

# An Audio Watermarking Scheme Based on Singular-Spectrum Analysis

Jessada Karnjana<sup>1,2(✉)</sup>, Masashi Unoki<sup>1</sup>, Pakinee Aimmanee<sup>2</sup>,  
and Chai Wutiwiwatchai<sup>3</sup>

<sup>1</sup> School of Information Science, Japan Advanced Institute of Science  
and Technology, 1-1 Asahidai, Ishikawa, Nomi 923-1292, Japan  
{jessada,unoki}@jaist.ac.jp

<sup>2</sup> Sirindhorn International Institute of Technology, Thammasat University,  
131 Moo 5, Tiwanont Road, Bangkadi, Pathumthani 12000, Muang, Thailand  
pakinee@siit.tu.ac.th

<sup>3</sup> National Electronics and Computer Technology Center, 112 Thailand Science Park,  
Phahonyothin Road, Pathumthani 12120, Khlong Luang, Thailand  
chai.wutiwiwatchai@nectec.or.th

**Abstract.** This paper proposes a blind audio watermarking scheme based on singular-spectrum analysis (SSA) which relates to several techniques based on singular value decomposition (SVD). SSA is used to decompose a signal into several additive oscillatory components where each component represents a simple oscillatory mode. The proposed scheme embedded a watermark into a host signal by modifying scaling factors of certain components of the signal. Test results show that the proposed scheme satisfies imperceptibility criterion suggested by IHC with the average ODG of 0.18. It is robust against many attacks, such as MP3 and MP4 compression, band-pass filtering, and re-sampling. This paper does not only propose a new watermarking scheme, it also discusses the singular value and reveals its meaning, which has been deployed and played an important role in all SVD-based schemes.

**Keywords:** Singular-spectrum analysis · Singular value decomposition · Singular value · Audio watermarking

## 1 Introduction

Since the late 20th century, the Internet has grown at unprecedented rate, and the spread of digital-multimedia transfer has increased exponentially, especially after we had advanced technologies in data compression, more powerful personal computers, and social communities. This communication channel is somehow a double-edge sword. On one hand, it is very useful and convenient to access gigantic source of data. On the other hand, it could be harmful when it is misused. For example, there are social concerns about authentication, copyright management, copy control, and the like of digital media. These issues have been aware

in music industry since 1990s [1]. One of the potential solutions solving this kind of problems in audio domain is to use an audio-watermarking technique.

Watermarking is a scheme of making information unnoticeable; therefore a user is not aware the existence of hidden information. This type of information hiding is called steganography. The other type is cryptography. Both are information hiding techniques; nevertheless, they have a different objective. While steganography deals with making information unnoticeable, the cryptography tries to make it unreadable. The use of watermarking for copyright control is claimed to be the original goal for audio watermark [2]. Besides copyright management and the like, audio watermarking could be used in many other applications where each has its own requirements. There are at least three areas of application, as stated in [2], for audio watermarking: (1) copyright marking and copy control, (2) forensic watermarking, such as digital fingerprinting, and (3) information hiding for annotation and added value. Since the main objective of audio watermarking is to add information into an audio signal transparently, in general, especially for commercial purposes, there are four requirements for an audio-watermarking method: (1) *Inaudibility*: human auditory system should not be able to detect a watermark, thus sound quality comparing between original and watermarked signals should be equal. (2) *Robustness*: it is an ability to survive of a watermark after attacks, such as re-sampling or compression, are applied to the watermarked signal. (3) *Blindness*: it is a property of extracting hidden information from a watermarked signal without the presence of the original in extracting processes. (4) *Confidentiality*: it is a property of concealment of hidden data [3]. In [2, 4], the authors have added an additive requirement which is a capacity or quantity of information embedded in the original signal. For example, for the purpose of value adding, increasing capacity is an important issue since the hidden information is not just a short serial number. Naturally, these requirements conflict each other's. The high robustness, for example, normally comes with the cost of low audio quality or semi-transparency. Work of name of author of [4] also concludes similar phenomenal that high capacity implies low robustness. Therefore, in addition to proposing a new effective technique many researches on audio watermarking have focused on how to compromise these conflicts. However, there is no work that can completely solve these problems. Some techniques are good at transparency or inaudibility, but not blind. Some techniques are good at high capacity, but not robust, especially for the certain attacks.

On classification of audio watermarking schemes, name of the author in [5] had categorized watermarking schemes into three categories. The first embeds information in time domain by mostly changing the least significant bit (LSB) of an original signal. The second introduces an echo, and the third embeds information in a certain transform domain such as frequency or wavelet domain. However, classification depends upon a set of criteria. For example, it could be classified into two categories: the one that deploys properties of the human auditory system and the one that does not. Some techniques deploy characteristics of human auditory system as a guideline for hiding information [6]. The others are

based on mathematical manipulation and do not much rely on special characteristics of human auditory system [7]. In this paper, we investigate the later kind and propose a watermarking scheme based on singular-spectrum analysis (SSA), which relates to singular value decomposition-based (SVD-based) watermarking techniques. These techniques rely on a mathematical method of extracting algebraic features called *singular values* (SVs) from a two-dimensional matrix.

The SVD-based watermarking has a lot of advantages [7–23]. The advantages are mainly due to properties of SVs, such as the invariance of SVs under common signal processing operations. For example, when a small modification is applied to the original signal represented by a two-dimensional matrix, its SVs changes unnoticeably. This property of SVs makes the SVD-based techniques robust against the common signal processing. Besides, it has a low computational complexity comparing with other transform-based methods.

The SVD-based audio watermarking is originally proposed by Özer et al. in 2005 [8] and based on the watermarking technique employed in the image domain [11]. From our survey, the SVD-based methods could be categorized into two frameworks based on the stage in extracting processes that it is employed information of a watermark which is from the embedding stage. For example, methods described in [8–10, 21] use information of a watermark in the extracting stage, therefore they are necessary non-blind methods. Those described in [7, 12–15, 17–20, 22, 23] do not use information of a watermark in extracting processes. They can be blind or non-blind. For example, [7, 15, 22, 23] are blind while the others are non-blind. Even though there are two frameworks, both share a common concept of SVD-based watermarking, i.e. SVs are modified slightly according to embedding rules. We can also use the positions of modified SVs as a criterion for categorization. For example, methods described in [12–14, 18–20] modify only the largest SV, but the method described in [23] modifies only some small SVs. Methods in [7, 15, 17, 21, 22] modify all SVs. Section 5 shows the effect of the position of modified SVs on watermarked-sound quality. It is important to note here that Lamarche et al. have pointed out that the robustness of those in the first framework are highly likely due to false positive rate [11]. Therefore, to avoid such kind of problems, our proposed method is designed to be blind.

Interestingly, to the best of our knowledge, there is not much discussion about the essence or meaning of SVs in previously SVD-based audio-watermarking techniques. When SV is modified, which audio feature is exactly modified. The answer does not depends only on the domain representing a signal, we believe, but also on how a matrix is created. From the view point of SSA, the meaning of SVs could be clearly explained. We are inspired by the robustness of SVD-based methods and intrigued by the question of physical meaning of SV in the hope that knowing the meaning and its relation with other physical features could help us to overcome the conflicts in requirements of audio watermarking.

The rest of this paper is organized as follows. Section 2 introduces the background of SSA and SVD. Our proposed method is detailed in Sect. 3. Performance evaluation and experimental results are given in Sect. 4. Remarks are made in Sect. 5. Section 6 summarizes this paper.

## 2 Singular-Spectrum Analysis

Singular-spectrum analysis (SSA) is a method of identifying and extracting oscillatory components from a signal [24]. There are many types of SSAs. The SSA we are going to describe in this section is called Basic SSA of which our proposed method is based on. SSA is used to decompose a signal (or time series) of interest into several additive oscillatory components. Each oscillatory component represents a simple oscillatory mode. We hypothesize that the relationship between SSA and oscillatory components is somehow similar to that between Empirical Mode Decomposition (EMD) and intrinsic mode functions.

SSA consists of two stages which involve analysis and synthesis. The decomposition stage has two decomposition steps which are *embedding* and *singular value decomposition* (SVD). It should be noted that the name of the first step has nothing to do with embedding a watermark. It is the SSA terminology. In reconstruction stage, there are also two steps which are *grouping* and *diagonal averaging*.

In the *embedding* step, a signal  $X = (f_0, f_1, \dots, f_{N-1})^T$  of length  $N$  is mapped to a trajectory matrix  $\mathbf{X}$  of size  $L \times K$ .

$$\mathbf{X} = \begin{bmatrix} f_0 & f_1 & f_2 & \cdots & f_{K-1} \\ f_1 & f_2 & f_3 & \cdots & f_K \\ f_2 & f_3 & f_4 & \cdots & f_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{L-1} & f_L & f_{L+1} & \cdots & f_{N-1} \end{bmatrix} \quad (1)$$

Each column vector of  $\mathbf{X}$  is called a lagged vector. The  $i$ th column of  $\mathbf{X}$ ,  $X_i$ , is defined as  $X_i = (f_{i-1}, f_i, \dots, f_{i+L-2})^T$ , where  $L$  is the *window length* which has a maximum value of  $N$ . Therefore, the matrix  $\mathbf{X}$  consists of  $K = N - L + 1$  lagged vectors, i.e.  $\mathbf{X} = [X_1 X_2 \dots X_K]$ . Note that  $L$  is the only parameter of Basic SSA. Since the trajectory matrix  $\mathbf{X}$  has equal elements on the diagonals, i.e.  $x_{i,j} = x_{i-1,j+1}$  where  $x_{i,j}$  is an element at  $i$ th row and  $j$ th column of  $\mathbf{X}$ , it is considered to be a Hankel matrix.

For the second step, *SVD* is a step that decomposes a matrix  $\mathbf{X}$  into a product of three matrices  $\mathbf{U}$ ,  $\mathbf{D}$ , and  $\mathbf{V}$  with the following relationship.

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T \quad (2)$$

where  $\mathbf{X}$  is a matrix being decomposed (in this case, it is the trajectory matrix from the first step),  $\mathbf{U}$  and  $\mathbf{V}$  are orthogonal matrices, i.e.  $\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{V}\mathbf{V}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I}$  where  $\mathbf{I}$  is the identity matrix, and  $\mathbf{D}$  is a diagonal matrix whose element is called singular value (SV). Columns of  $\mathbf{U}$  and  $\mathbf{V}$ ,  $U_i$  and  $V_i$ , which are sorted in descending order of corresponding eigenvalues, are eigenvectors of  $\mathbf{X}\mathbf{X}^T$  and  $\mathbf{X}^T\mathbf{X}$  respectively. Then, the elements of  $\mathbf{D}$  are the square root of the eigenvalues. If the eigenvalues of  $\mathbf{X}\mathbf{X}^T$  (or  $\mathbf{X}^T\mathbf{X}$ ) is denoted by  $\lambda_1, \lambda_2, \dots$ , and  $\lambda_L$ , then the trajectory matrix  $\mathbf{X}$  can be written as

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \dots + \mathbf{X}_d \quad (3)$$

where  $\mathbf{X}_i = \sqrt{\lambda_i}U_iV_i^T$  and  $d = \max\{i, \text{ such that } \lambda_i > 0\}$ . Each  $\mathbf{X}_i$  represents a simple oscillatory component of the signal  $X$ .

The third step is *grouping*. In this step, the set of indices  $\{1, 2, \dots, d\}$  obtained from the previous step is partitioned into  $m$  disjoint subsets  $I_1, I_2, \dots, I_m$ . Then,  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_d$  are grouped into  $m$  groups.

$$\mathbf{X} = \mathbf{X}_{I_1} + \mathbf{X}_{I_2} + \dots + \mathbf{X}_{I_m} \tag{4}$$

Since the purpose of this step is to separate the time series into meaningful additive sub-series such as trend or noise, according to separability conditions, which is not our watermarking purpose, the step is not included in our proposed method.

The last step is *diagonal averaging*. This last step transforms (*hankelizes*) each matrix  $\mathbf{X}_{I_j}$  of the grouped decomposition into a new series of length  $N$ . The hankelization of matrix  $\mathbf{Y}$  of dimension  $L \times K$  to the series  $Y = (g_0, g_1, \dots, g_{N-1})^T$  is defined as follows.

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} y_{m,k-m+2}^* & \text{for } 0 \leq k < L^* - 1 \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m,k-m+2}^* & \text{for } L^* - 1 \leq k < K^* \\ \frac{1}{N-k} \sum_{m=k-K^*+2}^{N-K^*+1} y_{m,k-m+2}^* & \text{for } K^* \leq k < N \end{cases} \tag{5}$$

where  $L^* = \min(L, K)$ ,  $K^* = \max(L, K)$ ,  $y_{ij}^* = y_{ij}$  if  $L < K$ , and  $y_{ij}^* = y_{ji}$  if  $L \geq K$ . In our proposed method,  $\mathbf{Y}$  is a watermarked trajectory matrix, which is a trajectory matrix after its SVs are modified.

### 3 Proposed Method

#### 3.1 Embedding Process

The embedding process is illustrated in Fig. 1 (left). First, an audio signal is segmented into non-overlapping frames. One bit of the watermark will be embedded into one frame. This also implies that the frame length determines embedding capacity. Then, trajectory matrices representing each frame are created, and SVD is applied on each matrix. A watermark bit is embedded by modifying SVs according to certain rules. In this work, we use simple rules similar to quantization index modulation. The rules can be summarized as follows.

Let  $\{\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_d}\}$  be a set of SVs in descending order, where  $d = \max\{i, \text{ such that } \lambda_i > 0\}$ , and  $\epsilon$  be a small real positive number. We use the following criterion to modify values of SVs. If the watermark bit is 0, then  $\sqrt{\lambda_u}, \sqrt{\lambda_{u+1}}, \dots$ , and  $\sqrt{\lambda_l}$  are replaced with  $(1 + \epsilon)\sqrt{\lambda_{l+1}}$ , and if the watermark bit is 1, then  $\sqrt{\lambda_u}, \sqrt{\lambda_{u+1}}, \dots$ , and  $\sqrt{\lambda_l}$  are replaced with  $(1 - \epsilon)\sqrt{\lambda_{u-1}}$  given that  $\sqrt{\lambda_u}$  is greater than  $\sqrt{\lambda_l}$ .

After modifying SVs, each modified trajectory matrix is hankelized. Finally, the watermarked signal is obtained by adding those hankelized frames.

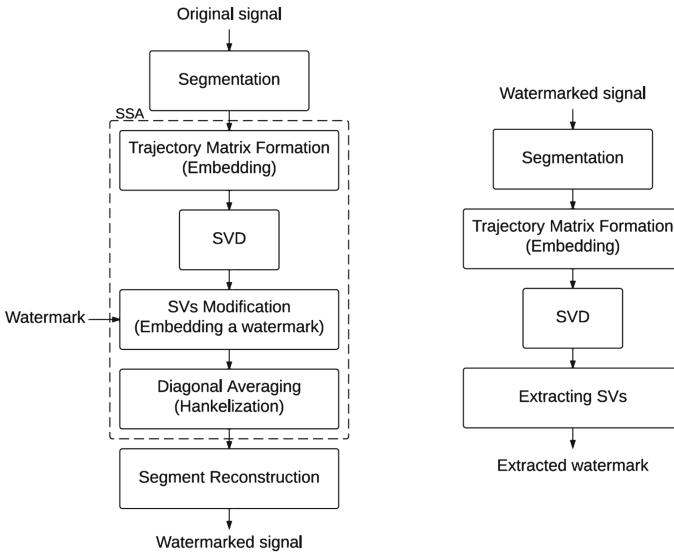


Fig. 1. Embedding and extracting processes

### 3.2 Extracting Process

The proposed extracting process is shown in Fig. 1 (right). The watermarked signal is segmented into non-overlapping frames. In the same way as embedding process described above, each frame is mapped to a trajectory matrix, and use SVD to extract SVs. The watermark bit is decoded by determining the value of  $\sqrt{\lambda_m}$ , where  $\sqrt{\lambda_m}$  is the median of  $\{\sqrt{\lambda_u}, \sqrt{\lambda_{u+1}}, \dots, \sqrt{\lambda_l}\}$ . If  $\sqrt{\lambda_{u-1}} - \sqrt{\lambda_m}$  is greater than  $\sqrt{\lambda_m} - \sqrt{\lambda_{l+1}}$ , the watermark bit is 1. Otherwise, the watermark bit is 0.

## 4 Evaluation

Following the evaluation criteria suggested by the committee of Information Hiding and its Criteria (IHC) [25], this work is evaluated in two major dimensions which are objective sound-quality tests of watermarked signals and robustness tests. The perceptual evaluation of audio quality (PEAQ) which is recommended by ITU-R-BS.1387-1 is used to measure the objective different grade (ODG). PEAQ measures the degradation of the watermarked signal being evaluated comparing with the original one and covers a scale from  $-5$  (worst) to  $0$  (best). IHC suggests that the ODG should be greater than  $-2.5$ .

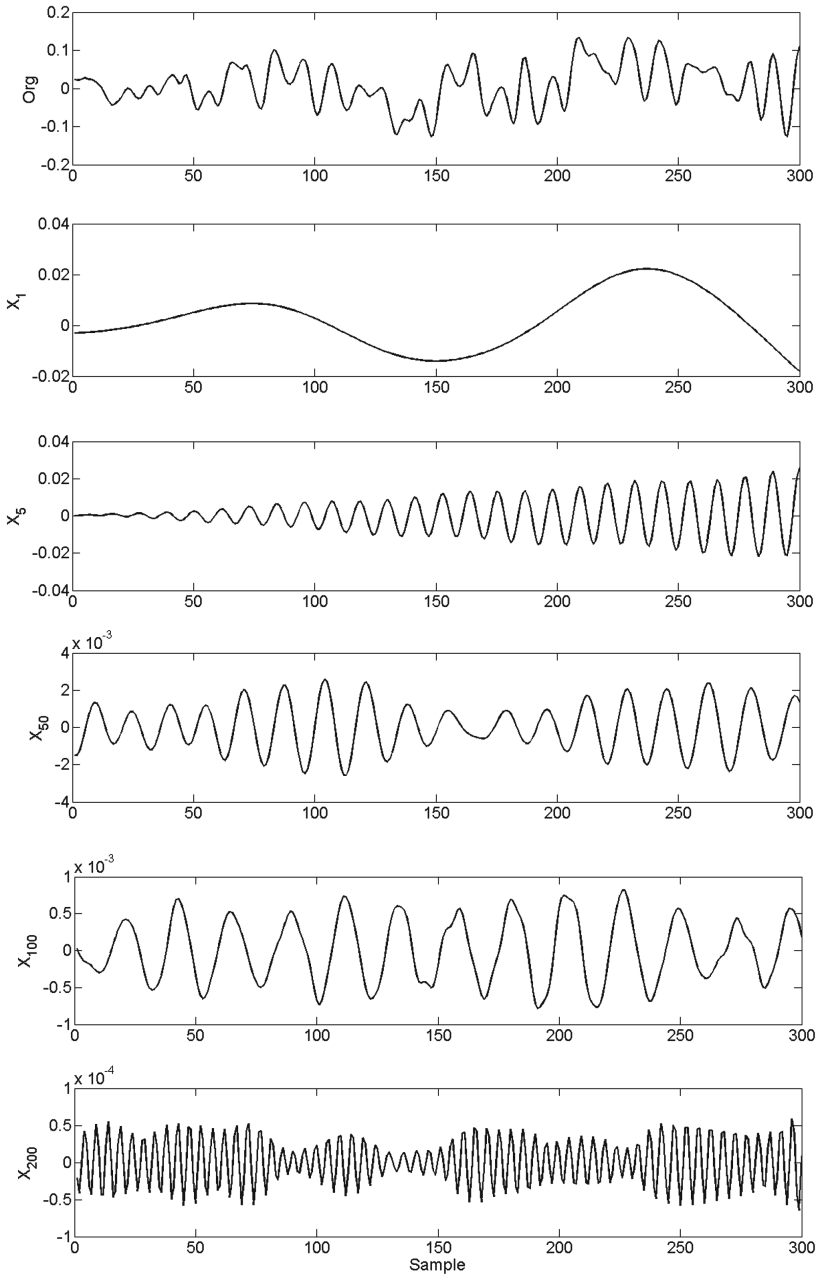
Twelve host signals from RWC music-genre database [26] (Track No. 01, 07, 13, 28, 37, 49, 54, 57, 64, 85, 91, and 100) used in IHC's 2012 audio watermarking competition were used in our experiments. All has a sampling rate of 44.1 kHz, 16-bit quantization, and two channels (stereo). Ninety-bit payloads per 15 seconds of the host signal are embedded into the host. This allows a maximum of 6 bps of embedding capacity.

For robustness evaluation, five attacks were applied to watermarked signals: Gaussian-noise addition with average SNR of 36 dB, re-sampling with 16 and 22.05 kHz, band-pass filtering with 100-6000 Hz and  $-12$  dB/Oct, MP3 compression with 128 kbps joint stereo, and MP4 compression with 96 kbps. We represent extraction precision in term of bit error rate (BER), which is defined by the number of bit errors divided by the total number of embedded bits. Remark that IHC suggests that the BER should be lower than 10%.

#### 4.1 Singular-Spectrum Analysis and Synthesis

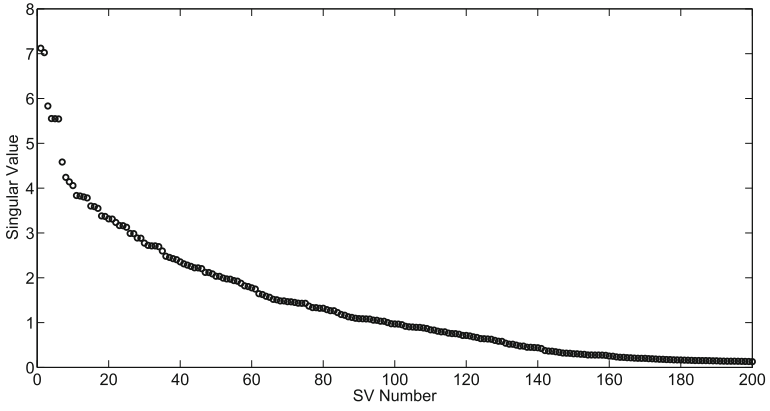
As mentioned in Section 2, SSA can be used to decompose a signal into oscillatory components. In our experiments, the *window length* is set to 500. Figure 2 shows an example of using SSA to decompose a signal. The first panel labeled *Org* is an original signal which is zoomed to observe 300 samples. The second to sixth panels show examples of five oscillatory components corresponding to the 1st, 5th, 50th, 100th, and 200th SVs, respectively. Specifically, the waveform  $X_1$  shown in the second panel is a result of hankelization of the matrix  $\mathbf{X}_1 = \sqrt{\lambda_1}U_1V_1^T$ , and the waveform  $X_5$  shown in the third panel is a result of hankelization of the matrix  $\mathbf{X}_5 = \sqrt{\lambda_5}U_5V_5^T$ , and so on. Therefore, SV could be interpreted as a scale factor of each oscillatory component. The lower the component order, the more contribution to the signal. This is because SVs are sorted in descending order. Actually, there are more than 200 components since there are more than 200 SVs that are greater than zero as shown in Fig. 3. The first panel of Fig. 4 shows a waveform of reconstructed signal comparing to the original. The reconstructed signal is constructed from only the first 100 oscillatory components. The second panel shows a residual signal or the difference between the original and reconstructed signals. The more components are added, the smaller the residual signal is. Therefore, it is possible to modify the high-order oscillatory components without affecting sound quality significantly. In this sense, our simple rules are corresponding to changing scale factors of certain components. Figure 5 shows an example of waveform when embedding bit 1 and bit 0 into  $X_{35}$ . It can be seen clearly from this example that what is modified is the scale factor of the component waveform.

Although the residual signal is very small when all oscillatory components are added up, it might exists. Thus, we first check ODGs of synthesis signals comparing to the originals. The result is shown in Fig. 6 (light gray). The sound quality of synthesis signals is not different from that of originals.

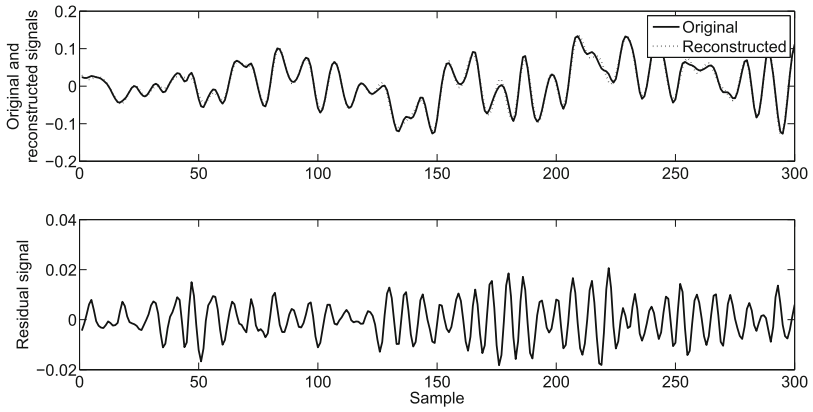


**Fig. 2.** An example of using SSA to decompose a signal

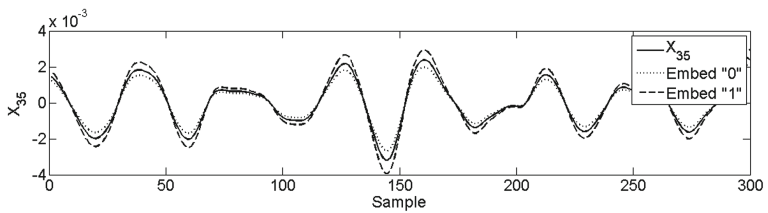




**Fig. 3.** Singular spectrum (The rst 200 SVs)



**Fig. 4.** Original and reconstructed signals (top), Residual signal or the difference between original and reconstructed signals (bottom)



**Fig. 5.** Embed “0” and “1” into the component  $X_{35}$

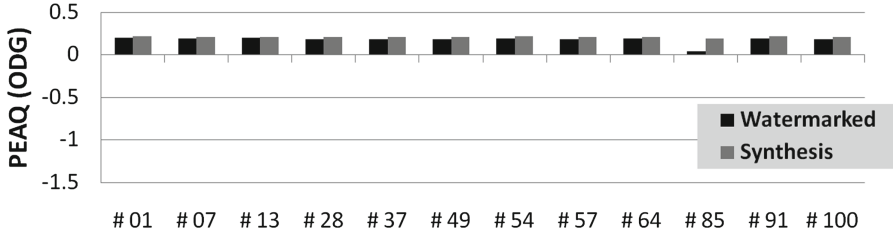


Fig. 6. ODGs of watermarked and synthesis signals

In addition to PEAQ, the log-spectral distance (LSD) and the signal-to-error ratio (SER) were also performed. LSD is a distance or distortion measure between two spectra, which is defined as the following formula given  $P(\omega)$  and  $\hat{P}(\omega)$  are power spectra of original and synthesis signals respectively.

$$\text{LSD} = \sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} \left[ 10 \log \frac{P(\omega)}{\hat{P}(\omega)} \right]^2 d\omega} \quad (6)$$

SER is the power ratio between a signal and the error. Given amplitudes  $A_{\text{sig}}(n)$  and  $A_{\text{syn}}(n)$  of original and synthesis signals respectively, SER is defined as follows.

$$\text{SER} = 10 \log \frac{\sum_n [A_{\text{sig}}(n)]^2}{\sum_n [A_{\text{sig}}(n) - A_{\text{syn}}(n)]^2} \quad (7)$$

The Eqs. (6) and (7) imply that the lower LSD, the power spectrum of a synthesis signal is more similar to that of an original, and the higher SER, the lower error. Therefore, for a perfect reconstruction, the ideal value of LSD is zero, and the ideal value of SER is infinity. The results from our experiment confirm that LSD between any pair of original and a synthesis signals is zero, and SER is inf. Thus, the framework of singular-spectrum analysis-synthesis could be used in audio watermarking.

## 4.2 Objective Sound-Quality Test

The following parameters are chosen for the proposed embedding rules:  $\epsilon = 0.1$ ,  $u = 21$ , and  $l = 49$ , i.e. the 21st to 49th SVs are modified with respect to a watermark bit as described in Sect. 3. Figure 6 shows ODGs of watermarked signals together with ODGs of synthesis signals. From the viewpoint of PEAQ, we can say that our proposed method introduces very small distortion to the sound quality. LSD and SER between original and watermarked signals are shown in Fig. 7. The results indicate that our proposed method satisfies inaudibility criterion.

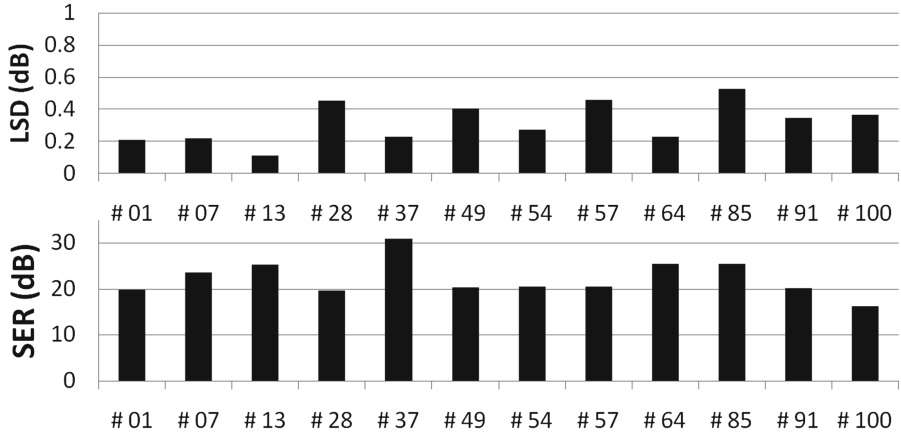


Fig. 7. Log-spectral distance and Signal-to-Error ratio of watermarked signals

### 4.3 Robustness Test

The results of robustness tests are shown in Table 1. The proposed method is robust against MP3 and MP4 compression, Gaussian-noise addition, band-pass filtering and re-sampling. Furthermore, in order to compare robustness against certain attacks, we implemented the scheme based on [15], which will refer as a conventional method in this paper. The proposed method is more robust against MP3 compression and band-pass filtering than the conventional method.

## 5 Discussion

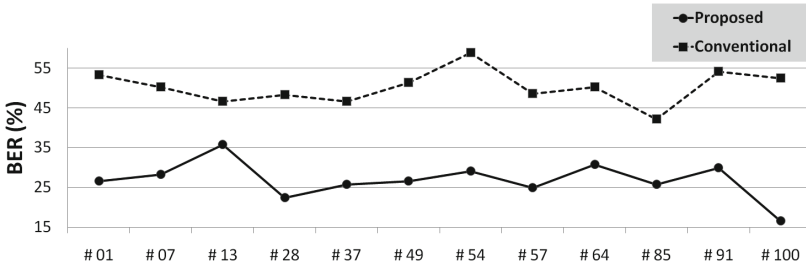
The results from our investigation suggest that the robustness could be improved if lower-order oscillatory components are modified instead of the higher-order ones. For example, the robustness against MP3 attack of the proposed scheme increases, as shown in Fig. 9, if values of the 2nd to 9th SVs, instead of the 21st to 49th SVs, are modified to embed a watermark. However, the sound quality of watermarked signals decreases as shown in Fig. 10. It reveals a trade-off between robustness and inaudibility.

Besides the position of modified SVs, the other point we would like to discuss here is the number of SVs which ranges from  $u$  to  $l$ . If the number of modified SVs increases, then the robustness increases. It is possible that we can keep ODG while BER is decreasing, especially in the case of high orders. For example, when we modify the 51st to 89th SVs, we have an average BER of 13.26% for MP3 attack. If we set  $l$  to 99, we will have the average BER of 10.83%, and for both cases the average ODG is around 0.18. Therefore, one way to improve the proposed scheme is choosing these parameters appropriately.

We also evaluated the robustness against single-echo addition. It is not a general attack of watermark. We found that both conventional and proposed methods are not robust against this attack as shown in Fig. 8. Because both

**Table 1.** Comparing BERs of the proposed and conventional methods when attacks are applied

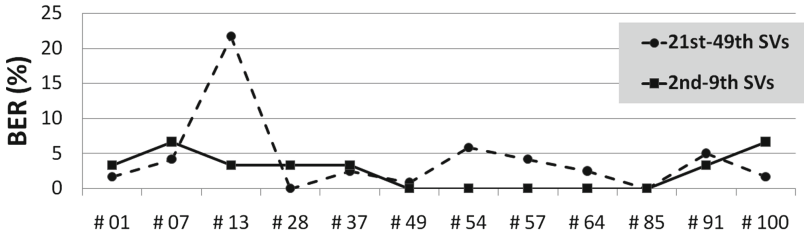
	No Attack		MP3 Attack		AWGN		MP4 Attack		Re-sampling		BP Filtering	
	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
#01	0.00	0.00	1.67	47.22	0.00	0.00	0.00	0.28	0.42	3.19	1.67	25.83
#07	3.33	0.00	4.17	98.33	3.33	0.00	4.17	0.00	3.75	0.42	4.17	48.61
#13	1.67	0.00	21.67	70.56	1.67	0.00	15.00	13.89	16.67	25.69	36.67	47.78
#28	0.00	0.00	0.00	87.78	0.00	0.00	0.00	0.00	0.00	0.00	0.83	43.61
#37	0.00	0.00	2.50	82.50	0.00	0.00	1.67	0.00	2.50	0.00	5.00	56.67
#49	0.00	0.00	0.83	33.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	21.39
#54	2.50	0.00	5.83	41.94	2.50	0.00	3.33	0.00	2.08	0.00	4.17	40.56
#57	4.17	0.00	4.17	33.61	4.17	0.00	5.83	0.00	2.92	0.00	10.00	28.89
#64	0.83	0.00	2.50	95.83	0.83	0.00	1.67	0.00	1.25	0.28	2.50	62.22
#85	0.00	0.00	0.00	18.89	0.00	0.00	0.00	0.00	0.00	0.83	0.83	0.83
#91	2.50	0.00	5.00	74.72	2.50	0.00	4.17	0.00	1.67	0.14	5.83	50.56
#100	0.83	0.00	1.67	23.61	0.83	0.00	0.00	0.28	1.25	1.53	0.83	30.00
AV	1.32	0.00	4.17	59.00	1.32	0.00	2.99	1.20	2.70	2.67	6.04	38.08
SD	1.48	0.00	5.83	29.04	1.48	0.00	4.30	3.40	4.56	7.31	10.05	17.30

**Fig. 8.** Comparing BERs of the proposed and conventional methods when single-echo addition with delay time of 100 ms,  $-6$  dB is applied

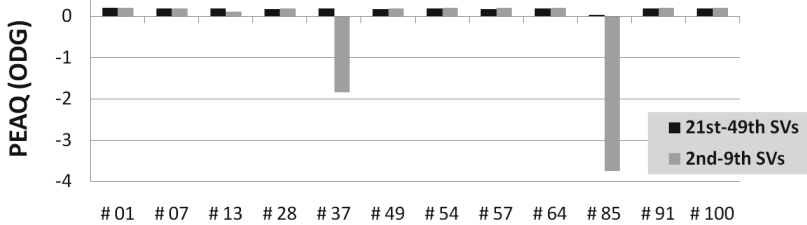
are time-domain based methods where matrices are created directly from the waveform, thus time-domain processing affects the elements of those matrices in a significant way.

We have verified that our proposed scheme can extend capacity from 6 to 18 bps without affecting both sound quality and a detection rate. The maximum capacity of the proposed scheme given preferred sound quality is determined by  $N$ . It could be increased if the parameters such as  $L$ ,  $u$ ,  $l$ , and  $\epsilon$  are chosen appropriately similar to a way to improve robustness.

Finally, this is a frame-based method, thus there also exists a problem of frame synchronization. The current proposed scheme assumes to know where to look for hidden information. One possible solution to this problem is using synchronization codes [27,28]. Moreover, in our experiments, we do not claim that the parameters, such as  $N$ ,  $L$ ,  $u$ ,  $l$ , and  $\epsilon$ , are the optimal ones. These are the subjects of our future works.



**Fig. 9.** Comparing BERs, after MP3 attack is applied, when lower-order components are modified to hide information



**Fig. 10.** Comparing ODGs when lower-order components are modified to hide information

## 6 Conclusion

This paper presented our investigation on deploying SSA for hiding information and proposed audio watermarking scheme based on it. SSA was used to decompose a signal into oscillatory components. Controlling the scale factors which can be done by modifying SVs of some components is the important procedure to embed a watermark. In this SSA view point, we can interpret the physical meaning of SV. We discover that SSA is a perfect analysis-synthesis tool. Although the proposed scheme is robust against many attacks especially MP3 and MP4 compression and satisfies inaudibility criterion, it still has a drawback that involves in audio watermark, for example, it is fragile to single-echo addition. The proposed method is a time-domain SSA-based method so that it seems to be robust against frequency domain processing. Our next interests are as follows: whether a frequency-domain (amplitude and phase spectra) SSA-based method will be robust against time-domain processing such as single-echo addition, and whether a hybrid SSA-based method of time- and frequency-domain will offer a solution to inaudible and robust audio watermarking.

**Acknowledgments.** This work was supported by a Grant-in-Aid for Scientific Research (B) (No. 23300070) and an A3 foresight program made available by the Japan Society for the Promotion of Science, and under a grant in the SIIT-JAIST-NECTEC Dual Doctoral Degree Program.

## References

1. Dutoit, T., Marques, F.: *Applied Signal Processing: A MATLAB (TM)-Based Proof of Concept*. Springer, New York (2009)
2. Craver, S., Wu, M., Liu, B.: What can we reasonably expect from watermarks? In: *IEEE Workshop on Application of Signal Processing to Audio and Acoustics*, pp. 223–226. IEEE Press, New York (2001)
3. Unoki, M., Miyauchi, R.: Method of digital-audio watermarking based on cochlear delay characteristics. In: Kondo, K. (ed.) *Multimedia Information Hiding Technologies and Methodologies for Controlling Data*, pp. 42–70. Information Science Reference, United States (2013)
4. Fallahpour, M., Megias, D.: High capacity audio watermarking using the high frequency band of the wavelet domain. *J. Multimed. Tools. Appl.* **52**(2–3), 485–498 (2011)
5. Ercelebi, E., Batakci, L.: Audio watermarking scheme based on embedding strategy in low frequency components with a binary image. *Digit. Signal. Process.* **19**(2), 265–277 (2009)
6. Unoki, M., Hamada, D.: Method of digital-audio watermarking based on cochlear delay characteristics. *Int. J. Innov. Comput. I.* **6**(3), 1325–1346 (2010)
7. Bhat, V., Sengupta, I., Das, A.: A new audio watermarking scheme based on singular value decomposition and quantization. *Circ. Syst. Signal. Pr.* **30**(5), 915–927 (2011)
8. Özer, H., Sankur, B., Memon, N.: An SVD-based audio watermarking technique. In: *7th Workshop on Multimedia and Security*, pp. 51–56. IEEE Press, New York (2005)
9. El-Samie, F.: An efficient singular value decomposition algorithm for digital audio watermarking. *Int. J. Speech. Technol.* **12**(1), 27–45 (2009)
10. Vongpraphip, S., Ketcham, M.: An intelligence audio watermarking based on DWT-SVD using ATS. In: *WRI Global Conference on Intelligence Systems*, pp. 150–154. IEEE Press, New York (2009)
11. Lamarche, L., Liu, Y., Zhao, J.: Flaw in SVD-based watermarking. In: *Canadian Conference on Electrical and Computer Engineering*, pp. 2082–2085. IEEE Press, New York (2006)
12. Al-Haj, A., Twal, C., Mohammad, A.: Hybrid DWT-SVD audio watermarking. In: *5th International Conference on Digital Information Management*, pp. 525–529. IEEE Press, New York (2010)
13. Al-Haj, A., Mohammad, A.: Digital audio watermarking based on the discrete wavelets transform and singular value decomposition. *Eur. J. Sci. Res.* **39**(1), 6–21 (2010)
14. Al-Yaman, M., Al-Tae, M., Alshammas, H.: Audio-watermarking based ownership verification system using enhanced DWT-SVD technique. *International Multi-Conference on Systems, Signals and Devices*, pp. 1–5. IEEE Press, New York (2012)
15. Bhat, V., Sengupta, I., Das, A.: An audio watermarking scheme using singular value decomposition and dither-modulation quantization. *J. Multimed. Tools. Appl.* **52**(2–3), 269–283 (2011)
16. Zhang, J.: Analysis on audio watermarking algorithm based on SVD. In: *2nd International Conference on Computer Science and Network Technology*, pp. 1986–1989. IEEE Press, New York (2012)

17. Singhal, A., Chaubey, A., Prakkash, C.: Audio watermarking using combination of multilevel wavelet decomposition, DCT and SVD. In: International Conference on Emerging Trends in Networks and Computer Communications, pp. 239–243. IEEE Press, New York (2011)
18. Dhar, P., Shimamura, T.: Audio watermarking in transform domain based on singular value decomposition and quantization. In: 18th Asia-Pacific Conference on Communications, pp. 516–521. IEEE Press, New York (2012)
19. Dhar, P., Shimamura, T.: An audio watermarking scheme using discrete fourier transformation and singular value decomposition. In: 35th International Conference on Telecommunications and Signal Processing, pp. 789–794. IEEE Press, New York (2012)
20. Dhar, P., Shimamura, T.: A DWT-DCT-based audio watermarking method using singular value decomposition and quantization. *J. Signal. Process.* **17**(3), 69–79 (2013)
21. Suresh, G.: An efficient and simple audio watermarking using DCT-SVD. In: International Conference on Devices, Circuits and System, pp. 177–181. IEEE Press, New York (2012)
22. Karimimehr, S., Samavi, S., Kaviani, H., Mahdavi, M.: Robust audio watermarking based on HWD and SVD. In: 20th Iranian Conference on Electrical Engineering, pp. 1363–1367. IEEE Press, New York (2012)
23. Jiang, W.: Fragile audio watermarking algorithm based on SVD and DWT. In: International Conference on Intelligence Computing and Integrated Systems, pp. 83–86. IEEE Press, New York (2010)
24. Golyandina, N., Nekrutkin, V., Zhigljavsky, A.: Analysis of Time Series Structure: SSA and related techniques. Chapman and Hall/CRC, United States (2001)
25. Information Hiding and its Criteria for evaluation. <http://www.ieice.org/emm/ihc/en/>
26. RWC Music Database. <https://staff.aist.go.jp/m.goto/RWC-MDB/>
27. Wu, S., Huang, J.: Efficiently self-synchronized audio watermarking for assured audio data transmission. *IEEE Trans. Broadcast.* **51**(1), 69–76 (2005)
28. Lei, B.: Robust SVD-based audio watermarking scheme with differential evolution optimization. *IEEE Trans. Audio, Speech, Language Process* **21**(11), 2368–2378 (2013)