



A novel SVD-based adaptive robust audio watermarking algorithm

Xiangyi Liu¹ · Xiaojie Li² · Canghong Shi¹ · Xianhua Niu¹ · Ling Xiong¹

Received: 26 July 2023 / Revised: 19 September 2023 / Accepted: 19 January 2024 /
Published online: 31 January 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

To solve the copyright problem of audio data, many singular value decomposition (SVD)-based audio watermarking schemes have been proposed, however, most SVD-based schemes cannot improve the imperceptibility and robustness while guaranteeing a certain embedding capacity. Therefore, we propose a new SVD-based adaptive robust audio watermarking method. In this method, after framing the host audio signal, a discrete wavelet transform (DWT) is performed on each frame, and then the obtained DWT coefficients are divided into two segments using a sub-sampling operation, and the SVD is performed on these two segments and the mean value of the two singular values is calculated. Then the watermark bits are embedded by modifying the singular values of the two segments using differential embedding method. In the above watermark embedding process, the proposed adaptive method generates different sizes of embedding parameters according to the original signal features of each frame to minimize the degradation of perceived quality. During the watermark extraction process, the watermark can still be correctly extracted without the original audio signal and embedding parameters. The experimental results show that the scheme is more robust than existing audio watermarking schemes under various attacks with a certain embedding capacity guaranteed.

✉ Xiangyi Liu
xiangyiliu18@163.com

✉ Canghong Shi
canghongshi@163.com

Xiaojie Li
lixiaojie000000@163.com

Xianhua Niu
niuxh@mail.xhu.edu.cn

Ling Xiong
lingdonghua99@163.com

¹ School of Computer and Software Engineering, Xihua University, Chengdu 610039, 100190 Chengdu, People's Republic of China

² The College of Computer Science, Chengdu University of Information Technology, Chengdu 610225, People's Republic of China

Keywords Audio watermarking · Discrete wavelet transform · Singular value decomposition · Differential embedding · Adaptive

1 Introduction

As artificial intelligence (AI) develops at a rapid pace and apps such as ChatGPT become more popular, social issues arising from AI are beginning to emerge[1]. ChatGPT generates content that is almost "fake", which is a great challenge for intellectual property protection.

Currently, the discrimination of AI-generated content can be done mainly by two technical ways: 1) identifying the features of the content generated by the AI model through algorithms, so as to identify whether the corresponding content is generated by AI; 2) adding a specific identifier to the AI-generated content, we can distinguish whether the corresponding content is generated by AI. Therefore, digital watermarking may be an effective solution.

Digital watermarking embed some identifying information (e.g., author ID and company Logo) into a digital carrier without compromising its original value. In case of copyright disputes, the identification information extracted is used to prove the copyright ownership. Digital watermarking can be applied to images [2–4], video [5–7], audio [8–10] and other digital objects. Embedding watermarks in digital audio signals is more difficult than embedding watermarks in digital images, mainly because the human auditory system has a higher sensitivity compared to the visual system, and in this paper, audio watermarking is investigated.

The audio watermarking algorithm mainly has the following characteristics: 1) imperceptibility: the audio signal after adding the watermark is imperceptible to the human ear. Generally, subjective quality evaluation and objective quality evaluation are used to evaluate. 2) Security: The watermarked information should be secure and difficult to tamper with or forge, and only legally authorized people should be able to detect the watermark. 3) Robustness: The watermark can still be correctly extracted after some attacks. 4) Payload: The payload of an audio signal refers to the number of watermarks embedded in the signal per unit of time, usually in bits per second (bps), while satisfying certain imperceptibility and robustness.

In recent years, researchers have used different techniques for audio watermarking scheme design, such as Spread Spectrum(SS), patchwork and Singular Value Decomposition(SVD) techniques.

In the SS-based paper [8], the spread spectrum sequence is embedded as watermark bits in the audio segment, and then the spread spectrum sequence is correlated with the watermark to extract the watermark. The SS-based watermarking method has a simple structure, but there is the problem of host signal interference, which affects the correct extraction of the watermark. To solve the problem of host signal interference, researchers have designed different watermark embedding functions [9–11]. In addition, a new SS-based algorithm [12] is proposed to adaptively adjust the amplitude of PN sequences to maximize the perceptual quality according to the audio segment characteristics, and then the corresponding PN sequence is embedded into a pair of sub-segments with similar properties in the audio segment. This scheme further reduces the interference of host signals on watermark extraction and thus enhances the robustness, compared with [9–11], the algorithm in [12] has a higher embedding capacity, but it has low computational efficiency and does not eliminate the interference problem of host signals. Based on [12], a new watermarking scheme[13] is proposed by the same author, which makes the host signal and the watermark component have the same polarity in extracting the watermark, thus eliminating the host signal interference. Compared with [12], this method has lower computational complexity. However, many audio watermarking methods based on SS have limited embedding capacity.

Audio watermarking algorithms based on patchwork have also received widespread attention. In paper [14], the authors divide the DCT coefficients into multiple pairs of frames and then select the appropriate DCT frame pairs using the proposed criterion, and finally embed the watermark into these pairs of frames by changing the relevant DCT coefficients under the control of the PN sequence. The selection criteria used by the algorithm in the embedding phase can also be used in the decoding phase. Based on [14], a watermarking scheme that can prevent de-synchronization attacks based on patchwork is proposed in paper [15]. The watermark is first inserted into the DCT coefficients of the audio signal, and subsequently, a set of synchronization sequences are embedded into the logarithmic DCT (LDCT) coefficients. When extracting the watermark, the position of synchronization bit in LDCT domain is analyzed to determine whether the received signal suffers from a de-synchronization attack. A multi-layer watermarking scheme based on patchwork technique is proposed in Paper [16]. The method can repeatedly embed the watermark by overlay in any order, without affecting the watermarking in other layers. The proposed DCT coefficient ordering ensures good imperceptibility, and the introduction of error buffer improves the algorithm's ability to withstand various conventional attacks. Liu et al. [17] proposed a new audio signal feature. Through analysis, the author concluded that the residuals of two groups of the frequency-domain coefficient logarithmic mean (FDLM) features are very resistant to some audio attacks. The residuals of FDLM features of two adjacent groups are represented as RFDLM features, and the watermark is embedded into RFDLM features to obtain watermarked signals.

The SVD technique is widely utilized by audio watermarking researchers because of the stability of its singular values after being attacked. In paper [18], an SVD-based adaptive audio watermarking scheme is proposed, which adaptively embeds the watermark into the singular values of each wavelet block by quantization index modulation (QIM). Although this adaptive method ensures the perceptual transparency of the watermarking scheme, it is not robust enough against echo addition and resampling attacks and has a low watermark capacity (45.9 bps). In paper [19], a watermarking algorithm based on SVD is proposed. After LWT/DWT transform and DCT transform are performed on the host signal to get the DCT coefficient, SVD operation is further performed to obtain the singular value, and the singular values are embedded in the watermark by adaptive DM quantization. The method enhances the robustness of the algorithm by the SVD technique and effectively solves the conflict between robustness and imperceptibility by DE optimization. A new SVD watermarking scheme based on entropy and log-polar transform (LPT) is proposed in [20]. The method segments the low-frequency DCT coefficients of each frame and calculates the entropy value of each segment, selects the DCT segment with the largest entropy value, and then embeds the watermark into the Cartesian component with the largest singular value of the segment. This scheme has high embedding capability, but is weak against resampling and MP3 compression attacks. In Paper [21], an SVD-based dual-domain watermarking scheme is proposed. After segmenting the original audio and dividing each segment into equal length frames, the most energetic voiced frame is extracted. In this voiced frame, a time-domain implicit synchronization mechanism (ISM) method is proposed to search for a suitable embedding region, and DCT and SVD are applied to this region. Finally, watermark embedding is achieved by quantizing the obtained singular values. In paper [22], the authors propose a new watermarking scheme based on the ratio of frequency singular value coefficients. The algorithm divides each frame signal into two segments and performs DCT on both segments separately, after selecting the DCT coefficients of the mid-frequency segment for SVD processing and then modifies the ratio of the singular values of the two segments to embed the watermark bits.

Previous watermarking algorithms based on SVD usually modify only a single singular value to embed the watermark [18–21] or embed the ratio of two singular values [22], and

do not guarantee high imperceptibility and robustness with a certain capacity. However, we propose a new watermarking scheme based on SVD, which utilizes the mean of two singular values for algorithm design to improve the robustness of the scheme, and an adaptive method to maximize the perceptual quality while still having good embedding capacity.

In this scheme, the original signal is firstly divided into frames and each frame is transformed by three-layer DWT to obtain the corresponding DWT coefficient. Second, the DWT coefficients of each frame are sub-sampled to obtain two equal-length coefficient vectors, and the average of the two singular values is calculated after the corresponding singular values are obtained by performing SVD operations on the two vectors. Third, the obtained mean value and the adaptively generated embedding parameters are used to modify the two singular values to realize watermark embedding.

This paper has the following advantages:

1. Previous SVD-based algorithms modify only one singular value or the ratio of two SVD segments. We use the mean value of two singular values for watermark embedding algorithm design, which makes the algorithm have good robustness.
2. The proposed adaptive watermark embedding method can generate different size embedding parameters according to different characteristics of each frame signal, which helps to improve the perceptual quality.
3. This scheme has no need for the original audio signal and embedding parameters during watermark extraction.

The rest of the paper is summarized below. Section 2 provides a preliminary introduction to DWT and SVD. Section 3 presents a detailed description of the proposed audio watermarking algorithm. In Section 4, the simulation results of the proposed algorithm are demonstrated. Finally, Section 5 gives the conclusion.

2 Preliminaries

2.1 Discrete wavelet transform (DWT)

DWT is a signal analysis method with multi-resolution capability in temporal and frequency domain, which has been widely used in digital watermarking direction. For a one-dimensional signal, the signal S is passed through a low-pass filter G and a high-pass filter H , respectively, and then a down-sampling operation can be performed to obtain the approximation coefficients cA^1 (low frequency) and the detail coefficients cD^1 (high frequency), respectively, as shown in Fig. 1, where n denotes the number of sampling points. After decomposing to obtain the approximate coefficients cA^1 and the detail coefficients cD^1 , cA^1 can be further decomposed to approximation component cA^2 and the detail component cD^2 . Similarly, cA^2 can be decomposed into approximation coefficients cA^3 and detail coefficients cD^3 . Figure 2 illustrates the three-level DWT decomposition. By continuously decomposing the approximate components, the signal can be decomposed into many low-resolution components.

In this paper, the decomposition level of DWT is determined by the frequency f_s , because after n -order wavelet decomposition of audio with sampling frequency f_s , the frequency range of wavelet approximation coefficient is $[0 - f_s/2^n]$ [23], while the frequency distribution of audio signals that can be perceived by the human ear is between 2-5kHz. For an audio signal with a sampling frequency of 44.1kHz, the approximation coefficient will contain the most audio energy when a 3-layer wavelet transform is applied. We also tested several commonly used wavelets, such as "haar", "db" and "mexh", and the different wavelet bases

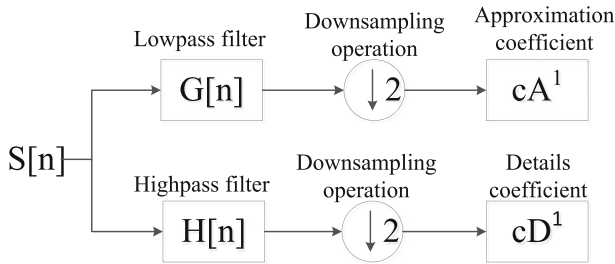


Fig. 1 One level DWT decomposition

had no significant effect on the experimental performance. When the reader has other data sets, the haar wavelet can meet our needs. Therefore, we use the "haar" wavelet and choose $n=3$ for the experiment.

2.2 Singular value decomposition (SVD)

Singular Value Decomposition provides a very convenient way of matrix decomposition that can be used to extract the essential information from digital signals. For any $m \times n$ size matrix A , we define its SVD transformation as:

$$A = U \Sigma V^T \tag{1}$$

Where U and V are unitary matrices of size $m \times m$ and $n \times n$, respectively, i.e., satisfying $U^T U = I, V^T V = I$. Σ is a matrix of size $m \times n$ with all zeros except on the main diagonal, where the numbers on the main diagonal are called singular values, and the superscript T denotes the transpose operation.

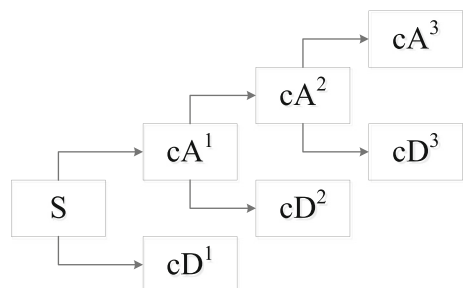
3 Proposed scheme

3.1 Watermark encryption

Considering the security requirements of the watermarking algorithm, we use Logistic chaotic mapping to encrypt the watermarking image. Equation (2) gives the definition of Logistic:

$$y(i + 1) = \mu \times y(i)[1 - y(i)], \quad 1 < i < m \times n \tag{2}$$

Fig. 2 3 level DWT decomposition



where $y(i)$ is the i -th sample point in y , $m \times n$ is the size of the watermark image. There are three main initial logical chaotic mapping parameters, y_0 , μ and T , and the logistic mapping of states is chaotic when the mapping equations satisfy the two conditions $0 < y_0 < 1$ and $3.5699456 < \mu \leq 4$. So in this paper, we randomly set $y_0 = 0.78$ and $\mu = 3.5821656$. With y_0 , μ and formula 2, obviously, we have $y(i) \in (0, 1)$.

Then using (3) we can get a binary sequence $z(i)$.

$$z(i) = \begin{cases} = 0, & y(i) < T \\ = 1, & otherwise \end{cases} \quad 1 < i < m \times n \tag{3}$$

where $z(i)$ is the i -th sample point in z . T is a predefined threshold, and from (3), we can see that the range of the threshold value T must be in the range of $(0,1)$, so we randomly select the value of T , here $T = 0.6$. In this paper, y_0 , μ and T are used as keys to ensure the security of the proposed method.

Convert a binary watermark image I to a one-dimensional sequence w , $w = \{w(i), 1 \leq i \leq m \times n\}$, finally, $w(i)$ is encrypted with $z(i)$ to obtain the encrypted watermark by the following formula:

$$\bar{w}(i) = w(i) \oplus z(i) \tag{4}$$

where $\bar{w}(i)$ is the i -th encrypted watermark bit and \oplus is the XOR operation.

3.2 Watermark embedding

Let I be a gray scale image of size $m \times n$. The scrambled binary sequence \bar{w} of size L_w is used as a watermark for watermark embedding. The watermark embedding flowchart is shown in Fig. 3, and the detailed steps of the embedding process are as follows:

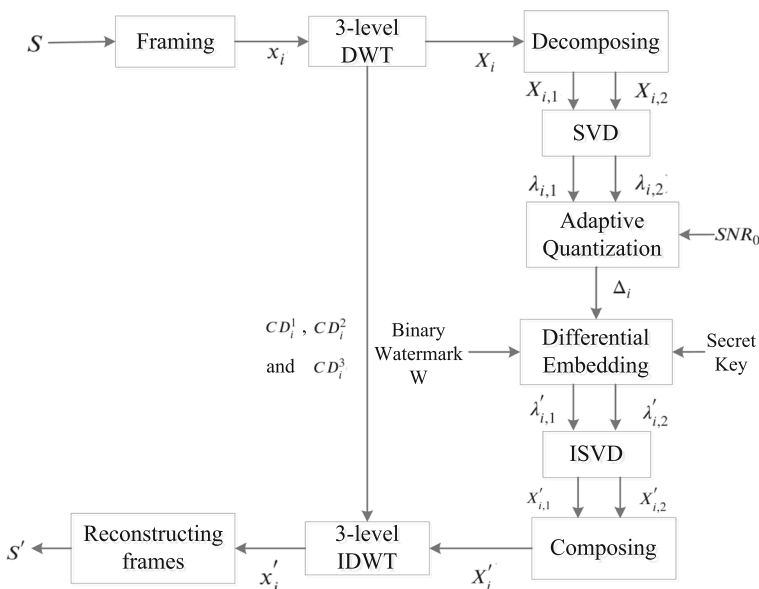


Fig. 3 Watermark embedding process

- (1) The original speech signal $S(n)$, ($1 \leq n \leq L$) is decomposed into frames, where L is the length of the audio. To facilitate the calculation, each frame length l_s is calculated according to (5).

$$l_s = 16 * \lfloor L / (16 * N) \rfloor \tag{5}$$

where the symbol $\lfloor \cdot \rfloor$ is the round down operation and N is the number of frames.

- (2) For each frame $x_i(n)$ ($1 \leq i \leq N, 1 \leq n \leq l_s$), a 3-level DWT transform is applied to generate the approximation coefficients ($CA_i^3(j)$) and a set of detail coefficients ($CD_i^3(j), CD_i^2(j), CD_i^1(j)$), ($1 \leq i \leq N, 1 \leq j \leq l_s / (2^3)$). To ensure that the algorithm is resistant to various attacks, we choose the high-energy approximation coefficient $CA_i^3(j)$ for the watermark embedding, naming $CA_i^3(j)$ as X_i and performing the following operations.
- (3) The approximate coefficient vector $X_i(j)$, ($1 \leq i \leq N, 1 \leq j \leq l_s / (2^3)$), of each frame is decomposed into two related sub-vectors $X_{i,1}$ and $X_{i,2}$, using the following sub-sampling operation:

$$\begin{cases} X_{i,1}(k) = X_i(2k - 1); \\ X_{i,2}(k) = X_i(2k). \end{cases} \tag{6}$$

where $k = 1, \dots, l_s / (2^4)$.

- (4) Performing SVD on $X_{i,1}$ and $X_{i,2}$, we can obtain

$$X_{i,j} = U_{i,j} \Lambda_{i,j} V_{i,j}^T \tag{7}$$

Where $X_{i,j}$ denotes the approximate coefficient of the j -th sub-vector of frame i , ($1 \leq i \leq N, j = 1, 2$), $U_{i,j}$ and $V_{i,j}$ are unitary matrices, $\Lambda_{i,j}$ is a diagonal matrix, and the superscript T denotes transpose operation. The values on the main diagonal of $\Lambda_{i,j}$ are singular values. since each frame of the signal is a one-dimensional vector, $\Lambda_{i,j}$ will have only one singular value. We denote the singular values of each frame sub-vector as $\lambda_{i,1}, \lambda_{i,2}$.

- (5) $\bar{w}(i)$ ($1 \leq i \leq N$) represents the i -th scrambled watermark bit, and the singular values $\lambda_{i,1}, \lambda_{i,2}$ of each frame sub-vector are embedded in the watermark using the differential embedding technique [24]. Then two modified singular values $\lambda'_{i,1}, \lambda'_{i,2}$ are obtained and the embedding rules are shown below: When $\bar{w}(i) = 1$, execute (8)

$$\begin{cases} \lambda'_{i,1} = \frac{1}{2}(\lambda_{i,1} + \lambda_{i,2}) + \Delta_i, \\ \lambda'_{i,2} = \frac{1}{2}(\lambda_{i,1} + \lambda_{i,2}) - \Delta_i. \end{cases} \tag{8}$$

When $\bar{w}(i) = 0$, execute (9)

$$\begin{cases} \lambda'_{i,1} = \frac{1}{2}(\lambda_{i,1} + \lambda_{i,2}) - \Delta_i, \\ \lambda'_{i,2} = \frac{1}{2}(\lambda_{i,1} + \lambda_{i,2}) + \Delta_i. \end{cases} \tag{9}$$

where Δ_i denotes the embedding parameter of the i -th frame, which is calculated by the adaptive method proposed in this paper, which will be given in Section 3.4

- (6) After obtaining the modified singular values, the original singular values $\lambda_{i,1}, \lambda_{i,2}$ are replaced by the modified singular values $\lambda'_{i,1}$ and $\lambda'_{i,2}$ to obtain the modified

diagonal matrices $\Lambda'_{i,1}$ and $\Lambda'_{i,2}$. Then the sub-vector $X'_{i,j}$ of the modified approximate coefficient of i -frame is obtained by SVD inverse operation:

$$X'_{i,j} = U_{i,j} \Lambda'_{i,j} V_{i,j}^T \tag{10}$$

(7) The two modified sub-vectors $X'_{i,1}$ and $X'_{i,2}$ are connected using the inverse of the sub-sampling described above to generate the modified approximate coefficient vector X'_i for the i -th frame.

$$\begin{cases} X'_i(2k - 1) = X'_{i,1}(k) \\ X'_i(2k) = X'_{i,2}(k) \end{cases} \tag{11}$$

where $k = 1, \dots, ls/(2^4)$.

(8) The i -th modified frame x'_i is obtained by performing the inverse 3-level DWT transform on the modified approximation components X'_i and the original set of detail components.

(9) Connect all modified frames x'_i to get watermarked audio S' .

3.3 Watermark extraction

It should be noted that this scheme does not require the original audio signal nor embedding parameters in the watermark extraction process. The watermark extraction process is shown in Fig. 4, and the specific implementation process is as follows:

We let S^* represent the audio signal with watermark, through the previous four steps in Section 3.2, we can get the singular value $\lambda^*_{i,1}$, $\lambda^*_{i,2}$ of the two sub-vectors of frame i and calculate the difference between them:

$$d(i) = \lambda^*_{i,1} - \lambda^*_{i,2} \tag{12}$$

Then the i -th encrypted watermark bit can be extracted using the following equation:

$$\begin{cases} \hat{w}(i) = 1, \text{ if } d(i) \geq 1; \\ \hat{w}(i) = 0, \text{ otherwise.} \end{cases} \tag{13}$$

$\hat{w}(i)$ is the encrypted watermark bits extracted from the i -th frame. After getting the encrypted watermark $\hat{w}(i)$, use formula (2) and formula (3), and the secret keys y_0, μ, T to

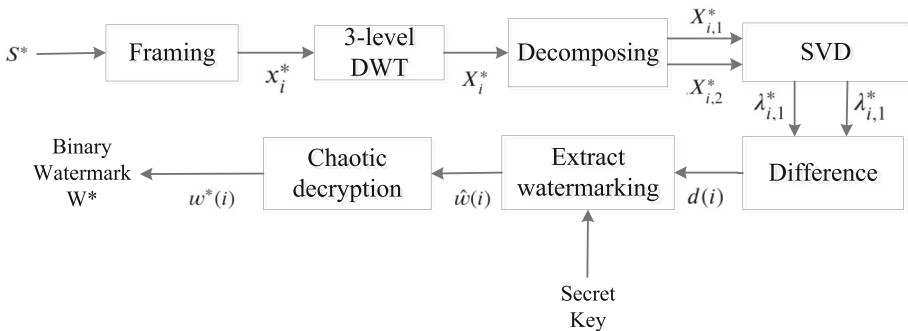


Fig. 4 Watermark extracting process

generate the binary sequence $z(i)$. Finally, the encrypted watermark sequence is decrypted by (14):

$$w^*(i) = z(i) \oplus \hat{w}(i) \tag{14}$$

Finally, the binary sequence $w^*(i)$ is converted to a matrix of size $m \times n$ to obtain the watermarked image.

3.4 Adaptive quantization method

Most existing audio watermarking schemes use fixed quantization parameters without considering the characteristics of the audio signal, which degrades the perceptual quality to some extent. The proposed adaptive method uses SNR value and the characteristics of each frame audio signal to obtain different sizes of quantization values, which improves the perceptual quality.

$$SNR = 10 \log_{10} \left(\frac{\sum_{i=1}^L s^2(i)}{\sum_{i=1}^L (s(i) - s'(i))^2} \right) \tag{15}$$

where $s(i)$ and $s'(i)$ represent the original audio signal and the watermarked audio signal, respectively.

Let the initial SNR be denoted as SNR_0 , and after the 3-level DWT, the energy change satisfies the following equation:

$$SNR_0 = 10 \log_{10} \left(\frac{\sum_{i=1}^{L/8} c^2(i)}{\sum_{i=1}^{L/8} (c(i) - c'(i))^2} \right) \tag{16}$$

where $c(i)$ and $c'(i)$ denote the approximate components of the original and embedded watermarked audio signals, respectively.

After sub-sampling to obtain the two-segment vector and performing the SVD operation on the two-segment vector, according to the conclusion in paper [22], the power of the coefficient vector is equal to the square of the singular value, so we can obtain:

$$SNR_0 = 10 \log_{10} \left(\frac{\lambda_{i,j}^2}{(\lambda_{i,j} - \lambda'_{i,j})^2} \right) \tag{17}$$

where $\lambda_{i,j}$ and $\lambda'_{i,j}$ represent the original and watermarked singular values of the j -th sub-vector of frame i , respectively

According to the embedding rules above, for the watermark to be extracted correctly, we need to ensure that:

$$|\lambda_{i,j} - \lambda'_{i,j}| \leq \Delta_i \tag{18}$$

where Δ_i denotes the embedding parameter of the i -th frame.

Using (17) and (18), we can obtain:

$$10^{-SNR_0/10} * \lambda_{i,j}^2 = (\lambda_{i,j} - \lambda'_{i,j})^2 \leq \Delta_i^2 \tag{19}$$

From (19), we can obtain:

$$\sqrt{10^{-SNR_0/10} * \lambda_{i,j}^2} \leq \Delta_i \tag{20}$$

It can be seen from the above discussion that SNR_0 is used to control the noise of watermark, and the quantization step of each frame signal is generated adaptively according to the

characteristics of each frame signal to maximize the perceived quality. In the embedding process, the specific adaptive formula is as follows:

$$\Delta_i = \sqrt{10^{-SNR_0/10} * (\lambda_{i,1}^2 + \lambda_{i,2}^2)} \quad (21)$$

where $\lambda_{i,1}$ and $\lambda_{i,2}$ are the singular values of the two sub-vectors of the i -th frame, respectively, and in the experiment we set $SNR_0 = 30$.

4 Experimental results

The experiments were executed on a Windows 10 laptop with Intel Core-i5-8250 U 1.60 GHz processor and 8.00 GB RAM, MATLAB 2021a and Adobe Audition CC 2018 were used to implement the proposed watermarking algorithm and attack the watermarked signal. We downloaded different audio signals including dance, folk, pop, English, French and Spanish from the Internet to simulate the performance of the proposed scheme. We changed the stereo to mono and cropped the selected audio file into an audio segment of about 10s, sampled at 44.1 kHz, and quantized using 16 bits in waveform format. The experiments use a binary image of size $32 \times 32 = 1024$ bits as a watermark embedded in the audio signal. The initial logistic chaotic mapping parameters y_0, μ, T are 0.78, 3.5821656, 0.6, respectively. The quantization step is determined by the proposed adaptive method and the level of DWT is 3.

4.1 Data payload

The data payload of an audio watermarking algorithm is defined as the number of bits per unit time embedded in the original signal, usually in bits per second (bps) units, as shown in (22):

$$P = \frac{B}{D} \quad (22)$$

P represents the watermark capacity, B represents the number of bits embedded into the original audio signal, and D represents the time duration of the original audio signal. The duration of the original audio used in the proposed scheme is 10s, and the embedded watermark image size is $32 \times 32 = 1024$ bit, so the watermark payload of the proposed scheme is 102.4bps, which meets the requirements of IFBI (Watermark capacity over 20 bps)[25].

4.2 Imperceptibility

Imperceptibility refers to the fact that the watermark is embedded in the original signal and is not easily perceived, i.e., it does not impair the quality of the original audio. In the experiments, we use both subjective quality evaluation and objective quality evaluation to assess the quality of the audio signal with watermark.

- i) Subjective Difference Grade (SDG) is a common subjective evaluation metric in which the tester is provided with both the original and watermarked audio for listening, and the testers are asked to rate the degree of difference between the two. As shown in Table 1, the SDG is generally divided into 5 levels, and the closer the SDG value is to 0, the smaller the difference between the audio before and after embedding the watermark, and the better the imperceptibility.

Table 1 Subjective and objective different grades

SDG	ODG	Description	Quality
0	0	Imperceptible	Excellent
-1	-1	Perceptible but not annoying	Good
-2	-2	Slightly annoying	Fair
-3	-3	Annoying	Poor
-4	-4	Very annoying	Bad

ii) Signal to Noise Ratio (SNR) is one of the objective evaluation criteria commonly used in audio watermarking technology, which is defined in (1).

$$SNR = 10 \log_{10} \left(\frac{\sum_{i=1}^L s^2(i)}{\sum_{i=1}^L (s(i) - s'(i))^2} \right) \quad (23)$$

Where $s(i)$ and $s'(i)$ represent audio signals before and after watermark embedding respectively. SNR reflects the overall distortion of watermarked audio. The larger the value, the smaller the distortion of audio signal and the better the imperceptibility of watermark.

iii) The Objective Difference Grade (ODG) is obtained through the Perceptual Evaluation of Audio Quality (PEAQ) [26, 27] algorithm using artificial neural networks, which takes into account the characteristics of the HAS (Human Auditory System) and compensates for the shortcomings of SNR which ignores the auditory perception of the human ear. The ODG value is between -4 and 0, and the closer its value is to 0 means that the algorithm has better transparency, which is shown in Table 1. We used a Matlab program published by the TSP Lab at McGill University[28] to implement the PEAQ metrics for imperceptibility testing.

Table 2 gives the measured values of different evaluation metrics for different types of audio signal. From the figure, we can see that all audio signals have the SNR values greater than 20db and can reach up to 26.9154db, meeting the requirements of IFPI. The ODG values are all greater than -1, indicating that the proposed scheme exhibits good imperceptibility. Furthermore, the SDG is close to 0, indicating that the testers had a difficult perceptual distinction between original and watermarked audio.

Figures 5 and 6 show the time domain representations of different types of original audio signals and watermarked audio signals and the differences between them, respectively. As can be seen from these two graphs, the waveform plots of the original and watermarked signals

Table 2 SNR, SDG and ODG for different audio signals using the proposed algorithm

Type	Audio	SNR(dB)	ODG	SDG
Music	Dance	24.3454	-0.3825	-0.1
	Folk	26.9154	-0.6969	-0.4
	Pop	24.6004	-0.6769	-0.3
Speech	English	24.8237	-0.5634	-0.3
	French	25.5323	-0.6601	-0.1
	Spanish	23.0419	-0.6461	-0.2

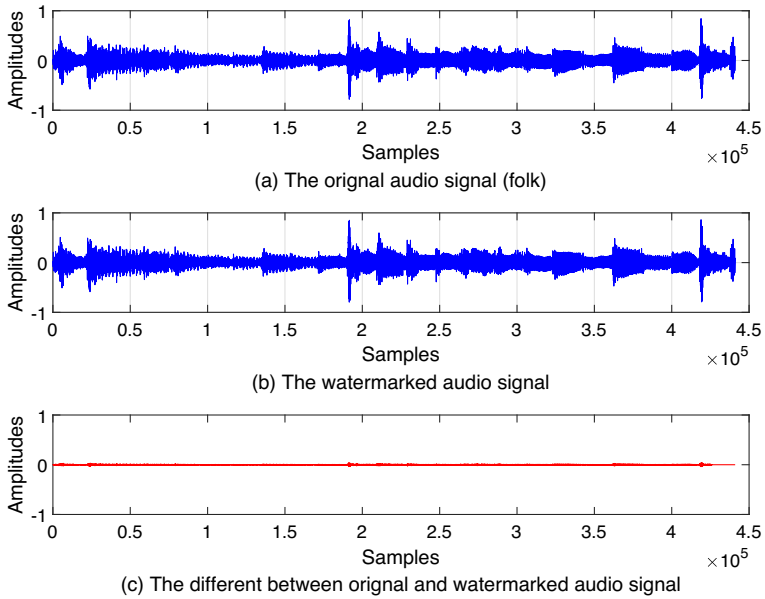


Fig. 5 Original, watermarked and difference waveforms of audio signal "folk"

are very similar with very small differences, that the human eye can barely distinguish them, indicating the good transparency of our algorithm.

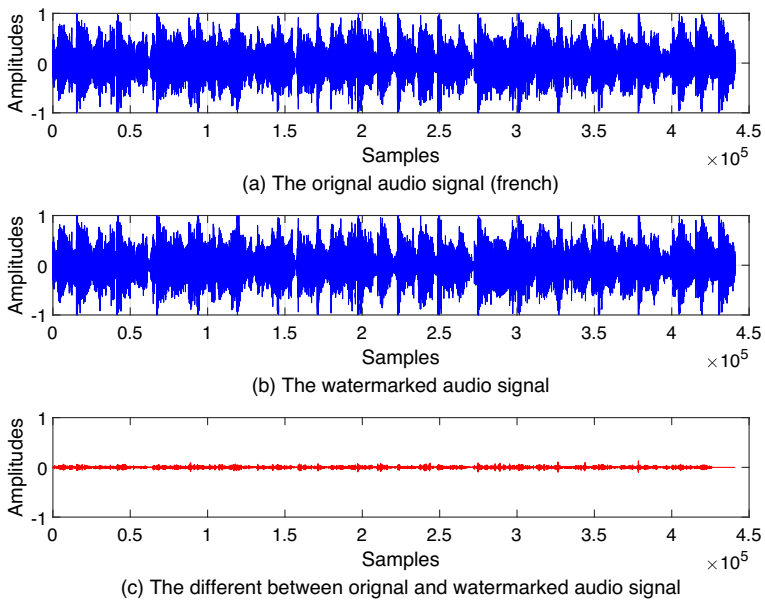


Fig. 6 Original, watermarked and difference waveforms of audio signal "french"

4.3 Robustness

Robustness is used to evaluate the ability of the watermarking algorithm to extract the watermark correctly even after it has been subjected to different attacks. Bit Error Rate (BER) is one of the evaluation criteria for robustness of audio watermarking algorithms and its definition is as follows:

$$BER = \frac{B_{ERR}}{B_T} \times 100\% \quad (24)$$

B_{ERR} indicates the number of bits extracted incorrectly and B_T indicates the total number of bits embedded. The smaller value of BER indicates the better attack resistance of the proposed scheme.

Researchers commonly use Normalized Cross-Correlation (NC) to measure the similarity between the original watermarked image and the extracted watermarked image. Since a binary image is embedded as a watermark in this experiment, we use the NC value as another indicator of robustness, which is defined as follows:

$$NC = \frac{\sum_{i=1}^m \sum_{j=1}^n W(i, j) W^*(i, j)}{\sqrt{\sum_{i=1}^m \sum_{j=1}^n W^2(i, j)} \sqrt{\sum_{i=1}^m \sum_{j=1}^n W^{*2}(i, j)}} \quad (25)$$

where $m \times n$ is the size of the watermark. $W(i, j)$ and $W^*(i, j)$ respectively represent the original and extracted watermark images.

We use some common signal processing operations as shown below to attack the watermarked signals:

- AWGN: White Gaussian noise with a signal-to-noise ratio of 30dB is added to the watermarked signal.
- Re-sampling: The watermarked audio signal is first down-sampled to half the original sampling rate, and then up-sampled to the original sampling rate.

Table 3 The BER values and the NC values under different attacks (the test signals are music)

Attack type	BER(%)			NC		
	Dance	Folk	Pop	Dance	Folk	Pop
No attack	0	0	0	1	1	1
AWGN	1.56	2.15	1.66	0.9856	0.9802	0.9850
Re-sampling	0	0	0.29	1	1	0.9973
Re-quantization	0.09	0	2.93	0.9991	1	0.9741
echo	2.93	0	1.46	0.9728	1	0.9866
Amplitude +10%	0	0	0.29	1	1	0.9973
Amplitude -10%	0	0	0.2	1	1	0.9982
Amplitude +20%	0	0	0.29	1	1	0.9973
Amplitude -20%	0	0	0.2	1	1	0.9982
Mp3 128kbps	0	0	0.2	1	1	0.9982
Mp3 64kbps	0	0	0.2	1	1	0.9982
AAC 128kbps	0	0	0.2	1	1	0.9982
Cropping 3000	0	0	0.29	1	1	0.9973

Table 4 The BER values and the NC values under different attacks (the test signals are speech)



























Attack type	BER(%)			NC		
	English	French	Spanish	English	French	Spanish
No attack	0	0	0	1	1	1
AWGN	0.09	0.39	0	0.9991	0.9964	1
Re-sampling	0	0	0	1	1	1
Re-quantization	0	0	0	1	1	1
echo	1.46	3.32	0.98	0.9866	0.9692	0.9910
Amplitude +10%	0	0	0	1	1	1
Amplitude -10%	0	0	0	1	1	1
Amplitude +20%	0	0	0	1	1	1
Amplitude -20%	0	0	0	1	1	1
Mp3 128kbps	0	0	0	1	1	1
Mp3 64kbps	0	0	0	1	1	1
AAC 128kbps	0	0	0	1	1	1
Cropping 3000	0.09	0	0.09	0.9991	1	0.9991

- Re-quantization: Quantize each watermark signal sample from 16 bits to 8 bits, and then to 16 bits again.
- Amplitude scaling: Scale the amplitude of the watermarked audio signal by $\pm 10\%$ and $\pm 20\%$.
- Echo addition: Add the echo signal of the host signal to the watermark signal with an amplitude of 20% and a delay of 0.5s.
- MP3 compression: The watermarked signal is compressed by MPEG-1 Layer-III, and its format is converted from wav to mp3 and back to wav with compressed bit rates are 128kbps and 64kbps respectively.
- AAC attack: The watermarked signals are compressed using a compression technique based on MPEG-4 advanced audio coding with a compression bit rate of 128 kbps.
- Cropping: The number of samples in the watermarked audio is randomly set to zero.

Tables 3 and 4 show the BER and NC values of different audio signals after the attacks. As can be seen from the figure, the proposed algorithm is less robust to AWGN and echo attacks, but the maximum BER value is 2.93% and the minimum NC value is 0.9728 in Table 3 and the maximum BER value is 3.32% and the minimum NC value is 0.9692 in Table 4, which is far below the requirement of IFPI (BER less than 20%). While under other attacks, most of the BER values are 0 and NC values are 1, indicating that the watermark can be extracted accurately and without errors. Especially, it is more robust to resampling, re-quantization, amplitude scaling, and Mp3 compression attacks.



























Tables 5, 6 and 7 show the robustness of the proposed scheme in comparison with the schemes in in [29, 30] and [31] for different types of audio signals subjected to different attacks, respectively, where the symbol "-" indicates that the experimental results under these attacks do not satisfy the IFPI criterion (BER over 20%) and the corresponding failed NC values. It can be seen from Table 5 that the proposed scheme is slightly less robust than scheme [29] under AWGN attack. However, the scheme in scheme [29] is less robust under echo attack and fails completely in the extraction under amplitude scaling attack. In contrast, the proposed scheme has a large number of NC values close to 1 and the extracted

Table 5 Comparison of the robustness of the proposed scheme and the scheme in reference [29] for 'dance'.

Attack type	BERs of(%)		NCs of		Detected watermark	
	[29]	Proposed	[29]	Proposed	[29]	Proposed
No attack	0	0	1	1		
AWGN	0.29	1.56	0.9973	0.9856		
Re-sampling	0	0	1	1		
Re-quantization	0	0.09	1	0.9991		
echo	19.63	2.93	0.8112	0.9728		
Amplitude +10%	-	0	-	1		
Amplitude -10%	-	0	-	1		
Amplitude +20%	-	0	-	1		
Amplitude -20%	-	0	-	1		
Mp3 128kbps	0	0	1	1		
Mp3 64kbps	0	0	1	1		
AAC 128kbps	0	0	1	1		
Cropping 3000	0.09	0	0.9991	1		



























watermarked image can still identify the original shape. In Table 5, the signal 'dance' with the lowest snr in the type of music with watermark is selected for testing, and its SNR value is 24.3454dB. Experimental results show that the test effect of the proposed scheme is still better than that of the comparative literature under low SNR. In Table 6 and 7, the BER of the proposed scheme is almost all zero, and the maximum BER value are less than 4% and 1%, respectively, while the original shape is basically invisible under echo attack in scheme [30] and under AWGN and echo attack in scheme [31], which indicates that the proposed scheme is more robust compared to scheme [30] and [31].

Table 6 Comparison of the robustness of the proposed scheme and the scheme in reference [30] for 'French'

Attack type	BERs of(%)		NCs of		Detected watermark	
	[30]	Proposed	[30]	Proposed	[30]	Proposed
No attack	0	0	1	1		
AWGN	0.09	0.39	0.9991	0.9964		
Re-sampling	0	0	1	1		
Re-quantization	0	0	1	1		
echo	-	3.32	-	0.9692		
Amplitude +10%	0	0	1	1		
Amplitude -10%	0	0	1	1		
Amplitude +20%	0	0	1	1		
Amplitude -20%	0	0	1	1		
Mp3 128kbps	0	0	1	1		
Mp3 64kbps	0	0	1	1		
AAC 128kbps	0	0	1	1		
Cropping 3000	1.17	0	0.9891	1		

Tables 8 and 9 show the average BER and average N values of the proposed algorithm and the compared algorithms under different attacks. The data in Table 8 shows that the proposed scheme is not well resistant to AWGN, re-quantization and echo attacks with the maximum BER value of 1.79%. However, the BER values under AWGN attack are also much better than those of methods [30, 31] and [32]. While the compared scheme [30] are not robust enough against the echo attacks, schemes [29] and [32] even fails completely under the magnitude scaling and echo attacks ,as for scheme [31], it is completely unable to resist AWGN, resampling and echo attacks. Also from Table 9, it can be seen that methods [29] and [32] are very vulnerable to amplitude scaling and echo attacks. As for method [30] and

Table 7 Comparison of the robustness of the proposed scheme and the scheme in reference [31] for 'Spanish'.

Attack type	BERs of(%)		NCs of		Detected watermark	
	[31]	Proposed	[31]	Proposed	[31]	Proposed
No attack	0	0	1	1		
AWGN	-	0	-	1		
Re-sampling	-	0	-	1		
Re-quantization	-	0	-	1		
echo	-	0.98	-	0.9910		
Amplitude +10%	0.78	0	0.9928	1		
Amplitude -10%	0.20	0	0.9982	1		
Amplitude +20%	5.96	0	0.9438	1		
Amplitude -20%	0.20	0	0.9982	1		
Mp3 128kbps	0.2	0	0.9982	1		
Mp3 64kbps	0.2	0	0.9982	1		
AAC 128kbps	0.2	0	0.9982	1		
Cropping 3000	0.78	0.09	0.9928	0.9991		

[31], it also fails to resist the echo attack because its BER is also greater than 20% and the extracted watermark image is basically unrecognizable. In contrast, the BER of the proposed method is mostly 0, and the maximum is 1.92%, while methods [29, 30] and [32] are very susceptible to amplitude scaling and echo attacks, method [31] is even susceptible to AWGN, resampling, and re-quantization attacks. In order to test the effect of audio sampling rate on the proposed method, we modified the "folk" signal with a sampling rate of 16k for the test. As mentioned in Section 2.1, the number of dwt decomposition layers in the proposed method depends on the sampling frequency, so when the signal with a sampling frequency of 16k is used, the number of decomposition layers in the proposed method becomes 2. Table 10

Table 8 Robustness comparisons between the proposed algorithm and the algorithms in reference [29–32] for music signals

Attack type	Average BER(%)				Average NC				Proposed	
	[29]	[30]	[31]	[32]	Proposed	[29]	[30]	[31]		[32]
No attack	0	0	0	0	0	1	1	1	1	1
AWGN	0.35	9.18	-	17.29	1.79	0.9964	0.9144	-	0.8452	0.9836
Re-sampling	0	0.98	19.95	0.55	0.10	1	0.9913	0.7933	0.9949	0.9991
Re-quantization	0	0	-	0.81	1.01	1	1	-	0.9925	0.9911
echo	-	-	-	-	1.46	-	-	-	-	0.9865
Amplitude +10%	-	0.72	0.81	-	0.10	-	0.9936	0.9918	-	0.9991
Amplitude -10%	-	0.81	0.35	18.62	0.07	-	0.9927	0.9967	0.8249	0.9994
Amplitude +20%	-	0.72	1.82	-	0.10	-	0.9936	0.9831	-	0.9991
Amplitude -20%	-	0.62	0.35	-	0.07	-	0.9944	0.9967	-	0.0667
Mp3 128kbps	0	0.71	0.35	0	0.07	1	0.9936	0.9967	1	0.9994
Mp3 64kbps	0	0.71	0.35	0	0.07	1	0.9936	0.9967	1	0.9994
AAC 128kbps	0	0.71	0.35	0	0.07	1	0.9936	0.9967	1	0.9994
Cropping 3000	0.16	1.82	0.95	1.17	0.10	0.9985	0.9833	0.9913	0.9891	0.9991

Table 9 Robustness comparisons between the proposed algorithm and the algorithms in reference [29–32] for speech signals

Attack type	Average BER(%)				Average NC					
	[29]	[30]	[31]	[32]	Proposed	[29]	[30]	[31]	[32]	Proposed
No attack	0	0	0	0	0	1	1	1	1	1
AWGN	0	2.21	-	15.40	0.16	1	0.9796	-	0.8565	0.9985
Re-sampling	0	0.13	-	0.75	0	1	0.9988	-	0.9931	1
Re-quantization	0	0	-	0.4533	0	1	1	-	0.9958	1
echo	-	-	-	-	1.92	-	-	-	-	0.9834
Amplitude +10%	-	0	3.42	-	0	-	1	0.9684	-	1
Amplitude -10%	-	0	1.44	-	0	-	1	0.9868	-	1
Amplitude +20%	-	0.23	6.28	-	0	-	0.9979	0.9414	-	1
Amplitude -20%	-	0	1.44	-	0	-	1	0.9868	-	1
Mp3 128kbps	0	0	1.37	0.03	0	1	1	0.9874	0.9997	1
Mp3 64kbps	0	0	1.37	0.03	0	1	1	0.9874	0.9997	1
AAC 128kbps	0	0	1.37	0.03	0	1	1	0.9874	0.9997	1
Cropping 3000	0.16	1.14	1.95	1.20	0.06	0.9985	0.9894	0.9821	0.9888	0.9994

Table 10 Robustness comparisons between the proposed algorithm and the algorithms in reference [29–32] for "Folk"(16k)

Attack type	BER(%)													
	[29]	[30]	[31]	[32]	Proposed	[29]	[30]	[31]	[32]	Proposed	[29]	[31]	[32]	Proposed
No attack	0	0	0	0	0	1	1	1	1	0	1	1	1	1
AWGN	0	16.21	-	18.46	7.52	1	0.8453	-	-	0.8302	1	-	0.8302	0.9298
Re-sampling	0	0	-	0.59	0	1	1	-	-	0.9946	1	-	0.9946	1
Re-quantization	0	0	-	0.49	0	1	1	-	-	0.9955	1	-	0.9955	1
echo	-	-	-	16.89	0.29	-	-	-	-	0.8402	-	-	0.8402	0.9973
Amplitude +10%	7.03	0	0	7.32	0	0.9346	1	1	1	0.9318	1	1	0.9318	1
Amplitude -10%	7.03	0	0	2.73	0	0.9346	1	1	1	0.9750	1	1	0.9750	1
Amplitude +20%	-	0	0	-	0	-	1	1	1	-	-	1	-	1
Amplitude -20%	-	0	0	-	0	-	1	1	1	-	-	1	-	1
Mp3 32kbps	0	0	0	0.29	0	1	1	1	1	0.9973	1	1	0.9973	1
Cropping 3000	1.17	1.86	0.68	1.66	0	0.9891	0.9828	0.9937	0.9846	0.9846	0.9846	0.9846	0.9846	1

gives the BER and NC values of the proposed scheme and schemes [29–32] under different attacks, where the symbol "-" indicates that the experimental results under these attacks do not comply with the IFPI standard (the BER exceeds 20%) and the corresponding failed NC values. From the table 10, it can be seen that the proposed scheme outperforms the schemes [31, 32] under re-sampling and re-quantization attacks, and significantly outperforms the schemes [29, 31] under amplitude scaling attacks. The BER values of the proposed schemes are mostly 0, although the largest ber value is 7.52%, it is also significantly better than the schemes [30–32] under AWGN attack. This scheme better extracts the reversible robustness features by means of global embedding and adaptive parameter selection, therefore, this scheme can be better applied in practice to protect users' digital audio copyrights.

5 Conclusion

In this paper, an adaptive robust audio watermarking algorithm based on svd is proposed. In this method, after applying the DWT transform to each frame, we divide the signal into two segments using sub-sampling operation and calculate the singular values of the two segments separately, then we generate different size of quantization parameters according to different features of the host signal using the proposed adaptive quantization method to maximize the perceptual quality of the algorithm, and finally we use differential embedding for watermark embedding. The watermark can still be extracted correctly when the original signal and embedding parameters are not available. The experimental results show that the proposed scheme is robust to some common attacks, especially the amplitude scaling attack, while guaranteeing a certain watermarking capacity compared to recent algorithms.

It is worth noting that although the proposed algorithm has good performance against conventional signal processing attacks, it cannot resist de-synchronization attacks and some novel attacks such as recapturing attacks. In the future, we may incorporate deep learning networks to improve the resistance of the proposed scheme against these attacks.

Acknowledgements This study was supported by the Sichuan Science and Technology program (Grant nos.2023NSFSC0470, 2022YFG0152, 2021YFQ0053), and the National Natural Science Foundation of China (NSFC) program (No.62171387, No.62202390).

Data Availability Statement Data can be provided if the request is reasonable.

Declarations

Conflict of Interest The authors declare that they have no conflicts of interest that might influence the work in this paper.

References

1. Wang Y, Pan Y, Yan M, Su Z, Luan TH (2023) A survey on ChatGPT: AI-generated contents, challenges, and solutions
2. Hu R, Xiang S (2021) Lossless robust image watermarking by using polar harmonic transform. *Signal Process* 179:107833
3. Chen Y, Jia Z-G, Peng Y, Peng Y-X, Zhang D (2021) A new structure-preserving quaternion qr decomposition method for color image blind watermarking. *Signal Process* 185:108088

4. Wang X-Y, Shen X, Tian J-L, Niu P-P, Yang H-Y (2022) Statistical image watermark decoder using high-order difference coefficients and bounded generalized gaussian mixtures-based hmt. *Signal Process* 192:108371
5. Asikuzzaman M, Pickering MR (2018) An overview of digital video watermarking. *IEEE Trans Circuits Syst Video Technol* 28(9):2131–2153. <https://doi.org/10.1109/TCSVT.2017.2712162>
6. Asikuzzaman M, Mareen H, Moustafa N, Choo K-KR, Pickering MR (2022) Blind camcording-resistant video watermarking in the dtcwt and svd domain. *IEEE Access* 10:15681–15698. <https://doi.org/10.1109/ACCESS.2022.3146723>
7. Chen S, Malik A, Zhang X, Feng G, Wu H (2023) A fast method for robust video watermarking based on zernike moments. *IEEE Trans Circ Syst Video Technol* 1–1. <https://doi.org/10.1109/TCSVT.2023.3281618>
8. Malvar HS, Florêncio DA (2003) Improved spread spectrum: A new modulation technique for robust watermarking. *IEEE Trans Signal Process* 51(4):898–905
9. Valizadeh A, Wang ZJ (2010) Correlation-and-bit-aware spread spectrum embedding for data hiding. *IEEE Trans Inf Forensics Secur* 6(2):267–282
10. Zhang P, Xu S-Z, Yang H-Z (2012) Robust audio watermarking based on extended improved spread spectrum with perceptual masking. *Int J Fuzzy Syst* 14(2)
11. Zhang X, Wang ZJ (2013) Correlation-and-bit-aware multiplicative spread spectrum embedding for data hiding. In: 2013 IEEE International workshop on information forensics and security (WIFS), pp 186–190. IEEE
12. Xiang Y, Natgunanathan I, Rong Y, Guo S (2015) Spread spectrum-based high embedding capacity watermarking method for audio signals. *IEEE/ACM Trans Audio Speech Lang Process* 23(12):2228–2237
13. Xiang Y, Natgunanathan I, Peng D, Hua G, Liu B (2017) Spread spectrum audio watermarking using multiple orthogonal pn sequences and variable embedding strengths and polarities. *IEEE/ACM Trans Audio Speech Lang Process* 26(3):529–539
14. Natgunanathan I, Xiang Y, Rong Y, Zhou W, Guo S (2012) Robust patchwork-based embedding and decoding scheme for digital audio watermarking. *IEEE Trans Audio Speech Lang Process* 20(8):2232–2239
15. Xiang Y, Natgunanathan I, Guo S, Zhou W, Nahavandi S (2014) Patchwork-based audio watermarking method robust to de-synchronization attacks. *IEEE/ACM Trans Audio Speech Lang Process* 22(9):1413–1423
16. Natgunanathan I, Xiang Y, Hua G, Beliakov G, Yearwood J (2017) Patchwork-based multilayer audio watermarking. *IEEE/ACM Trans Audio Speech Lang Process* 25(11):2176–2187
17. Liu Z, Huang Y, Huang J (2018) Patchwork-based audio watermarking robust against de-synchronization and recapturing attacks. *IEEE Trans Inf Forensics Secur* 14(5):1171–1180
18. Vivekananda BK, Sengupta I, Das A (2010) An adaptive audio watermarking based on the singular value decomposition in the wavelet domain - sciencedirect. *Digital Signal Process* 20(6):1547–1558
19. Lei B, Soon IY, Tan EL (2013) Robust svd-based audio watermarking scheme with differential evolution optimization. *IEEE Trans Audio Speech Lang Process* 21(11):2368–2378
20. Dhar PK, Shimamura T (2014) Blind svd-based audio watermarking using entropy and log-polar transformation. *J Inform Sec Appl* 20(C):74–83
21. Wu Q, Qu A, Huang D (2020) Robust and blind audio watermarking algorithm in dual domain for overcoming synchronization attacks. *Math Probl Eng* 2020:1–15
22. Zhao J, Zong T, Xiang Y, Gao L, Zhou W, Beliakov G (2021) Desynchronization attacks resilient watermarking method based on frequency singular value coefficient modification. *IEEE/ACM Trans Audio Speech Lang Process* 29:2282–2295. <https://doi.org/10.1109/TASLP.2021.3092555>
23. Jiang W, Huang X, Quan Y (2019) Audio watermarking algorithm against synchronization attacks using global characteristics and adaptive frame division. *Signal Process* 162
24. Benoraira A, Benmahammed K, Boucenna N (2015) Blind image watermarking technique based on differential embedding in dwt and dct domains. *Eurasip J Adv Signal Process* 2015(1):55
25. Saadi S, Merrad A, Benziane A (2019) Novel secured scheme for blind audio/speech norm-space watermarking by arnold algorithm. *Signal Process* 154(JAN):74–86
26. Bernardi G, Van Waterschoot T, Wouters J, Moonen M (2018) Subjective and objective sound-quality evaluation of adaptive feedback cancellation algorithms. *IEEE/ACM Trans Audio Speech Lang Process* 26(5):1–1
27. Torcoli M, Kastner T, Herre J (2021) Objective measures of perceptual audio quality reviewed: An evaluation of their application domain dependence. *arXiv e-prints*
28. Kabal P, et al (2002) An examination and interpretation of itu-r bs. 1387: Perceptual evaluation of audio quality. TSP Lab Technical Report, Dept. Electrical & Computer Engineering, McGill University, 1–89

29. Wang X, Wang P, Zhang P, Xu S, Yang H (2013) A norm-space, adaptive, and blind audio watermarking algorithm by discrete wavelet transform. *Signal Processing*
30. Li J-F, Wang H-X, Wu T, Sun X-M, Qian Q (2018) Norm ratio-based audio watermarking scheme in dwt domain. *Multimed Tools Appl* 77(12):14481–14497
31. Budiman G, Suksmono AB, Danudirdjo D (2020) Wavelet-based hybrid audio watermarking using statistical mean manipulation and spread spectrum. In: 2020 27th international conference on telecommunications (ICT), pp 1–5 . <https://doi.org/10.1109/ICT49546.2020.9239581>
32. Dhar PK (2015) A blind audio watermarking method based on lifting wavelet transform and qr decomposition. In: 2014 8th international conference on electrical and computer engineering (ICECE)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Xiangyi Liu received the B.S. degree in Computer Science and Technology from Xichang college, Xichang, China, in 2021, and she is now pursuing the M.S. degree at School of Computer Science and Technology, Xihua University, China. Her current researches focus on multimedia information security and digital watermarking.

Xiaojie Li (Member, IEEE) received the Ph.D. degree in computer science and engineering from the College of Computer Science, Sichuan University, Chengdu, China, in 2015. She is currently an Associate Professor with the College of Computer Science, Chengdu University of Information Technology, Chengdu. Her research interests include machine learning, neural networks, and data mining.

Canghong Shi received the B.S. degree in mathematics and applied mathematics from Hebei Normal University, Shijiazhuang, China, in 2009, the M.S. degree in basic mathematics from the Chengdu University of Information Technology, Chengdu, China, in 2014, and the Ph.D. degree in information and communication engineering from Southwest Jiaotong University, Chengdu, China, in 2020. He is currently a Lecturer with the school of computer and software engineering, Xihua University, Chengdu. His current researches focus on multimedia information security, digital audio signal forensics, machine learning, and digital watermarking. He has published 10 peer research articles and registered 2 patents.

Xianhua Niu (Member, IEEE) received the B.S. degree in communication engineering and the Ph.D. degree in information security from Southwest Jiaotong University, Chengdu, China, in 2006 and 2012, respectively. She is currently a Professor with the School of Computer and Software Engineering, Xihua University, and a Post-Doctoral Member with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu. Her research interests include sequence design, coding theory and information safety.