark sequence in DCT domain in terms of remaining numbers. Experimental results show that our proposed method is better than other previous methods from transparency, robustness and performance, as well as capacity viewpoints.

II. RELATED WORKS

So far, several audio watermarking methods have been proposed. Generally, these methods can be divided into two general categories including time domain and transform domain. Although time domain methods are easier and faster than transform domain methods but they have lower resistant against signal processing attacks. Among time domain methods, we can point to AM [6, 7], Echo [8-11], LSB [12], Patchwork [13, 14], QIM [15, 16], Interpolation [17] and Spline Interpolation [18] methods.

In the transform domain methods, embedding operation is performed in a special domain such as Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Singular Value Decomposition (SVD) and Cestrum. For this purpose, first, the original signal is divided into a set of blocks and then watermark information is embedded in transformed blocks. The DFT transform is one of the practicable frequency domains applied in many signal processing areas, well. So far, several DFT-based audio watermarking methods have been presented [19-21].

Such methods are capable to insert watermark sequence by using amplitude and phase information of original signal. The DCT is another important transform that has lower complexity compared to other transforms, and is applied in the compression area, well. I.J. Cox et al. (1997) proposed a non-blind audio watermarking in which a watermark sequence is embedded in the largest DCT coefficients, except DC coefficient [22]. Hung-HSU Tsai et al. (2003) presented an intelligent audio watermarking method in terms of characteristics of the human auditory model in the DCT domain which applied an Artificial Neural Networks (ANN) to learn relationship between original signal and watermark signal [23]. P.K. Dhar et al. (2010) proposed a method in which the watermark sequence is embedded in the maximum energy of absolute frame coefficients values [24].

Because of outstanding features of the Discrete Wavelet Transform (DWT), it has been used successfully in the field of signal processing. Wang et al. (2004), has presented a DTW-based method in which the watermark information is inserted in the low frequency coefficients [25]. Retrieving watermark sequence is blind. Wu et al (2005) presented an algorithm in terms of QIM technique to embed watermark sequence in the low frequencies [26]. Also, E. Ercelebia and L. Batakc (2009) suggested low frequency-based blind method in which Lifting-Based Wavelet Transform (LBWT) technique is used for embedding and extraction operations, rapidly [27].

S. Kirbiz et al. (2009) presented a new DTW-based blind method in which the extraction operation is carried out in terms of pattern recognition, by helping GMM technique [28]. The SVD transform is countered as a useful tool in linear algebra that has numerous applications such as compression, encryption, and signal processing operations. This tool is able to decompose a matrix in a manner that the square error to be minimum. Recently, SVD transform is successfully applied in the audio watermarking.

Ozer et al. (2005) proposed SVD-based audio watermarking so that watermark information is inserted in the SVD of original signal spectrogram [29]. In the presented method by C. Change et al. (2005), the watermark sequence is inserted in the largest quantized SV coefficients [30]. Afterward, in reference [31], an improved method suggested in which the norm of SV coefficients is used instead of the largest quantized SV coefficients. Ali Al-Haj and Ahmad Mohammad (2010) presented a new non-blind method based on SVD-DTW [32]. In this method, the watermark sequence is not directly inserted into the wavelet coefficients rather it is inserted in the singular values of wavelet coefficients.

V. Bhat K. et al. (2010) proposed a SVD-DTW based blind method in such a way that the watermark is included in the QIM of SVD Eigen values in the wavelet domain[33]. Bai Ying Lei et al. (2011) offered a method in which the watermark sequence is put in the high-frequency of SVD-DCT and the extraction operation is blind [34]. Wang et al. (2011) presented a new method based on Reversed SVD (RSVD) [35]. The watermark extraction process is done blindly without needing original signal. In addition, efforts on the subject of audio watermarking methods in cepstral domain have been presented [36-38]. These methods try to embed watermark information by using manipulating cepstral coefficients of original signal blocks.

III. DISCRETE COSINE TRANSFORM (DCT)

The DCT is a famous transform that is able to show samples of an audio signal in terms of summation of cosine functions in the different frequencies. The DCT transform is defined as followings [39]:

$$X(k) = w(k)\sum_{n=0}^{N-1} x(n)\cos\left(\frac{\pi(2n-1)(k-1)}{2N}\right) \tag{1}$$

$$k = 0,1,...,N-1$$

$$w(k) = \begin{cases} \dfrac{1}{\sqrt{N}}, & k = 0 \\ \sqrt{\dfrac{2}{N}}, & k = 1,2,...,N-1 \end{cases} \tag{2}$$

where, $x(n)$ is original audio signal and N is the number of samples. One of the most important obvious features of DCT transform is energy congestion in the few samples. This feature can be used in order to decrease distortion of original signal in the audio watermarking process.

IV. THE PROPOSED METHOD

In this section, we explain the details of embedding and extraction algorithms of the proposed method.

A. The Embedding Algorithm

Wherefrom audio signals are non-stationary, therefore, in order to apply signal-processing rules, it is necessary to analyze them in the short time intervals (frames). Block diagram of the proposed embedding method has been shown in Figure 1.

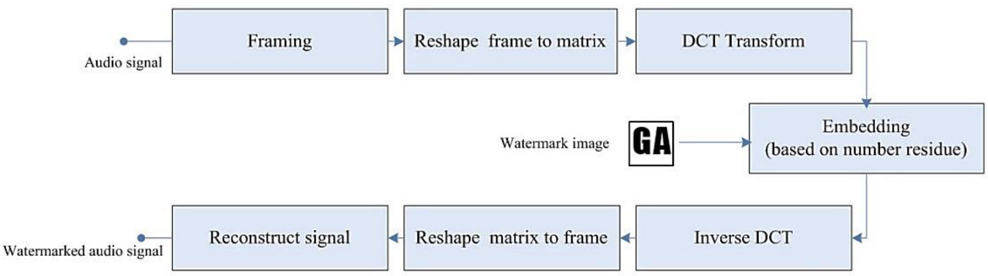

Figure 1. Block diagram of the proposed embedding method

In the following, the steps of the proposed embedding algorithm for embedding watermark sequence into audio signal are presented. It should be mentioned that watermark sequence length equal with *M*.

1. Divide audio signal into equal-sized frames based on the number of watermark bits *M*.
2. Convert the obtained frames to 2-dimention square matrixes $A_k = \{a_{ij}\}_k, 1 \le k \le M, 1 \le i, j \le N$.
3. Apply the DCT transform from the matrixes, $B_k = \text{DCT}\{A_k\},\ 1 \le k \le M$.
4. Select the highest DCT frequency coefficient for embedding watermark bit $X = b^k_{\sqrt{N},\sqrt{N}}$.
5. Convert decimal value of *X* into a 12-bit binary form.
6. Set $Y \leftarrow X(4-8)$ whereas $0 \le Y \le 31$.
7. Let m_1 and m_2 be the pair-wise co-prime numbers. Here, m_1 =3 and m_2=13 are considered.
8. Set $r_1 \leftarrow Y \mod m_1$, $r_2 \leftarrow Y \mod m_2$, $T = \max(m_1, m_2)$ and $d = |r_1 - r_2|$.
9. In order to embed watermark bit '1', the required condition is $d > \frac{T}{2}$. If it is not satisfied, *Y* value must modify as following:

candidates $=$ [];
differences $=$ [];
cnt $= 1$;
for $i = 1$ to 31 do
begin
    $r_1 = i \bmod m_1;\ r_2 = i \bmod m_2;\ d = \text{abs}|r_1 - r_2|$
    if $d \ge \frac{T}{2}$ then
    begin
        candidates(cnt) $= i$
        cnt $=$ cnt $+ 1$;
    end;
end;
differences $=$ abs(candidates[:]-*Y*)
Y $=$ candidate(min(differences[:]))

10. In order to embed watermark bit '0', the required condition is $d \le \frac{T}{2}$. If it is not satisfied, *Y* value must modify as following:

candidates $=$ [];
differences $=$ [];
cnt $= 1$;
for $i = 1$ to 31 do
begin
    $r_1 = i \bmod m_1;\ r_2 = i \bmod m_2;\ d = \text{abs}|r_1 - r_2|$
    if $d < \frac{T}{2}$ then
    begin
        candidates(cnt) $= i$
        cnt $=$ cnt $+ 1$;
    end;
end;
differences $=$ abs(candidates[:]-*Y*)
Y $=$ candidate(min(differences[:]))

11. Retrieve new value of *X* value via *Y* value.
12. Reconstruct DCT block by means of modified *X* value.
13. Apply inverse DCT.
14. Repeat 3-13 steps for remaining blocks.
15. End.

B. The Extraction Algorithm

In order to extract embedded watermark sequence, we only need two numbers m_1 and m_2 as keys. The block diagram of the proposed extractor has been shown in Figure 2.

The steps of extraction process are as followings:

1. Divide watermarked audio signal into equal-sized frames.
2. Convert the obtained frames to 2-Dimention square matrixes $A_k = \{a_{ij}\}_k, 1 \le k \le M, 1 \le i, j \le N$.
3. Apply the DCT transform from the matrixes, $B_k = \text{DCT}\{A_k\}, 1 \le k \le M$.
4. Select the highest DCT frequency coefficient for embedding watermark bit $X = b^k_{\sqrt{N},\sqrt{N}}$.
5. Convert decimal value of *X* into a 12-bit binary form.

6. Set $Y \leftarrow X(4-8)$ whereas $0 \le Y \le 31$.
7. Let m_1 and m_2 be the pair-wise co-prime numbers. Here, m_1 =3 and m_2=13 are considered.
8. Set $r_1 \leftarrow Y \mod m_1$, $r_2 \leftarrow Y \mod m_2$, $T = \max(m_1, m_2)$ and $d = |r_1 - r_2|$.
9. If $d > \frac{T}{2}$ then watermark bit='1', otherwise watermark bit ='0'.
10. Repeat 3-8 steps for remaining blocks.
11. End.

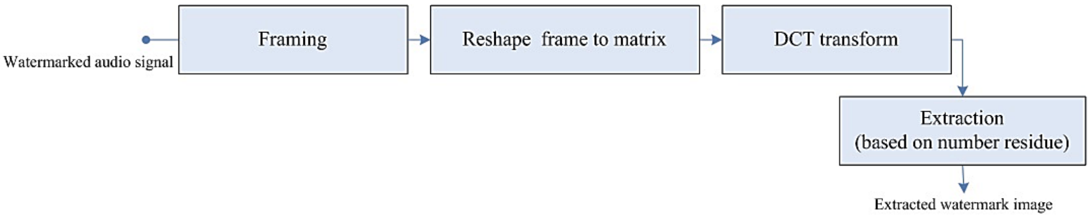

Figure 2. Block diagram of extraction process

V. EVALUATION MEASURES

In this section, different measures to evaluate the proposed method are studied.

A. Perceptual Quality

A watermark method is perceptually transparent if it is able to embed watermark sequence in the original signal so that changes of watermarked signal as compared with original signal are unnoticeable. In order to evaluate perceptual quality of the proposed method, we have applied subjective and objective tests.

A.1. Objective Test

The objective test is performed by several measures that among them, we can address to *SNR* and SNR_{seg} measures [40].

- *SNR* measure: *SNR* is a statistical metric that is introduced by International Federation Phonographic Industry (IFPI) to compute likelihood between watermarked and original signals. The *SNR* measure is computed by:

$$SNR = 10\log_{10} \frac{\sum S^2[n]}{\sum_n \left[S[n] - S'[n]\right]^2} \tag{3}$$

where, $S[n]$ and $S'[n]$ correspond to the original and watermarked signals, respectively. It should be mentioned, according to IFPI standard, and to satisfy perceptually transparent, the least *SNR* must be 20 dB.

- SNR_{seg} measure: SNR_{seg} is defined as the average of *SNR* values over small segments of watermarked signal that it is countered as one important objective metric. It is computed as following:

$$SNR_{seg} = \frac{1}{M}\sum 10\log_{10}\left[\sum_{n=N\times(j-1)+1}^{N\times j} \frac{S^2[n]}{\left[S[n]-S'[n]\right]^2}\right] \tag{4}$$

where, *M* is the number of frames and *N* is the number of samples per frame. The mentioned classic objective measures don't consider to human auditory system characteristics. Audio perceptual evaluation is computed in terms of PEAQ metric that is according to human physiological model and it has been introduced by ITU Recommendation Standard BS.1387 [41].

The output of PEAQ measure is ODG value that is evaluated by EAQUAL software, The ODG values are in the interval [-4 0] that is illustrated in Table 1.

Table 1. The ODG and SG values [41]

ODG	SG	Impairment	Quality
0.0	5.0	Imperceptible	Excellent
-1.0	4.0	Perceptible	Good
-2.0	3.0	Slightly annoying	Fair
-3.0	2.0	Annoying	Poor
-4.0	1.0	Very annoying	Bad

A.2. Subjective Test

The subjective test is important, because that the final judgment is achieved by human auditory system. For this purpose, the five listeners are provided with original and watermarked audio signals and are asked them to classify difference by SG values between [1 5], according to Table 1. The higher SG indicates better transparency [42].

B. Robustness and Performance

Robustness: Resistance of an audio watermarking method can be considered as its ability to extracting embedded watermark sequence from received signal, after signal processing attacks such as compression, adding noise, filtering and geometric transforms and etc. In the following, we introduce two important measures Normalized Correlation (*NC*) and Bit Error Rate (*BER*).

B.1. Normalized Correlation (*NC*)

Calculating Normalized Correlation (*NC*) between original and watermarked signal is obtained by Equation (5).

$$NC(W,\bar{W}) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{M} W(i,j).\bar{W}(i,j)}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{M} W^2(i,j)}.\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{M} \bar{W}^2(i,j)}} \tag{5}$$

where, W and $\bar{W}$ are original and extracted watermark sequences, respectively. A *NC* value close to '1' indicates that the correlation between original and extracted watermark is very high, and a value close to '0' indicates that the correlation is very low.

B.2. Bit Error Rate (*BER*)

Bit error rate is calculated by Equation (6) as:

$$BER(W,\bar{W}) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{M} W(i,j) \oplus \bar{W}(i,j)}{M \times M} \quad (6)$$

where, the symbol $\oplus$ is exclusive or (XOR) operator.

- Performance: True Positive (*TP*) and False Positive (*FP*) errors are two important measures which are used in order to evaluate performance audio watermarking systems.
- True Positive (*TP*): This parameter indicates the number of watermark bits '1' that the system has properly detected them as watermark bits '1'.
- False Positive (*FP*): This parameter indicates the number of watermark bits '0' that the system has wrongly detected them as watermark bits '1'.

In this case, watermark extraction problem is considered as a hypothesis testing problem where two hypothesizes H_0 and H_1 as well as two sub-hypothesis from hypothesis H_1 (H_{1a} and H_{1b}) are defined as following [28]:

H_0: The audio frame under test does not host watermark.
H_1: The audio frame under test hosts the watermark.
H_{1a}: The audio frame under test is watermark by '1'.
H_{1b}: The audio frame under test is watermarked by '0'.

Therefore, *TP* and *FP* rates can be computed for hypothesis H_i in terms of conditional probability density functions as followings:

$$TP(H_i) = P(H_i \mid H_i), \\ FP(H_i) = P(H_i \mid H_j) + P(H_i \mid H_k) \quad (7)$$

where, $i \neq j \neq k$ and $i,j,k = 0,1a,1b$.

C. Capacity

The data capacity indicates the number of bits that are embedded in the audio signal per unit time. It should be mentioned that the least payload must be 20 bps. The capacity is measured in the unit of bps (bit per second) according to Equation (8).

$$C = \frac{F_s}{N}\text{bps} \quad (8)$$

where, F_s is sampling frequency and N is the number of samples per frame.

VI. EXPERIMENTAL RESULTS

The proposed method has been implemented by MATLAB software, completely. In order to evaluate the proposed method, four different audio files including Blues, Electronic, Rock and Jazz are used where all of them are 16-bit mono audio signal in WAVE format, sampled at 44100 Hz. The watermark information is an image with size 48×48 bits that is illustrated in Figure 3.

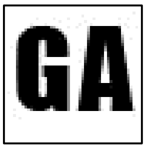

Figure 3. Binary watermark

A. Perceptual Transparency Tests

In this section, the obtained results of objective test, including *SNR*, SNR_{seg} and *ODG*, as well as the subjective tests, including the average of SGs (where called Mean Opinion Score '*MOS*') , on different audio files have been illustrated in the Table 2.

Table 2. The obtained results of objective and subjective tests of proposed method on different audio files

Audio Signals	Objective Tests		Subjective Test
	SNR	SNR_{seg}	*MOS*
Blues	31.62	47.49	4.68
Electronic	38.1	55.10	4.87
Rock	24.34	44.70	4.5
Jazz	38.06	54.70	4.87
Average	33.03	50.49	4.73

Comparing transparency of the proposed method with other methods has been shown in Table 3.

Table 3. Comparison results of transparency the proposed method vs former works

Reference	*SNR*	SNR_{Seg}	*MOS*
Erfani. Y and Siapoush. S, 2009[38]	22.7	N/A	4.7
Bhat. V and et al, 2010[33]	24.37	N/A	4.46
Wang. J and et al, 2011[35]	27.23	27.15	N/A
B.Y. Lei and et al, 2011[34]	32.53	34.81	4.71
The proposed method	33.01	50.49	4.73

B. Robustness and Performance Tests vs Stirmark Attacks

In order to evaluate resistance of the proposed method versus signal processing attacks, we have used Stirmark software. The Stirmark software is a standard tool including several different attacks. Tables 4 to 7 show the obtained results of the proposed method robustness against many Stirmark attacks, for several different audio files.

Table 4. *BER* and *NC* of the proposed method versus many Stirmark attacks for Blues audio file

Attack	*BER* (%)	*NC*
Original	0	1
addbrumm_10100	0	1
Dynnoise	0.3472	0.99
Addnoise	0	1
Addsinus	0	1
Compressor	1.64	0.9871
Extrastereo 70	0	1
fft_invert	0	1
fft_real_reverse	0	1
Invert	0	1
Normalize	0	1
Lsbzero	0	1
rc-highpass	0	1
Zerocross	5.42	0.9575

Also, the average *BER* and *NC* of the proposed method versus Stirmark attacks, for different audio files, have been illustrated in Table 8. Also, the performance of the proposed method, in terms of TP and FP metrics, against many Stirmark attacks for different audio signals has been shown in Tables (9-12).

Table 5. *BER* and *NC* of the proposed method versus many Stirmark attacks for Electronic audio file

Attack	*BER* (%)	*NC*
Original	0	1
addbrumm_10100	6.07	0.9519
Dynnoise	6.46	0.9484
addnoise_900	0.65	0.9949
Addsinus	0.82	0.9936
Compressor	13.28	0.8916
extrastereo70	0	1
fft_invert	0.13	0.9990
fft_real_reverse	0.99	0.9922
Invert	0.3	0.9976
Normalize	0	1
Lsbzero	0	1
rc-highpass	0.17	0.9986
Zerocross	0	1

Table 6. *BER* and *NC* of the proposed method versus many Stirmark attacks for Rock audio file

Attack	*BER* (%)	*NC*
Original	0	1
addbrumm_10100	0	1
Dynnoise	0	1
addnoise_900	0.04	0.9997
Addsinus	0	1
Compressor	0	1
extrastereo70	0	1
fft_invert	0	1
fft_real_reverse	0	1
Invert	0	1
Normalize	0	1
Lsbzero	0	1
rc-highpass	0	1
Zerocross	6.7	0.9459

Table 7. *BER* and *NC* of the proposed method versus many Stirmark attacks for Jazz audio file

Attack	*BER* (%)	*NC*
Original	0	1
addbrumm_10100	6.59	0.9476
Dynnoise	7.29	0.9417
addnoise_900	0.52	0.9959
Addsinus	0.91	0.9929
Compressor	10.72	0.9129
extrastereo70	0	1
fft_invert	0	1
fft_real_reverse	1.25	0.9902
Invert	0.08	0.9993
Normalize	0	1
Lsbzero	0	1
rc-highpass	0	1
Zerocross	0.04	0.9997

Table 8. The average of *BER* and *NC* of the proposed method vs. Stirmark attacks for several audio files

Attack	*BER* (%)	*NC*
Original	0	1
addbrumm_10100	3.16	0.9748
Dynnoise	3.52	0.9718
addnoise_900	0.30	0.99762
Addsinus	0.43	0.99662
Compressor	6.41	0.9479
extrastereo70	0	1
fft_invert	0.032	0.9997
fft_real_reverse	0.56	0.9956
Invert	0.09	0.99922
Normalize	0	1
Lsbzero	0	1
rc-highpass	0.04	0.9996
Zerocross	3.04	0.9757
Average	1.25	0.9899

Table 9. *TP* and *FP* of the proposed method versus many Stirmark attacks for Blues audio file

Attack	*TP* (%)	*FP* (%)
no attack	100	0
addbrumm_10100	100	0
Dynnoise	99.45	0
Addnoise	100	0
Addsinus	100	0
Compressor	97.43	0
extrastereo70	100	0
fft_invert	100	0
fft_real_reverse	100	0
Invert	100	0
Normalize	100	0
Lsbzero	100	0
rc-highpass	100	0
Zerocross	91.55	0

Table 10. *TP* and *FP* of the proposed method versus many Stirmark attacks for Electronic audio file

Attack	*TP* (%)	*FP* (%)
no attack	100	0
addbrumm_10100	92.97	4.37
Dynnoise	91.15	2.18
Addnoise	99.39	0.72
Addsinus	99.32	1.09
Compressor	82.3	4.73
extrastereo70	100	0
fft_invert	99.86	0.12
fft_real_reverse	99.05	1.09
Invert	99.59	0.12
Normalize	100	0
Lsbzero	100	0
rc-highpass	99.79	0.12
Zerocross	100	0

Table 11. *TP* and *FP* of the proposed method versus many Stirmark attacks for rock audio file

Attack	*TP* (%)	*FP* (%)
no attack	100	0
addbrumm_10100	100	0
Dynnoise	100	0
addnoise_900	99.39	0
Addsinus	100	0
Compressor	100	0
extrastereo70	100	0
fft_invert	100	0
fft_real_reverse	100	0
Invert	100	0
Normalize	100	0
Lsbzero	100	0
rc-highpass	100	0
Zerocross	89.46	0

Table 12. *TP* and *FP* of the proposed method versus many Stirmark attacks for Jazz audio file

Attack	*TP* (%)	*FP* (%)
no attack	100	0
addbrumm_10100	99.02	4.25
Dynnoise	90.14	2.67
addnoise_900	99.39	0.36
Addsinus	99.18	1.09
Compressor	84.26	1.70
extrastereo70	100	0
fft_invert	100	0
fft_real_reverse	98.58	0.97
Invert	99.93	0.12
Normalize	100	0
Lsbzero	100	0
rc-highpass	100	0
Zerocross	99.93	0

The average *TP* and *FP* metrics for diffident audio signals has been shown in Table 13. The performance comparison of proposed method and Kirbiz et al. method [28] from *TP* view point has been shown in Table 14.

Table 13. The average performance of the proposed method against many Stirmark attacks in terms of *TP* and *FP* metrics

Attack	*TP* (%)	*FP* (%)
no attack	100	0
addbrumm_10100	97.99	2.15
Dynnoise	95.18	1.21
addnoise_900	99.54	0.27
Addsinus	99.62	0.54
Compressor	90.99	1.60
extrastereo70	100	0
fft_invert	99.96	0.03
fft_real_reverse	99.40	0.51
Invert	99.88	0.06
Normalize	100	0
Lsbzero	100	0
rc-highpass	99.94	0.03
Zerocross	95.23	0
Average	98.41	0.45

Table 14. The *TP* comparison of the proposed method with SVM and GMM decoders based presented method [28]

Attack	Proposed method	GMM-based	SVM-based
Original	100	95.40	94.15
addbrumm_10100	97.99	95.26	94.02
Dynnoise	95.18	91.20	89.23
addnoise_900	99.54	88.34	86.39
Addsinus	99.62	95.39	94.12
Compressor	90.99	94.28	93.06
extrastereo70	100	95.40	94.15
fft_invert	99.96	31.52	31.97
fft_real_reverse	99.40	95.40	94.15
Invert	99.88	31.52	31.97
Lsbzero	100	95.40	94.15
rc-highpass	99.94	95.38	94.17
Zerocross	95.23	90.59	87.78
Average	98.41	84.23	83.02

As can be seen, the performance of the proposed method is about 15% better than GMM decoder based method [28] from *TP* rate viewpoint. Also, *FP* rate of the proposed method is under 0.5% while according to reference [28], the *FP* rate is about 10%.

C. Capacity Test

The data capacity of proposed method according to Equation (8) is 261 bps. A comparison between our method and several recent methods is given in Table 15.

Table 15. Comparing capacity of the proposed our scheme with several methods

Reference	Payload (bps)
Wu and et al., 2005 [26]	172
Kang et al., 2008 [14]	43
M. Fan and H. Wang , 2009 [43]	86
V. Bhat and et al., 2010 [33]	45.9
J. Wang and et al., 2011 [35]	187
B.Y. Lei. and et al., 2011 [34]	43
V. Bhat and et al., 2011 [31]	196
Proposed method	261

VII. CONCLUSIONS

This paper presents a new audio watermarking method in terms of residue numbers in DCT transform domain. First, the proposed method chooses a portion of number related to the highest DCT coefficient of each transformed frame. Then, according to remaining difference of the selected number on two co-prime numbers, the watermark bits are embedded. The watermark extracting process is blind. In order to evaluate the proposed scheme, several tests, including subjective and objective tests, robustness and performance tests and also capacity test against many Stirmark attacks have been carried out on several different audio files. The obtained results show that the efficiency of the proposed method is higher than other methods.

REFERENCES

[1] I. Cox, M. Millar, J. Bloom, J. Fridrich, T. Kalker, "Digital Watermarking and Steganography", Morgan Kaufmann, USA, 2007.

[2] M. Arnold, "Audio Watermarking - Features, Applications and Algorithms", IEEE International Conference on Multimedia and Exp. (ICME 2000), Vol. 2, pp. 1013-1016, New York, USA, 30 July - 2 August 2000.

[3] F. Hartung, M. Kutter, "Multimedia Watermarking Techniques", Proceedings of the IEEE, Vol. 87, No. 7, pp. 1079-1107, 1999.

[4] S. Katzenbeisser, F. Petitolas, "Information Hiding Techniques for Steganography and Digital Watermarking", Artech, Boston, 2000.

[5] J. Blackledge, F. Omar, "Audio Data Verification and Authentication Using Frequency Modulation Based Watermarking, International Society for Advanced Science and Technology", Journal of Electronics and Signal Processing, Vol. 3, Issue 2, pp. 51-63, 2008.

[6] W. Lie, L. Chang, "Robust and High Quality Time-Domain Audio Watermarking Subject to Psychoacoustic Masking", IEEE International Symposium on Circuits and Systems (ISCAS2001), Vol. 2, pp. 45-48, Sydney, Australia, May 2001.

[7] P. Bassia, I. Pitas, N. Nikolaidis, "Robust Audio Watermarking in the Time Domain", IEEE Transactions on Multimedia, Vol. 3, No. 2, pp. 232-241, June 2001.

[8] D. Gruhl, A. Lu, W. Bender, "Echo Hiding", Proc. Information Hiding, Springer, pp. 295-315, 1996.

[9] H.O. Oh, J.W. Seok, J.W. Hong, D.H. Youn, "New Echo Embedding Technique for Robust and Imperceptible Audio Watermarking", IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 3, pp. 1341-1344, 2001.

[10] B.S. Ko, R. Nishimura, Y. Suzuki, "Time-Spread Echo Method for Digital Audio Watermarking", IEEE Transactions on Multimedia, Vol. 7, No. 2, pp. 212-221, 2005.

[11] H.J. Kim, Y.H. Choi, "A Novel Echo-Hiding Scheme with Backward and Forward Kernels", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, No. 8, pp. 885-889, 2003.

[12] S.J. Lee, S.H. Jung, "A Survey of Watermarking Techniques Applied to Multimedia", IEEE International Symposium on Industrial Electronics, pp. 272-277, 2001.
[13] W. Bender, D. Gruhl, N. Morimoto, A. Lu, "Techniques for Data Hiding", IBM Systems Journal, Vol. 35, No. 3.4, pp. 313-336, 1996.
[14] H. Kang, K. Yamaguchi, B. Kurkoski, K. Kobayashi, "Full Index Embedding Patchwork Algorithm for Audio Watermarking", IEICE Transactions on Information & Systems, Vol. 91, No. 11, pp. 2731, 2008.
[15] B. Chen, G.W. Wornell, "Quantization Index Modulation Methods for Digital Watermarking and Information Embedding of Multimedia", Journal of VLSI Signal Processing, Vol. 27, No. 1, pp. 7-33, 2001.
[16] B. Chen, G.W. Wornell, "Quantization Index Modulation - A Class of Provably Good Methods for Digital Watermarking and Information Embedding", IEEE Transactions on Information Theory, Vol. 47, No. 4, pp. 1423-1443, 2001.
[17] V. Martin, M. Chabert, B. Lacaze, "An Interpolation Based Watermarking Scheme", Signal Processing, Vol. 88, Issue 3, No. 3, pp. 539-557, 2008.
[18] R. Fujimoto, M. Lwaki, T. Kiryu, "A Method of High Bit-Rate Data Hiding in Music Using Spline Interpolation", International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 11-14, 2006.
[19] P.K. Dhar, J.M. Kim, "Digital Watermarking Scheme Based on Fast Fourier Transformation for Audio Copyright Protection", International Journal of Security and its Applications, Vol. 5, No. 2, pp. 33-48, 2011.
[20] M. Fallahpour, D. Megias, "High Capacity Audio Watermarking Using FFT Amplitude Interpolation", IEICE Electronics Express, Vol. 6, No. 14, pp. 1057-1063, 2006.
[21] M.Q. Fan, H.X. Wang, "Statistical Characteristic Based Robust Audio Watermarking for Resolving Playback Speed Modification", Digital Signal Processing, Vol. 21, No. 1, pp. 110-117, 2011.
[22] I.J. Cox, J. Kilian, F.T. Leighton, T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia", IEEE Transactions on Image Processing, Vol. 6, No. 12, pp. 1673-1687, 1997.
[23] H.H. Tsai, J.S. Cheng, P.T. Yu, "Audio Watermarking Based on HAS and Neural Networks in DCT Domain", EURASIP Journal on Applied Signal Processing, Vol. 3, pp. 252-263, 2003.
[24] P.K. Dhar, M.I. Khan, S. Ahmad, "A New DCT Based Watermarking Method for Copyright Protection of Digital Audio", International Journal of Computer Science and Information Technology, Vol. 2, Issue 5, No. 5, pp. 91-101, 2010.
[25] R. Wang, D. Xu, J. Chen, C. Du, "Digital Audio Watermarking Algorithm Based on Linear Predictive Coding in Wavelet Domain", The 7th International Conference in Signal Processing, Vol. 3, pp. 2393-2396, 31 August - 4 September 2004.
[26] S. Wu, J. Haung, D. Huang, Y.Q Shi, "Efficiently Self-Synchronized Audio Watermarking for Assured Audio Data Transmission", IEEE Transactions on Broadcasting, Vol. 51, Issue 1, No. 1, pp. 69-76, 2005.
[27] E. Ercelebi, L. Batakc, "Audio Watermarking Scheme Based on Embedding Strategy in Low Frequency Components with a Binary Image", Digital Signal Processing, Vol. 19, Issue 2, No. 2, pp. 265-277, 2009.
[28] S. Kirbiz, Y. Ulker, B. Gunsel, "A Pattern Recognition Framework to Blind Audio Watermark Decoding", AEU International Journal of Electronics and Communications, Vol. 63, Iss. 2, No. 2, pp. 92-102, 2009.
[29] H. Ozer, B. Sankur, N. Memon, "An SVD Based Audio Watermarking Technique", Multimedia and Security Workshop, pp. 51-56, 2005.
[30] C.C. Chang, P. Tsai, C.C. Lin, "SVD Based Digital Image Watermarking Scheme", Pattern Recognition Letters, Vol. 26, Issue 10, No. 10, pp. 1577-1586, 2005.
[31] V. Bhat K., et al., "An Audio Watermarking Scheme Using Singular Value Decomposition and Dither-Modulation Quantization", Multimedia Tools and Applications, Vol. 52, Issue 2, No. 2, pp. 369-383, 2011.
[32] A. Ali, M. Ahmad, "Digital Audio Watermarking Based on the Discrete Wavelets Transform and Singular Value Decomposition", Europe Journal of Science Research, Vol. 39, Issue 1, No. 1, pp. 6-21, 2010.
[33] V. Bhat K., I. Sengupta, A. Das, "An Adaptive Audio Watermarking Based on the Singular Value Decomposition in the Wavelet Domain", Digital Signal Processing, Vol. 20, Issue 6, No. 6, pp. 1547-1558, 2010.
[34] B.Y. Lei, "Blind and Robust Audio Watermarking Scheme Based on SVD-DCT", Signal Processing, Vol. 91, Issue 8, No. 8, pp. 1973-1984, 2011.
[35] J. Wang, R. Healy, J. Timoney, "A Robust Audio Watermarking Scheme Based on Reduced Singular Value Decomposition and Distortion Removal", Signal Processing, Vol. 91, Issue 8, No. 8, pp. 1693-1708, 2011.
[36] S.K. Lee, Y.S. Ho, "Digital Audio Watermarking in the Cestrum Domain", IEEE Transactions on Consumer Electronics, Vol. 46, Issue 3, No. 3, pp. 744-750, 2000.
[37] C.T. Hsieh, P.Y. Sou, "Blind Cestrum Domain Audio Watermarking Based on Time Energy Features", The 14th International Conference on Digital Signal Processing, Vol. 2, pp. 705-708, 2002.
[38] Y. Erfani, S. Siahpoush, "Robust Audio Watermarking Using Improved TS Echo Hiding", Digital Signal Processing, Vol. 19, Issue 5, No. 5, pp. 809-814, 2009.
[39] A. Prochazka, N.G. Kingsbury, P.J.W. Payner, J. Uhlir, "Signal Analysis and Prediction", Birkhauser Boston, 1998.
[40] J.R. Deller, J.G. Proakis, J.L. Hansen, "Discrete Time Processing of Speech Signals", Macmillan, 1993.
[41] A. Lerch, "Zplane Development, Equal-Evaluate Audio Quality", Version 0.1.3alpha.
[42] C. Neubauer, J. Herre, "Digital Watermarking and Its Influence on Audio Quality", The 105th AES Convention, 1998.
[43] M. Fan, H. Wang, "Chaos-Based Discrete Fractional Sine Transform Domain Audio Watermarking Scheme", Computers and Electrical Engineering, Vol. 35, Issue 3, No. 3, pp. 506-516, 2009.

BIOGRAPHIES

Mohammad Mosleh received the B.Sc. degree in Computer Hardware Engineering from Dezful Branch, Islamic Azad University, Dezful, Iran in 2003, the M.Sc. degree in Architecture of Computer Systems as well as the Ph.D. degree in Computer Engineering from Science and Research Branch, Islamic Azad University, Tehran, Iran in 2006 and 2010, respectively. He is an Assistant Professor in the Department of Computer Engineering at the Dezful Branch, Islamic Azad University. His main research interests are in the areas of speech processing, audio watermarking, machine learning and intelligent systems.

Najmeh Hosseinpour received the B.Sc. degree in Electronics Engineering from Dezful Branch, Islamic Azad University, Dezful, Iran in 2009. She also received the M.Sc. degree in Architecture of Computer Systems from the same university in 2012. Her main research activities cover the areas of intelligent systems, speech recognition and speaker recognition.