

Article

A Novel Robust Audio Watermarking Algorithm by Modifying the Average Amplitude in Transform Domain

Qiuling Wu ^{1,2,*} and Meng Wu ¹

¹ College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China; wum@njupt.edu.cn

² College of Zijin, Nanjing University of Science and Technology, Nanjing 210046, China

* Correspondence: redpond2000@163.com; Tel.: +86-139-5103-2810

Received: 18 April 2018; Accepted: 1 May 2018; Published: 4 May 2018



Featured Application: This algorithm embeds a binary image into an audio signal as a marker to prove the ownership of this audio signal. With large payload capacity and strong robustness against common signal processing attacks, it can be used for copyright protection, broadcast monitoring, fingerprinting, data authentication, and medical safety.

Abstract: In order to improve the robustness and imperceptibility in practical application, a novel audio watermarking algorithm with strong robustness is proposed by exploring the multi-resolution characteristic of discrete wavelet transform (DWT) and the energy compaction capability of discrete cosine transform (DCT). The human auditory system is insensitive to the minor changes in the frequency components of the audio signal, so the watermarks can be embedded by slightly modifying the frequency components of the audio signal. The audio fragments segmented from the cover audio signal are decomposed by DWT to obtain several groups of wavelet coefficients with different frequency bands, and then the fourth level detail coefficient is selected to be divided into the former packet and the latter packet, which are executed for DCT to get two sets of transform domain coefficients (TDC) respectively. Finally, the average amplitudes of the two sets of TDC are modified to embed the binary image watermark according to the special embedding rule. The watermark extraction is blind without the carrier audio signal. Experimental results confirm that the proposed algorithm has good imperceptibility, large payload capacity and strong robustness when resisting against various attacks such as MP3 compression, low-pass filtering, re-sampling, re-quantization, amplitude scaling, echo addition and noise corruption.

Keywords: audio watermarking; robustness; blind extraction; discrete wavelet transform; Discrete cosine transform

1. Introduction

With the rapid development of the Internet, multimedia data stored in digital form can be easily replicated and destroyed by illegal users, so protection against intellectual property infringement increasingly becomes an important issue. There are two primary methods to overcome the above problems, which are digital signature [1,2] and digital watermarking [3,4]. Digital signature is a kind of number string which can be used as the secret key for both senders and receivers [5], and it easily stimulates the desire of illegal users to destroy the multimedia data. Digital watermarking technology conceals the watermarks into the multimedia data and later extracts such watermarks to prove the owner of multimedia data, so it is an efficient approach to protect the media contents, widely used for copyright protection, broadcast monitoring, fingerprinting, data authentication, and medical

safety, and it has become a hot topic in the field of communication and information security in recent years [6–8].

According to the different application carriers, digital watermarking technology can be divided into image watermarking technology [9,10], video watermark technology [11], audio watermark technology, and so on. Compared with the image watermarking technology, it is harder to develop the audio watermarking technology, mainly because the human auditory system is more sensitive than the visual system. With the widespread presence of audio media on the Internet, more and more people are beginning to pay attention to the research about audio watermarking technology, and more and more research algorithms have appeared. An audio watermarking algorithm generally takes into consideration four aspects including imperceptibility, robustness, security and payload capacity [12]. These four indexes are in conflict with each other. The increase of one index may cause a decrease in other indexes. An excellent audio watermarking algorithm should not only guarantee that the watermarked audio has good imperceptibility, but also provide enough payload capacity to accommodate necessary information. In addition, it should have a strong robustness to resist various signal processing attacks in practical application, so as to ensure the security of the extracted watermark. At present, most audio watermarking algorithms have some shortcomings, such as low capacity, poor robustness, and serious decline of the carrier audio quality. The audio watermarking algorithm for copyright protection must have good imperceptibility and strong robustness in resisting most common signal processing attacks that the audio may suffer during the transmission process to ensure that the watermark can be extracted accurately.

This paper presents an audio watermarking algorithm in the discrete wavelet transform (DWT) and discrete cosine transform (DCT) domains. The multi-resolution analysis characteristic of DWT renders excellent robustness against most attacks, and DCT has the merits of high energy compaction in low-frequency coefficients, which makes it suitable for compression applications, so these two methods are widely used in signal processing [13–15]. In this study, our aim is to explore all useful properties of the DWT and DCT for audio watermarking so that the issues of robustness, imperceptibility, security and payload capacity can be improved as much as possible. In this algorithm, the carrier audio signal is first segmented into multiple audio fragments (the number of fragments should be larger than the length of watermark), and then each audio fragment is followed by DWT and DCT in order to obtain the two sets of TDC which are modified to embed the binary watermark. The major contributions of this algorithm are as follows. Firstly, a novel audio watermarking algorithm based on DWT and DCT is presented by modifying the average amplitudes of the transform coefficients to embed the watermark. Secondly, the algorithm has strong robustness when resisting against various common signal processing attacks, so it can be used to prove the owner of the audio media. Thirdly, the payload capacity of the algorithm reaches 172.27 bps which is higher than most related algorithms; thus, more watermarking information can be stored in the audio media. Finally, the algorithm does not require the participation of the original audio signal when extracting the watermark, which is very convenient for practical applications.

The remaining parts of this paper are organized as follows: We review some related works about audio watermarking algorithm in Section 2. Section 3 describes the principle of the proposed watermarking algorithm in detail. This section is divided into three subjects, including the principle of watermark embedding, the principle of watermark extracting and the impact of the embedding depth on algorithm performance. The implementation of the proposed algorithm is described in Section 4, including the process of embedding and extracting. Section 5 evaluates the performance of this algorithm, including imperceptibility, capacity, and robustness, and then compares such experimental results with other algorithms in recent years. Finally, Section 6 draws up the conclusions and gives the possible future research task.

2. Related Works

In this section, we recall some previous related works on audio watermarking algorithms. Over the past decades, many audio watermarking algorithms have appeared. Audio watermarking technology can be generally implemented in either the time domain [16–18] or transform domains. Lei [16] proposed an audio algorithm by modifying the group amplitude. This algorithm had low payload capacity because it utilized three fragments to present a one-bit watermark. Erfani [17] presented an audio watermarking method with less robustness based on the time spread echo. Basia [18] presented an audio watermarking algorithm for copyright protection by modifying the amplitude of each audio sample. In general, the time-domain algorithms can be implemented easily and require less computation, but are usually less robust to many kinds of digital signal processing attacks [12]. Compared with the time-domain algorithms, transform-domain algorithms are more robust because they take advantage of the audio signal characteristics and human auditory properties [19]. There are many transform domain algorithms, such as discrete Fourier transform (DFT) [20–22], DCT [6,23], DWT [8,19,24–28] and singular value decomposition (SVD) [29]. Asmara [30] compared the characteristics of DFT, DCT and DWT when they were applied to watermarking algorithm. Natgunanathan [20] presented a patchwork-based watermarking algorithm for stereo audio signals by exploiting the similarity of the two audio channels of stereo signals in the DFT domain. This algorithm had good robustness against several conventional attacks, but the payload capacity was not high. Megias [21] presented a blind watermarking algorithm for audio signal to resist against self-synchronization by using fast Fourier transform. The algorithm embedded synchronization signals in the time domain and watermarks in the frequency domain. However, the self-synchronization code in time domain was vulnerable to some attacks, which would lead to the watermark being unable to survive. Tewari [22] proposed a digital audio watermarking algorithm which modified the middle frequency-band of DCT coefficients to embed the watermarks. Natgunanathan [6] designed another patchwork-based audio watermarking method which embedded and extracted watermark bits in a multilayer framework by modifying the mean values of selected fragments in the DCT domain. The payload capacity was higher than that in paper [20]. Hu [23] presented a large capacity audio watermarking algorithm by developing perceptual masking in the DCT domain. The authors claimed that the payload capacity of their algorithm reached 848.08 bps because they embedded the watermarks into three DCT coefficients respectively. However, they did not verify the overall performance of the algorithm when embedding the watermarks into those three DCT coefficients simultaneously. Due to the multi-resolution characteristics of DWT, many audio watermarking algorithms used DWT to analyze the frequency components of the audio signal in order to improve the performance of the algorithm. A variable-dimensional vector modulation (VDVM) algorithm was presented in paper [8], this algorithm maximized the efficiency of the norm-space DWT-based audio watermarking algorithm to achieve higher payload capacity, but its robustness was not very satisfactory. Kumsawat [19] used a genetic algorithm to search the optimal quantization step in order to improve both audio quality and robustness. This approach achieved good robustness against most of the attacks except for low-pass filtering, but its payload capacity was very low. Li [24] proposed a content-dependent localized audio watermarking algorithm to combat random cropping and time-scale modification. The watermark bits were embedded into the steady high-energy local regions to improve the robustness. This algorithm could resist synchronization attacks, but it had poor robustness against conventional signal processing attacks, such as equalization, re-sample and echo. Chen [25] proposed an adaptive method by modifying the average values of the wavelet-based entropy to embed the watermarks, but the robustness to re-sampling and low-pass-filtering attacks was quite low. An audio watermarking algorithm was proposed based on DWT in paper [26]. The algorithm changed the energy values of the former and the latter part of each audio fragment to hide confidential information. It had good robustness against several common attacks, but the author did not verify the robust performance when resisting MP3 compression which was the most common format for audio media. Wu [27] presented a self-synchronized audio watermarking algorithm based on DWT by

embedding the watermarks into the low frequency-band. The algorithm had large payload capacity, but its robustness was poor against MP3 compression and noise corruption. Wu [28] proposed a self-synchronized audio watermarking method in which the synchronization code and the watermarks are embedded with the low-frequency sub-band in the DWT domain, but this algorithm has a high bit error rate (BER) against MP3 compression. Abd [29] utilized a twofold strategy to embed the image watermark into audio signal based on SVD. The algorithm blended the watermark with the diagonal matrix holding singular values and then performed the second SVD on the modified matrix after applying the first SVD to a 2-D matrix. The matrices which contained left- and right-singular vectors must be conserved in order to extract the watermark.

All the above algorithms were designed in a single transform domain. In recent years, there have been many audio watermarking algorithms in multiple transform domains. An audio watermarking algorithm against desynchronization attacks was proposed based on support vector regression in paper [31]. The algorithm used the support vector machines (SVM) theory to locate the optimal embedding positions, and embedded the watermarks into the statistical average value of low-frequency components in DWT and DCT domains. Wang [32] proposed an audio watermarking algorithm according to the multi-resolution characteristic of DWT and the energy compression capability of DCT, but its robustness was poor against low-pass filtering and amplitude scaling. Vivekananda [33] utilized DWT and SVD to propose an adaptive audio watermarking by applying a quantization index modulation (QIM) process on the SVD values in the DWT domain. Bhat [34] presented a SVD-DWT blind watermarking algorithm to embed watermark into the audio signals in which the quantization steps were determined by the statistical properties of the involved DWT coefficients. Lei [35] attempted to embed the watermark into the high frequency-band of the SVD-DCT block. They claimed the performance generally better than the previous SVD-based methods. Hu [12] integrated discrete wavelet packet transformation (DWPT), SVD and QIM to achieve an approach for blind audio watermarking. The SVD was employed to analyze the matrix formed by the DWPT coefficients and embedded the watermarks by controlling singular values subject to perceptual criteria. Vivekananda [36] proposed a robust and blind audio watermarking algorithm based on SVD and QIM. Apart from the above algorithms, there are other audio watermarking algorithms in papers [4,7]. Wang [4] integrated exponent moments (EMs) and QIM to achieve an audio watermarking algorithm which used EMs to improve the robustness and QIM to realize the blind extraction of watermark, but this algorithm was not robust enough to resist amplitude scaling. Xiang [7] presented a reversible audio hiding scheme by using non-causal prediction. The scheme used the minimum error power method to calculate the optimum order and the prediction correlation of the audio data which could be used to embed the watermarks.

It can be seen from the above introduction of the related works that the performance of the algorithms is not only related to the transform domain of signal processing, but also related to the embedding rules. Even though the algorithms use the same transform domain, their performance varies greatly due to the different embedding rules. The proposed algorithm in this paper combines the characteristics of DWT and DCT to process audio signal, and uses special embedding rules to embed the watermarks, which provides this algorithm with good robustness.

3. Principle of the Watermarking Algorithm in Transform Domain

3.1. Principle of Watermark Embedding

The multi-resolution analysis characteristic of DWT renders excellent robustness against most attacks, so DWT can be used in audio watermarking algorithm to decompose the audio signal into wavelet coefficients with different frequency bands used for carrying watermarks. The energy compression capability of DCT can concentrate the main energy of the audio signal on the low frequency coefficient of DCT. This study will take advantage of these two methods to develop a novel robust audio watermarking algorithm. Since the human auditory system is insensitive to minor

changes in the high-frequency components of the audio signal, the watermark information can be hidden in these high-frequency components obtained by DWT and DCT on the carrier audio signal.

Suppose that the carrier audio signal is A , which has K sample points, and it can be expressed by the following formula:

$$A = \{a(k), 1 \leq k \leq K\}, \quad (1)$$

where $a(k)$ is the k th sample value of this audio signal. A is divided into M audio fragments A_l ($1 \leq l \leq M$) with N sample points, and then the r -level DWT is performed on A_l to obtain the wavelet coefficient D_l showed in Formula (2), including the approximation coefficient $Ce(r)$ and the detail coefficients $De(i)$ ($i = 1, 2, \dots, r$):

$$D_l = DWT(A_l) = Ce(r) \oplus De(r) \oplus De(r-1) \cdots \oplus De(2) \oplus De(1), \quad (2)$$

where $Ce(r)$ is the r th level approximation coefficient decomposed by DWT, containing the lowest frequency component of the audio signal. The minor changes in $Ce(r)$ will cause a significant drop in the audio quality, so usually watermarks cannot be embedded into this frequency-band. $De(i)$ ($i = 1, 2, \dots, r$) is the i th level detail coefficient. The smaller i is, the higher the frequency component contained in $De(i)$ will be, and the smaller the impact of minor changes of $De(i)$ on the audio quality will be. Therefore, the watermarks may be embedded into $De(i)$. However, the high-frequency components are vulnerable to malicious attacks, so the detail components near the approximate components can be chosen to conceal watermark, which not only has little influence on the audio quality, but also can resist malicious attacks. In this study, the r th level detail coefficient $De(r, n)$ ($n = 1, 2, \dots, N/2^r$) can be selected as the embedding frequency band. Divide $De(r, n)$ into two packets respectively according to Formulas (3) and (4), including the former packet $De_1(r, j)$ and the latter packet $De_2(r, j)$ with the length of $N/2^{r+1}$:

$$De_1(r, j) = De(r, j) \quad j = 1, 2, \dots, N/2^{r+1}, \quad (3)$$

$$De_2(r, j) = De(r, \frac{N}{2^{r+1}} + j) \quad j = 1, 2, \dots, N/2^{r+1}. \quad (4)$$

Perform DCT on $De_1(r, j)$ and $De_2(r, j)$ to obtain two transform-domain coefficients $C_1(r, j)$ and $C_2(r, j)$ with the length of $N/2^{r+1}$, and then connect $C_1(r, j)$ and $C_2(r, j)$ to form an array $C(r, n)$ with the length of $N/2^r$. Calculate the average amplitudes of $|C(r, n)|$, $|C_1(r, j)|$ and $C_2(r, j)$ according to Formulas (5)–(7).

The average amplitude of $|C(r, n)|$ is

$$M_c = \frac{2^r}{N} \sum_{n=1}^{N/2^r} |C(r, n)|, \quad n = 1, 2, \dots, N/2^r. \quad (5)$$

The average amplitude of the former packet $C_1(r, j)$ is

$$M_{c1} = \frac{2^{r+1}}{N} \sum_{j=1}^{N/2^{r+1}} |C_1(r, j)|, \quad j = 1, 2, \dots, N/2^{r+1}. \quad (6)$$

The average amplitude of the latter packet $C_2(r, j)$ is

$$M_{c2} = \frac{2^{r+1}}{N} \sum_{j=1}^{N/2^{r+1}} |C_2(r, j)|, \quad j = 1, 2, \dots, N/2^{r+1}. \quad (7)$$

Suppose that the binary image watermark to be embedded is $W = \{w(q), 1 \leq q \leq L\}$, where $w(q) \in \{0, 1\}$, L is the length of the watermarks, $L \leq M$. The average amplitudes of the two packets are modified to embed the watermark. The embedding rules are as follows:

If $w(q) = 1$, modify $C_1(r, j)$ and $C_2(r, j)$ according to the following Formulas (8) and (9):

$$C'_1(r, j) = C_1(r, j) \times \frac{(1 + \lambda)M_c}{M_{c1}}, \quad (8)$$

$$C'_2(r, j) = C_2(r, j) \times \frac{(1 - \lambda)M_c}{M_{c2}}. \quad (9)$$

If $w(q) = 0$, modify $C_1(r, j)$ and $C_2(r, j)$ according to the following Formulas (10) and (11):

$$C'_1(r, j) = C_1(r, j) \times \frac{(1 - \lambda)M_c}{M_{c1}}, \quad (10)$$

$$C'_2(r, j) = C_2(r, j) \times \frac{(1 + \lambda)M_c}{M_{c2}}, \quad (11)$$

where λ is the embedding depth and its span is within the interval of (0,1). $C'_1(r, j)$ is the modified coefficient of the former packet, and $C'_2(r, j)$ is the modified coefficient of the latter packet. Perform the inverse DCT on $C'_1(r, j)$ and $C'_2(r, j)$ to obtain $De'_1(r, j)$ and $De'_2(r, j)$ respectively, and then recombine them into $De'(r, n)$. Replace $De(r)$ with $De'(r, n)$ in Formula (2) to get the watermarked coefficient D'_l . Finally, perform the inverse DWT on D'_l to reconstruct the watermarked audio fragment A'_l and then recombine the watermarked audio signal A' .

3.2. Principle of Watermark Extracting

When extracting the watermark, the watermarked audio A' is divided into M audio fragments A'_l ($1 \leq l \leq M$) with N sample points, and then perform r -level DWT on each audio fragment to obtain the wavelet coefficient $Ce'(r)$ and $De'(i)$ ($i = 1, 2, \dots, r$). Divide $De'(r)$ into the former packet $De'_1(r, j)$ and the latter packet $De'_2(r, j)$. Perform DCT on the two packets to obtain $C'_1(r, j)$ and $C'_2(r, j)$ and then calculate their average amplitudes respectively according to Formulas (6) and (7).

If $w(q) = 1$, according to Formulas (6)–(9), the average amplitude of $C'_1(r, j)$ is

$$\begin{aligned} M'_{c1} &= \frac{2^{r+1}}{N} \sum_{j=0}^{N/2^{r+1}} |C'_1(r, j)| \\ &= \frac{2^{r+1}}{N} \sum_{j=0}^{N/2^{r+1}} |C_1(r, j) \times \frac{(1+\lambda)M_c}{M_{c1}}| \\ &= \frac{(1+\lambda)M_c}{M_{c1}} \frac{2^{r+1}}{N} \sum_{j=0}^{N/2^{r+1}} |C_1(r, j)| \\ &= \frac{(1+\lambda)M_c}{M_{c1}} M_{c1} \\ &= (1 + \lambda)M_c, \quad j = 1, 2, \dots, N/2^{r+1}. \end{aligned} \quad (12)$$

The average amplitude of $C'_2(r, j)$ is

$$\begin{aligned} M'_{c2} &= \frac{2^{r+1}}{N} \sum_{j=0}^{N/2^{r+1}} |C'_2(r, j)| \\ &= \frac{2^{r+1}}{N} \sum_{j=0}^{N/2^{r+1}} |C_2(r, j) \times \frac{(1-\lambda)M_c}{M_{c2}}| \\ &= \frac{(1-\lambda)M_c}{M_{c2}} \frac{2^{r+1}}{N} \sum_{j=0}^{N/2^{r+1}} |C_2(r, j)| \\ &= \frac{(1-\lambda)M_c}{M_{c2}} M_{c2} \\ &= (1 - \lambda)M_c, \quad j = 1, 2, \dots, N/2^{r+1}. \end{aligned} \quad (13)$$

According to Formulas (12) and (13), when $\lambda > 0$, $M'_{c1} \geq M'_{c2}$.

Similar to the above analysis process, if $w(q) = 0$, according to Formulas (10) and (11), the average amplitude of $C'_1(r, j)$ is

$$M'_{c1} = \frac{2^{r+1}}{N} \sum_{j=0}^{N/2^{r+1}} |C'_1(r, j)| = (1 - \lambda)M_c, j = 1, 2, \dots, N/2^{r+1} \quad (14)$$

The average amplitude of $C'_2(r, j)$ is

$$M'_{c2} = \frac{2^{r+1}}{N} \sum_{j=0}^{N/2^{r+1}} |C'_2(r, j)| = (1 + \lambda)M_c, j = 1, 2, \dots, N/2^{r+1} \quad (15)$$

It can be seen from Formula (14) and (15), when $\lambda > 0$, $M'_{c1} \leq M'_{c2}$. Based on the above analysis, the binary watermark can be extracted from the audio fragment A'_l according to Formula (16):

$$w'(q) = \begin{cases} 1 & , \text{ if } M'_{c1} \geq M'_{c2} \\ 0 & , \text{ if } M'_{c1} < M'_{c2} \end{cases}, q = 1, 2, \dots, L \quad (16)$$

3.3. Impact of the Embedding Depth on Algorithm Performance

The principle of watermark embedding in Section 3.1 shows that a watermark is embedded by modifying the amplitudes of $C_1(r, j)$ and $C_2(r, j)$; the larger the variation of the amplitude is, the worse the quality of the carrier audio is. Otherwise, the watermark cannot be accurately extracted if the variation is too small. Therefore, the variation of the amplitude should be maintained within a certain range, which not only guarantees the imperceptibility of the algorithm, but also maintains good robustness. It can be seen from Formulas (8)–(11) that the modified $C'_1(r, j)$ and $C'_2(r, j)$ are related to the embedding depth, so the following experiment tests the impact of embedding depth on the performance of the algorithm. The tested carrier audio signal is a song downloaded from the Internet, lasting for 60 s approximately, sampled at 44,100 Hz and 16-bit quantization. The watermark bits were a series of random binary string, only including 1 and 0, and long enough to cover the entire carrier signal. In order to objectively evaluate the performance of this watermarking algorithm, the quality of the watermarked audio can be determined using signal-to-noise ratio (SNR) as the performance index formulated as

$$\text{SNR}(A, A') = 10 \lg \left\{ \frac{\sum_{k=1}^K A^2}{\sum_{k=1}^K (A' - A)^2} \right\}, \quad (17)$$

where A and A' denote the carrier audio signal and the watermarked audio signal respectively. The larger the SNR is, the smaller the decrease of the audio quality will be, and the better the imperceptibility of algorithm will be. Usually, when SNR is over 20 dB, the audio quality is good.

The robustness of the proposed algorithm to resist various attacks is evaluated using the bit error rate (BER), which is defined as

$$\text{BER}(w, w') = \frac{\sum_{q=1}^L w(q) \oplus w'(q)}{L} \times 100\%, \quad (18)$$

where $w(q)$ and $w'(q)$ denote the original watermark and the extracted watermark respectively, \oplus stands for the exclusive-OR operator, and L is the length of the watermark. Generally, the smaller value of BER implies that the algorithm has good robustness against attacks.

Normalized correlation (NC) coefficient can be used to compare the similarity between the original watermark and the extracted watermark represented as Formula (19). If NC is close to 1, $w(q)$ is very similar to $w'(q)$. On the other hand, $w(q)$ and $w'(q)$ will be very different when NC is close to zero:

$$NC(w, w') = \frac{\sum_{q=1}^L w(q) \times w'(q)}{\sqrt{\sum_{q=1}^L w(q)^2 \sum_{q=1}^L w'(q)^2}}. \quad (19)$$

When λ changes from 0 to 1, the experimental results of SNR and BER are shown in Figure 1. Figure 1a shows that the larger the λ is, the smaller the SNR of the watermarked audio signal is, and the worse the imperceptibility of the algorithm is. Figure 1b shows that the larger the λ is, the smaller the BER of the extracted watermark is, and the better the robustness of the algorithm is. Thus, imperceptibility and robustness are contradictory, a larger λ can be selected to improve the robustness of the algorithm under the premise of ensuring that the SNR is greater than 20 dB.

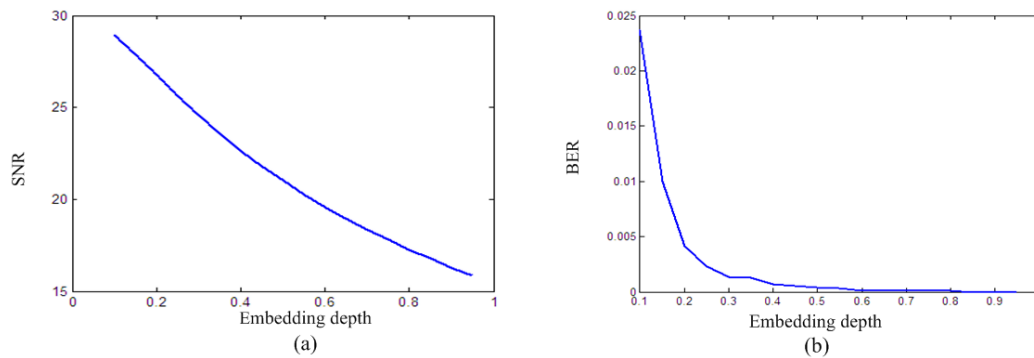


Figure 1. Performance evaluation under different embedding depth: (a) evaluation about SNR; (b) evaluation about BER.

4. Implementation of the Proposed Algorithm

In Section 3, the watermark embedding principle and extraction principle of the proposed algorithm are described. The detailed implementation steps are described in this section including two parts: embedding watermark and extracting watermark.

4.1. Procedure for Embedding Watermark

The watermark embedding diagram of this proposed algorithm is showed in Figure 2. The detailed embedding steps are described as follows:

- Step 1: Convert the image watermark into the binary bit stream $W = \{w(q), 1 \leq q \leq L\}$ with the length of L .
- Step 2: The carrier audio A is divided into M audio fragments A_l ($1 \leq l \leq M$) with the length of N after low-pass filtering. M is the number of the audio fragments, $M \geq L$.
- Step 3: When l changes from 1 to M , perform the r -level DWT on each fragment A_l to obtain the wavelet coefficients, and select $De(r)$ as the embedding frequency-band.
- Step 4: Divide $De(r)$ into the former packet $De_1(r, j)$ and the latter packet $De_2(r, j)$ with the length of $N/2^{r+1}$ according to Formulas (3) and (4).
- Step 5: Perform DCT on $De_1(r, j)$ and $De_2(r, j)$ to obtain $C_1(r, j)$ and $C_2(r, j)$ respectively.
- Step 6: Connect $C_1(r, j)$ and $C_2(r, j)$ to form an array $C(r, n)$ with the length of $N/2^r$.
- Step 7: Calculate the average amplitude of $|C(r, n)|$, $|C_1(r, j)|$ and $|C_2(r, j)|$ according to Formulas (5)–(7).

- Step 8: If $w(q) = 1$, embed one-bit watermark into $De(r)$ according to Formulas (8) and (9). If $w(q) = 0$, embed one-bit watermark into $De(r)$ according to Formulas (10) and (11).
- Step 9: Perform the inverse DCT on $C'_1(r, j)$ and $C'_2(r, j)$ to obtain $De'_1(r, j)$ and $De'_2(r, j)$.
- Step 10: Recombine $De'_1(r, j)$ and $De'_2(r, j)$ into $De'(r, n)$ and perform the inverse DWT to reconstruct the watermarked audio fragment A'_l .
- Step 11: Repeat Step 3 to Step 10 until all watermarks are embedded.
- Step 12: Recombine $A'_l (1 \leq l \leq M)$ as the watermarked audio signal A' .

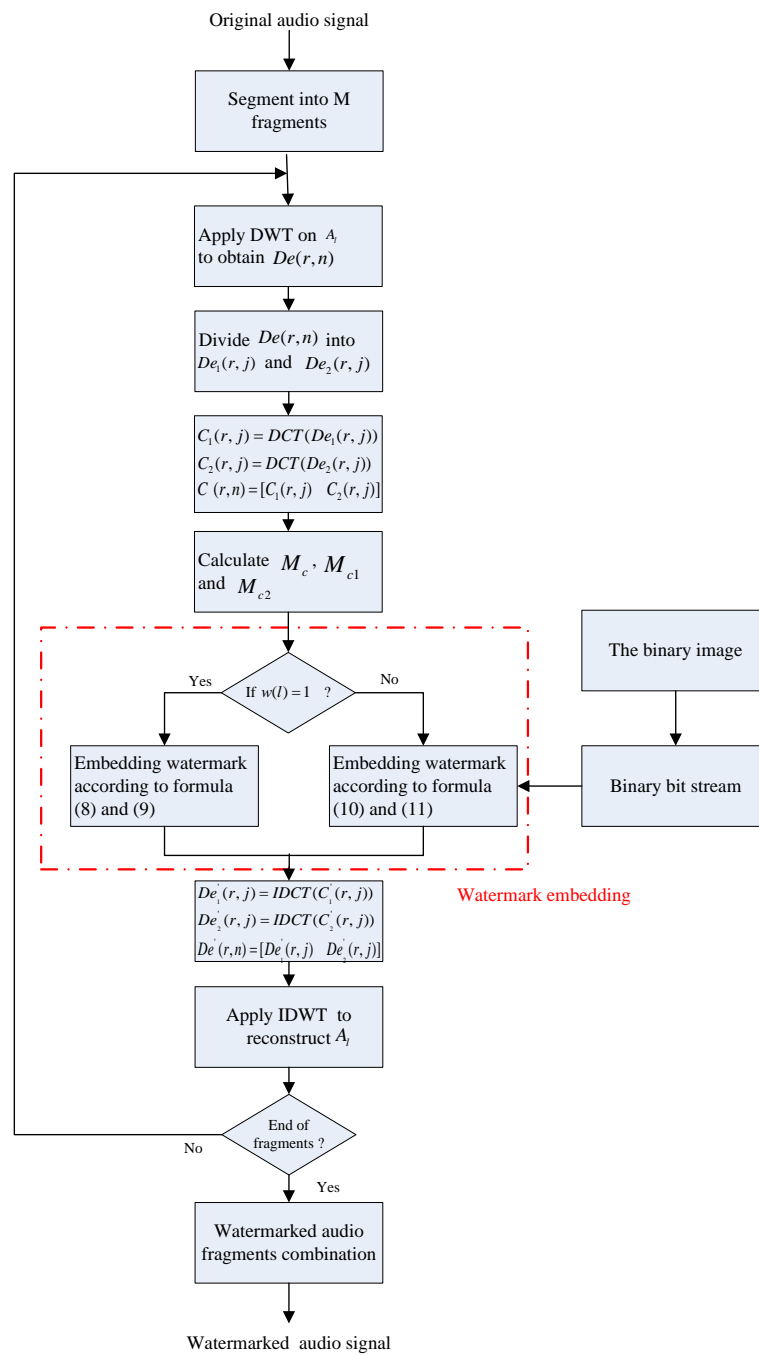


Figure 2. The watermark embedding diagram of this proposed algorithm. DWT: discrete wavelet transform; DCT: discrete cosine transform; IDWT: inverse discrete wavelet transforms.

4.2. Procedure for Extracting Watermark

The watermark extracting diagram of this proposed algorithm is showed in Figure 3. The detailed extracting steps are described as follows:

- Step 1: Segment the watermarked audio signal A' into M audio fragments A'_l with the length of N .
- Step 2: Perform r -level DWT on A'_l to obtain the wavelet coefficients $De'(r)$.
- Step 3: Divide $De'(r)$ into $De'_1(r, j)$ and $De'_2(r, j)$ with the length of $N/2^{r+1}$.
- Step 4: Perform DCT on $De'_1(r, j)$ and $De'_2(r, j)$ to obtain $C'_1(r, j)$ and $C'_2(r, j)$ respectively.
- Step 5: Calculate the average amplitudes of $|C_1(r, j)|$ and $|C_2(r, j)|$ to obtain M'_{c1} and M'_{c2} according to Formulas (6) and (7).
- Step 6: If $M'_{c1} > M'_{c2}$, the extracted binary information is '1', otherwise, it is '0'.
- Step 7: Repeat Step 2 to Step 6 until all binary watermarks are extracted.
- Step 8: Convert the extracted binary stream into binary image watermark.

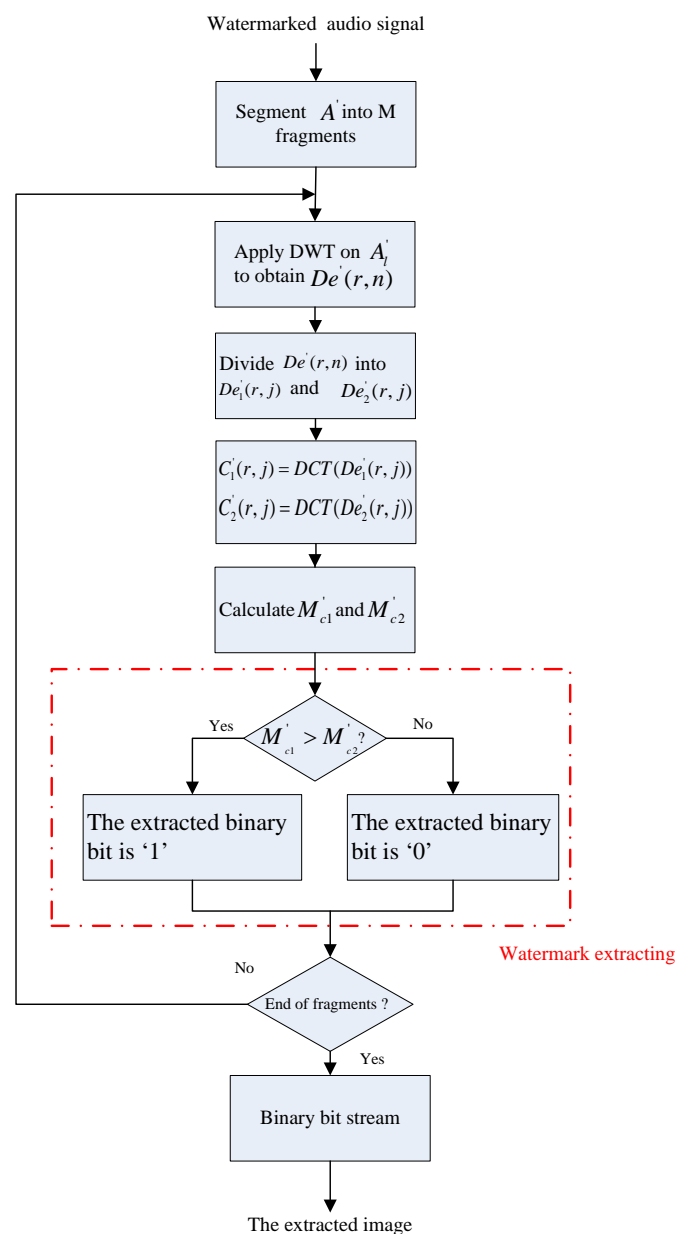


Figure 3. The watermark extracting diagram of this proposed algorithm.

5. Performance Evaluation

This section will use a large number of experiments to evaluate the performance of the proposed algorithm. The detailed experimental environment is described as follows: (1) Computer system: Microsoft Windows XP Professional; (2) Programming Language: MATLAB 6.5; (3) Software for processing audio signals: Cool Edit Pro V2.1. Experimental parameters are as follows: (1) The tested carrier audio signals consist of 20 songs downloaded from the Internet, sampled at 44,100 Hz and 16-bit quantization; (2) Four binary images are used as the watermark shown in Figure 4 respectively; (3) Perform four-level DWT on each audio fragment; (4) Embed the watermark into the wavelet coefficient $De(4)$; (5) The length of the audio fragment is 256; (6) The embedding depth is 0.4.

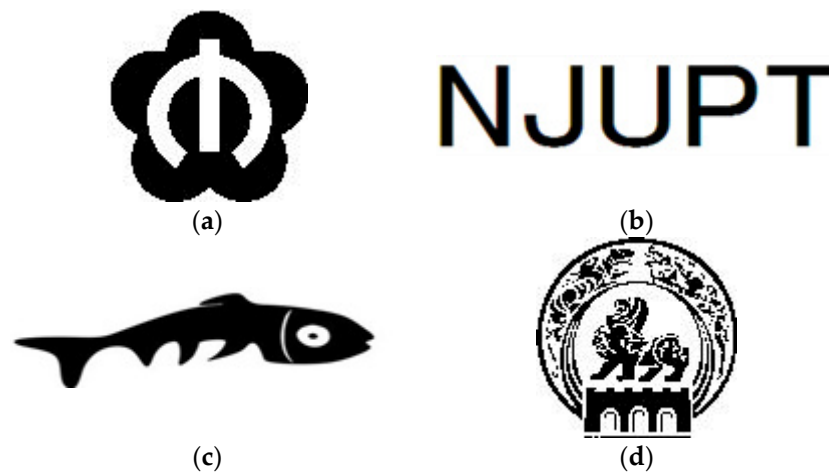


Figure 4. Four binary images to be embedded into the carrier audio signals: (a) The first image with dimensions of 100×96 ; (b) The second image with dimensions of 200×50 ; (c) The third image with dimensions of 200×50 ; (d) The fourth image with dimensions of 100×100 .

5.1. Imperceptibility and Payload Capacity

The test of each audio signal was repeated 10 times, so it takes 200 experiments to test all the audio signals. Based on the principle of watermark embedding in Section 3, one-bit watermark can be embedded for each audio fragment. Since the length of the audio fragment is 256, the payload capacity of this algorithm is $44,100/256 = 172.27$ bps.

The average results about the SNR of the audio signal, the NC and BER of the extracted watermark and the payload capacity are listed in Table 1. The experimental results showed that the payload capacity of the algorithm is the same as that of paper [25], higher than that of paper [4,12], and far higher than that of paper [16,19].

Table 1. Experimental results of the signal-to-noise ratio (SNR), normalized correlation (NC), bit error rate (BER) and payload capacity (no attack).

Items	Proposed	Paper [4]	Paper [16]	Paper [25]	Paper [12]	Paper [19]
SNR (dB)	23.49	N/A	21.37	18.42	20.32	26.79
Capacity(bps)	172.27	125	43.07	172.27	139.97	34.14
NC	1	N/A	N/A	N/A	N/A	1
BER (%)	0.00	0.00	N/A	N/A	0.12	0.00

Note that N/A means no report is found in the selected algorithm.

The average SNR of this proposed algorithm is 23.49 dB, which is higher than that in papers [12,16,25] but not paper [19], while the payload capacity of this proposed algorithm is five times that of the paper [19].

The images extracted without any attack by this algorithm are very similar to the original images because NC equals 1 and BER equals 0 as shown in Table 1. Figure 5 shows the waveform comparison of the carrier audio and the watermarked audio (only show a short fragment lasting about three seconds) without performing any attack. It can be seen from Figure 5 that two waveform figures have no obvious changes before and after the watermark was embedded into the carrier audio. Figure 6 shows that two spectrogram figures are slightly different at high frequency band, mainly because the watermarks are embedded into the high frequency band of the watermarked audio in Figure 6b, but human ears are not sensitive to these minor changes of the high frequency component. The experimental result of SNR in Table 1, Figures 5 and 6 all indicate the excellent imperceptibility of this algorithm.

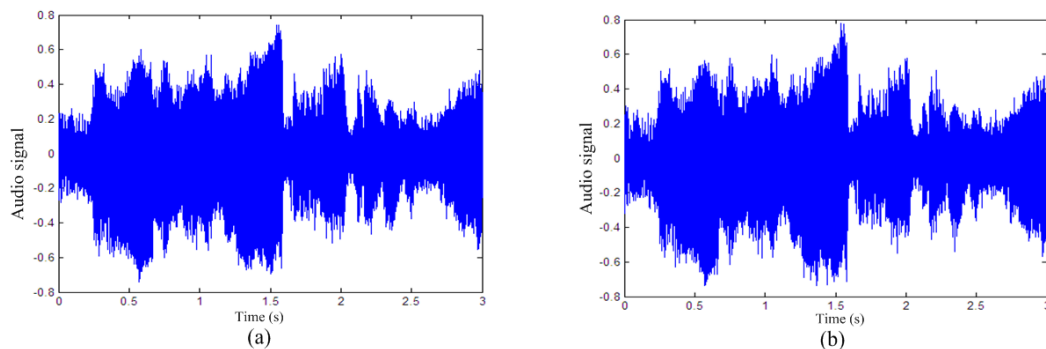


Figure 5. Waveform comparison of the carrier audio and the watermarked audio: (a) original audio signal; (b) watermarked audio signal.

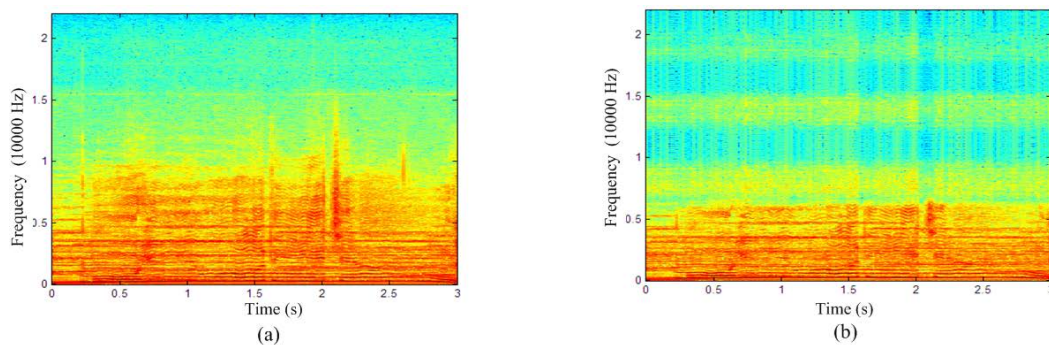


Figure 6. Spectrogram comparison of the carrier audio and the watermarked audio: (a) original audio signal; (b) watermarked audio signal.

5.2. Robustness

Robustness is an important index for evaluating the performance of the watermarking algorithm. This study examines the NC and BER between the carrier watermark and the extracted watermark to assess the robustness against various attacks. The attack types considered in the test are as follows:

- A. Low-pass filtering: applying low-pass filter with cutoff frequency of four kilohertz.
- B. Amplitude scaling: scaling the amplitude of the watermarked audio signal by 0.8.
- C. Amplitude scaling: scaling the amplitude of the watermarked audio signal by 1.2.
- D. Noise corruption: adding zero-mean Gaussian noise to the watermarked audio signal with 20 dB.
- E. Noise corruption: adding zero-mean Gaussian noise to the watermarked audio signal with 30 dB.
- F. Noise corruption: adding zero-mean Gaussian noise to the watermarked audio signal with 35 dB.

- G. MP3 compression: applying MP3 compression with 64 kbps to the watermarked audio signal.
- H. MP3 compression: applying MP3 compression with 128 kbps to the watermarked audio signal.
- I. Re-sampling: dropping the sampling rate of the watermarked audio signal from 44,100 Hz to 22,050 Hz and then rose back to 44,100 Hz.
- J. Re-quantization: quantizing the watermarked audio signal from 16-bit/sample to 8-bit/sample and then back to 16-bit/sample.
- K. Echo addition: adding an echo signal with a delay of 50 ms and a decay of five percent to the watermarked audio signal.

The extracted images and the average experimental results of NC are shown in Figures 7–10 respectively.

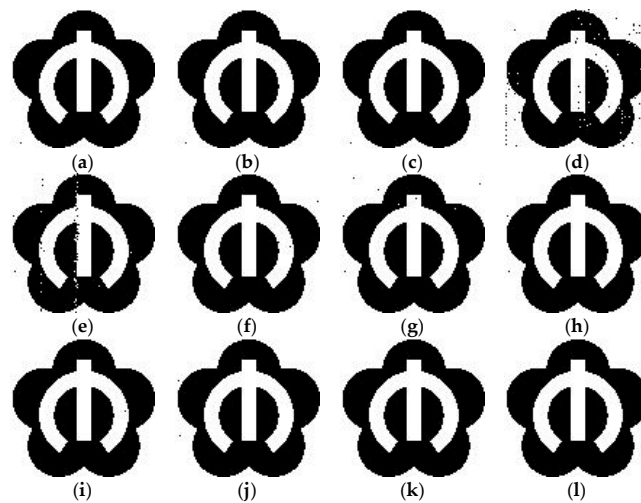


Figure 7. The extracted pictures of the first image under different attacks: (a) Low-pass filtering, NC = 1; (b) Amplitude scaling by 0.8, NC = 1; (c) Amplitude scaling by 1.2, NC = 1; (d) Noise corruption (20 dB), NC = 0.9665; (e) Noise corruption with 30 dB, NC = 0.9965; (f) Noise corruption with 35 dB, NC = 0.9992; (g) MP3 compression with 64 kbps, NC = 0.9998; (h) MP3 compression with 128 kbps, NC = 1; (i) Re-sampling, NC = 1; (j) Re-quantization, NC = 0.9982; (k) Echo addition, NC = 1; (l) no attack, NC = 1.

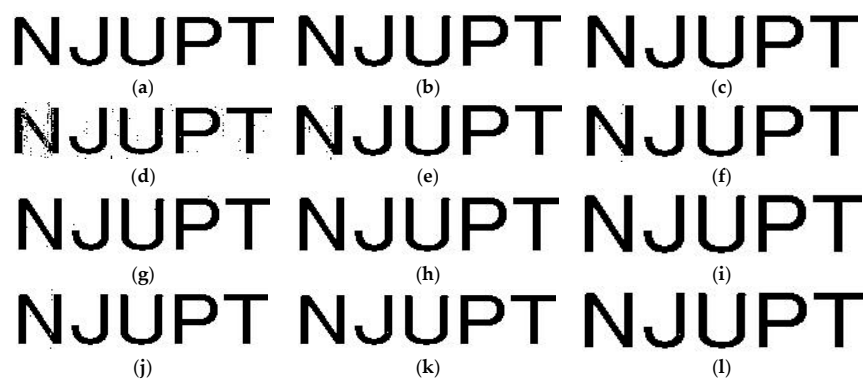


Figure 8. The extracted pictures of the second image under different attacks: (a) Low-pass filtering, NC = 1; (b) Amplitude scaling by 0.8, NC = 1; (c) Amplitude scaling by 1.2, NC = 1; (d) Noise corruption (20 dB), NC = 0.9954; (e) Noise corruption with 30 dB, NC = 0.9996; (f) Noise corruption with 35 dB, NC = 0.9999; (g) MP3 compression with 64 kbps, NC = 0.9999; (h) MP3 compression with 128 kbps, NC = 1; (i) Re-sampling, NC = 1; (j) Re-quantization, NC = 0.9998; (k) Echo addition, NC = 1; (l) no attack, NC = 1.

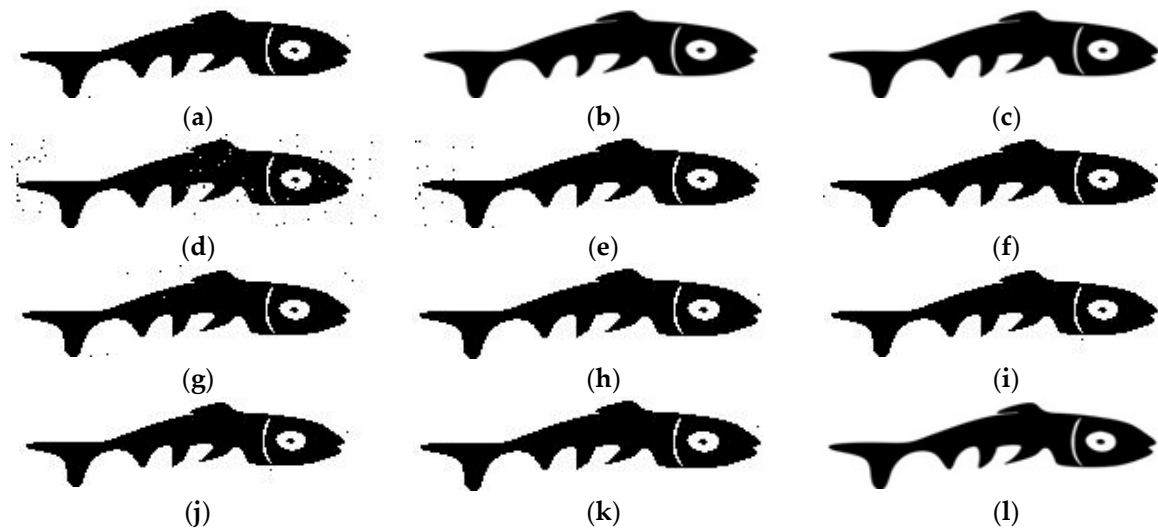


Figure 9. The extracted pictures of the second image under different attacks: (a) Low-pass filtering, NC = 1; (b) Amplitude scaling by 0.8, NC = 1; (c) Amplitude scaling by 1.2, NC = 1; (d) Noise corruption (20 dB), NC = 0.9911; (e) Noise corruption with 30 dB, NC = 0.9975; (f) Noise corruption with 35 dB, NC = 0.9992; (g) MP3 compression with 64 kbps, NC = 0.9998; (h) MP3 compression with 128 kbps, NC = 1; (i) Re-sampling, NC = 1; (j) Re-quantization, NC = 0.9986; (k) Echo addition, NC = 0.9999; (l) no attack, NC = 1.

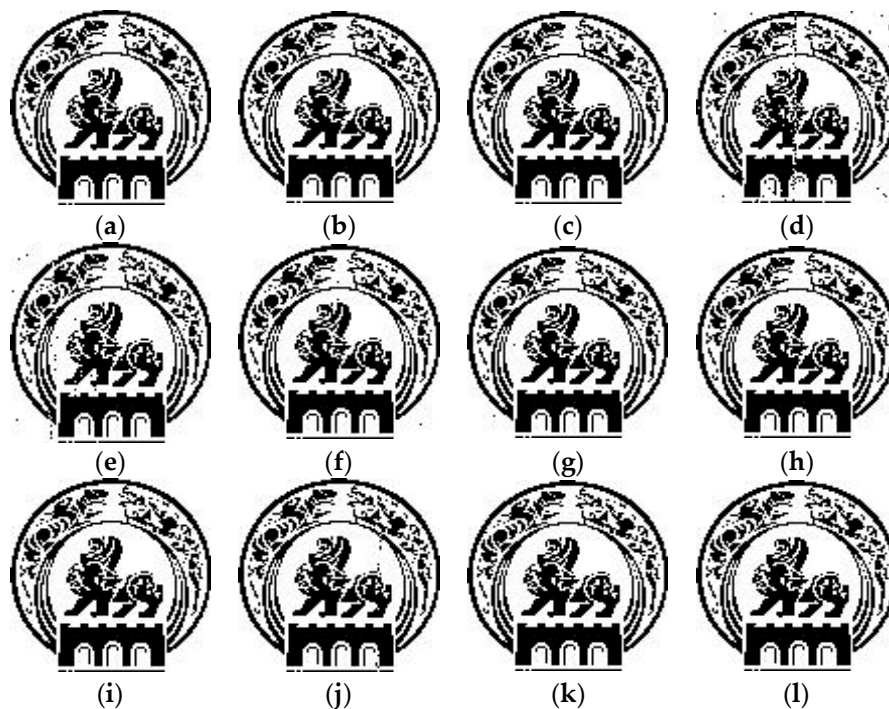


Figure 10. The extracted pictures of the first image under different attacks: (a) Low-pass filtering, NC = 1; (b) Amplitude scaling by 0.8, NC = 1; (c) Amplitude scaling by 1.2, NC = 1; (d) Noise corruption (20 dB), NC = 0.9856; (e) Noise corruption with 30 dB, NC = 0.9991; (f) Noise corruption with 35 dB, NC = 0.9998; (g) MP3 compression with 64 kbps, NC = 0.9998; (h) MP3 compression with 128 kbps, NC = 1; (i) Re-sampling, NC = 0.9998; (j) Re-quantization, NC = 0.9990; (k) Echo addition, NC = 1; (l) no attack, NC = 1.

It can be seen from Figures 7–10 that the extracted image watermarks are very similar to the original image watermarks when resisting against low-pass filter with cutoff frequency 4 kHz, amplitude scaling by 0.8 and 1.2, MP3 compression with 128 kbps and 64 kbps, re-sampling, re-quantization, echo addition with a delay of 50 ms and a decay of 5%, noise corruption in 30 dB and 35 dB. The extracted image watermarks shown in Figures 7d, 8d, 9d and 10d are relatively obscure when suffering attack from noise corruption in 20 dB, but with the decrease of white noise, they become more and more clear, as shown in Figure 7e,f Figure 8e,f Figure 9e,f and Figure 10e,f.

The average BER values of the extracted watermarks under different attacks are listed in Table 2. It can be seen that this proposed algorithm has excellent robustness against low-pass filter with cutoff frequency four kilohertz, amplitude scaling by 0.8, amplitude scaling by 1.2, MP3 compression with 128 kbps, MP3 compression with 64 kbps, re-sampling, re-quantization, echo addition with a delay of 50 ms and a decay of five percent, noise corruption in 30 dB and noise corruption in 35 dB, so it is far superior to the algorithms proposed in papers [4,16,25]. In particular, when this algorithm resists low-pass filter, BER is only 0.01%, which is much better than 21.975% in paper [16], 6.93% in paper [19], 28.250% in paper [25], 0.12% in paper [12] and 0.39% in paper [4].

Table 2. Average BER values (%) of the extracted watermarks under different attacks.

Attack	Proposed	Paper [4]	Paper [16]	Paper [25]	Paper [12]	Paper [19]
A	0.01	0.39	21.97	28.25	0.12	6.93
B	0.01	2.87	0.50	0.30	0.12	N/A
C	0.01	17.92	0.47	0.35	N/A	N/A
D	2.27	N/A	N/A	N/A	1.29	N/A
E	0.22	N/A	N/A	N/A	0.31	N/A
F	0.07	0.78	N/A	N/A	N/A	0.00
G	0.08	1.95	2.45	6.85	0.12	0.00
H	0.01	N/A	1.12	4.97	1.61	0.00
I	0.01	0.00	1.00	6.45	0.12	0.00
J	0.14	0.78	N/A	N/A	0.12	0.00
K	0.01	N/A	N/A	N/A	0.84	N/A

Note that N/A means no report is found in the selected algorithm.

The robustness is slightly inferior to that in paper [19] when resisting re-quantization, and also slightly inferior to that in paper [12] when resisting noise corruption in 20 dB. When suffering attack from noise corruption, the quality of the extracted watermark is slightly poor, which is mainly because the algorithm is achieved by comparing two sets of TDC obtained from the fourth level detail coefficient. When the additional noise is very loud, the fourth level wavelet coefficient will be affected by noise, thus reducing the accuracy of the extracted watermark. As the noise becomes smaller, the quality of watermark is significantly improved under the noise attack of 30 and 35 dB.

6. Conclusions

In this paper, a novel and blind audio watermarking algorithm with strong robustness is proposed according to the multi-resolution characteristic of DWT and the energy compression capability of DCT. The cover audio signal is first segmented into audio fragments, and then each audio fragment is performed by DWT and DCT in order to obtain two sets of TDC which are modified to embed the binary watermark. This proposed algorithm can realize blind extraction of digital watermark without the participation of carrier audio signal when extracting watermark, which is convenient for the practical application. The embedding depth is the key factor to determine the performance of the algorithm. In the case that the algorithm has good imperceptibility, the embedding depth should be increased to improve the robustness. The experimental results show that the average SNR of the carrier audio signals reaches 23.49 dB in the case of the payload capacity of 172.27 bps, which indicates that this proposed algorithm has large capacity and good imperceptibility. In addition, this proposed

algorithm has excellent robustness against white noise, low-pass filtering, re-sampling, re-quantization, echo addition, amplitude scaling and MP3 compression compared with other audio watermarking algorithms. Without synchronization signal, the watermark extraction begins with the first audio fragment, so this algorithm cannot combat the desynchronization attack, although it can effectively resist the conventional signal processing attacks generated during the use of audio media. In the next study, we will focus on the issue of how to add synchronization signals in the carrier audio, as well as the problem of combating other types of attacks, such as desynchronization attack, collusion attack type II and strong noise corruption.

Author Contributions: Q.W. conceived the algorithm and designed the experiments; Q.W. performed the experiments; Q.W. and M.W. analyzed the results and adjusted the experimental scheme; Q.W. drafted the manuscript; Q.W. and M.W. revised the manuscript. All authors read and approved the final manuscript.

Acknowledgments: This research work was supported by the National Natural Science Foundation of China for Youth, China (Grant No. 61602263), Research and Innovation Project for Graduate Students of Jiangsu Province, China (Grant No. KYLX_0815), the University Natural Science Foundation of Jiangsu Province, China (Grant No. 16KJB510015).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shiva, M.G.; D'Souza, R.J.; Varaprasad, G. Digital signature-based secure node disjoint multipath routing protocol for wireless sensor networks. *IEEE Sens. J.* **2012**, *12*, 2941–2949. [\[CrossRef\]](#)
2. Kai, C.; Kuo, W.C. A new digital signature algorithm based on chaotic maps. *Nonlinear Dyn.* **2013**, *74*, 1003–1012.
3. Zhou, X.; Zhang, H.; Wang, C.Y. A robust image watermarking technique based on DWT, APDCBT, and SVD. *Symmetry* **2018**, *10*, 77. [\[CrossRef\]](#)
4. Wang, X.Y.; Shi, Q.L.; Wang, S.M.; Yang, H.Y. A blind robust digital watermarking using invariant exponent moments. *Int. J. Electron. Commun.* **2016**, *70*, 416–426. [\[CrossRef\]](#)
5. Wang, C.Y.; Zhang, Y.P.; Zhou, X. Robust image watermarking algorithm based on ASIFT against geometric attacks. *Appl. Sci.* **2018**, *8*, 410. [\[CrossRef\]](#)
6. Natgunanathan, I.; Xiang, Y.; Hua, G.; Beliaikov, G.; Yearwood, J. Patchwork-Based multi-layer audio watermarking. *IEEE Trans. Audio Speech Lang. Process.* **2017**, *25*, 2176–2187. [\[CrossRef\]](#)
7. Xiang, S.J.; Li, Z.H. Reversible audio data hiding algorithm using noncausal prediction of alterable orders. *EURASIP J. Audio Speech Music Process.* **2017**, *4*. [\[CrossRef\]](#)
8. Hu, H.T.; Hsu, L.Y.; Chou, H.H. Variable-dimensional vector modulation for perceptual-based DWT blind audio watermarking with adjustable payload capacity. *Digit. Signal Process.* **2014**, *31*, 115–123. [\[CrossRef\]](#)
9. Singh, D.; Singh, S.K. DWT-SVD and DCT based robust and blind watermarking scheme for copyright protection. *Multimed. Tools Appl.* **2017**, *76*, 13001–13024. [\[CrossRef\]](#)
10. Lutovac, B.; Daković, M.; Stanković, S.; Orović, I. An algorithm for robust image watermarking based on the DCT and Zernike moments. *Multimed. Tools Appl.* **2017**, *76*, 23333–23352. [\[CrossRef\]](#)
11. Shukla, D.; Sharma, M. A new approach for scene-based digital video watermarking using discrete wavelet transforms. *Int. J. Adv. Appl. Sci.* **2018**, *5*, 148–160. [\[CrossRef\]](#)
12. Hu, H.T.; Chou, H.H.; Yu, C.; Hsu, L.Y. Incorporation of perceptually adaptive QIM with singular value decomposition for blind audio watermarking. *EURASIP J. Adv. Signal Process.* **2014**, *12*, 1–12. [\[CrossRef\]](#)
13. Glowacz, A. DC Motor Fault Analysis with the Use of Acoustic Signals, Coiflet Wavelet Transform, and K-Nearest Neighbor Classifier. *Arch. Acoust.* **2015**, *40*, 321–327. [\[CrossRef\]](#)
14. Li, Z.X.; Jiang, Y.; Hu, C.; Peng, Z. Recent progress on decoupling diagnosis of hybrid failures in gear transmission systems using vibration sensor signal: A review. *Measurement* **2016**, *90*, 4–19. [\[CrossRef\]](#)
15. Glowacz, A. Diagnostics of Direct Current machine based on analysis of acoustic signals with the use of symlet wavelet transform and modified classifier based on words. *Eksplot. Niezawodn. Maint. Reliab.* **2014**, *16*, 554–558.
16. Lei, W.N.; Chang, L.C. Robust and high-quality time-domain audio watermarking based on low-frequency amplitude modification. *IEEE Trans. Multimed.* **2006**, *8*, 46–59.

17. Erfani, Y.; Siahpoush, S. Robust audio watermarking using improved TS echo hiding. *Digit. Signal Process.* **2009**, *19*, 809–814. [[CrossRef](#)]
18. Basia, P.; Pitas, I.; Nikolaidis, N. Robust audio watermarking in the time domain. *IEEE Trans. Multimed.* **1998**, *3*, 232–241. [[CrossRef](#)]
19. Kumsawat, P. A genetic algorithm optimization technique for multiwavelet-based digital audio watermarking. *EURASIP J. Adv. Signal Process.* **2010**, *1*, 1–10. [[CrossRef](#)]
20. Natgunanathan, I.; Xiang, Y.; Rong, Y.; Peng, D. Robust patchwork-based watermarking method for stereo audio signals. *Multimed Tools Appl.* **2014**, *72*, 1387–1410. [[CrossRef](#)]
21. Megias, D.; Serra-Ruiz, J.; Fallahpour, M. Efficient self-synchronized blind audio watermarking system based on time domain and FFT amplitude modification. *Signal Process.* **2010**, *90*, 3078–3092. [[CrossRef](#)]
22. Tewari, T.K.; Saxena, V.; Gupta, J.P. A digital audio watermarking algorithm using selective mid band DCT coefficients and energy threshold. *Int. J. Audio Technol.* **2014**, *17*, 365–371.
23. Hu, H.T.; Hsu, L.Y. Robust transparent and high capacity audio watermarking in DCT domain. *Signal Process.* **2015**, *109*, 226–235. [[CrossRef](#)]
24. Li, W.; Xue, X.; Lu, P. Localized audio watermarking technique robust against time-scale modification. *IEEE Trans. Multimed.* **2006**, *8*, 60–90. [[CrossRef](#)]
25. Chen, S.T.; Huang, H.N.; Chen, C.J.; Tseng, K.K.; Tu, S.Y. Adaptive audio watermarking via the optimization point of view on the wavelet-based entropy. *Digit. Signal Process.* **2013**, *23*, 971–980. [[CrossRef](#)]
26. Wu, Q.L.; Wu, M. Novel Audio information hiding algorithm based on Wavelet Transform. *J. Electron. Inf. Technol.* **2016**, *38*, 834–840.
27. Wu, S.; Huang, J.; Huang, D.; Shi, Y.Q. Efficiently self-synchronized audio watermarking for assured audio data transmission. *IEEE Trans. Broadcast.* **2005**, *51*, 69–76. [[CrossRef](#)]
28. Chen, S.T.; Wu, G.D.; Huang, H.N. Wavelet-domain audio watermarking algorithm using optimisation-based quantisation. *IET Signal Process.* **2010**, *4*, 720–727. [[CrossRef](#)]
29. Abd, F.; Samie, E. An efficient singular value decomposition algorithm for digital audio watermarking. *Int. J. Audio Technol.* **2009**, *12*, 27–45.
30. Asmara, R.A.; Agustina, R.; Hidayatulloh. Comparison of Discrete Cosine Transforms (DCT), Discrete Fourier Transforms (DFT), and Discrete Wavelet Transforms (DWT) in Digital Image Watermarking. *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*, 245–249.
31. Wang, X.; Qi, W.; Liu, P. A new adaptive digital audio watermarking based on support vector regression. *J. Netw. Comput. Appl.* **2008**, *31*, 735–749. [[CrossRef](#)]
32. Wang, X.Y.; Zhao, H. A novel synchronization invariant audio watermarking algorithm based on DWT and DCT. *IEEE Trans. Signal Process.* **2006**, *54*, 4835–4840. [[CrossRef](#)]
33. Vivekananda, B.K.; Sengupta, I.; Das, A. An adaptive audio watermarking based on the singular value decomposition in the wavelet domain. *Digit. Signal Process.* **2010**, *20*, 1547–1558.
34. Dhar, P.K.; Shimamure, T. Blind audio watermarking in transform domain based on singular value decomposition and exponential-log operations. *Radioengineering* **2017**, *26*, 552–561. [[CrossRef](#)]
35. Lei, B.Y.; Soon, I.Y.; Li, Z. Blind and robust audio watermarking algorithm based on SVD-DCT. *Signal Process.* **2011**, *91*, 1973–1984. [[CrossRef](#)]
36. Vivekananda, B.K.; Sengupta, I.; Das, A. A new audio watermarking scheme based on singular value decomposition and quantization. *Circuits Syst. Signal Process.* **2011**, *30*, 915–927.

