Speech Watermarking Technique Based on Singular Spectrum Analysis and Automatic Parameter Estimation using Differential Evolution for Tampering Detection

Kasorn Galajit
School of Information Science
Japan Advanced Institute of Science and Technology
Ishikawa, Japan
kasorn.galajit@nectec.or.th

Mongkonchai Intarauksorn
Sirindhorn International Institute of Technology
Thammasat University
Pathum Thani, Thailand
5822792834@g.siit.tu.ac.th

Jessada Karnjana
NECTEC
National Science and Technology Development Agency
Pathum Thani, Thailand
jessada.karnjana@nectec.or.th

Pakinee Aimmanee
Sirindhorn International Institute of Technology
Thammasat University
Pathum Thani, Thailand
pakinee@siit.tu.ac.th

Masashi Unoki
School of Information Science
Japan Advanced Institute of Science and Technology
Ishikawa, Japan
unoki@jaist.ac.jp

***Abstract*—A singular spectrum analysis-based watermarking scheme is proposed to detect speech signal tampering. The watermark is embedded into the original signal by modifying a part of less-significant singular values of the original signal, and later the extracted watermark is compared with the original watermark to detect the tampering. Differential evolution is deployed to select a part of singular spectrum to be modified to balance between the robustness of the scheme and the sound quality of the watermarked signal. The experimental results show that the proposed method can detect the types and the position of the signal being altered. The performance of the scheme has been improved since the previous methods in terms of inaudibility resulting in the excellent sound quality of the watermarked signal. Because the robustness has to be traded off against inaudibility, the proposed method seems to be not robust. However, the robustness can be improved by revising the cost function.**

***Index Terms*—singular spectrum analysis, automatic parameter estimation, speech-tampering detection, semi-fragile watermarking, differential evolution**

I. Introduction

The development of advanced digital technology gives benefit to societies and communities. However, the inappropriate use of these technologies can cause problems. For example, there are many digital tools that can duplicate and easily alter speech signals and allow those modified signals to be used in such a way that appears to be authentic. In other words, those tools allow the speech signal to be modified without checking ownership or authentication. Thus, the legal issues concerned with unauthorized speech-signal modification and tampering have risen in number and played an important role, especially if the recorded speech signals contain vital information, for instance, the recorded speech used in the court or the recorded speech used in a criminal investigation. Speech watermarking can be a possible solution to solve such issues.

To detect the tampering in speech signals, the secret information is embedded into the host signal and can later be extracted and analyzed. The extracted watermark is compared with the original watermark to detect the tampering. The analysis of the extracted watermark can be used to check the modification of the speech signal and its integrity. The required properties of the watermarking scheme depend upon the goal to be achieved. For the purpose of tampering detection, there are two main required properties. The first is the semi-fragility: the watermark is robust enough not to be significantly altered by non-malicious signal processing but rather easily transformed by the attacks. The second is inaudibility: the human auditory system should not perceive the secret information. In short, the two mandatory requirements for tampering detection are the inaudibility and the semi-fragility.

In the literature, Yan et al. proposed the semi-fragile speech-watermarking scheme by using quantization of linear prediction parameters. However, the parameters used in the scheme were selected by trial and error [1]. Wu et al. proposed

a speech fragile-watermarking scheme. Their results were reasonable, but their work focused only on the tampering with the speech content [2]. Wang et al. proposed a speech watermarking method based on formant tuning [3], [4]. Their proposed scheme satisfied both inaudibility and semi-fragility. However, it was too fragile to some signal processing operations that should not be considered as attacks such as pitch shifting and echo addition.

Recently, we proposed tampering detection for the speech signals by semi-fragile watermarking based on the singular-spectrum analysis (SSA) [5]. The watermark was embedded into the host signal by changing a part of the singular spectrum of host signal with respect to the watermark bit. The scheme could identify the speech segment that was tampered with and the type of the attacks. However, the embedding rule was not flexible. The modified singular value depended only on the largest and the smallest singular values. The scheme has not been parameterized so that it is adjustable. The performance of watermarking scheme could be more efficient if the embedding parameters could be adjusted. Thus, we proposed the scheme with ad-hoc parameters by allowing its parameters to be fine-tuned depending on the characteristics of input signal [6]. The scheme with ad-hoc parameters gave better performance than the one with a fixed rule. However, the tuning parameters were selected by trial and error.

This work aims to improve the performance of the semi-fragile watermarking scheme for tampering detection. The idea of the scheme is that the watermark is embedded into the host signal, and the extracted watermark is compared with the original watermark to identified the tampering. From our studies, we discovered that the SSA-based watermarking scheme could be made robust, fragile, or semi-fragile depending on the modified part of the singular spectrum. Since the scheme for tampering detection requires the inaudibility and the semi-fragility, the watermark is embedded into the host signal by modifying a part of its less-significant singular values with respect to a watermark bit and an embedding rule. Since, the modifying affects both the sound quality of the watermarked signal and the robustness of the scheme, the modified part must be determined appropriately in order to balance the sound quality and the robustness. Thus, the differential evolution (DE) optimization is deployed to properly select a part of the singular spectrum to be modified. DE determines the optimum parameters with respect to a cost function for balancing between the inaudibility and the robustness against many attacks.

The rest of the paper is organized as follows. Section II describes the proposed scheme. The embedding process, extraction process, and tampering detection are detailed. Section III explains a method for parameter estimation by which the optimal parameters will be obtained automatically by incorporating differential evolution. The experiment results and performance evaluation are provided in Section IV. Discussion and conclusion are made in Section V and Section VI, respectively.

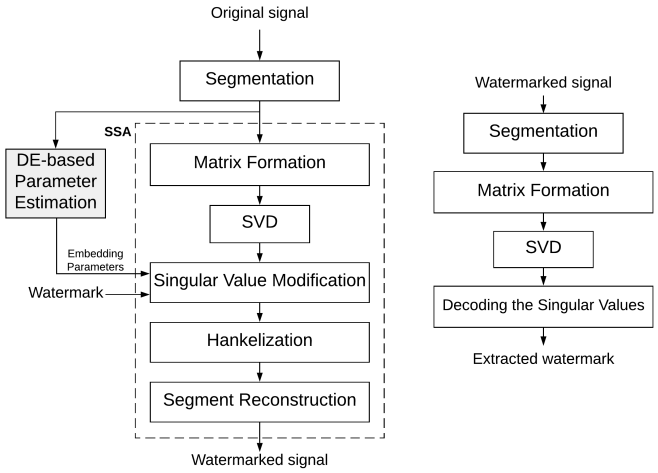

Fig. 1. Watermark embedding (left) and extraction (right) processes.

II. Proposed Watermarking Scheme

The first subsection describes the embedding process, and the second subsection describes the extraction process of the proposed scheme to achieve the inaudibility and the semi-fragility for tampering detection. The last subsection details a differential evolution optimizer for automatic parameters estimation.

A. Embedding Process

In the proposed watermarking scheme, one watermark bit will be embedded into one frame. There are six steps in the speech embedding process as shown in Fig. 1, which is detailed as follows. First, *segmentation*: a host signal is segmented into non-overlapping frames, where a total number of frames is equal to the number of the watermark bits. Second, *matrix formation*: each frame is used to construct the trajectory matrix $\mathbf{X}$ to represent the frame. An speech segment $F = [f_0 \;\; f_1 \; ... \; f_{N-1}]^\dagger$, where f_i for $i = 0, 1, ..., N-1$ denotes N samples of segment F. Segment F is mapped to a trajectory matrix $\mathbf{X}$ of size $L \times K$.

$$\boldsymbol{X} = \begin{bmatrix} f_0 & f_1 & \cdots & f_{K-1} \\ f_1 & f_2 & \cdots & f_K \\ \vdots & \vdots & \ddots & \vdots \\ f_{L-1} & f_L & \cdots & f_{N-1} \end{bmatrix}, \tag{1}$$

where L is a window length of matrix formation, $2 \leqslant L \leqslant N$, and $K = N - L + 1$.

Third, *singular value decomposition (SVD)*: SVD is performed on each trajectory matrix $\mathbf{X}$ to get a set of singular values of $\mathbf{X}$ in descending order. This set is called a singular spectrum of $\mathbf{X}$, which is denoted by $\{\sqrt{\lambda_0}, \sqrt{\lambda_1}, ..., \sqrt{\lambda_q}\}$, where $\sqrt{\lambda_i}$ for $i = 0, 1, 2..., q$ are the singular values, and $\sqrt{\lambda_q}$ is the smallest non-zero singular value.

Fourth, *sigular value modification*: how the singular spectrum is modified depending upon the objective to achieve in embedding. In the proposed scheme, the inaudibility and semi-fragility are the required properties in this scheme for

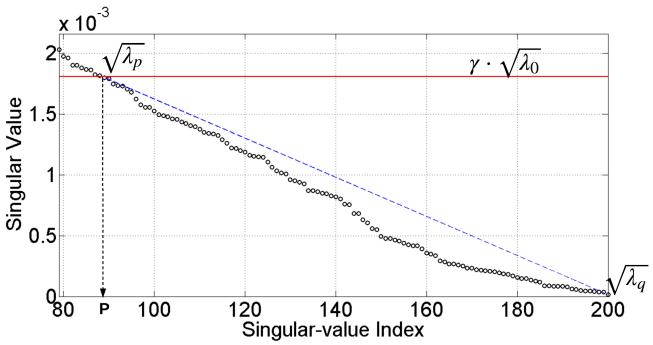

Fig. 2. Selected part of singular spectrum to be modified. The red line indicates the threshold level and the blue dashed line connects $\sqrt{\lambda_p}$ and $\sqrt{\lambda_q}$.

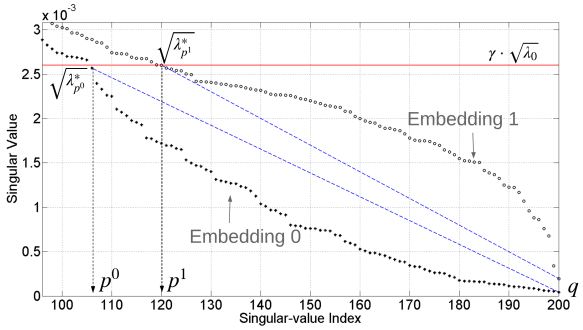

Fig. 3. Decoding the hidden watermark bit: most of singular values (circle) are under the threshold level and above the blue dashed line when a watermark bit is 1, and most of singular values (asterisks) are under the threshold level and also under the blue dashed line when a watermark bit is 0.

tampering detection. Thus, the less-significant part of the singular spectrum is selected to be modified in order to hide the watermark bit. The singular values of which its value is less than a threshold level are selected to be the modified part. We defined this threshold level by $\gamma \cdot \sqrt{\lambda_0}$, where $\sqrt{\lambda_0}$ is the highest singular value and the γ is a constant. The γ is adaptively determined by the differential evolution for each host signal.

Let $\sqrt{\lambda_p}$ denote the largest singular value that is less than the threshold level, $\sqrt{\lambda_i^*}$ is a modified singular value from index $i = p$ to q, and w is the embedded-watermark bit. The embedding rule is as follows.

$$\sqrt{\lambda_i^*} = \begin{cases} \sqrt{\lambda_i} + \alpha_i \cdot (\sqrt{\lambda_p} - \sqrt{\lambda_i}), & \text{if } w = 1, \\ \sqrt{\lambda_i} \quad \text{(i.e., unchanged)}, & \text{if } w = 0, \end{cases} \tag{2}$$

where p is an index of the largest singular value that is less than the threshold level, and α_i, called *embedding strength*, is normally distributed over the interval $[p, q]$ and has the maximum value of 1. Specifically, α_i is determined by the following equation

$$\alpha_i = e^{-\frac{(i-\mu)^2}{2(\sigma^2)}}, \tag{3}$$

where μ and σ are the mean and the standard deviation of the normal distribution.

The parameter set $\{\gamma, \mu, \sigma\}$ is the parameters of our embedding process, which should be optimized. The optimization of this parameter set is described in the next section. An example of a selected part of the singular spectrum is depicted in Fig. 2. Fifth, *Hankelization*: once the selected part of the singular spectrum of the frame is modified, the modified matrix is reconstructed by reversed SVD. Then, each matrix is transformed into the watermarked frame.

The final step, *segment reconstruction*: the watermarked signal is produced by stacking frames from the previous step.

B. Extraction and Tampering Detection Process

The extraction process takes the watermarked signal as an input for extracting the embedded watermark. The extraction process consists of five steps. The first three steps are the same as the first three steps of the embedding process, which are the *segmentation*, the *matrix formation*, and the *singular value decomposition*. The fourth step is *decoding the singular value*: the watermark bits are extracted in this step by decoding the singular spectra, and how the singular spectra are decoded depends on how the singular spectra are modified. The embedding rule must be known in the extraction process. To understand the idea behind the decoding step, let us start by considering Fig. 3. This figure shows two extracted singular spectra of one watermarked frame: $\{\sqrt{\lambda_{0^0}^*}, ..., \sqrt{\lambda_{p^0}^*}, ..., \sqrt{\lambda_{q^0}^*}\}$ and $\{\sqrt{\lambda_{0^1}^*}, ..., \sqrt{\lambda_{p^1}^*}, ..., \sqrt{\lambda_{q^1}^*}\}$. The superscripts 0 and 1 of the indices of singular values denote the embedded watermark bits. It can be noticed that most of the singular values (circles) under the red line are above the blue dashed line when a watermark bit 1 is embedded, and most of the singular values (asterisks) under the red line are also under the blue dashed line when a watermark bit 0 is embedded. Therefore, we can use the following condition to determine the embedded watermark bit $\hat{w}$.

$$\hat{w} = \begin{cases} 0, & \text{if } \sum_{i=p}^{q} \left(\sqrt{\lambda_i^*} - l(i) \right) < 0, \\ 1, & \text{if } \sum_{i=p}^{q} \left(\sqrt{\lambda_i^*} - l(i) \right) \geq 0, \end{cases} \tag{4}$$

where $l(i)$ is the corresponding values on the blue dashed line, which is defined by

$$l(i) = \left(\frac{\sqrt{\lambda_p^*} - \sqrt{\lambda_q^*}}{p - q} \right) \cdot (i - q) + \sqrt{\lambda_q^*}. \tag{5}$$

Once all hidden bits are extracted, the extraction precision rate is calculated to predict the degree of tampering.

III. DE for Parameter Estimation

The embedding rule (2) used in the embedding process involves three important parameters: γ, μ, and σ. These parameters are used to select a range of the singular spectrum to be modified. Since the singular spectrum of each speech segment is different from another, it is reasonable to set the parameters according to the speech segment. In other words,

a different set of parameters are required in order to minimize the distortion due to the embedding process for different speech segment.

Differential evolution (DE) is deployed for automatic parameter estimation. The detail of DE can be found in [12]. DE is used to balance between many requirements of the scheme. For example, The watermark is required to be imperceptible, and the scheme requires a high extraction precision. DE balances the requirements by iteratively improving candidate parameters by evaluating the cost function. There are three reasons that differential evolution is selected to be the optimizer in this work. First, it is a multipoint optimizer, i.e., the effect of the starting point problem can be mitigated. Second, it is a derivative-free approach, and we do not have to worry about whether the cost function is differentiable. Third, it has been proved that it is the fastest search algorithm in its computational class [12].

In this work, the cost function is calculated from many factors, which are log spectral distance (LSD), and the bit-error rate (BER). The log-spectral distance (LSD), sometimes called log-spectral distortion, is a distance measure (expressed in dB) between two spectra: the spectrum of the original signal and that of the watermarked signal. LSD is defined by

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

where $P(\omega)$ and $\hat{P}(\omega)$ are the spectrum of the original signal and that of the watermarking signal, respectively.

Let $w(i)$ is the embedded-watermark bit, and $\hat{w}(i)$ is the extracted-watermark bit for $i=1$ to M, the BER is defined by

$$\mathrm{BER} = \frac{1}{M}\sum_{i=1}^{M} w(i) \oplus \hat{w}(i), \tag{7}$$

where $\oplus$ is the bitwise XOR operator, and M is the total number of frames.

The cost function implemented in the proposed scheme is defined as follow.

$$\mathrm{Cost} = \sqrt{\mathrm{LSD}^2 + \overline{\mathrm{BER}}^2}, \tag{8}$$

where $\overline{\mathrm{BER}}$ is the average BER, which is defined by

$$\overline{\mathrm{BER}} = \beta_1 B1 + \beta_2 (B2 + B3 + B4) + \beta_3 B5, \tag{9}$$

where β_1, β_2, β_3 are constants, and $\beta_1 + 3\beta_2 + \beta_3 = 1$. $B1$ is BER of the extraction process without any attack performed on the watermarked signal, $B2$ is BER when G.711 was performed, $B3$ is BER when MP3 was performed, $B4$ is BER when MP4 was performed, and $B5$ is BER when G.726 was performed on the watermarked signal, respectively.

Let us consider the cost function. LSD represents a cost in terms of the objective measures of inaudibility, and $\overline{\mathrm{BER}}$ represents a cost in terms of the objective measures of semi-fragility. The DE optimizer finds the parameter set $\{\gamma, \mu, \sigma\}$ that minimizes the cost value. In our experiment, the maximum number of iterations was set to be 20, and other constants, such as a number of population and crossover constant, were set as suggested in [12]. The DE optimizer is shown in Fig. 4.

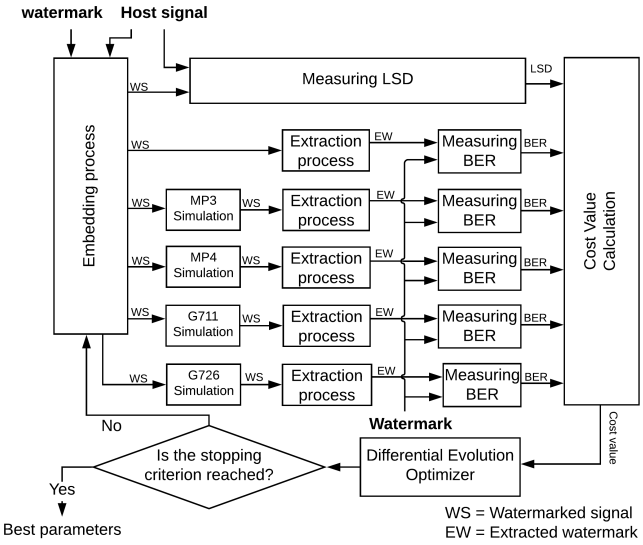

Fig. 4. DE optimizer for automatic parameter estimation.

IV. Results and Evaluation

In our experiment, twelve speech signals from the ATR database (B set) were used [8]. All signals have a sampling rate of 16 kHz, 16-bit quantization, and one channel. The watermark was embedded into the host signals starting from the initial frame. The frame size was 25 ms or 400 samples. In other words, the embedding capacity was 40 bps. One hundred and twenty bits were embedded into each signal in total, and the embedding duration of each signal was 3 seconds. The proposed scheme was evaluated in three aspects: the sound quality of watermarked signals, semi-fragility, and tampering detection. The evaluation results were compared with our previous work [5], [6] and three other conventional methods: a method based on embedding information into the least significant bit (LSB) [9], the cochlear-delay-based (CD-based) method [10], and the formant-enhancement based (FE-based) method [3], [4].

A. Sound Quality Evaluation

Three objective evaluations were conducted to evaluate the speech quality of watermarked signals: log-spectral distance (LSD), the perceptual evaluation of speech quality (PESQ), and the signal-to-distortion ratio (SDR). The LSD is a distance measure (expressed in dB), as expressed by (6). PESQ measures the degradation of the watermarked signal compared with the original signal. PESQ used in our simulation is recommended by ITU-T recommendation P.862, and it maps the degradation to a PESQ score, which ranges from very annoying (-0.5) to imperceptible (4.5) [13]. SDR is the power ratio between the signal and the distortion (expressed in dB), defined by

$$\mathrm{SDR} = 10\log\frac{\sum_n [A(n)]^2}{\sum_n \left[A(n) - \hat{A}(n)\right]^2}, \tag{10}$$

TABLE I
SOUND-QUALITY EVALUATIONS: THE PROPOSED SCHEME VS. THE OTHER METHODS.

	PESQ score	**LSD**(dB)	**SDR**(dB)
LSB-based method [9]	4.49	0.19	65.35
CD-based method [10]	∼3.1-4.3	∼0.6-0.8	-
FE-based method [3] [4]	∼3.9	∼0.4	-
SSA-based method (fixed rule) [5]	3.64	0.69	30.96
SSA-based method (with ad-hoc parameters) [6]	3.70	0.65	31.58
Proposed method	4.50	0.18	56.03

where $A(n)$ and $\hat{A}(n)$ are the amplitudes of the original and those of the watermarked signals, respectively.

In this work, we set the criteria for the acceptance of the good sound quality as follows. The LSD should be smaller than 1 dB, PESQ score should be higher than 3.0, and the SDR should be greater than 30 dB [4]. The results of the sound quality evaluation are shown in Table I. All watermarking methods satisfied the sound quality evaluation. Besides the LSB-based method, the proposed method was the best in terms of inaudibility, whereas the others were comparable. The proposed scheme performed better in all three measurements compared with the previously proposed methods.

B. Semi-fragility Evaluation

Watermarking scheme should be robust against normal speech processing (e.g., compression and speech codecs) and fragile to malicious attacks (e.g., Gaussian-noise addition and pitch shifting). The robustness can be indicated by the bit-error rate (BER), as defined in (7). In this work, we choose the BER of 10% as the criterion, and the lower BER indicates stronger robustness [4], [6]. If BER is higher than 20%, the speech signal is considered as being tampered with. A speech signal is presumably unintentionally modified or tampered with a low degree if its BER is between 10% and 20%.

We evaluated the fragility of the proposed scheme against tampering the speech signal with various possible attacks. In this experiment, ten signal-processing operations were performed on the watermarked signal: G.711 speech companding, G.726 companding, MP3 compression with 128 kbps, MP4 compression with 96 kbps, band-pass filtering (BPF) with 100-6000 Hz and -12 dB/Octave, Gaussian-noise addition (AWGN) with 15-dB and 40-dB signal-to-noise ratios (SNR), pitch shifting (PSH) by $\pm 4\%$, $\pm 10\%$, and $\pm 20\%$, single-echo addition with -6 dB and the delay time of 20 and 100 ms, replacing $1/3$ and $1/2$ of the watermarked signals with an unwatermarked segment, and $\pm 4\%$ speed changing (SCH).

The results are shown in Table II. The proposed method is fragile when the attacks are performed on the watermarked signal. However, it seems not to be robust compared with others and our previous methods. The semi-fragility of the proposed method needs to be improved, and this issue will be mentioned in the next section.

C. Tampering Detection

This section demonstrates how our proposed scheme can be used for tampering detection. A 75×15 bitmap image in Fig. 5 (a) was used as the watermark. In order to embed the complete image, 12 speech signals are repeatedly connected to construct a long host signal. After embedding the watermark, the middle segment of the watermarked signal was tampered with the malicious attacks listed in Table II. Replacing the watermarked signal with un-watermarked signal can be considered as the content replacement. Speeding up or slowing down the watermarked signal can be considered as modifying the duration and tempo of the signal. A pitch shift is to proportional-shift frequency component, and it can be referred to as manipulating the individuality of the speaker. The results are shown in Fig. 5. The proposed scheme not only can locate the position that has been tampered with but also can identify types of the attacks as well as the degree of the attacks.

V. DISCUSSION

There are three important issues that we would like to discuss here. The first issue involves a weight of the cost function of DE, which is shown in (8) and (9). Many factors contribute to the cost function, for example, the MP3 and MP4 attack simulations, LSD, G711 and G726 simulations, as depicted in Fig. 4. The different weight of those factors could result in different sets of the optimal parameters, and the weight also affects the performance of the scheme. For example, the proposed scheme gives the excellent sound quality of the watermarked signal, as shown in Table I, but the robustness of the scheme is slightly reduced, as shown in Table II. The relationship between the weight of the cost function and the performance of the scheme will be studied in the future study. The second issue is that the parameter γ is shared in the embedding process and the extraction process. Since the blindness watermarking should need only the watermarked signal for the watermark extraction, this sharing can cause the scheme to be considered as semi-blind. The future study may be needed to make the scheme blindness. The final issue is to evaluate method in terms of its ability to predict the types and degrees of attacks. It can be seen that the reconstructed images in Fig. 5. can indicate the attack degrees. If the speech segment is replaced by an unwatermarked segment, the tampered area can be noticed in the reconstructed image, as shown in Fig. 5(s) (replaced with 1/3 of a watermarked signal) and Fig. 5(t) (replaced with half of a watermarked signal). In the case of changing in speed of the watermarked signal, the difference between speeding up and speeding down can be distinguished, as shown in Fig. 5(q) and Fig. 5(r). In addition, in the case of pitch shifting, the reconstructed images were significantly different when different degrees of the attacks were performed on the watermarked signal. Thus, the proposed scheme is effective for predicting degrees of the attacks, as shown from Fig. 5(k) to Fig. 5(p).

VI. CONCLUSION

In this paper, the semi-fragile watermarking scheme has been proposed for tampering detection. The scheme is an SSA-based watermarking technique with automatic parameter estimation. The watermark is embedded into the host signal

TABLE II
BER(%): THE PROPOSED SCHEME VS. THE OTHER METHODS.

	LSB-based method [9]	**CD-based method [10]**	**FE-based method [3] [4]**	**SSA-based method (fixed rule) [5]**	**SSA-based method (with ad-hoc parameters) [6]**	**Proposed method**
No attack	0.00	∼0-1	0.00	0.49	0.36	9.44
Non-Malicious Signal Processing Operations						
G.711	0.00	∼4	0.00	0.49	0.36	9.44
G.726	51.77	∼10-25	0.00	27.66	21.07	36.11
MP3	50.49	-	-	3.69	5.39	18.40
MP4	49.53	-	-	32.79	34.19	38.25
Malicious Attacks						
BPF	50.83	-	-	50.23	50.46	41.38
AWGN (15, 40 dB)	50.70, 49.53	-	∼54	49.69, 24.53	48.67, 23.28	36.80, 37.04
PSH (-4%, -10%, -20%)	35.64, 35.33, 4.08	-	∼31, -, -	10.58, 22.03, 47.83	14.25, 36.16, 51.47	20.48, 22.01, 22.70
PSH ($+4\%$, $+10\%$, $+20\%$)	34.42, 34.36, 38.03	-	-	12.44, 15.33, 20.47	7.78, 10.92, 21.94	13.32, 14.16, 16.18
Echo (20, 100 ms)	50.18, 51.34	-,∼50	-,∼5	15.76, 20.33	9.22, 18.05	11.31, 17.98
Replace ($1/3$, $1/2$)	16.51, 24.97	-	∼57, -	17.08, 25.78	18.57, 26.25	22.56, 30.62
SCH (-4%, $+4\%$)	49.47, 48.72	-	∼20, -	47.00, 47.19	46.58, 46.94	22.15, 21.04

Fig. 5. Comparison of the watermark image between an original image (a) and the reconstructed images after performing the following signal-processing operations: (b) BPF, (c) G.711, (d) G.726, (e) AWGN (15 dB), (f) AWGN (40 dB), (g) Echo (20 ms), (h) Echo (100 ms), (i) MP3, (j) MP4, (k) PSH $+20\%$, (l) PSH -20%, (m) PSH $+10\%$, (n) PSH -10%, (o) PSH $+4\%$ (p) PSH -4%, (q) SCH $+4\%$, (r) SCH -4%, (s) Replace $(1/3)$, and (t) Replace $(1/2)$.

by modifying the selected part of the singular spectrum. The modified part should be appropriately selected in order to balance the two primary requirements: inaudibility and semi-fragility. A differential evolution algorithm is deployed to estimate the embedding parameters, in other words, to select the proper part for modification. The proposed method was improved from the previous methods regarding "inaudibility". As the results, in Table I, LSD and SDR of the proposed method are the outstanding performance. On the other hand, due to a trade-off between inaudibility and robustness, the proposed method seems to be not robust. BER under no attack condition in the proposed method is about 9%. So, other BERs under attack conditions are not good, in comparison with the previous methods. However, BERs can be improved by revising the cost function.

ACKNOWLEDGMENT

This work was supported under a grant in the SIIT-JAIST-NSTDA Dual Doctoral Degree Program. It was also supported by the Grant-in-Aid for Scientific Research (B) (No.17H01761) and I-O DATA foundation.

REFERENCES

[1] B. Yan, Z.M. Lu, S.H. Sun, and J.S. Pan, "Speech authentication by semi-fragile watermarking," In International Conference on Knowledge-Based and Intelligent Information and Engineering Systems , Springer, Berlin, Heidelberg, pp. 497–504, September 2005.

[2] C.P. Wu, and C.C. Kuo, "Fragile speech watermarking for content integrity verification," In Circuits and Systems, 2002 ISCAS 2002.,IEEE International Symposium , vol. 2, pp. II–II, 2012.

[3] S. Wang, M. Unoki, M., and N.S. Kim, "Formant enhancement based speech watermarking for tampering detection," In Fifteenth Annual Conference of the International Speech Communication Association, 2014.

[4] S. Wang, and M. Unoki, M. "Speech watermarking method based on formant tuning," IEICE TRANSACTIONS on Information and Systems, vol. 98.1, pp 29–37, 2015.

[5] J. Karnjana, M. Unoki, P. Aimmanee, and C. Wutiwiwatchai, "Tampering detection in speech signals by semi-fragile watermarking based on singular-spectrum analysis," In Advances in Intelligent Information Hiding and Multimedia Signal Processing, Springer, Cham, pp. 131–140, 2017.

[6] J. Karnjana, K. Galajit, P. Aimmanee, C. Wutiwiwatchai, and M. Unoki, "Speech watermarking scheme based on singular-spectrum analysis for tampering detection and identification," In Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), pp. 193–202, December 2017.

[7] J. Karnjana, M. Unoki, P. Aimmanee, and C. Wutiwiwatchai, "SSA-based audio-information-hiding scheme with psychoacoustic model," In Signal and Information Processing Association Annual Summit and Conference (APSIPA), pp. 1–10, December 2016.

[8] K. Takeda, Y. Sagisaka, S. Katagiri, and M. Abe, "Speech database users manual," ATR Interpreting Telephony Research Laboratories, TR-I-0028, 1988.

[9] P. Bassia and I. Pitas, "Robust audio watermarking in the time domain," In EUSIPCO, vol. 98, pp. 25–28, 1998.

[10] M. Unoki and R. Miyauchi, "Detection of tampering in speech signals with inaudible watermarking technique," In Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP), pp. 118–121, 2012.

[11] J. Karnjana, M. Unoki, P. Aimmanee, and C. Wutiwiwatchai, "Singular-Spectrum Analysis for Digital Audio Watermarking with Automatic Parameterization and Parameter Estimation," IEICE TRANSACTIONS on Information and Systems, vol. 99.8, pp. 2109–2120, 2016.

[12] R. Storn, and K. Price, "Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces," Journal of global optimization, vol. 11.4, pp. 341–359, 1997.

[13] Recommendation, ITU-T, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," Rec. ITU-T P. 862, 2001.