An Audio Watermarking Scheme Based on Automatic Parameterized Singular-Spectrum Analysis Using Differential Evolution

Jessada Karnjana*†, Pakinee Aimmanee*, Masashi Unoki† and Chai Wutiwiwatchai‡,
* Sirindhorn International Institute of Technology, Thammasat University,
131 Moo 5, Tiwanon Rd., Bangkadi, Muang, Pathumthani, 12000 Thailand
E-mail: pakinee@siit.tu.ac.th
† School of Information Science, Japan Advanced Institute of Science and Technology,
1-1 Asahidai, Nomi, Ishikawa, 923-1292 Japan
E-mail: {jessada,unoki}@jaist.ac.jp
‡ National Electronics and Computer Technology Center,
112 Thailand Science Park, Phahonyothin Rd., Khlong Luang, Pathumthani, 12120 Thailand
E-mail: chai.wutiwiwatchai@nectec.or.th

***Abstract*—This paper proposes an audio watermarking scheme based on singular-spectrum analysis (SSA) and differential evolution. In our framework, a watermark is embedded into an audio signal by modifying the amplitude of some oscillatory components which are decomposed by SSA, and a parameter set for the modification is determined by differential evolution. Test results showed that, although there is a trade-off between inaudibility and robustness, the sound quality of watermarked signal could be improved considerably while the bit error rate could be satisfied. Our proposed scheme is inaudible and robust. Furthermore, based on analyzing the second derivative of singular spectrum, it was found that our proposed scheme can be completely blind.**

I. Introduction

The popularity and convenience of multimedia transfer via the Internet recently raise public concerns about authentication, copyright management, copy control, and the like. Besides encryption techniques, another potential solution solving the problems is to use watermarking, which is a technique of making information unnoticeable [1]. The media interested in this work is audio. To deal with such concerns we propose a novel and effective audio-watermarking scheme based on singular-spectrum analysis (SSA) and differential evolution.

In general, there are five requirements for an audio-watermarking scheme [2]. (1) *Inaudibility*: human auditory system should not be able to detect a watermark. (2) *Robustness*: it is an ability to protect a watermark when attacks are applied to a watermarked signal. (3) *Blindness*: it is a property of extracting hidden information from a watermarked signal without the original signal in extraction processes. (4) *Confidentiality*: it is a property of concealment of hidden data. (5) *Capacity*: quantity of information embedded in the original signal. Naturally, these requirements conflict each other's. The high capacity, for example, implies low robustness [2]. Based on our survey, most published results showed trade-offs of these requirements.

There have been several audio watermarking techniques proposed by researchers, and different techniques were proposed with different objectives. For example, Fallahpour and Megias [2] focused the work on large capacity while keeping inaudibility in an acceptable range. Unoki and Hamada [3] proposed an audio watermarking scheme based on cochlear delay characteristic of human ears. The concept of this proposed scheme is that human auditory system cannot distinguish between a normal sound and the sound that is almost the same except some of its low frequency components are delayed. It is inaudibility and robust, however, it is non-blind. We can classify all published audio watermarking techniques into two groups: the first one based on human perceptual model, e.g. [2]-[5], and another based solely on mathematical manipulation, e.g. [6]-[21]. This research investigated the latter, especially the techniques relying on singular value decomposition (SVD), i.e. a mathematical method of extracting algebraic features called singular value. It was shown experimentally that SVD-based audio watermarking has a lot of advantages [6]-[21]. The advantages are basically derived from properties of singular values, such as invariance of singular values under common signal processing, i.e. if small modification occurs in the original signal represented by a matrix, its singular values will vary very slightly [8]. When the hidden information is embedded into the singular values, therefore, this property of singular values makes the SVD-based techniques robust against common signal processing attacks.

Based on our survey, all SVD-based watermarking techniques [6]-[22] embed information by making small changes to singular values according to a watermark bit and embedding rules. Some techniques modify only the largest singular value [12][13][18]. Some techniques modify all singular values [14][16][20]. Robustness of the latter is a bit better than that of the former. Also, there is the technique that modifies a group of small singular values [21]. This technique is said to be fragile audio watermarking for application such as tempering detection. Most SVD-based techniques are non-blind [6][12][17], a few are blind [10][14][20]. Lamarche *et al.* had pointed out that the robustness of some non-blind methods is subject to high false positive rate [22]. To avoid such a problem, we focused our work on the blind technique.

Recently, we proposed an audio watermarking scheme based on singular-spectrum analysis (SSA) [23], which is one of the SVD-based watermarking techniques, and demonstrated benefits of deploying SSA over SVD. Our embedding rule was very simple. We force an interval of singular values flat, and there are two levels of flatness according to a watermark bit as shown in Fig. 1. Since it is core technique of this work, it is described briefly in Section II.

The test results showed that the proposed scheme satisfied imperceptibility criterion and was robust against many attacks, such as MP3 and MP4 compression, band-pass filtering, and re-sampling. The objective evaluation of sound quality of watermarked signal from our previous method was very good, however, when subjective tests were used to evaluate, some clips had quite low subjective difference grade (SDG). We analyzed the singular spectrum of the poor-SDG clips and discovered that a singular-value gap between modified and unmodified singular values of those clips, during the embedding process, is larger than some threshold value as illustrated in Fig. 2. This large gap is the cause of perceptual distortion. According to the singular-value-modification rule of our previous work, singular values are modified based on a predetermined set of parameters. Since the singular spectrum varies from clip to clip, it is reasonable to justify that the parameter set depends upon the clip as well. Therefore, adaptive parameters are needed in order to reduce the sound distortion of watermarked signal due to embedding.

There are three approaches to achieve this adaptive parameters. First, analyzing the singular-spectrum pattern. From our exploratory investigation, the pattern of singular spectrum of an audio frame itself could be used to adjust the parameter set. Second, combining a psychoacoustic model to the scheme. To follow this approach, it is necessary that physical meaning of singular value is known so as to establish a complete link between singular value and the model. In SSA-based framework, as discussed in [23], singular values could be interpreted their physical meanings. Last, optimal parameters are obtained by using an artificial intelligence approach. In this work, the third approach is focused and explored to improve the performance of SSA-based audio watermarking scheme.

II. Previous Framework

Our previous framework consists of two main processes: embedding and extracting, which are detailed as follows.

A. Embedding Process

The embedding process is shown in Fig. 3 (left). Our method used singular-spectrum analysis as the main tool in hiding information. Singular-spectrum analysis (SSA) is a powerful technique of identifying and extracting useful information, e.g. oscillatory components, seasonality components, or trends, from a signal [24], and it can be used for solving various problems such as finding trends of different resolution, extraction of periodicities with varying amplitudes, and finding structure in short time series [25]. There are many types of SSA. Our proposed scheme is based on the basic SSA. The embedding process consists of four following steps.

1) Segmentation and Trajectory Matrix Formation: First, an audio signal is segmented into several non-overlapping frames. Then, trajectory matrices $\boldsymbol{X}$ of size $L \times K$ representing each frame $X = (f_0, f_1, ..., f_{N-1})^T$ of length N are created.

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

Each column vector of $\boldsymbol{X}$ is called lagged vector, and a lagged vector X_i is defined as $X_i = (f_{i-1}, f_i, ..., f_{i+L-2})^T$, where $i = 1, 2, ..., K$, and L is a *window length* with $2 \leq L \leq N$. Therefore, the matrix $\boldsymbol{X}$ consists of K lagged vectors, $[X_1 X_2 ... X_K]$. Since the lagged vector X_i is constructed by a one-sample lag of X_{i-1}, the element $x_{i,j}$ is equal to the element $x_{i-1,j+1}$, where $x_{i,j}$ is an element at ith row and jth column of $\boldsymbol{X}$. From this property, we say that the trajectory matrix $\boldsymbol{X}$ is a Hankel matrix.

2) Singular Value Decomposition: After obtaining the trajectory matrix $\boldsymbol{X}$, we perform SVD to it. SVD is used to decompose a matrix $\boldsymbol{X}$ into a product of three matrices $\boldsymbol{U}$, $\boldsymbol{D}$, and $\boldsymbol{V}$ with the following relationship.

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

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal matrices, and $\boldsymbol{D}$ is a diagonal matrix whose element is called singular value.

Let $\{\sqrt{\lambda_1}, \sqrt{\lambda_2}, ..., \sqrt{\lambda_d}\}$ denote a set of singular values of the matrix $\boldsymbol{X}$ in descending order, called singular spectrum of $\boldsymbol{X}$, where λ_i for $i = 1, 2, ..., d$ is the eigenvalue of $\boldsymbol{X}\boldsymbol{X}^T$ (or $\boldsymbol{X}^T\boldsymbol{X}$) and $d = \arg\max_i(\lambda_i > 0)$. The trajectory matrix $\boldsymbol{X}$ can be rewritten as

$$\boldsymbol{X} = \boldsymbol{X}_1 + \boldsymbol{X}_2 + ... + \boldsymbol{X}_d, \quad (3)$$

where $\boldsymbol{X}_i = \sqrt{\lambda_i} U_i V_i^T$. Note that the expansion (3) is uniquely defined if and only if all eigenvalues are distinct. Each $\boldsymbol{X}_i$ represents a simple oscillatory component of the signal X. Therefore, the singular value can be interpreted as a scaling factor of each simple oscillatory component [23].

3) Singular Value Modification: To hide information, one bit of a watermark will be embedded into one frame by modifying singular values according to certain rule, i.e., physically, the amplitude of some oscillatory components connected with those singular values will be changed. Given a singular spectrum of $\boldsymbol{X}$, the rule can be summarized as follows.

If the watermark bit is "0", then $\sqrt{\lambda_{u+1}}, \sqrt{\lambda_{u+2}}, ...,$ and $\sqrt{\lambda_{l-1}}$ are replaced with $\sqrt{\lambda_l} + \epsilon \cdot (\sqrt{\lambda_u} - \sqrt{\lambda_l})$, and if the

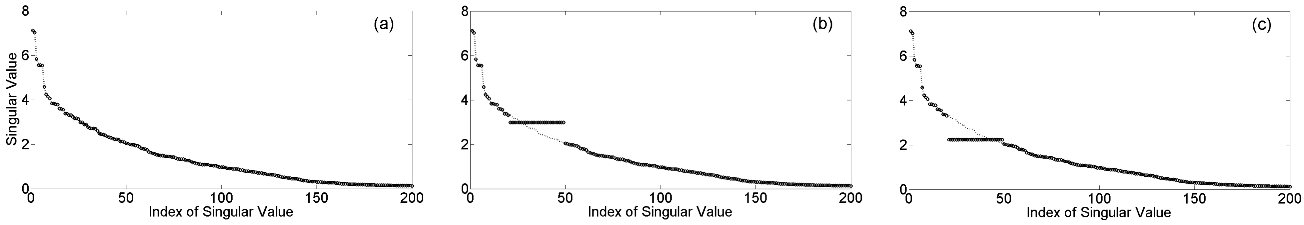

Fig. 1: (a) Singular spectrum, (b) its modification after embedding “1”, and (c) its modification after embedding “0”.

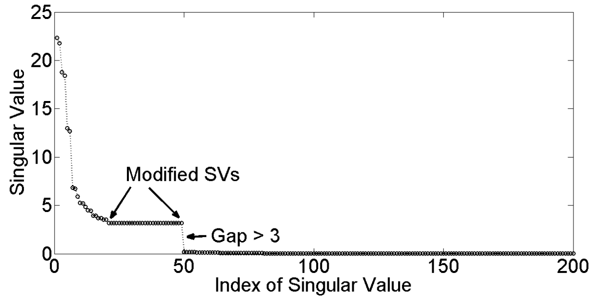

Fig. 2: Singular spectrum of a poor-SDG, watermarked frame.

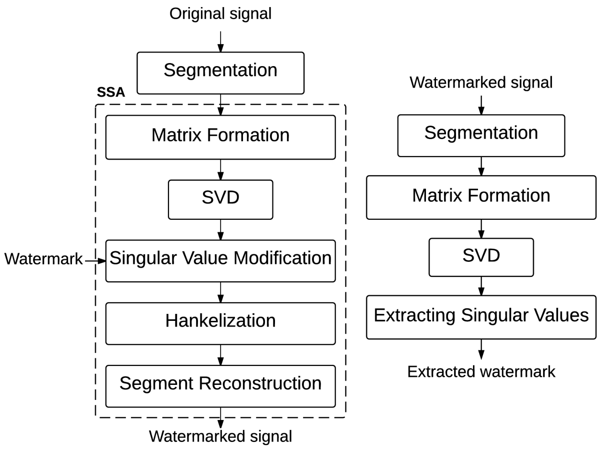

Fig. 3: The embedding (left) and extracting (right) processes.

watermark bit is “1”, then $\sqrt{\lambda_{u+1}}, \sqrt{\lambda_{u+2}}, ...,$ and $\sqrt{\lambda_{l-1}}$ are replaced with $\sqrt{\lambda_l} + (1-\epsilon)\cdot(\sqrt{\lambda_u} - \sqrt{\lambda_l})$ given that ϵ is a real positive number $\in [0, 0.5]$, and $\sqrt{\lambda_u}$ is greater than $\sqrt{\lambda_l}$, or, in other words, $u < l$.

In our experiments, we set $u = 20$, $l = 50$, and $\epsilon = 0.1$.

4) Hankelization and Segment Reconstruction: After modifying singular values with regard to the rule, the modified matrix is transformed into a new series of length N by reversing SVD or hankelization operation.

The hankelization of matrix $\boldsymbol{Y}$ of size $L \times K$ to a series $Y = (g_0, g_1, ..., g_{N-1})^T$ is defined as follows.

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} y^*_{m,k-m+2} & 0 \le k < L^* - 1 \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y^*_{m,k-m+2} & L^* - 1 \le k < K^* \\ \frac{1}{N-k} \sum_{m=k-K^*+2}^{N-K^*+1} y^*_{m,k-m+2} & K^* \le k < N, \end{cases} \tag{4}$$

where $L^* = \min(L, K)$, $K^* = \max(L, K)$, $y^*_{ij} = y_{ij}$ if $L < K$, and $y^*_{ij} = y_{ji}$ if $L \ge K$. In our proposed method, $\boldsymbol{Y}$ is a watermarked matrix, which is a trajectory matrix after its singular values are modified with respect to a watermark bit. The series Y is a watermarked frame. Finally, the watermarked signal is obtained by combining those frames.

B. Extracting Process

The extracting process is shown in Fig. 3 (right). As in the embedding process, the watermarked signal is segmented into several non-overlapping frames. Then, each frame is mapped to a trajectory matrix. To extract singular values, SVD is used.

Then, the watermark bit is decoded by determining the value of $\sqrt{\lambda_m}$, where $\sqrt{\lambda_m}$ is the median of $\{\sqrt{\lambda_{u+1}}, \sqrt{\lambda_{u+2}}, ..., \sqrt{\lambda_{l-1}}\}$. If $\sqrt{\lambda_m}$ is closer to $\sqrt{\lambda_u}$ than to $\sqrt{\lambda_l}$, the watermark bit is “1”. Otherwise, the watermark bit is “0”.

III. Improved Framework

The main structure of improved framework is the same as the previous framework. However, there are two major differences which involve singular-value modification and the extracting rule. As mentioned in Section II, the embedding rule obviously has three important parameters (u, l, and ϵ) which affect sound quality of watermark signal and robustness of the scheme. To improve inaudibility and robustness simultaneously, we deploy differential evolution to find the optimal parameters. The details are as follows.

A. Differential Evolution-based Singular Value Modification

Differential evolution is a parallel direct search method that optimizes a problem by iteratively improving candidate solutions with regard to an objective function and some constraints. It is a member of evolutionary algorithm which is a subset of evolutionary computation in artificial intelligence. It proved to be the fastest search method in this family [26].

Differential evolution utilizes NP D-dimensional parameter vectors $x_{i,G}$, called target vector, where $i = 1, 2, ..., NP$ as a population for generation G. The algorithm consists of four processes as illustrated in Fig. 4. First, the initial vector population $\{x_{i,1}\}$ is generated randomly, and it should cover entire parameter space. In the case that a priori knowledge (or a preliminary solution) is available, members of the first population might be generated by adding normally distributed noise to the preliminary solution. In our experiment, we used the parameter set from [23] as the preliminary solution.

Second, for each target vector $x_{i,G}$, a mutant vector $v_{i,G+1}$ is created by the following formula.

$$v_{i,G+1} = x_{r_1,G} + F \cdot (x_{r_2,G} - x_{r_3,G}), \tag{5}$$

where r_1, r_2, and r_3 are randomly chosen from $\{1, 2, ..., NP\}$ and mutually different, and they must be different from the index i as well. This condition makes the minimum number of members of population is four. F is a user-defined factor ranging from 0 to 2. It controls the convergence by scaling the difference of two vectors. Note that NP in the differential evolution is constant, it does not change during the evolutionary process.

Third, a trial vector $u_{i,G+1}$ is generated by using a pair of the target vector $x_{i,G}$ and its mutant vector $v_{i,G+1}$.

$$u_{i,G+1} = (u_{1i,G+1}, u_{2i,G+1}, ..., u_{Di,G+1}) \tag{6}$$

$$u_{ji,G+1} = \begin{cases} v_{ji,G+1} & \text{if } (\mathrm{randb(j)} \leq CR) \text{ or } \mathrm{j} = \mathrm{rnbr(i)} \\ x_{ji,G} & \text{otherwise,} \end{cases} \tag{7}$$

where $\mathrm{randb(j)}$ is the jth evaluation of a uniform random number generator with result $\in [0, 1]$. CR is the user-defined crossover constant $\in [0, 1]$. A large CR speeds convergence. $\mathrm{rnbr(i)}$ is chosen randomly from $\{1, 2, ..., D\}$ which ensures that the trial vector $u_{i,G+1}$ will get at least one parameter from the mutant vector $v_{i,G+1}$.

The fourth process in differential evolution cycle is selection. The trial vector $u_{i,G+1}$ is compared to the target vector $x_{i,G}$ by using a greedy criterion to decide which vector will be a member for next generation $G+1$. If a cost value of $u_{i,G+1}$ is smaller than that of $x_{i,G}$, then $x_{i,G+1}$ is set to $u_{i,G+1}$; otherwise, $x_{i,G+1}$ retains the same vector $x_{i,G}$. Differential evolution strategy, therefore, has NP competitions in one generation. After all NP members for generation $G+1$ are obtained, mutant vectors of members of this generation will be generated by the mutation process again. The evolutionary cycle of mutation, crossover and selection iteratively continues until a stopping criterion is reached, and the solution is the vector from the last generation that yields the lowest cost.

The optimization deployed in our proposed scheme is shown in Fig. 5. The cost function is defined as follows.

$$\text{Cost Value} = \sqrt{\alpha \left(\mathrm{LSD} + \left(1 - \mathrm{Sig}\left(\mathrm{SER}\right)\right)\right)^2 + \beta \overline{BER}^2}, \tag{8}$$

where LSD, $\mathrm{Sig(SER)}$ and $\overline{BER}$ are the log-spectral distance (LSD), the sigmoid function of signal-to-error ratio (SER) and the average bit-error rate (BER), respectively.

LSD is defined as the following formula given $P(\omega)$ and $\hat{P}(\omega)$ are power spectra of original and watermarked signals respectively.

$$\mathrm{LSD} = \sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} \left[10 \log \frac{P(\omega)}{\hat{P}(\omega)}\right]^2 d\omega} \tag{9}$$

Based on our previous work, we conducted listening tests with 30 normal-hearing subjects. The SDG scores are shown in Table I. We found that, in our work, the relationship between SDG and LSD is stronger than that between SDG and the objective difference grade (ODG). Specifically, we can predict the perceptually-annoying-SDG if LSD is greater than 0.35 dB. Based on this finding, we decide to use LSD instead of ODG in (8).

SER is the power ratio between a signal and the error. Given amplitudes $A_{\mathrm{org}}(n)$ and $A_{\mathrm{wmk}}(n)$ of original and watermarked signals respectively, SER is defined as follows.

$$\mathrm{SER} = 10 \log \frac{\sum_n \left[A_{\mathrm{org}}(n)\right]^2}{\sum_n \left[A_{\mathrm{org}}(n) - A_{\mathrm{wmk}}(n)\right]^2} \tag{10}$$

The equations (9) and (10) imply that the lower LSD, the power spetrum of a watermarked signal is more similar to that of an original, and the higher SER, the lower error. It is clear that we want to minimize LSD and to maximize SER. From our previous results [23], SER was greater than 0. Therefore, we presume that $(1 - \mathrm{Sig(SER)})$ is positive. In short, the term $\mathrm{LSD} + (1 - \mathrm{Sig(SER)})$ in (8) represents inaudibility.

$\overline{BER}$ is the average of BER, the number of error bits divided by the total number of embedded bits. Thus, it represents robustness of the scheme.

Two user-defined constants α and β are weighting factors with $\alpha + \beta = 1$ which control balance between inaudibility and robustness. The ideal cost value is zero, and the zero-cost value is our stopping criterion. In our experiment, there was three parameters that we want to optimize, hence the number of dimension of parameter vector, D, was 3. We set the other parameters for differential evolution as follows: $\alpha = \beta = 0.5$, $NP = 10D = 30$ (as suggested in [26]), $F = 0.8$, $CR = 0.7$, and the maximum iteration was 20. That is, if we cannot reach zero cost, the evolution cycle will stop after 20th iteration.

B. New Extracting Rule

In the previous work, we used only one singular value, which is the median of modified singular values, to decode an embedded bit. In this work, we used all information of singular

TABLE I: Comparing LSD, ODG, and SDG of watermarked signals from our previous method. (The numbers in the first row show the track number of twelve host signals used in the experiment.)

	#01	#07	#13	#28	#37	#49	#54	#57	#64	#85	#91	#100
LSD	0.21	0.22	0.11	0.45	0.23	0.41	0.27	0.46	0.23	0.53	0.35	0.37
ODG	0.20	0.19	0.20	0.18	0.19	0.18	0.19	0.18	0.20	0.04	0.19	0.18
SDG	-1.00	-0.72	-1.10	-2.27	-1.05	-1.30	-0.85	-1.83	-1.39	-1.25	-2.22	-2.93

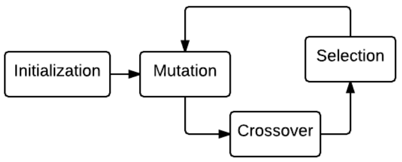

Fig. 4: Differential evolution processes.

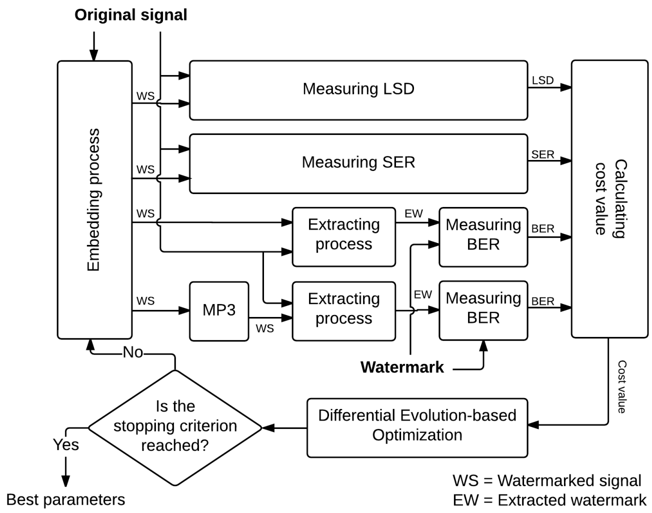

Fig. 5: Differential evolution optimization.

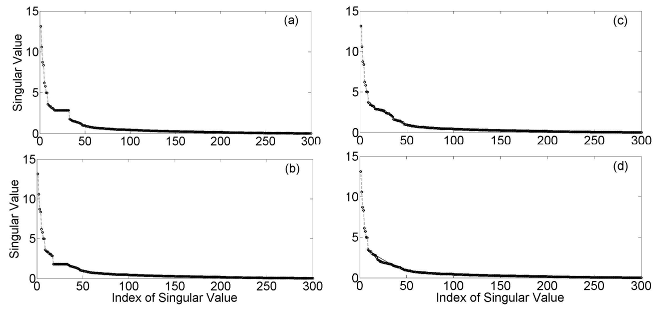

Fig. 6: Example of singular spectra in embedding and detection processes in proposed method: (a) when embedding "1" and (b) when embedding "0" in the embedding process, (c) singular spectrum of the frame embedded "1" and (d) singular spectrum of the frame embedded "0" in the extracting process.

values on $[u+1, l-1]$ to predict the watermark bit hoping to get a better result. We discover that when we reconstruct a watermarked frame and analyze its singular spectrum, the flatness resulting from rounding up or down of a sequence of singular values, according to the embedding rule, becomes distorted as shown in Fig. 6. However, the distortion shows patterns, as illustrated in Fig. 7. That is, if singular values on such a range is rounded down toward the lower bound, a convex upward curve is obtained; or else, a concave downward curve. Therefore, in this improved framework, we proposed to use concavity and convexity of singular spectrum on such interval to predict the watermark bit by the following scheme.

All singular values on $[u+1, l-1]$ are fitted on a degree-two polynomial, $y(x) = ax^2 + bx + c$ where y is singular value and x is index of the singular value. The coefficient a of this quadratic formula has played an important role as it indicates the rate of change of the singular values. Thus, a sign of coefficient a of the polynomial can be used to predict the watermark bit. That is, a minus sign indicates concavity or bit "1", and a plus sign indicates convexity or bit "0". An example of polynomial fitting is shown in Fig. 8, where the watermark bit "1" is embedded into a frame by forcing the singular values on $[18, 32]$ toward the upper bound, the 17th singular value.

In this work, we assume to know u and l because, at this stage, our goal is to improve the sound quality. However, as we will discuss in Section V, this set could be extracted blindly and automatically from singular spectrum.

IV. Evaluations

Twelve host signals from RWC music-genre database (Track No. 01, 07, 13, 28, 37, 49, 54, 57, 64, 85, 91, and 100) were used in our experiments. These tracks were used in 2012 IHC audio watermarking competition. All has a sampling rate of 44.1 kHz, 16-bit quantization, and two channels. Ninety-bit payloads per 15 seconds are embedded into the host. We compare our proposed approach with our previous work [23] and the conventional SVD-based approach [8]. We chose [8] as our reference for two reasons. One, it is one of a few blind SVD-based techniques, and, second, the result from [8] is promising.

The perceptual evaluation of audio quality (PEAQ) and LSD were used to measure the objective sound quality of watermarked signal. PEAQ measures the degradation of the watermarked signal being evaluated comparing with the original and covers a scale from -4 (worst) to 0 (best). LSD is a distance measure between two spectra, as defined in (9).

For robustness evaluation, six attacks were applied to watermarked signals: Gaussian-noise addition with average SNR of

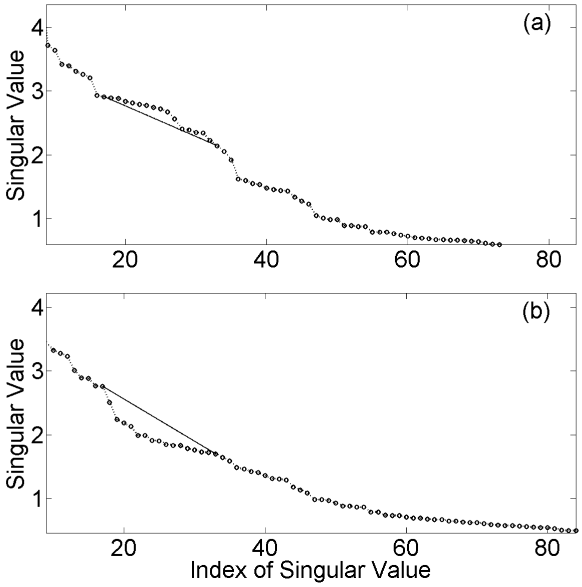

Fig. 7: (a) On the interval $[u+1, l-1]$, the singular spectrum curve is concave downward if bit "1" is embedded. (b) The singular spectrum curve is convex upward if bit "0" is embedded. In this example, u and l are 17 and 33 respectively.

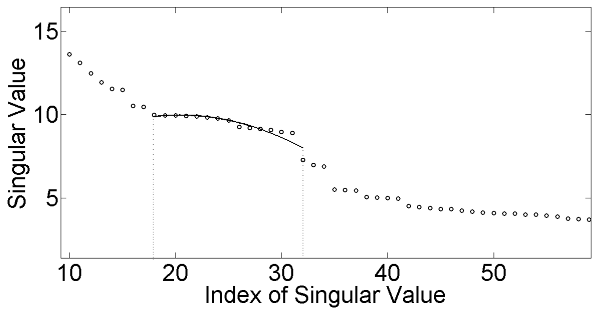

Fig. 8: The singular values on $[18, 32]$ are fitted on a quadratic equation $y(x) = ax^2 + bx + c$, where $a = -0.0146$, $b = 0.598$, and $c = 3.853$. Since the value of a is negative, the graph is concave. Therefore, the watermark bit is "1".

36 dB, re-sampling with 16 and 22.05 kHz, band-pass filtering with 100-6000 Hz and -12 dB/Oct, $\pm 4\%$ pitch shifting, MP3 compression with 128 kbps joint stereo, and MP4 compression with 96 kbps. We represent extraction precision in term of bit error rate (BER), and the BER should be lower than 10%.

A. Sound-quality Tests

The PEAQ and LSD of watermarked signal comparing between our current and previous methods are shown in Fig. 9 and Fig. 10 respectively. The results show that there is not much difference in ODG. However, LSD drops considerably, i.e. the sound quality of watermarked signal improves in a sense. Moreover, as shown in Table I, the correlation between LSD and SDG is stronger than that between ODG and SDG.

Besides, the results from listening tests with 35 normal-hearing subjects indicate that the sound quality of clips no. 28, 49, 57, 85, 91 and 100 improves greatly, i.e. from average SDG of -1.49 to -0.67, as shown in Table II.

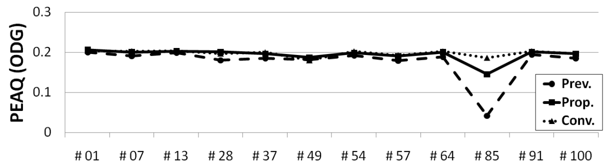

Fig. 9: ODGs comparison of proposed, previous and conventional methods.

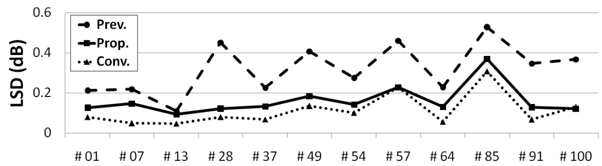

Fig. 10: LSDs comparison of proposed, previous and conventional methods.

B. Robustness Tests

The results of robustness tests are shown in Fig. 11-17. The BERs from the current and previous methods are compared. The BERs from conventional SVD-based method [8] are included as well. Our proposed and previous methods outperform the conventional SVD-based method in robustness when MP3, band-pass filtering, and pitch shifting are applied. However, when there is no attack, or MP4 or re-sampling attacks are applied, the conventional method is more robust than the proposed ones. Nevertheless, our average BER is under 10% and is smaller than that of the conventional method, as summarized in Table III. Therefore, we consider that our proposed schemes are more robust against various attacks than the conventional method.

As discussed in [8], the robustness of these schemes are due to the characteristics of singular values. That is, the singular spectrum resists change. In other words, the singular spectrum tries to return to its original shape after signal processing is applied. An example of this phenomenon is shown in Fig. 18. This figure shows singular spectra of one frame from our experiment. The singular spectrum of the original signal is represented by the dot line. The watermark bit "1" is embedded by modifying singular values on $[18, 32]$, where $\epsilon = 0.025$. Thus, we can see the concavity of the singular spectrum, marked by "$\circ$", of the watermarked frame on that range. When MP3 compression is applied to this frame, the singular values changes slightly as represented by "$\times$". But the shape of the singular spectrum remains the same. Thus, this can be accounted for the robustness of SSA-based scheme.

The average BER of current and previous schemes are 4.40% and 2.84%, respectively. Even though the robustness drops, we can consider this result as a trade-off between inaudibility and robustness. However, it is not a linear trade.

TABLE II: SDGs comparison of proposed, previous, and conventional methods.

	#01	#07	#13	#28	#37	#49	#54	#57	#64	#85	#91	#100
Prop.	-0.88	-0.36	-0.78	-0.45	-0.53	-0.94	-0.84	-0.49	-0.72	-0.35	-0.86	-0.77
Prev.	-1.00	-0.72	-1.10	-2.27	-1.05	-1.30	-0.85	-1.83	-1.39	-1.25	-2.22	-2.93
Conv.	-2.37	-0.97	-1.30	-0.48	-0.43	-0.58	-0.90	-1.38	-1.00	-1.78	-0.88	-1.70

TABLE III: Comparing average BERs (%) from our proposed, previous, and conventional methods.

	Prev.	Prop.	Conv.
No attack	1.31	1.24	0.00
MP3 attack	4.16	6.24	59.00
MP4 attack	2.98	6.52	1.20
Gaussian noise addition	1.36	1.24	0.00
Re-sampling	2.70	3.71	2.70
Band-pass filtering	6.04	10.62	38.10
Pitch shifting	1.31	1.24	7.20
Average	2.84	4.40	15.46
Standard deviation	1.77	3.58	23.53

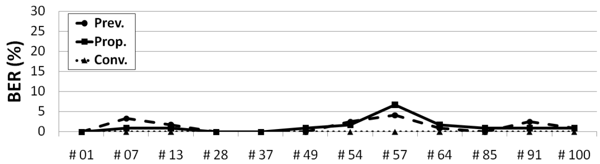

Fig. 11: BERs comparison of proposed, previous, and conventional methods when there is no attack.

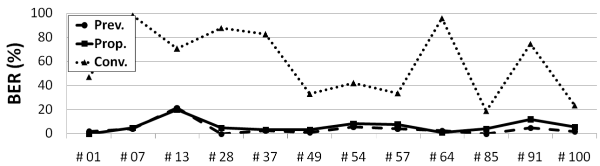

Fig. 12: BERs comparison of proposed, previous, and conventional methods when MP3 attack is applied.

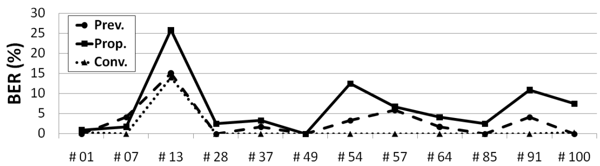

Fig. 13: BERs comparison of proposed, previous, and conventional methods when MP4 attack is applied.

Because while the robustness drops very little, the inaudibility improves considerably.

V. Discussion

There are two points we would like to discuss here. First, we found that the parameter set obtained from differential evolution is good at some specific tasks which included in differential evolution optimization model. For example, when MP3 attack is removed from the model illustrated in Fig. 5, the

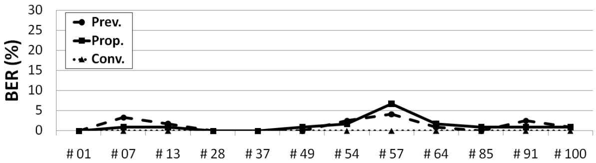

Fig. 14: BERs comparison of proposed, previous, and conventional methods when Gaussian noise is added.

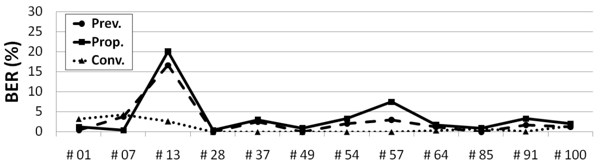

Fig. 15: BERs comparison of proposed, previous, and conventional methods when re-sampling is applied.

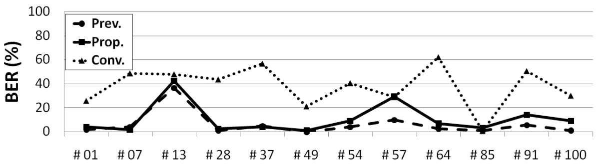

Fig. 16: BERs comparison of proposed, previous, and conventional methods when band-pass filtering is applied.

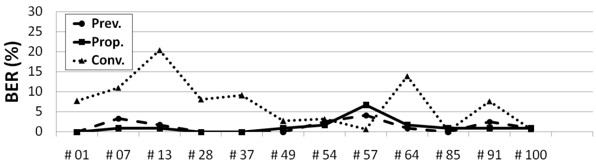

Fig. 17: BERs comparison of proposed, previous, and conventional methods when pitch shifting is applied.

watermarked clip will be fragile to MP3 compression. Thus, the practical model should include as many attacks as possible in order to find the best parameter set.

Second, in the extraction process, we assumed to know $\{u, l\}$. However, this set can be extracted automatically from singular spectrum. We discovered that, when no information is embedded, the second derivative of singular spectrum looks like an underdamped harmonic oscillator. If a watermark bit is embedded, it causes one or two, depending on the watermark bit and $\{u, l\}$, abrupt changes in slope of singular spectrum. Therefore, a spike, which is caused by the abrupt change in slope of singular spectrum, presents in the second derivative as shown in Fig. 19. By detecting the spike (labeled "l")

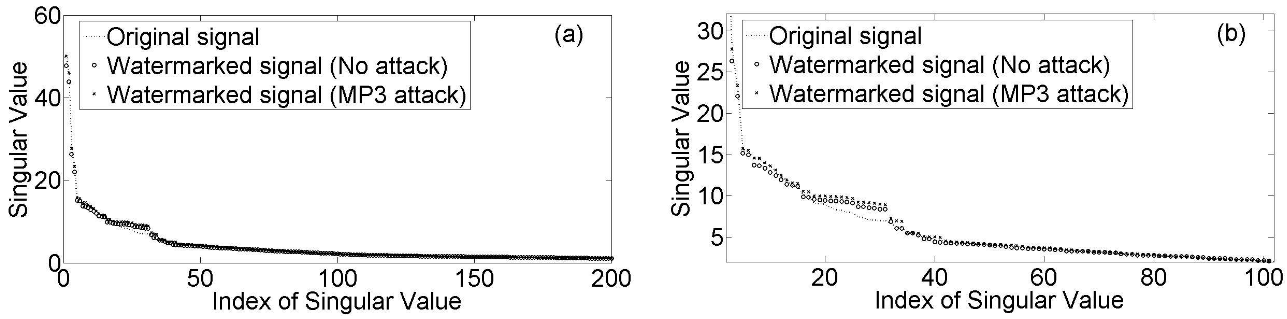

Fig. 18: (a) Singular spectra of an original signal, watermarked signals when no attack and MP3 compression are applied, (b) their closeup, in order to investigate singular values.

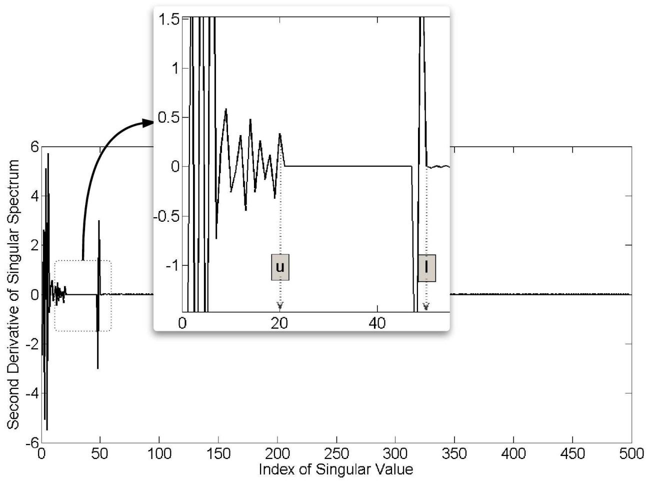

Fig. 19: Second derivative of a singular spectrum given the watermark bit is 1.

and calculating the distance between the spike and the point at which the oscillation stops (labeled "u"), we can have the set $\{u, l\}$. Analyzing the second derivative of singular spectrum can also help us to avoid false positive detection. Because it is highly likely that when there is no spike, there is no information hidden in a frame. However, there is some difficulty due to the distortion of flatness. In other words, if the coefficient a of polynomial fitted by singular values on $[u{+}1, l{-}1]$ is very close to 0, it is not easy to detect the spike. This issue will be investigated further.

VI. Conclusions

This paper presented the improvement in inaudibility property of our proposed watermarking scheme based on automatic parameterized SSA using differential evolution. Differential evolution was used to search for the best parameter set suitable for embedding a watermark into a clip. We successfully showed that utilizing differential evolution could enhance the sound quality of watermarked signal and maintain the robustness of the scheme at the same time. We also showed that, by analyzing the singular spectrum and its second derivative, it is possible to extract the embedding parameter set and, hence, a watermark bit from a watermarked frame blindly. In addition, the same strategy can be used to avoid false positive detection.

Acknowledgment

This work was supported by an A3 foresight program made available by the Japan Society for the Promotion of Science and partially supported by a Grant-in-Aid for Scientific Research (A) (No. 25240026). It was also under a grant in the SIIT-JAIST-NECTEC Dual Doctoral Degree Program.

References

[1] S. Craver, M. Wu, and B. Liu, "What can we reasonably expect from watermarks?," in Proc. *IEEE WASPAA*, pp. 223–226, 2001

[2] M. Fallahpour and D. Megias, "High capacity audio watermarking using the high frequency band of the wavelet domain," *J. Multimed. Tools. Appl.*, vol. 52, no. 2-3, pp. 485–498, 2011

[3] M. Unoki and D. Hamada, "Method of Digital-Audio Watermarking Based on Cochlear Delay Characteristics," *Int. J. Innov. Comput. I.*, vol. 6, no. 3, pp. 1325–1346, 2010

[4] X. Wen, X. Ding, J. Li, L. Gao, and H. Sun, "An Audio Watermarking Algorithm Based on Fast Fourier Transform," in Proc. *ICIII*, pp. 363–366, 2009

[5] E. Ercelebi and L. Batakci, "Audio Watermarking Scheme Based on Embedding Strategy in Low Frequency Components with a Binary Image," *Digit. Signal. Process.*, vol. 19, no. 2, pp. 265–277, 2009

[6] H. Ozer, B. Sankur, and N. Memon, "An SVD-based Audio Watermarking Technique," in Proc. *ACM MMSec*, pp. 51–56, IEEE Press, New York, 2005

[7] A. Al-Haj and A. Mohammad, "Digital Audio Watermarking Based on the Discrete Wavelets Transform and Singular Value Decomposition," *Eur. J. Sci. Res. 39(1)*, pp. 6–21, 2010

[8] V. Bhat K., I. Sengupta, and A. Das, "A New Audio Watermarking Scheme Based on Singular Value Decomposition and Quantization," *Circ. Syst. Signal. Pr.*, vol. 30, no. 5, pp. 915–927, 2011

[9] P. Dhar and T. Shimamura, "A DWT-DCT-Based Audio Watermarking Method Using Singular Value Decomposition and Quantization," *J. Signal. Process.*, 17(3), pp. 69-79, 2013

[10] Fathi E. Abd El-Samie, "An Efficient Singular Value Decomposition Algorithm for Digital Audio Watermarking," *Int. J. Speech. Tech.*, vol. 12, issue 1, pp. 27–45, 2009

[11] S. Vongpraphip and M. Ketcham, "An Intelligence Audio Watermarking Based on DWT-SVD Using ATS," in Proc. *WRI GCIS*, Xiamen, pp. 150–154, 2009

[12] A. Al-Haj, C. Twal and A. Mohammad, "Hybrid DWT-SVD Audio Watermarking," in Proc. *ICDIM*, Thunder Bay, ON, pp. 525–529, 2010

[13] M.S. Al-Yaman, M.A. Al-Taee and H.A. Alshammas, "Audio-Watermarking Based Ownership Verification System Using En-hanced DWT-SVD Technique," in Proc. *IEEE SSD*, Chemnitz, pp. 1–5, 2012

[14] V. Bhat K., I. Sengupta and A. Das, "An Audio Watermarking Scheme Using Singular Value Decomposition and Dither-Modulation Quantization," *J. Multimed. Tools. Appl.*, vol. 52, nos. 2-3, pp. 269–283, 2011

[15] J. Zhang, "Analysis on Audio Watermarking Algorithm based on SVD," in Proc. *ICCSIT*, Changchun, China, pp. 1986–1989, 2012

[16] A. Singhal, A. N. Chaubey and C. Prakkash, "Audio Watermarking Using Combination of Multilevel Wavelet Decomposition, DCT and SVD," in Proc. *ETNCC*, Udaipur, pp. 239–243, 2011

[17] P. K. Dhar and T. Shimamura, "Audio Watermarking in Transform Domain Based on Singular Value Decomposition and Quantization," in Proc. *APCC*, Jeju Island, pp. 516–521, 2012

[18] P. K. Dhar and T. Shimamura, "An Audio Watermarking Scheme Using Discrete Fourier Transformation and Singular Value Decomposition," in Proc. *TSP*, Prague, pp. 789–794, 2012

[19] G. Suresh, "An Efficient and Simple Audio Watermarking Using DCT-SVD," in Proc. *ICDCS*, Coimbatore, pp. 177–181, 2012

[20] S. Karimimehr, S. Samavi, H. R. Kaviani and M. Mahdavi, "Robust Audio Water-marking based on HWD and SVD," in Proc. *ICEE*, Tehran, pp. 1363–1367, 2012

[21] W. Jiang, "Fragile Audio Watermarking Algorithm Based on SVD and DWT," in Proc. *ICISS*, Guilin, pp. 83–86, 2010

[22] L. Lamarche, Y. Liu and J. Zhao, "Flaw in SVD-based Watermarking," in Proc. *CCECE*, Ottawa, Ont., pp. 2082-2085, 2006

[23] J. Karnjana, M. Unoki, P. Aimmanee and C. Wutiwiwatchai, "An Audio Watermarking Scheme Based on Singular-Spectrum Analysis," in Proc. *IWDW*, pp. 1–15, 2014

[24] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky, *Analysis of Time Series Structure: SSA and related techniques*, Chapman and Hall/CRC, 2001

[25] H. Hassani, "Singular spectrum analysis: Methodology and comparison," *J. Data. Sci.*, vol. 5, no. 2, pp. 239–257, 2007

[26] R. Storn and K. Price, "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces," *J. Global. Optim.*, vol. 11, no. 4, pp. 341–359, 1977