Int J Speech Technol (2009) 12: 27–45
DOI 10.1007/s10772-009-9056-2

An efficient singular value decomposition algorithm for digital audio watermarking

Fathi E. Abd El-Samie

Received: 7 July 2009 / Accepted: 17 October 2009 / Published online: 12 November 2009

Abstract The singular value decomposition (SVD) mathematical technique is utilized, in this paper, for audio watermarking in time and transform domains. Firstly, the audio signal in time or an appropriate transform domain is transformed to a 2-D format. The SVD algorithm is applied on this 2-D matrix, and an image watermark is added to the matrix of singular values (SVs) with a small weight, to guarantee the possible extraction of the watermark without introducing harmful distortions to the audio signal. The transformation of the audio signal between the 1-D and 2-D formats is performed in the well-known lexicographic ordering method used in image processing. A comparison study is presented in the paper between the time and transform domains as possible hosting media for watermark embedding. Experimental results are in favor of watermark embedding in the time domain if the distortion level in the audio signal is to be kept as low as possible with a high detection probability. The proposed algorithm is utilized also for embedding chaotic encrypted watermarks to increase the level of security. Experimental results show that watermarks embedded with the proposed algorithm can survive several attacks. A segment-by-segment implementation of the proposed SVD audio watermarking algorithm is also presented to enhance the detectability of the watermark in the presence of severe attacks.

Keywords Audio watermarking · SVD · DFT · DST · DCT · DWT

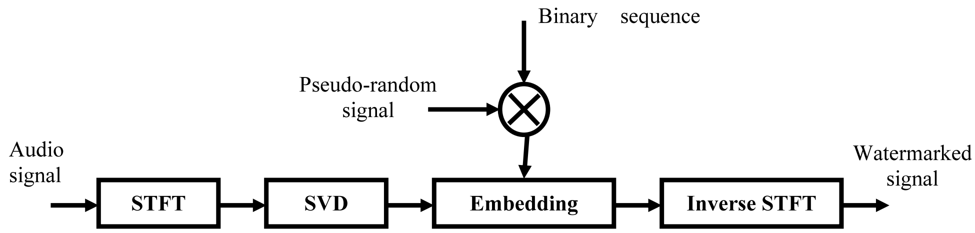
F.E. Abd El-Samie (✉)
Department of Electronics and Electrical Communications, Faculty of Electronic Engineering, Menoufia University, Menouf, Egypt
e-mail: fathi_sayed@yahoo.com

1 Introduction

Watermarking is a growing field of research, because of its importance for several applications, such as information hiding, copyright protection, fingerprinting and authentication (Macq et al. 2004). Digital image and video watermarking has received a great interest from researchers and several algorithms have been proposed for this purpose (Macq et al. 2004; Lu et al. 2005; Kim and Lee 2003; Chu 2003; Ghouti et al. 2006). Image watermarking algorithms can be classified to two main categories; algorithms that can achieve efficient embedding and detection of the watermark only and algorithms that allow the detection and extraction of the watermark. For audio watermarking, most of the existing algorithms are designed to achieve the efficient detection of the watermark without extracting meaningful information from the watermarked audio signal (Xiang and Huang 2007; Liu and Inoue 2003; Lemma et al. 2003; Li et al. 2006; Erküçük et al. 2006).

Several audio watermarking algorithms have been proposed in recent years. Watermark embedding through a quantization process is one of such algorithms (Wang et al. 2007; Özer and Sankur 2005). It is a very simple algorithm; some little information is embedded in audio signals through the quantization of the samples. It is clear from the nature of this algorithm that a simple filtering or a noise attack may lead to a great difficulty in the watermark detection. Another algorithm is based on the spread spectrum technique and is implemented by adding pseudo-random sequences to the small segments of the audio signal. This algorithm depends on the orthogonality between the pseudo-random sequences, if they are of sufficiently long length (Liu and Inoue 2003). Unfortunately, the orthogonality is destroyed if the lengths of the sequences used are short.

Fig. 1 The SVD audio watermarking method proposed by Özer and Sankur



Away from the audio watermarking algorithms, which are directed towards the embedding and detection of the watermark only, there are a few attempts to embed meaningful information in audio signals. This meaningful information can be in the form of images, and if a higher level of security is required, they can be encrypted images. Some research work has been performed for conveying meaningful information through audio watermarking using the support vector regression algorithm (Wang et al. 2007). In fact, there is no image transmission in this algorithm. What happens exactly is a direct modification in the audio signal characteristics that can be detected at the receiver. If the modification is detected successfully, a trained neural system could be used to generate the image required from a database at the receiver. Although this idea is good, audio signals can be distorted in the communication channel, and hence the generation of the required image with the neural system becomes difficult.

In this paper, a new audio watermarking algorithm, which allows image embedding in audio signals, is proposed. This algorithm can be implemented on audio signals in time domain or in another appropriate transform domain. The audio signal is transformed into a 2-D format and the SVs of the resulting 2-D matrix are used for watermark embedding. Both normal images and encrypted images can be embedded in audio signals with this algorithm. One of the most appropriate image encryption algorithms that can be utilized for watermark embedding is the 2-D chaotic Baker map algorithm, because it is a permutation based algorithm, which has no accumulation of error due to attacks (Han et al. 2006; Fridrich 1997; Qian et al. 2008; Huang and Lei 2008; Koduru and Chandrasekaran 2008; Usman et al. 2007).

The rest of the paper is organized as follows. Section 2 briefly presents a previous attempt to apply the SVD algorithm in audio watermarking. Section 3 presents the proposed SVD audio watermarking algorithm. Section 4 presents the proposed audio watermarking algorithm. Section 5 explains the discrete transforms that can be utilized for watermark embedding in audio signals. Section 6 gives an explanation of the chaotic encryption algorithm, which can be used for watermark encryption. Section 6 presents a different implementation of the proposed SVD audio watermarking algorithm on a segment-by-segment basis. Section 7 gives the audio quality metrics that can be used to evaluate the quality of audio signals after watermark embedding. Section 8 introduces the experimental results and Sect. 9 gives the concluding remarks.

2 Traditional SVD audio watermarking

The SVD technique is exploited in image watermarking, because of its sophisticated properties (Liu and Tan 2002; Sun et al. 2008; Zhu et al. 2006). This technique provides an elegant way for extracting algebraic features from a 2-D matrix. The main properties of the matrix of the SVs can be exploited in watermarking. When a small perturbation happens to the original data matrix, no large variations occur in the matrix of the SVs, which makes this technique robust against attacks (Liu and Tan 2002; Sun et al. 2008; Zhu et al. 2006).

A previous attempt for the implementation of the SVD technique in audio watermarking was presented by Özer and Sankur (2005). A block diagram of this method is shown in Fig. 1. This method is based on the SVD of the spectrogram of the audio signal. The SVD of the spectrogram is modified adaptively according to the watermark bits, which are obtained by the multiplication of a certain binary sequence and a pseudo-random signal. Although this method is simple, it is restricted to embedding binary watermarks, only. In addition, the effect of the traditional attacks like filtering, additive noise and cropping has not been studied in this method. The next section will present an efficient implementation of the SVD technique for audio watermarking which allows embedding meaningful images in audio signals.

3 The proposed SVD audio watermarking algorithm

This section presents a new SVD audio watermarking algorithm, which allows embedding images in audio signals and the extraction of these images in the watermark detection process. The steps of the proposed embedding algorithm are summarized as follows:

1. The audio signal is either used in time domain or transformed to an appropriate transform domain.
2. The 1-D obtained signal is transformed into a 2-D matrix (A matrix).

3. The SVD is performed on the A matrix.

$$A = USV^T \tag{1}$$

4. The chaotic encrypted watermark (W matrix) is added to the SVs of the original matrix.

$$D = S + KW \tag{2}$$

5. The SVD is performed on the new modified matrix (D matrix).

$$D = U_w S_w V_w^T \tag{3}$$

6. The watermarked signal in 2-D format (A_w matrix) is obtained using the modified matrix of SVs (S_w matrix).

$$A_w = U S_w V^T \tag{4}$$

7. The 2-D A_w matrix is transformed again into a 1-D signal.
8. If watermarking is performed in a transform domain, an inverse of this transform is performed.

To extract the possibly corrupted watermark from the possibly distorted watermarked audio signal, given the U_w, S, V_w matrices and the possibly distorted audio signal, the above steps are reversed as follows:

1. If watermarking is performed in a transform domain, this transform is performed.
2. The 1-D obtained signal is transformed into a 2-D matrix A_w^*. The $*$ refers to the corruption due to attacks.
3. The SVD is performed on the possibly distorted watermarked image (A_w^* matrix).

$$A_w^* = U^* S_w^* V^{*T} \tag{5}$$

4. The matrix that includes the watermark is computed.

$$D^* = U_w S_w^* V_w^T \tag{6}$$

5. The possibly corrupted encrypted watermark is obtained.

$$W^* = (D^* - S)/K \tag{7}$$

6. The obtained matrix W^* is decrypted.
7. The correlation coefficient between the decrypted matrix and the original watermark is estimated. If this coefficient is higher than a certain threshold, the watermark is present.

4 Discrete transforms for audio watermarking

In this section, the discrete wavelet transform (DWT), the discrete cosine transform (DCT), the discrete sine transform (DST), and the discrete Fourier transform (DFT) are briefly revised for the purpose of watermark embedding. All these transforms are invertible, which enables the extraction of the embedded watermark.

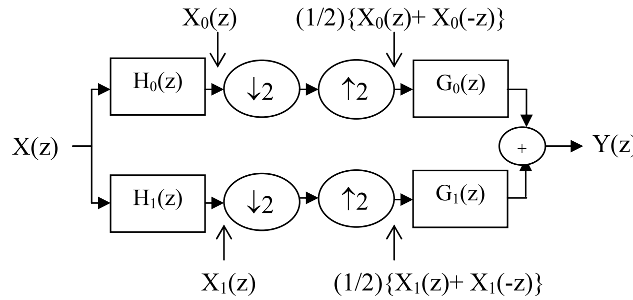


Fig. 2 The two band decomposition-reconstruction wavelet filter bank

4.1 The DWT

The idea of the DWT is to represent a signal as a series of an approximation and details. Figure 2 shows the two band decomposition-reconstruction wavelet filter bank. The signal is lowpass filtered with $H_0(z)$ to give an approximation signal and highpass filtered with $H_1(z)$ to give a detail signal. The wavelet basis function is chosen such that a perfect reconstruction can be achieved. For this filter bank to achieve perfect reconstruction, the following two equations must be satisfied (Lim 1990; Pratt 1991; Walker 1999; Guillemain and Martinet 1996; Prochazka et al. 1998; Wornell 1996).

$$\{H_0(z)G_0(z) + H_1(z)G_1(z)\} = 2 \tag{8}$$

$$H_1(z) = z^{-k}G_0(-z) \quad \text{and} \quad G_1(z) = z^k H_0(-z) \tag{9}$$

For the Haar wavelet transform, the transfer functions of the filters are given by:

$$H_0(z) = \frac{1}{2}(1 + z^{-1}) \tag{10}$$

$$G_0(z) = (z + 1) \tag{11}$$

$$G_1(z) = \frac{1}{2}(z - 1) \tag{12}$$

$$H_1(z) = (z^{-1} - 1) \tag{13}$$

For audio watermarking, the approximation and detail components can be concatenated to give a 1-D vector of the same length as the audio signal. This vector can be used for watermark embedding with the proposed algorithm.

4.2 The DCT

The DCT expresses the samples of the audio signal in terms of a sum of cosine functions oscillating at different frequencies. The DCT is defined by the following equation (Lim 1990; Pratt 1991; Walker 1999; Guillemain and Martinet 1996; Prochazka et al. 1998; Wornell 1996):

$$X(k) = w(k) \sum_{n=0}^{N-1} x(n) \cos\left(\frac{\pi(2n-1)(k-1)}{2N}\right)$$
$$k = 0, \ldots, N-1 \tag{14}$$

where $x(n)$ is the audio signal with length N samples.

$$w(k) = \begin{cases} \frac{1}{\sqrt{N}}, & k = 0 \\ \sqrt{\frac{2}{N}}, & k = 1, \ldots, N-1 \end{cases} \tag{15}$$

The DCT has a sophisticated characteristic of energy compaction by collecting most of the signal energy in a few samples leaving the other samples very small in magnitude. This characteristic can be exploited in audio watermarking to reduce the deterioration in the audio signal due to watermarking.

4.3 The DST

The DST is similar to the DCT, but it uses sine functions oscillating at different frequencies rather than cosine functions. The DST is defined by the following equation (Lim 1990; Pratt 1991; Walker 1999; Guillemain and Martinet 1996; Prochazka et al. 1998; Wornell 1996):

$$X(k) = \sum_{n=0}^{N-1} x(n) \sin\left(\frac{\pi kn}{N+1}\right) \quad k = 0, \ldots, N-1 \tag{16}$$

It has also an energy compaction property that can be utilized for audio watermarking.

4.4 The DFT

The DFT comprises both the DCT and the DST by implementing a complex basis function as follows (Lim 1990; Pratt 1991; Walker 1999; Guillemain and Martinet 1996; Prochazka et al. 1998; Wornell 1996):

$$X(k) = \sum_{n=0}^{N-1} x(n) \mathrm{e}^{-2j\pi nk} \quad k = 0, \ldots, N-1 \tag{17}$$

The Fourier transform gives complex valued samples, and so both the Fourier transform magnitude and phase can be used for watermark embedding.

5 Chaotic encryption

Chaotic encryption of the watermark image is performed using the chaotic Baker map. The Baker map is a chaotic map that generates a permuted version of a square matrix. In its discretized form, the Baker map is an efficient tool to randomize a square matrix of data. The discretized map can be represented for an $M \times M$ matrix as follows (Han et al. 2006; Fridrich 1997; Qian et al. 2008; Huang and Lei 2008; Koduru and Chandrasekaran 2008; Usman et al. 2007):

$$B(r, s) = \left[\frac{M}{n_i}(r - M_i) + s \bmod \left(\frac{M}{n_i}\right), \right.$$
$$\left. \frac{n_i}{M}\left(s - s \bmod \left(\frac{M}{n_i}\right)\right) + M_i\right] \tag{18}$$

where $B(r, s)$ are the new indices of the data item at (r, s), $M_i \le r < M_i + n_i, 0 < s < M$ and $M_i = n_1 + n_2 + \cdots + n_i$.

In steps, the chaotic encryption is performed as follows:

1. An $M \times M$ square matrix is divided into k vertical rectangles of height M and width n_i.
2. These vertical rectangles are stretched in the horizontal direction and contracted vertically to obtain an $n_i \times M$ horizontal rectangle.
3. These rectangles are stacked as shown in Fig. 3(a), where the left one is put at the bottom and the right one at the top.
4. Each $n_i \times M$ vertical rectangle is divided into n_i boxes of dimensions $M = n_i \times n_i$, containing exactly M points.
5. Each of these boxes is mapped column by column into a row as shown in Fig. 3(b).

Figure 3 shows an example of the chaotic encryption of an (8×8) square matrix. The secret key, $S_{\text{key}} = (n_1, n_2, n_3) = (2, 4, 2)$.

6 The proposed segment-by-segment algorithm

The proposed SVD audio watermarking algorithm concentrates on embedding a single watermark in the audio signal as a whole. If a watermark is embedded several times on a segment-by-segment basis, it is expected that the detectability of this watermark will be enhanced and its robustness against attacks will be increased. Dividing the audio signal into small segments then embedding the watermark in the SVs of each segment, separately, gives the chance that one or more of the watermarks will survive the attacks and a high correlation coefficient in the detection will be obtained.

6.1 Watermark embedding

In the proposed segment-by-segment algorithm, the original audio signal is divided into non-overlapping segments. The encrypted watermark is embedded in the SVs (S matrix) of each segment after the transformation of the segment into a small 2-D matrix. An SVD is performed on each of these new matrices to get the S matrices of the segments. Then, these SV matrices are used to build the watermarked segments.

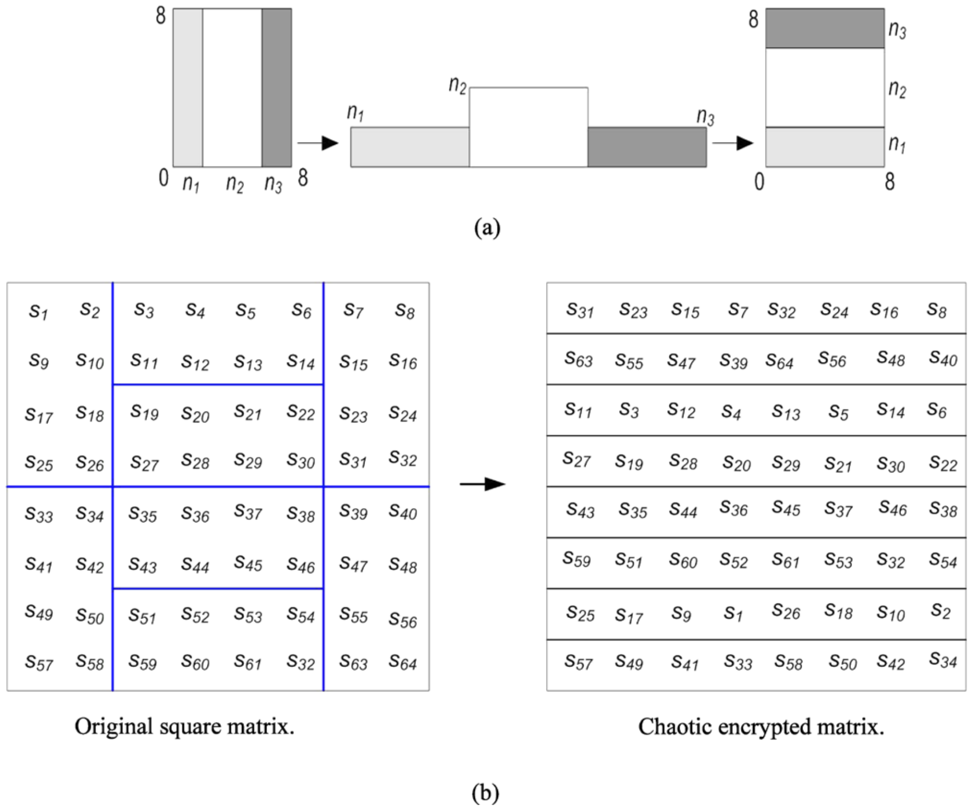


Fig. 3 Chaotic encryption. (**a**) Discretized Baker map. (**b**) Chaotic encryption of an 8×8 matrix

The steps of the embedding process can be summarized as follows:

1. The audio signal is either used in time domain or transformed to a certain transform domain.
2. The obtained signal is divided into non-overlapping segments and each segment is transformed into a 2-D matrix.
3. The SVD is performed on the 2-D matrix of each segment (B_i matrix) to obtain the SVs (S_i matrix), where $i = 1, 2, 3, \ldots, N$, and N is the number of segments.

$$B_i = U_i S_i V_i^T \tag{19}$$

4. The encrypted watermark (W matrix) is added to the S_i matrix of each segment.

$$D_i = S_i + KW \tag{20}$$

5. The SVD is performed on each D_i matrix to obtain the SVs of each one (S_{wi} matrix).

$$D_i = U_{wi} S_{wi} V_{wi}^T \tag{21}$$

6. The S_{wi} matrices are used to build the watermarked segments in the time domain.

$$B_{wi} = U_i S_{wi} V_i^T \tag{22}$$

7. The watermarked segments are transformed into the 1-D format.
8. The watermarked segments are rearranged back into a 1-D signal.
9. If watermarking is performed in a transform domain, an inverse of this transform is performed.

6.2 Watermark detection

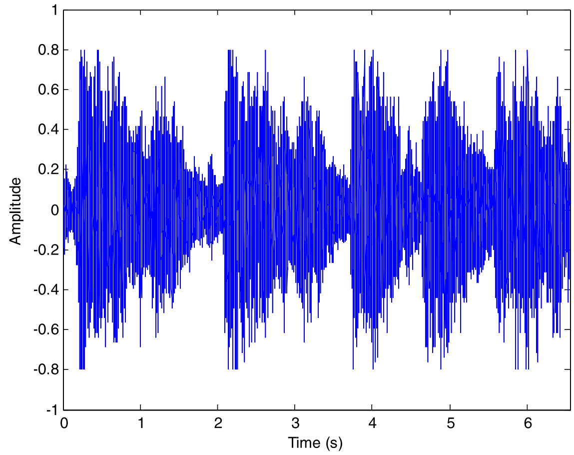
Having U_{wi}, V_{wi}, S_i, matrices and the possibly distorted audio signal, the steps mentioned below can be followed to get the possibly corrupted watermark.

1. If watermarking is performed in a transform domain, this transform is performed.
2. The obtained signal is divided into small segments having the same length used in the embedding process and these segments are transformed into the 2-D format.
3. The SVD is performed on each possibly distorted watermarked segment (B_{wi}^* matrix) to obtain the SVs of each one (S_{wi}^* matrix).

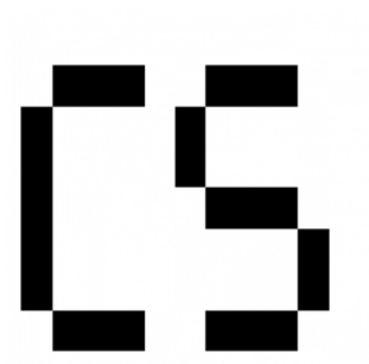
$$B_{wi}^* = U_i^* S_{wi}^* V_i^{*T} \tag{23}$$

4. The matrices that may contain the watermark are obtained using the U_{wi}, V_{wi}, S_{wi}^*, matrices.

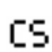
$$D_i^* = U_{wi} S_{wi}^* V_{wi}^T \tag{24}$$



(a) Original audio signal.



(b) Original watermark in the proposed SVD method.



(c) Original watermark in the segment-by-segment SVD method.

Fig. 4 Original signal and original watermarks

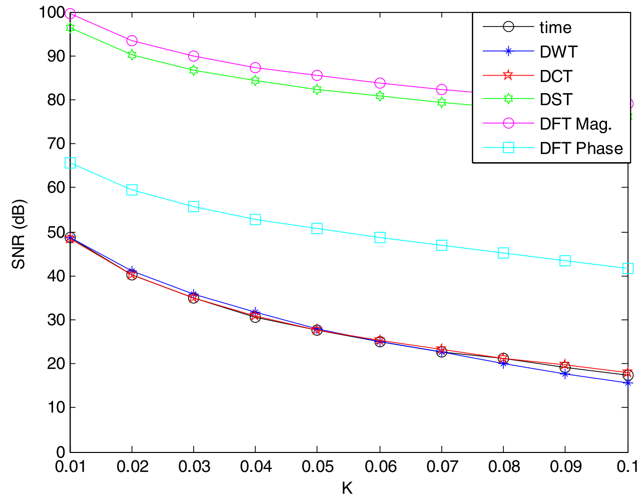
5. The possibly corrupted encrypted watermarks (W_i^* matrices) are extracted from the D_i matrices.

$$W_i^* = (D_i^* - S_i)/K \tag{25}$$

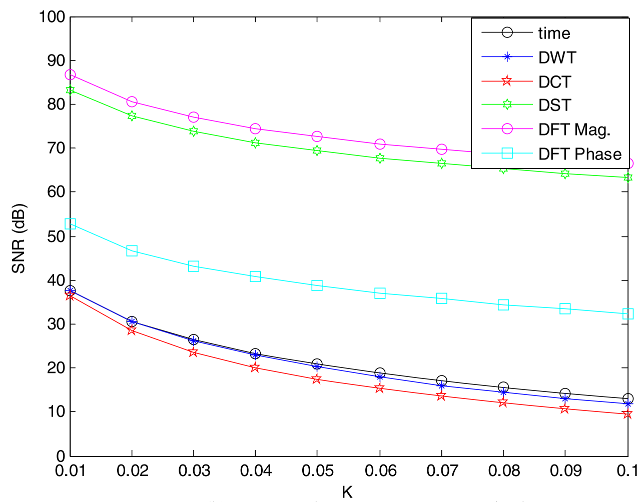
6. The obtained matrix W_i^* matrices are decrypted.
7. The correlation coefficient between each decrypted matrix W_i^* and the original watermark is estimated. If at least one of the coefficients is higher than a certain threshold, the watermark is present.

7 Objective quality metrics for audio signals

The proposed audio watermarking algorithm has two main objectives; hiding of an encrypted image in the audio signal



(a) SVD method.

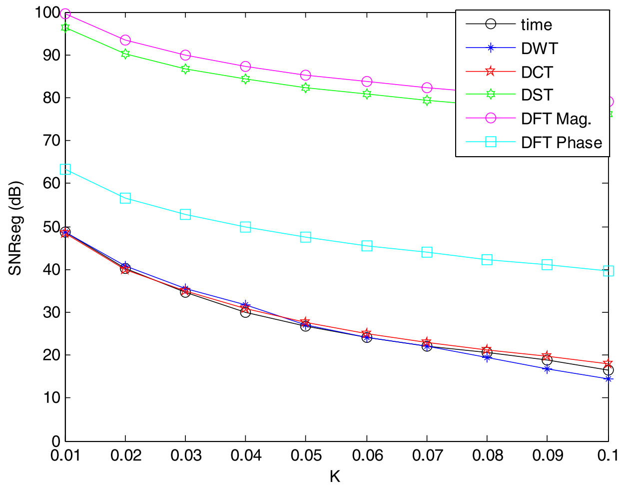


(b) Segment-by-segment SVD method.

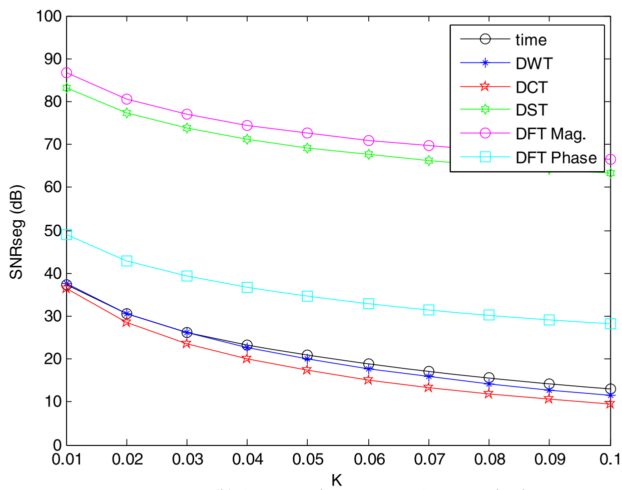
Fig. 5 Variation of the SNR of the watermarked audio signal with the watermark strength in the absence of any attacks

and maintaining the perceptual quality of the audio signal. So, there is a need of some metrics to assess the perceptual quality of the watermarked audio signal. Several approaches, based on subjective and objective metrics, have been adopted in the literature for this purpose (Kubichek 1993; Wang et al. 1992; Yang et al. 1998; McDermott et al. 1978; Crochiere et al. 1980). Concentration in this paper will be on the objective metrics.

Objective metrics are generally divided into intrusive and non-intrusive metrics. Intrusive metrics can be classified into three main groups. The first group includes time domain metrics such as the traditional signal-to-noise ratio (SNR) and segmental signal-to-noise ratio (SNRseg). The second group includes linear predictive coefficients (LPCs) metrics, which are based on the LPCs of the audio signal and its derivative parameters such as the linear reflection coefficients (LRC), the log likelihood ratio (LLR), and the cep-

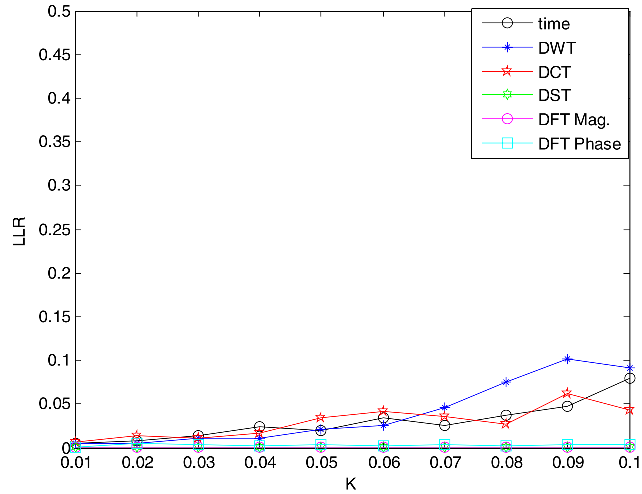


(a) SVD method.

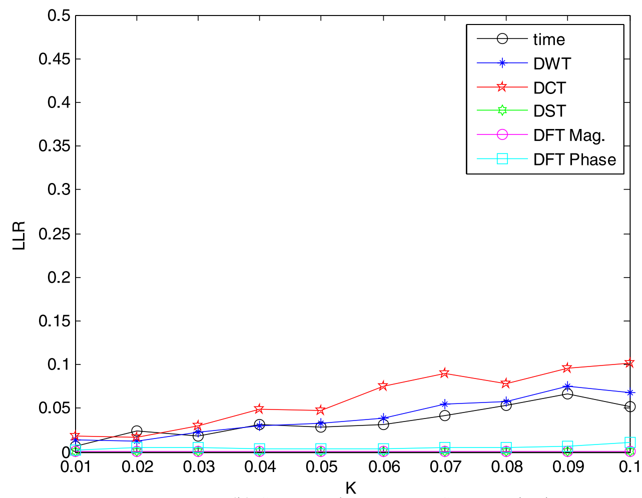


(b) Segment-by-segment SVD method.

Fig. 6 Variation of the SNRseg of the watermarked audio signal with the watermark strength in the absence of any attacks



(a) SVD method.



(b) Segment-by-segment SVD method.

Fig. 7 Variation of the LLR of the watermarked audio signal with the watermark strength in the absence of any attacks

stral distance (CD). The third group includes the spectral domain metrics, which are based on the comparison between the power spectrum of the original signal and the processed signal. An example of such metrics is the spectral distortion (SD) (Kubichek 1993; Wang et al. 1992; Yang et al. 1998; McDermott et al. 1978; Crochiere et al. 1980).

7.1 The signal-to-noise ratio

The SNR is defined as follows (Kubichek 1993; Wang et al. 1992; Yang et al. 1998; McDermott et al. 1978; Crochiere et al. 1980):

$$\mathrm{SNR} = 10\log_{10}\frac{\sum_{i=1}^{N} x^2(i)}{\sum_{i=1}^{N}(x(i)-y(i))^2} \tag{26}$$

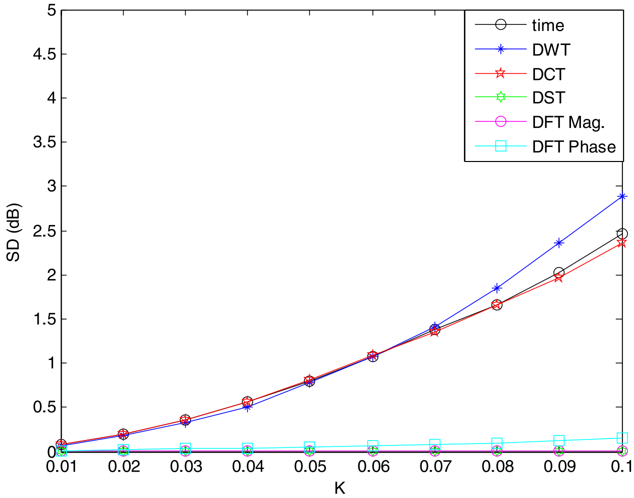
where $x(i)$ is the original audio signal, $y(i)$ is the watermarked audio signal, i is the sample index, and N is the number of samples of the watermarked audio signal.

7.2 The segmental signal-to-noise ratio

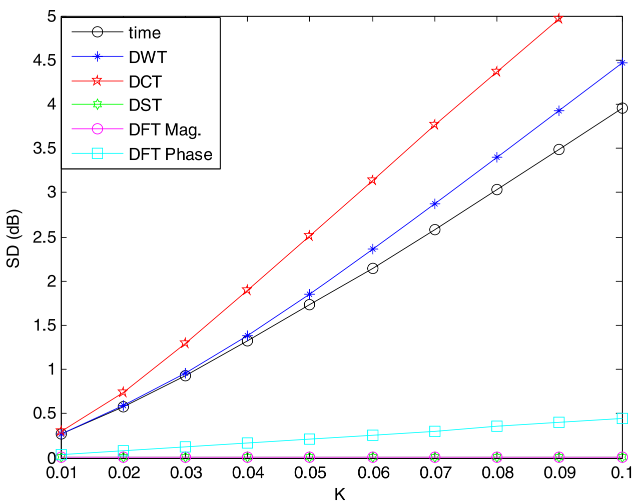
The most popular one of the time-domain metrics is the segmental signal-to-noise ratio (SNRseg). SNRseg is defined as the average of the SNR values of short segments of the watermarked signal. It is a good estimator for audio signal quality. It is defined as follows (Kubichek 1993; Wang et al. 1992; Yang et al. 1998; McDermott et al. 1978; Crochiere et al. 1980):

$$\mathrm{SNRseg} = \frac{10}{L}\sum_{m=0}^{M-1}\log_{10}\sum_{i=Nm}^{Nm+N-1}\left(\frac{x^2(i)}{(x(i)-y(i))}\right)^2 \tag{27}$$

where L is the number of segments in the watermarked audio signal.

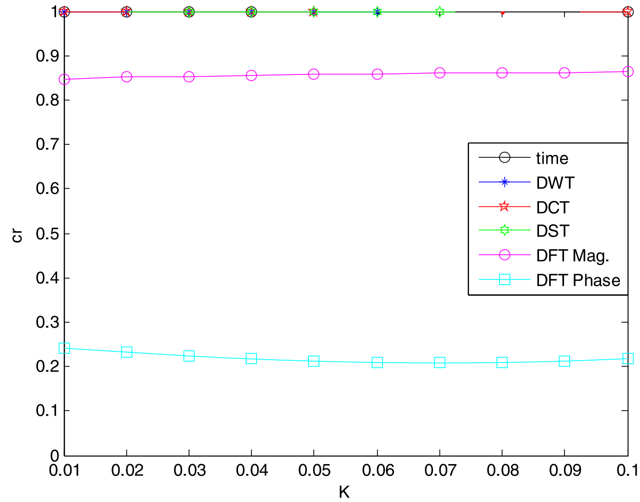


(a) SVD method.

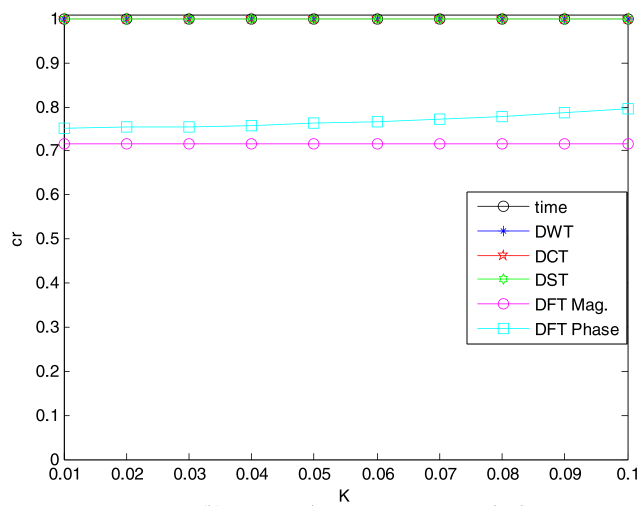


(b) Segment-by-segment SVD method.

Fig. 8 Variation of the SD of the watermarked audio signal with the watermark strength in the absence of any attacks



(a) SVD method.



(b) Segment-by-segment SVD method.

Fig. 9 Variation of the correlation coefficient c_r with the watermark strength in the absence of any attacks

7.3 The log likelihood ratio

The LLR metric for an audio segment is based on the assumption that the segment can be represented by a pth order all-pole linear predictive coding model of the form (Kubichek 1993; Wang et al. 1992; Yang et al. 1998; McDermott et al. 1978; Crochiere et al. 1980):

$$x(n) = \sum_{m=1}^{p} a_m x(n-m) + G_x u(n) \tag{28}$$

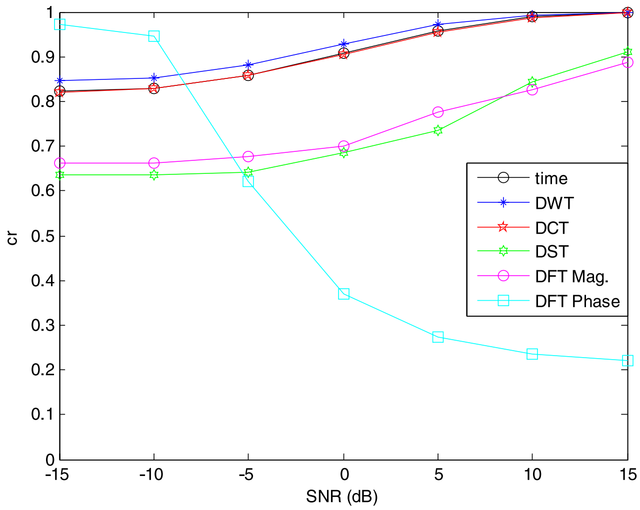
where $x(n)$ is the nth audio sample, a_m (for $m = 1, 2, \ldots, p$) are the coefficients of an all-pole filter, G_x is the gain of the filter and $u(n)$ is an appropriate excitation source for the filter. The audio signal is windowed to form frames of 15 to 30 ms length. The LLR metric is then defined as (Crochiere et al. 1980):

$$\mathrm{LLR} = \left| \log\left(\frac{\vec{\mathbf{a}}_{\mathbf{x}} \overline{\mathbf{R}}_{\mathbf{y}} \vec{\mathbf{a}}_{\mathbf{x}}^{\mathbf{T}}}{\vec{\mathbf{a}}_{\mathbf{y}} \overline{\mathbf{R}}_{\mathbf{y}} \vec{\mathbf{a}}_{\mathbf{y}}^{\mathbf{T}}} \right) \right| \tag{29}$$

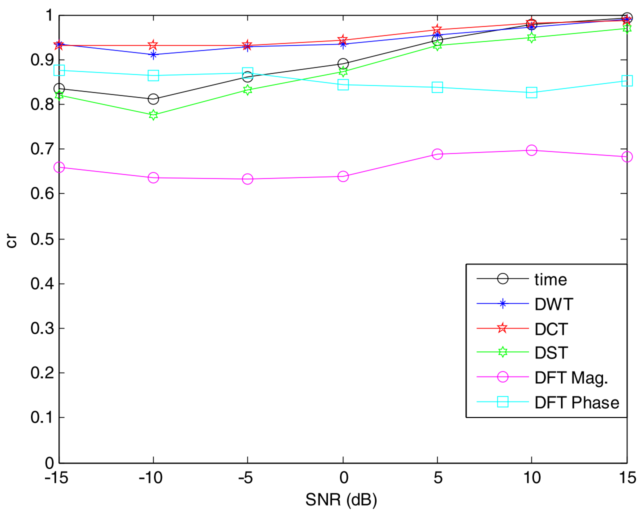
where $\vec{\mathbf{a}}_{\mathbf{x}}$ is the LPCs coefficient vector $[1, a_x(1), a_x(2), \ldots, a_x(p)]$ for the original audio signal $x(n)$, $\vec{\mathbf{a}}_{\mathbf{y}}$ is the LPCs coefficient vector $[1, a_y(1), a_y(2), \ldots, a_y(p)]$ for the watermarked audio signal $y(n)$, and $\overline{\mathbf{R}}_{\mathbf{y}}$ is the autocorrelation matrix of the watermarked audio signal. The closer the LLR to zero, the higher is the quality of the watermarked audio signal.

7.4 The spectral distortion

The SD is a form of metrics that is implemented in frequency domain on the frequency spectra of the original and

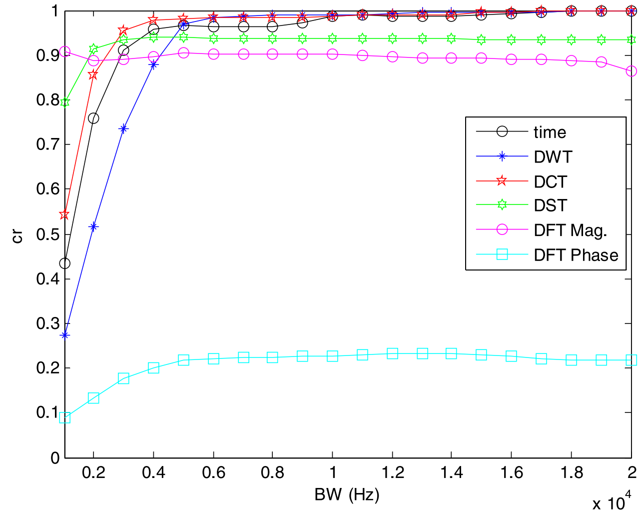


(a) SVD method.

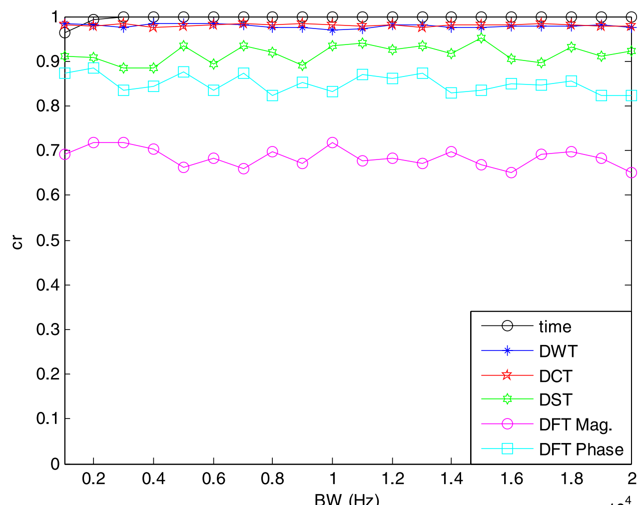


(b) Segment-by-segment SVD method.

Fig. 10 Variation of the correlation coefficient c_r with the SNR in the presence of white noise attack



(a) SVD method.



(b) Segment-by-segment SVD method.

Fig. 11 Variation of the correlation coefficient c_r with the filtering BW in the presence of the filtering attack

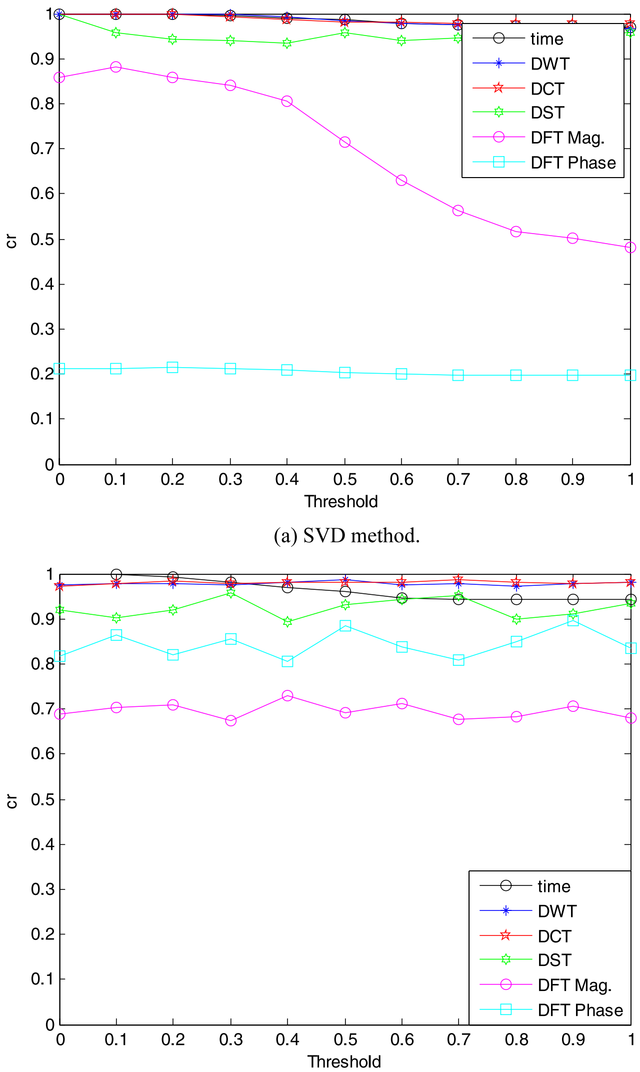
the watermarked signals. It is calculated in dB to show how far is the spectrum of the watermarked signal from that of the original signal. The SD can be calculated as follows (Kubichek 1993; Wang et al. 1992; Yang et al. 1998; McDermott et al. 1978; Crochiere et al. 1980):

$$\mathrm{SD} = \frac{1}{L}\sum_{m=0}^{M-1}\sum_{i=Nm}^{Nm+N-1}|V_x(i) - V_y(i)| \tag{30}$$

where $V_x(i)$ is the spectrum of the original audio signal in dB for a certain segment, $V_y(i)$ is the spectrum of the watermarked audio signal in dB for the same segment. The smaller the SD, the better is the quality of the audio watermarked signal.

8 Experimental results

In this section, several experiments are carried out to test the performance of the proposed SVD audio watermarking algorithm. Time and transform domains are used for watermark embedding. Both the proposed SVD watermarking algorithm and the segment-by-segment algorithm are simulated. The CS image is used as a watermark to be embedded in the Handel signal available in Matlab. The original Handel signal with the watermarks used in all experiments are shown in Fig. 4. In all experiments, the above-mentioned audio quality metrics are used for the evaluation of the quality of the watermarked audio signal and the correlation coefficient c_r is used to measure the closeness of the obtained watermark to the original watermark. Experiments are first



(a) SVD method.

(b) Segment-by-segment SVD method.

Fig. 12 Variation of the correlation coefficient c_r with the compression threshold in the presence of the wavelet compression attack

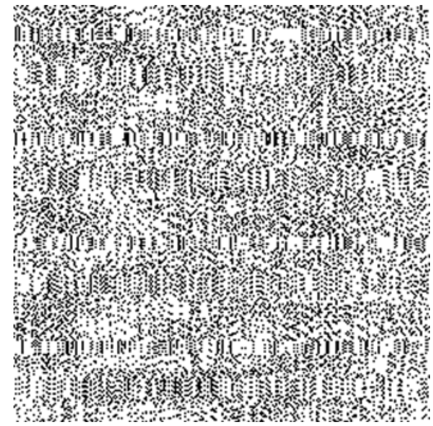


Fig. 13 Chaotic encrypted watermark for the SVD method. $c_r = 0.0181$

performed without encryption to choose the appropriate domain for watermark embedding.

In the first experiment, the effect of the watermark strength K used to add the watermark to the matrix of SVs of the audio signal is studied. The results of this experiment in the absence of attacks are shown in Figs. 5 to 9. It is clear from this experiment that in the absence of attacks, watermark embedding in the DFT magnitude or the DST domain achieves the lowest distortion level in the audio signal but the DST domain is preferred to the DFT magnitude in the detection process. It is also clear that segment-by-segment audio watermarking causes more distortion in the audio signal but achieves more success in the detection in the presence of attacks, as will be shown in the next experiments.

In the second experiment, the robustness of both the SVD audio watermarking method and the segment-by-segment method is studied in the presence of an additive white Gaussian noise (AWGN) attack. Figure 10 shows that watermark embedding in the DWT domain, the DCT domain, or the time domain achieves the highest detection correlation coefficient at low SNR values. From Fig. 8, it is clear that watermark embedding in the time domain achieves the smallest distortion as compared to the DCT domain and the DWT domain, especially for the segment-by-segment method. It is also clear that the segment-by-segment method increases the correlation coefficient of approximately all cases of watermarking.

In the third experiment, the robustness of both the SVD audio watermarking method and the segment-by-segment method is studied in the presence of a lowpass filtering attack. A third order Butterworth filter is used in this attack. Figure 11 shows that watermark embedding in the time domain achieves the highest detection correlation coefficient for the segment-by-segment method and a sufficiently high correlation coefficient values for the SVD method. It is also clear that the segment-by-segment method increases the correlation coefficient of approximately all cases of time and transform domain watermarking in the case of low bandwidth filtering, which is a severe case.

In the fourth experiment, the robustness of both the SVD audio watermarking method and the segment-by-segment method is studied in the presence of a wavelet compression attack. The results of this experiment are given in Fig. 12. Although, watermark embedding in the time domain is not the best case in correlation coefficient values for this attack, the time domain can be chosen as the most appropriate domain for watermark embedding due to the lowest spectral distortion, the sufficiently high values of the detection correlation coefficient and the ability to survive attacks.

Several other experiments are carried out to test the performance of the proposed SVD audio watermarking algorithm for embedding encrypted watermarks in the time domain. The Chaotic Baker map is used to encrypt the water-

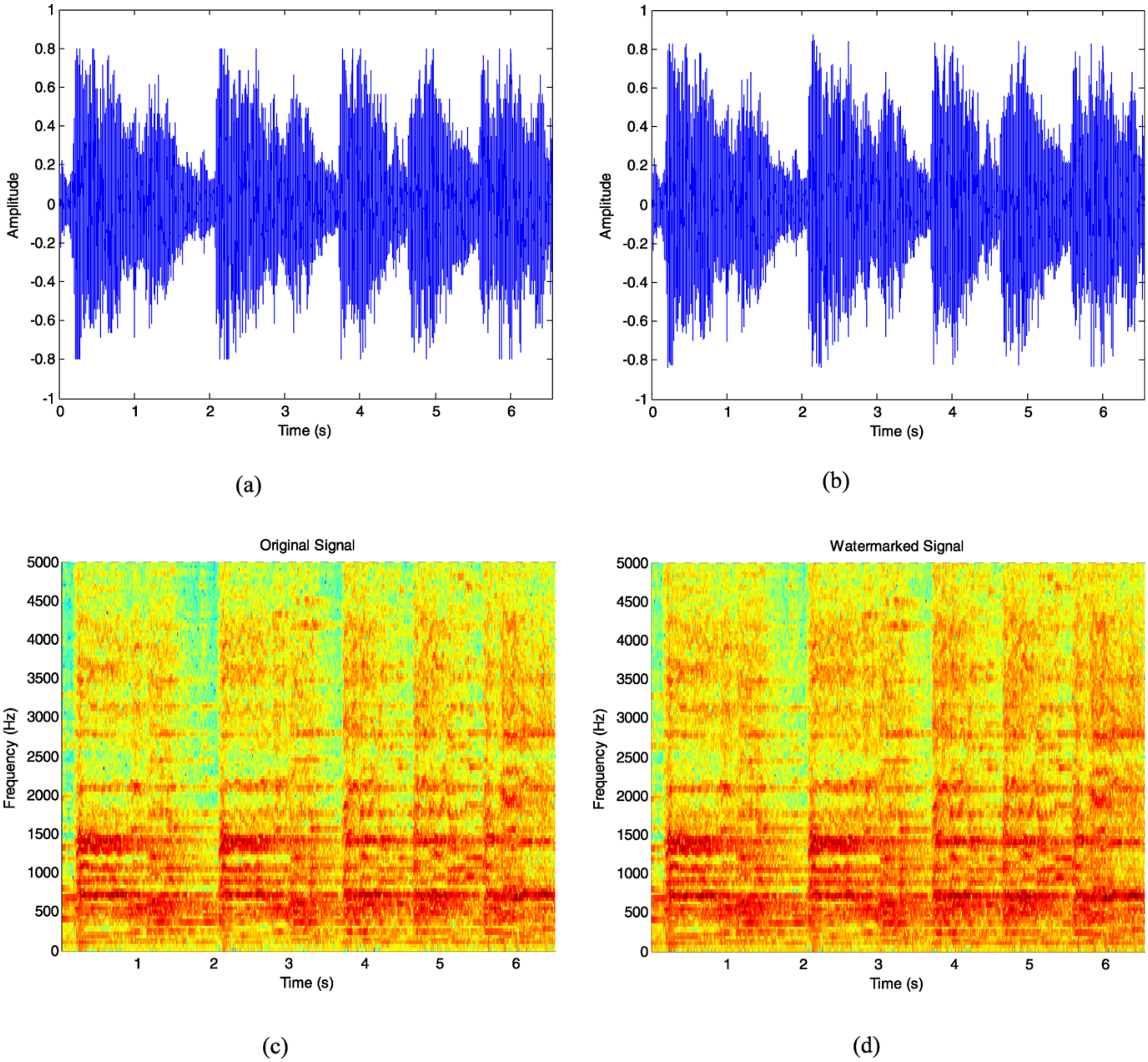


Fig. 14 (**a**) Original audio signal. (**b**) Watermarked audio signal, SNR = 27.1282 dB, SNRseg = 26.3137 dB, LLR = 0.0234, SD = 0.8373 dB. (**c**) Spectrogram of the original signal. (**d**) Spectrogram of the watermarked signal

mark image as shown in Fig. 13. The proposed SVD audio watermark embedding and extraction processes are performed in the absence of attacks and the results are shown in Figs. 14 and 15. It is clear from these results that the SVD audio watermarking with chaotic encrypted images doesn't degrade the quality of the watermarked audio signal. It is also clear that the watermark is perfectly reconstructed in the absence of any attacks.

Then, four attacks on the watermarked audio signal are studied; an additive white Gaussian noise attack, a lowpass filtering attack, a cropping attack and a wavelet compression attack. The results of watermark embedding and ex-

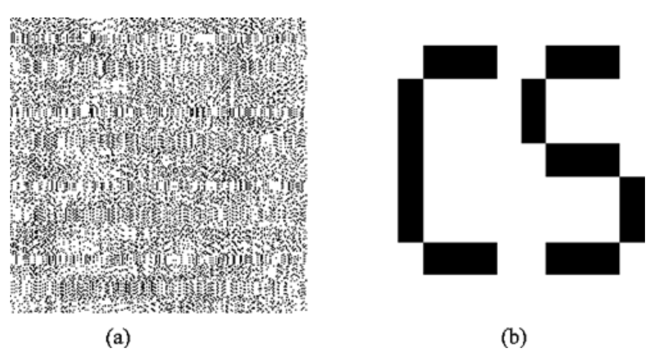


Fig. 15 Extracted watermark without attacks. (**a**) Encrypted watermark $c_r = 0.0181$. (**b**) Decrypted watermark. $c_r = 1$

Fig. 16 Results of the noise attack. (**a**) Noisy watermarked audio signal, SNR $= -10.2559$ dB, SNRseg $= -10.2899$ dB, LLR $= 0.3399$, SD $= 26.0165$ dB. (**b**) Spectrogram of the noisy watermarked signal. (**c**) Extracted encrypted watermark $c_r = -0.0065$. (**d**) Decrypted watermark after extraction. $c_r = 0.2609$

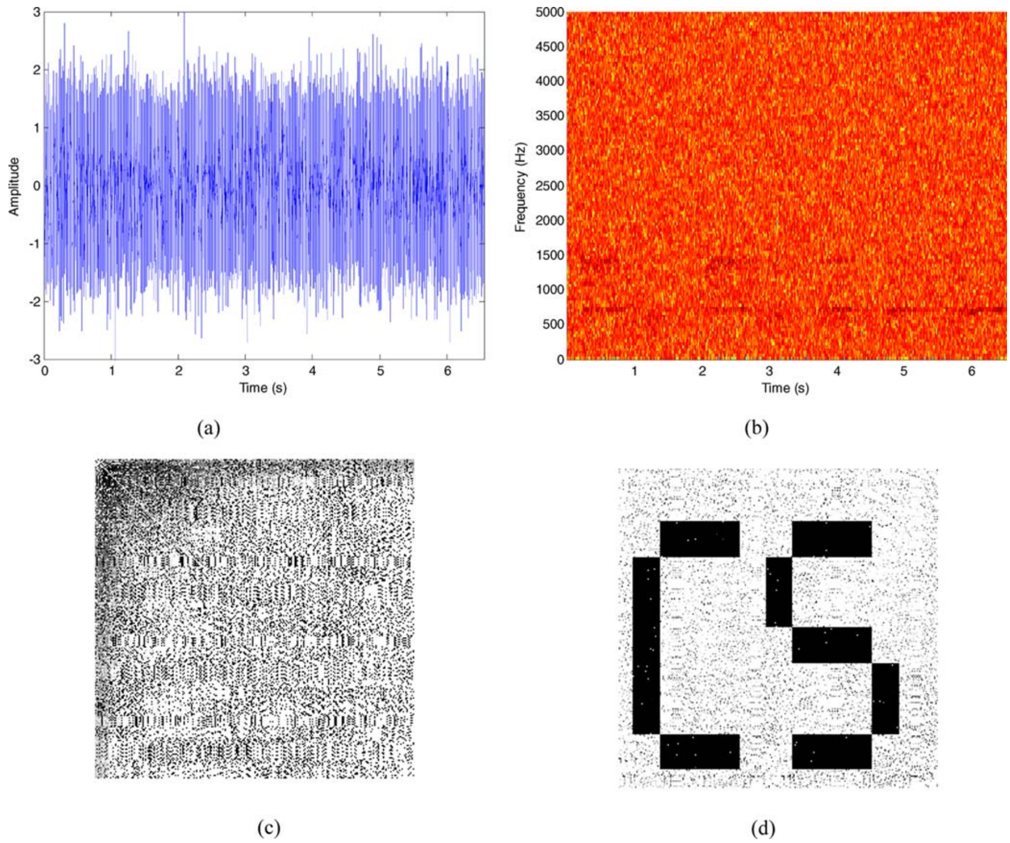
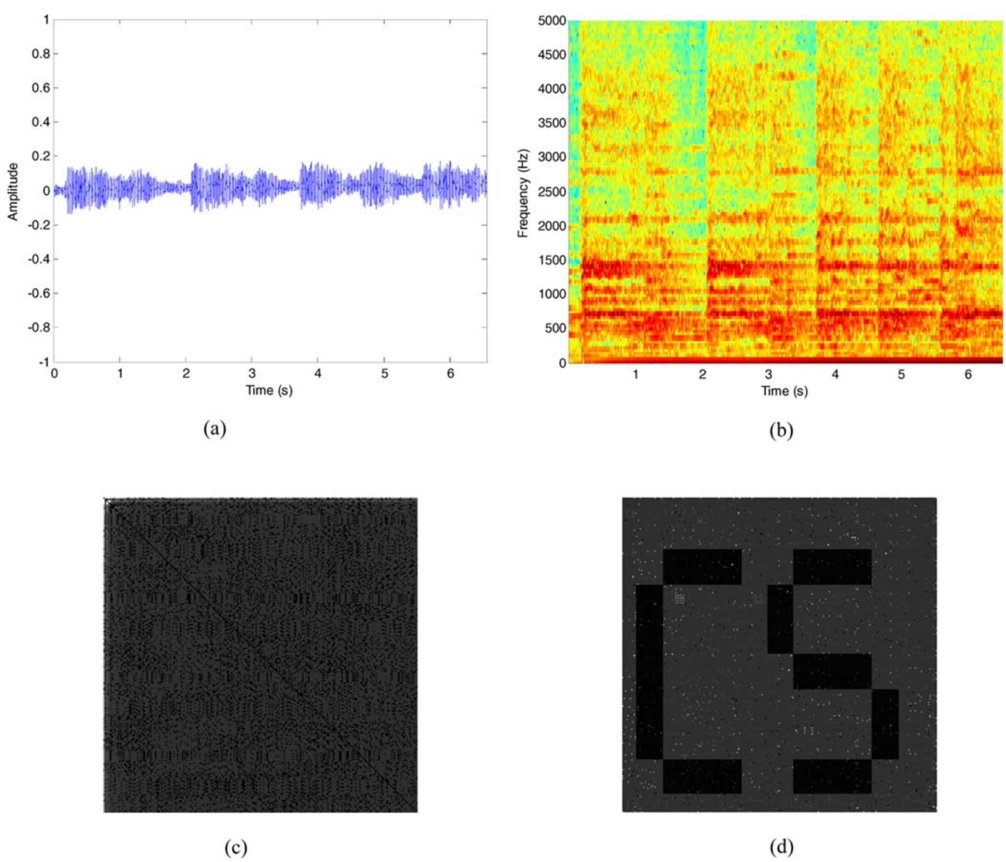


Fig. 17 Results of the lowpass filtering attack. (**a**) Watermarked audio signal, SNR $= 1.5625$ dB, SNRseg $= 1.5615$ dB, LLR $= 0.3919$, SD $= 34.6958$ dB. (**b**) Spectrogram of the watermarked signal. (**c**) Extracted encrypted watermark. $c_r = -0.0006$. (**d**) Decrypted watermark after extraction. $c_r = 0.0232$



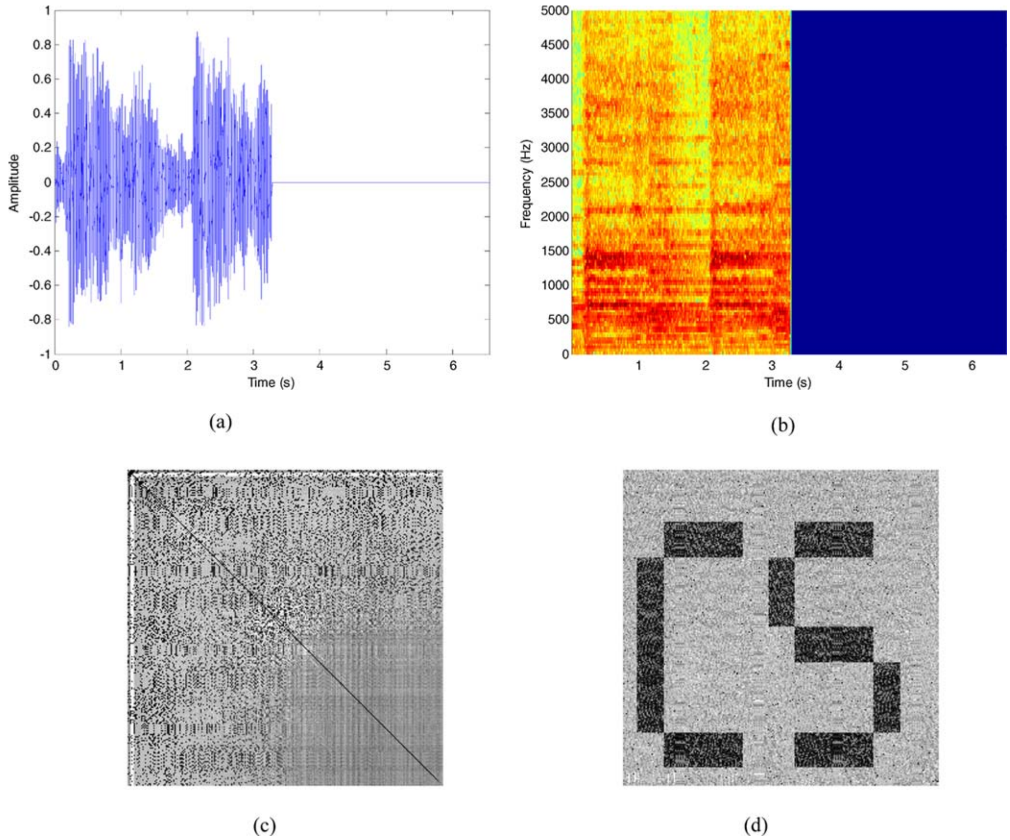


Fig. 18 Results of the cropping attack. (**a**) Watermarked audio signal, SNR = 3.0339 dB, SNRseg = 3.0006 dB, LLR = 0.2723, SD = 11.9626 dB. (**b**) Spectrogram of the watermarked signal. (**c**) Extracted encrypted watermark. $c_r = -0.0062$. (**d**) Decrypted watermark after extraction. $c_r = 0.1609$

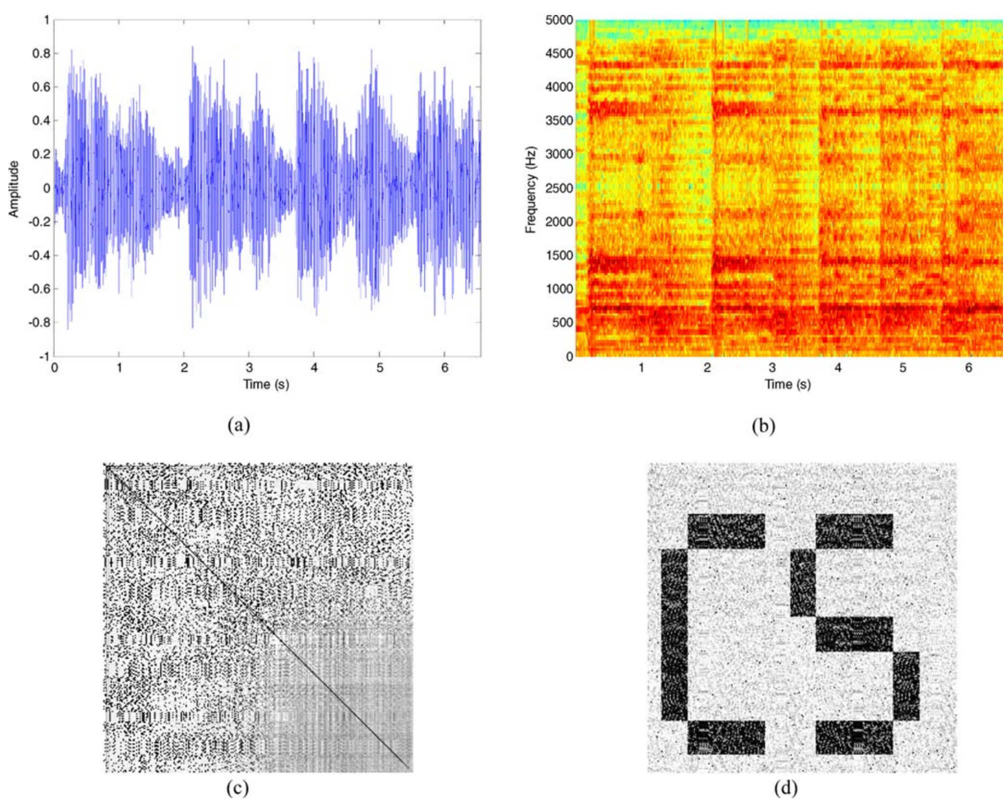


Fig. 19 Results of the lossy wavelet compression attack. (**a**) Watermarked audio signal, SNR = 9.1000 dB, SNRseg = 9.0417 dB, LLR = 0.1190, SD = 6.0649 dB. (**b**) Spectrogram of the watermarked signal. (**c**) Extracted encrypted watermark. $c_r = 0.0154$. (**d**) Decrypted watermark after extraction. $c_r = 0.5356$

traction for these four attacks are shown in Figs. 16 to 19. It is clear that in the presence of attacks, the correlation coefficient values for the extracted watermarks get lower but the watermarks are still visible after decryption.

The performance of the proposed segment-by-segment method with encrypted watermarks is tested and compared to embedding the encrypted watermark in the signal as whole. The chaotic encrypted watermark for this case is shown in Fig. 20. The results of the segment-by-segment method in the absence and presence of attacks are shown in Figs. 21 to 26. From these results, it is clear that the segment-by-segment method is similar in performance to the SVD method in the absence of attacks. It is also clear that the distortion level of the audio signals in the segment-by-segment method is still low. Another observation is that the segment-by-segment method increases the correlation coefficient between the extracted watermark and the original watermark in the presence of all attacks and enables a perfect reconstruction in the presence of the cropping attack.

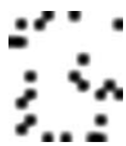


Fig. 20 Encryption of the block watermark used for segment-by-segment watermarking. $c_r = -0.0572$

Figures 27 to 29 show a comparison between the SVD watermarking method and the segment-by-segment method in the presence of the noise attack, the filtering attack and the wavelet compression attack, respectively. From the re-

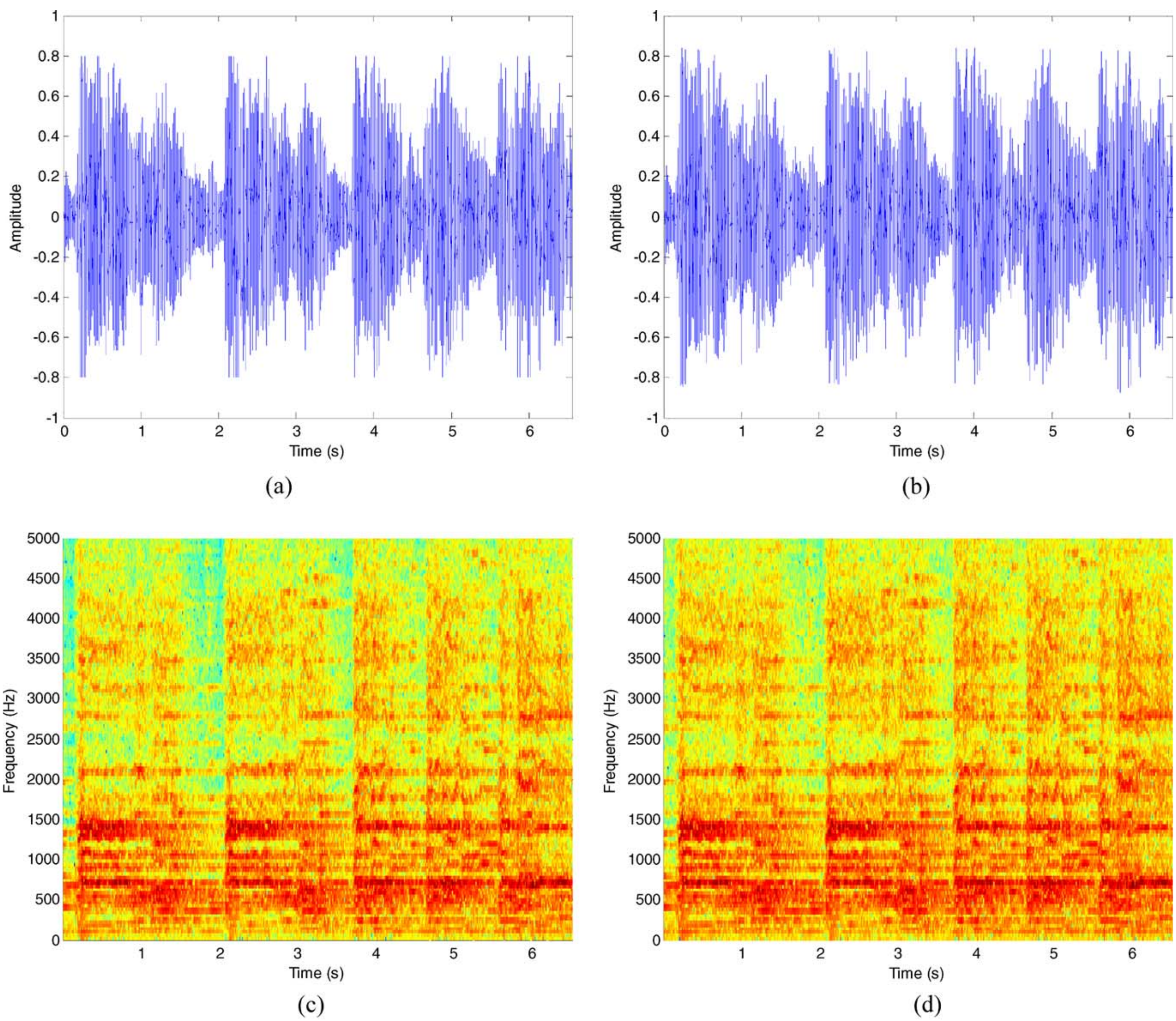


Fig. 21 (**a**) Original audio signal. (**b**) Segment-by-segment watermarked audio signal, SNR = 21.3723 dB, SNRseg = 21.2889 dB, LLR = 0.0432, SD = 1.6345 dB. (**c**) Spectrogram of the original signal. (**d**) Spectrogram of the watermarked signal

sults in these figures, it is clear that for a low SNR environment, the segment-by-segment method is preferred because it increases the detection probability of the watermark. For a filtering attack, although the extracted watermark has a low correlation coefficient because the filter removes most of the signal details, the segment-by-segment method has a better performance than the SVD method. For the wavelet compression attack, it is clear that as the threshold below which, wavelet coefficients are neglected increases, the segment-by-segment method achieves a better performance than the SVD method. This is attributed to the high probability that at least one of the several watermarks will not be affected by the compression process.

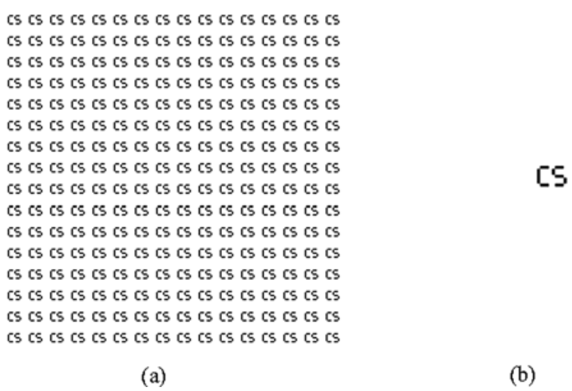


Fig. 22 Extracted watermarks without attacks. (**a**) Watermarks after decryption. (**b**) Decrypted watermark which has $c_{r\,\max} = 1$

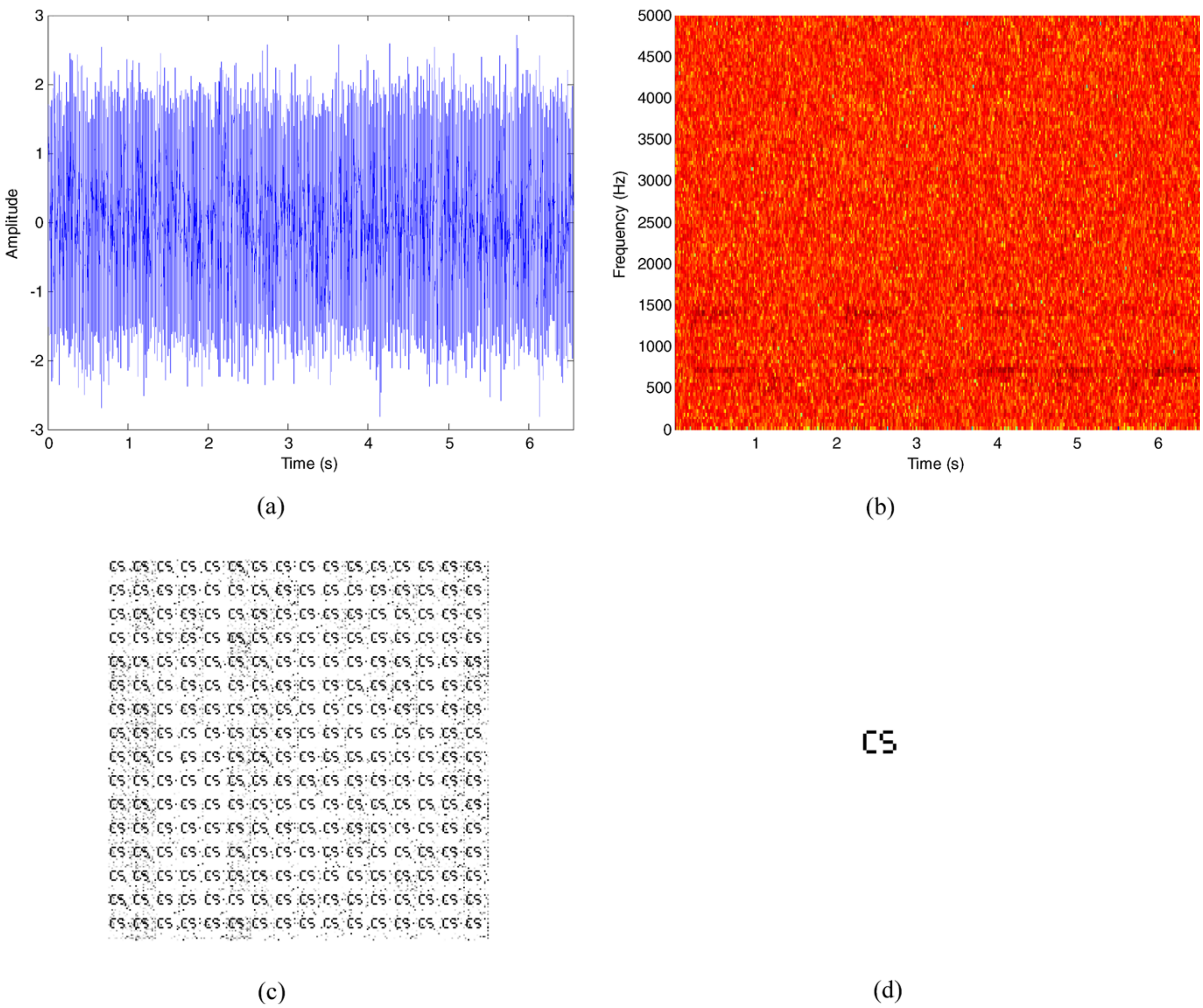


Fig. 23 Results of the noise attack. (**a**) Noisy watermarked audio signal, SNR = −10.5382 dB, SNRseg = −10.5728 dB, LLR = 0.3355, SD = 26.5315 dB. (**b**) Spectrogram of the noisy watermarked signal. (**c**) Extracted watermarks. (**d**) Extracted watermark which has $c_{r\,\max} = 0.3355$

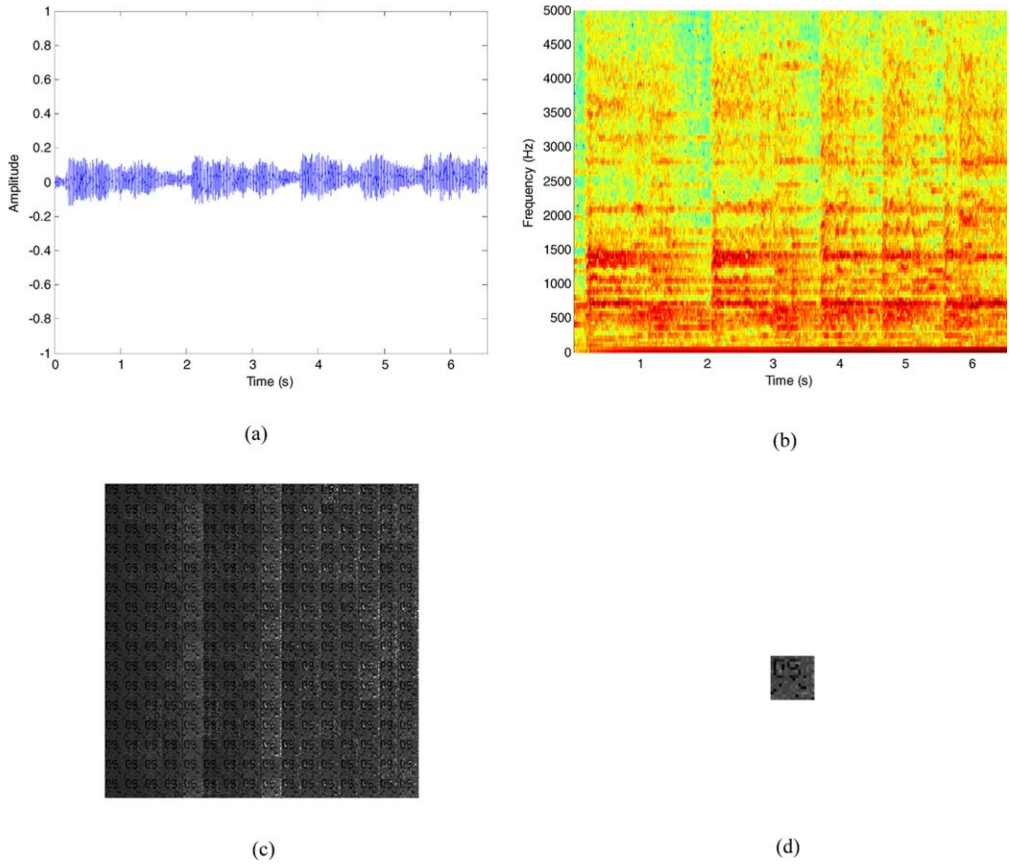


Fig. 24 Results of the filtering attack. (**a**) Watermarked audio signal, SNR = 1.6045 dB, SNRseg = 1.6029 dB, LLR = 0.3891, SD = 34.0978 dB. (**b**) Spectrogram of the watermarked signal. (**c**) Extracted watermarks. (**d**) Extracted watermark which has $c_{r\,\max} = 0.0743$

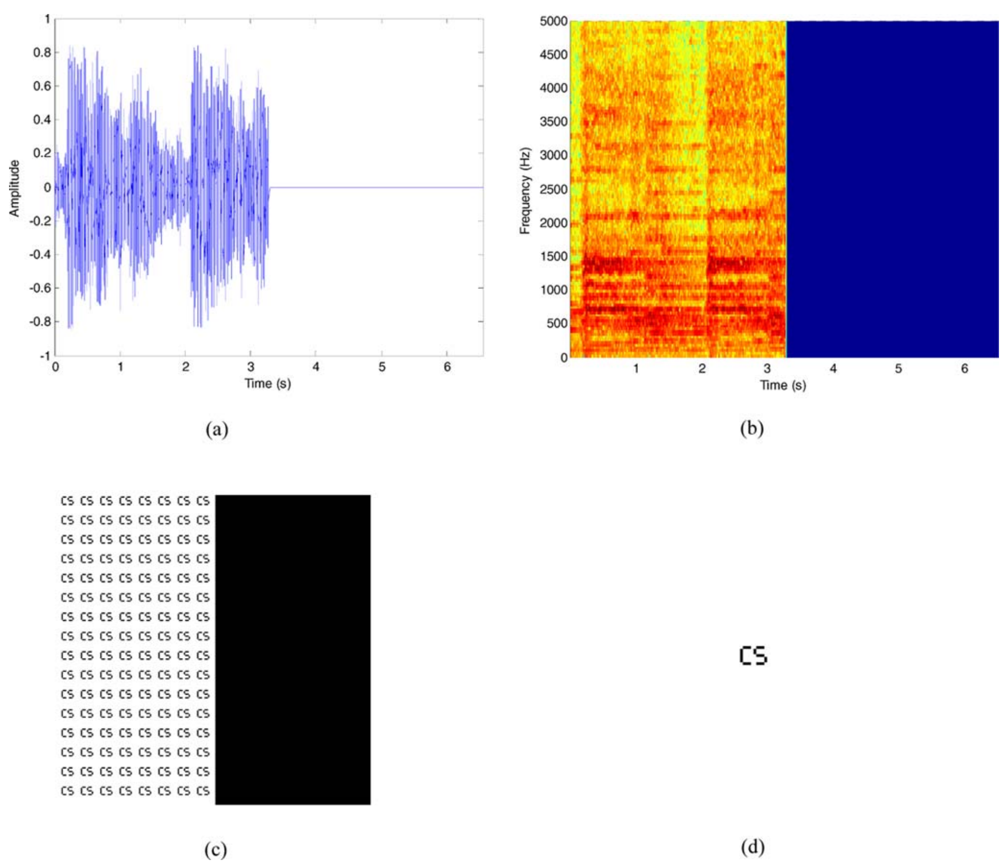
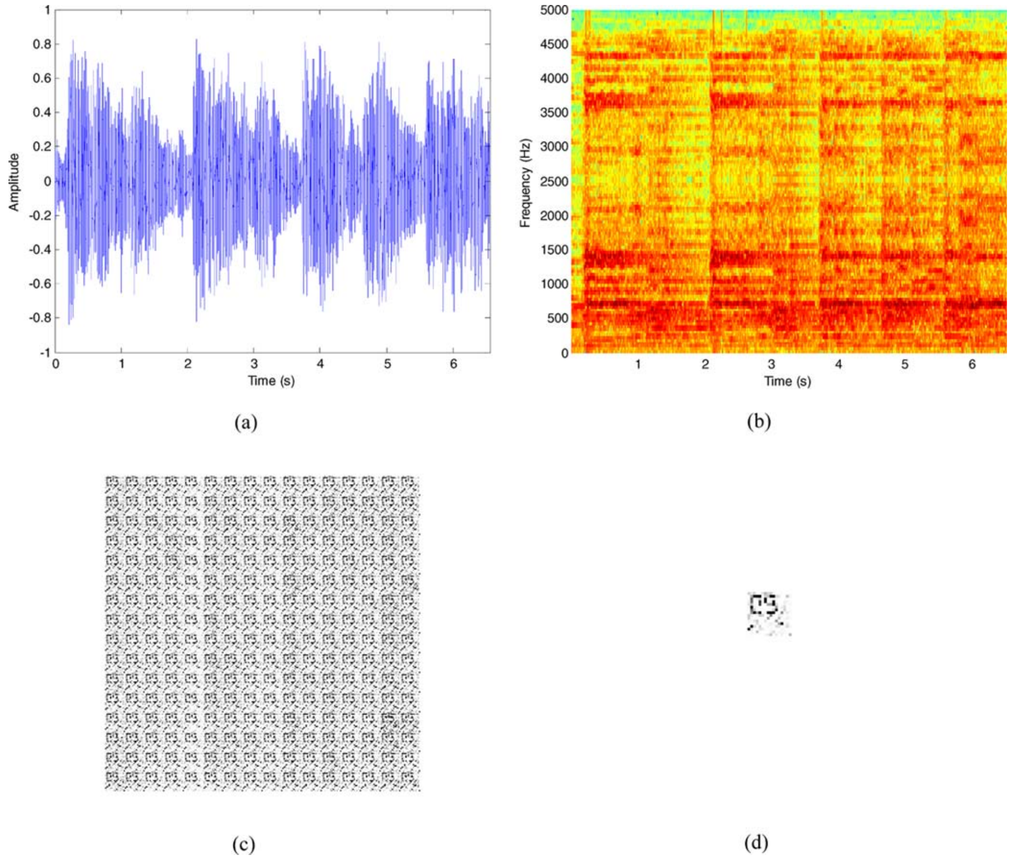


Fig. 25 Results of the cropping attack. (**a**) Watermarked audio signal, SNR = 3.0096 dB, SNRseg = 2.9767 dB, LLR = 0.2602, SD = 11.7413 dB. (**b**) Spectrogram of the watermarked signal. (**c**) Extracted watermarks. (**d**) Extracted watermark which has $c_{r\,\max} = 1$

Fig. 26 Results of the wavelet compression attack. (**a**) Watermarked audio signal, SNR = 8.9637 dB, SNRseg = 8.9100 dB, LLR = 0.1583, SD = 6.0448 dB. (**b**) Spectrogram of the watermarked signal. (**c**) Extracted watermarks. (**d**) Extracted watermark which has $c_{r\,\max} = 0.7157$



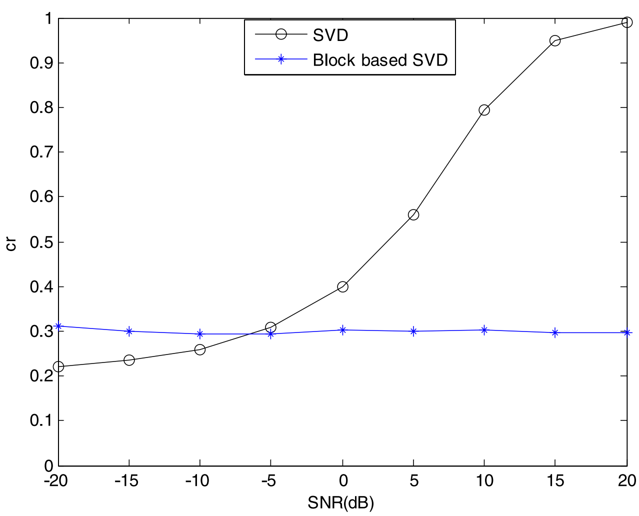


Fig. 27 Correlation coefficient between the extracted watermark and the original watermark vs. the SNR for both the SVD and the segment-by-segment SVD audio watermarking methods in the presence of a noise attack

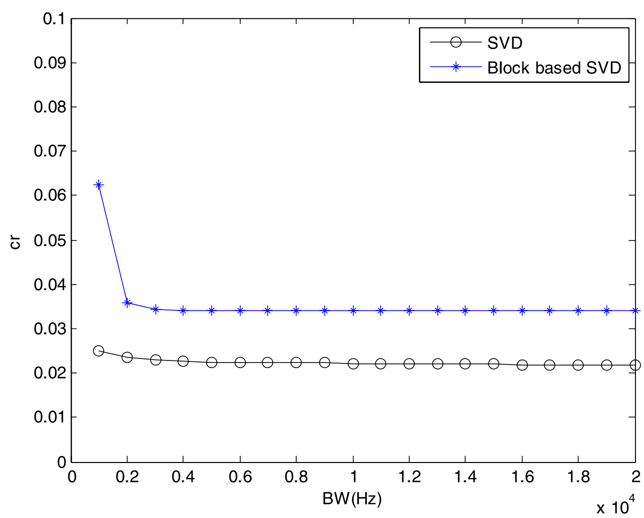


Fig. 28 Correlation coefficient between the extracted watermark and the original watermark vs. the filter BW for both the SVD and the segment-by-segment SVD audio watermarking methods in the presence of a filtering attack

9 Conclusions

This paper presented an efficient SVD audio watermarking algorithm, which can be implemented on audio signals in time domain or in an appropriate transform domain. The time domain is the preferred medium for watermark embedding because of the lowest distortion level associated with this domain and the high ability to extract watermarks in the presence of attacks. In this algorithm, encrypted images are embedded as watermarks in audio signals to achieve a high

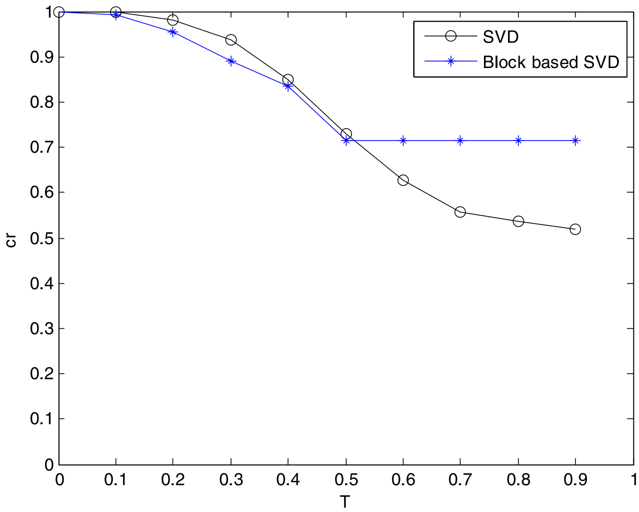


Fig. 29 Correlation coefficient between the extracted watermark and the original watermark vs. the wavelet compression threshold for both the SVD and the segment-by-segment SVD audio watermarking methods in the presence of a wavelet compression attack

degree of security. The proposed audio watermarking algorithm can be implemented on the audio signal as a whole or on a segment-by-segment basis. Experimental results proved that watermark embedding in the proposed algorithm does not deteriorate the audio signals. It has been clear through experiments that the chaotic Baker map encryption algorithm is an efficient algorithm for watermark encryption, because it is a permutation based algorithm. A comparison study between the two implementations of the proposed audio watermarking algorithm has shown that the segment-by-segment implementation achieves a higher detectability than the implementation on the signal as a whole.

References

Chu, W. C. (2003). DCT-based image watermarking using subsampling. *IEEE Transactions on Multimedia*, *5*(1), 34–38.

Crochiere, R. E., Tribolet, J. E., & Rabiner, L. R. (1980). An interpretation of the log likelihood ratio as a measure of waveform coder performance. *IEEE Transactions on Acoustics, Speech, And Signal Processing*, *28*(3), 318–323.

Erküçük, S., Krishnan, S., & Glu, M. Z. (2006). A robust audio watermark representation based on linear chirps. *IEEE Transactions on Multimedia*, *8*(5), 925–936.

Fridrich, J. (1997). Image encryption based on chaotic maps. In *Proceedings of the IEEE international conference on systems, man, and cybernetics* (pp. 1105–1110).

Ghouti, L., Bouridane, A., Ibrahim, M. K., & Boussakta, S. (2006). Digital image watermarking using balanced multiwavelets. *IEEE Transactions on Signal Processing*, *54*(4), 1519–1536.

Guillemain, P., & Martinet, R. K. (1996). Characterization of acoustic signals through continuous linear time-frequency representations. *Proceedings of the IEEE*, *84*(4), 561–585.

Han, F., Yu, X., & Han, S. (2006). Improved Baker map for image encryption. In *Proceedings of the first international symposium on systems and control in aerospace and astronautics, ISSCAA 2006* (pp. 1273–1276).

Huang, F., & Lei, F. (2008). A novel symmetric image encryption approach based on a new invertible two-dimensional map. In *Proceedings of the international conference on intelligent information hiding and multimedia signal processing IIHMSP 2008* (pp. 1340–1343).

Kim, H. S., & Lee, H. K. (2003). Invariant image watermark using Zernike moments. *IEEE Transactions on Circuits and Systems for Video Technology*, *13*(8), 766–775.

Koduru, S. C., & Chandrasekaran, V. (2008). Integrated confusion-diffusion mechanisms for chaos based image encryption. In *IEEE 8th international conference on computer and information technology workshops* (pp. 260–263).

Kubichek, R. (1993). Mel-cepstral distance measure for objective speech quality assessment. In *Proceedings of the IEEE pacific rim conference on communications, computers and signal processing* (pp. 125–128).

Lemma, A. N., Aprea, J., Oomen, W., & de Kerkhof, L. V. (2003). A temporal domain audio watermarking technique. *IEEE Transactions on Signal Processing*, *51*(4), 1088–1097.

Li, W., Xue, X., & Lu, P. (2006). Localized audio watermarking technique robust against time-scale modification. *IEEE Transactions on Multimedia*, *8*(1), 60–69.

Lim, J. S. (1990). *Two-dimensional signal and image processing*. New York: Prentice Hall.

Liu, Z., & Inoue, A. (2003). Audio watermarking techniques using sinusoidal patterns based on pseudorandom sequences. *IEEE Transactions on Circuits and Systems for Video Technology*, *13*(8), 801–812.

Liu, R., & Tan, T. (2002). An SVD-based watermarking scheme for protecting rightful ownership. *IEEE Transactions on Multimedia*, *4*(1), 121–128.

Lu, Z. M., Xu, D. G., & Sun, S. H. (2005). Multipurpose image watermarking algorithm based on multistage vector quantization. *IEEE Transactions on Image Processing*, *14*(6), 822–831.

Macq, B., Dittmann, J., & Delp, E. J. (2004). Benchmarking of image watermarking algorithms for digital rights management. *Proceedings of the IEEE*, *92*(6), 971–984.

McDermott, B. J., Scaglia, C., & Goodman, D. J. (1978). Perceptual and objective evaluation of speech processed by adaptive differential PCM. In *Proceedings of the IEEE international conf. on acoustic, speech and signal processing (ICASSP)* (pp. 581–585).

Özer, H., & Sankur, B. (2005). An SVD based audio watermarking technique. In *Proceedings of the IEEE 13th conference on signal processing and communications applications* (pp. 452–455).

Pratt, W. K. (1991). *Digital image processing*. New York: Wiley.

Prochazka, A., Uhlir, J., Rayner, P. J. W., & Kingsbury, N. J. (1998). *Signal analysis and prediction*. Basel: Birkhauser.

Qian, Q., Chen, Z., & Yuan, Z. (2008). Video compression and encryption based-on multiple chaotic system. In *Proceedings of the 3rd international conference on innovative computing information and control (ICICIC'08)*.

Sun, X., Liu, J., Sun, J., Zhang, Q., & Ji, W. (2008). A robust image watermarking scheme based-on the relationship of SVD. In *Proceedings of the international conference on intelligent information hiding and multimedia signal processing*.

Usman, K., Juzojil, H., & Nakajimal, I. (2007). Medical image encryption based on pixel arrangement and random permutation for transmission security. In *Proceedings of the 9th international conference on e-health networking, application and services* (pp. 244–247).

Walker, J. S. (1999). *A primer on wavelets and their scientific applications*. Boca Raton: CRC Press.

Wang, S., Sekey, A., & Gersho, A. (1992). An objective measure for predicting subjective quality of speech coders. *IEEE Journal on Selected Areas in Communication*, *10*(5), 819–829.

Wang, X., Qi, W., & Niu, P. (2007). A new adaptive digital audio watermarking based on support vector regression. *IEEE Transactions on Audio, Speech, and Language Processing*, *15*(8), 2270–2277.

Wornell, G. W. (1996). Emerging applications of multirate signal processing and wavelets in digital communications. *Proceedings of the IEEE*, *84*(4), 586–603.

Xiang, S., & Huang, J. (2007). Histogram-based audio watermarking against time-scale modification and cropping attacks. *IEEE Transactions on Multimedia*, *9*(7), 1357–1372.

Yang, W., Benbouchta, M., & Yantorno, R. (1998). Performance of the modified bark spectral distortion as an objective speech quality measure. In *Proceedings of the IEEE international conf. on acoustic, speech and signal processing (ICASSP)*, Washington, USA (Vol. 1, pp. 541–544).

Zhu, X., Zhao, J., & Xu, H. (2006). A digital watermarking algorithm and implementation based on improved SVD. In *Proceedings of the IEEE 18th international conference on pattern recognition (ICPR'06)*.