# Hybridization of speech information hiding and encryption for double-layer security in speech communication

Kasorn Galajit*†‡, Jessada Karnjana†, Pakinee Aimmanee‡ and Masashi Unoki*
* Japan Advanced Institute of Science and Technology, Ishikawa, Japan
E-mail: {kasorn,unoki}@jaist.ac.jp Tel/Fax: +81-76-1511237
† NECTEC, National Science and Technology Development Agency, Pathum Thani, Thailand
E-mail: jessada.karnjana@nectec.or.th Tel/Fax: +66-2-5646900
‡ Sirindhorn International Institute of Technology, Thammasat University, Pathum Thani, Thailand
E-mail: pakinee@siit.tu.ac.th Tel/Fax: +66-2-9869009

***Abstract*—Speech transmitted over communication systems must be highly secured since they are vulnerable to attacks. In this study, we use the singular spectrum analysis (SSA)-based information hiding method with the transformation method to boost security for speech communication. Arnold transformation is performed on watermark signals to provide secured watermarks. The secured watermarks are then embedded into the host signal using SSA-based information hiding to obtain a watermarked signal. The watermarked signal is encrypted before being sent through a communication channel. The experimental results show that effective encryption and information hiding are feasible. The key sensitivity and the discrepancy between the watermarked speech signal and its encrypted version measured using the correlation coefficient and the signal-to-noise ratio suggest that only the authorized person who has the encryption key will be able to access the speech's contents. A secured watermark signal's imperceptibility and a high bit error rate without a watermark key indicate hidden information is limited in access. This hybridized system provides speech security and selectively grants access to data at varying levels.**

## I. Introduction

Advancements in digital technology, particularly the internet and communications engineering, have made sharing multimedia data more accessible and easier. Speech and audio, which is one form of multimedia data, has been widely used in various systems and applications, e.g., audio and video conferencing, a voice-activated command for system control, remote medical treatment [1], [2]. However, transmitting sensitive and important information over a communication network or system presents security risks since the communication systems can be vulnerable to attacks [3]. Thus, it is necessary to provide secure communication for such multimedia data.

This work focuses on the security of speech signals. Cryptography is a classical method that provides security by concealing speech signals to prevent them from being stolen and modified. However, cryptography does not protect speech signals once the content has been decrypted [4]. Our research aims to provide an extra level of security for speech signals by combining information hiding and encryption. Encryption allows only authorized persons to access the original contents of the speech signals, and information hiding can be used to either provide a secret communication channel or to track fraud or modification of the speech signals.

Several techniques have been previously developed to secure speech signals. For example, RSA, a widely used public-key cryptosystem, has been used for speech data encryption and decryption [5]. The chaotic algorithm had a shorter computation time, but there was a trade-off with security level [6]. Encryption of speech signals with multiple secret keys and uses Discrete Cosine Transform (DCT) or the Discrete Sine Transform (DST) to remove the signal intelligibility [7]. The system based on Advanced Encryption Standard (AES) provides security enhancement for speech signal [8]. These techniques were based on a cryptography system. Therefore, there is a crucial disadvantage: once the content has been decrypted, speech signals do not be protected anymore. Besides, information hiding is one solution to protect speech signals. Information hiding uses hidden information to detect tampering in speech signal with different hiding techniques [9], [10], [11], [12], [13]. However, the information hiding does not protect the content of the cover speech. There were attempts to cooperate two techniques together, called *hybridized system*. For example, multiple scrambling was applied to strengthen information hiding [14]. However, this work focused only on protecting the hidden information, but anyone can access the cover's contents. The audio encryption algorithm using an elliptical curve and Arnold transformation was evaluated to determine its suitability for information hiding, but it did not include the implementation or evaluation of an information hiding scheme [15]. The hybrid domain was applied in audio watermarking with chaotic encryption, but the encryption was only applied to the hidden information [16].

In this proposed method, the Arnold scrambling algorithm is deployed in conjunction with the singular spectrum analysis (SSA)-based information hiding method, resulting in a hybridized system that enhances the security of speech communication. The advantages of the hybridized system are

as follows. First, the watermarks are transformed before they are hidden in the host signal; thus, this can be used as a secure communication channel accessible only by a person with keys to access hidden information. Second, the hidden watermarks provide the ability to detect any speech signal modification, as was done in Karnjana et al.'s study [10]. Third, the watermarked signals are encrypted before they are sent to the communication channel, so only authorized persons who have a key are allowed to access the speech signal's contents. Therefore, the main advantage of a hybridized system over the cryptography method is that after the content is decrypted, the information hiding technique can still provide security for the speech signals. For example, information hiding provides a secret communication using hidden information, or hidden information can track fraud or modification of decrypted cover speech. In this hybridized system, access to the data can be selectively granted to authorized persons at varying levels; for example, a person who has one key can only access the contents of the speech signal, while those who have two keys can access both the content and hidden information. The hybridized system can be used to establish a secure communication channel and a system with different levels of access.

## II. Background

### A. Singular spectrum analysis-based Information hiding

SSA-based information hiding was proposed by Karnjana et al. [17]. The scheme used basic SSA to analyze host signals and extract the singular spectra, and the watermark signal was hidden in a part of the spectra. The scheme consisted of embedding and extraction processes.

In embedding, the host speech signal is segmented into non-overlapping frames. One watermark bit is embedded into one frame. Thus, the number of frames is equal to the number of the watermark bits to be embedded. Then the trajectory matrix $\mathbf{F}$ which represents each frame $F$ is constructed. Singular value decomposition (SVD) is performed on each trajectory matrix $\mathbf{F}$ to obtain each frame's singular spectra. The singular spectra are modified to hide the watermark bit (0 or 1), and the part of the singular spectra to be modified depends on the requirement of the information hiding application. The modified trajectory matrix $\mathbf{Y}$ is constructed by SVD reversion and then hankelized. The hankelization of a modified trajectory matrix $\mathbf{Y}$ yields a signal $G$, where $G$ is a frame of the watermarked signal. The frames are stacked to reconstruct the watermarked signal.

In extraction, the watermarked signal is first segmented into non-overlapping frames, and the trajectory matrix is constructed in the same way as in the embedding process. Then SVD is performed on the trajectory matrix to obtain the singular spectra. The singular spectra of the signal of each frame are typically convex; however, the watermark bit embedded into an interval of the singular spectrum of a host frame results in a concave part on the interval of the singular spectrum of the reconstructed, watermarked frame.

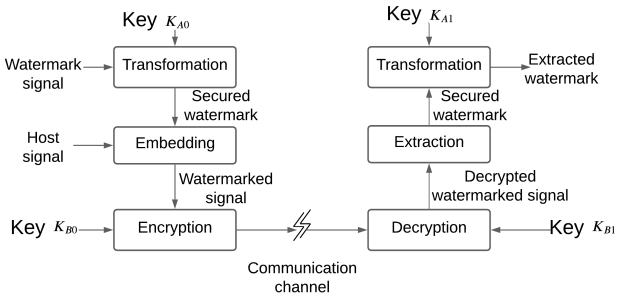


Fig. 1. Proposed scheme: Emitter (left), and receiver (right).

This property can be utilized to extract the watermark bit from each frame.

### B. Arnold scrambling algorithm

The Arnold scrambling algorithm, or Arnold transformation, describes a discrete mapping from site $(x_t,y_t)$ to site $(x_{t+1},y_{t+1})$ with circumference $N$, where $(0 \leq t < N)$ and $\mathrm{mod}$ is a modulo function [18], [19], [20].

$$\begin{bmatrix} x_{t+1} \\ y_{t+1} \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} \bmod N. \quad (1)$$

Arnold transformation is used to alter a matrix $X$ of dimension $N$x$N$ into a matrix $X^{'}$ to decrease the correlation coefficient between the matrices. Arnold transformation is cyclical, and iterated. The scrambling key is needed as a secret key to identify the number of iterations during the transformation process to bring back the original matrix. In the proposed method, Arnold transformation is applied to the watermark signal to provide a secured watermark signal, and it is, in turn, applied to the watermarked signal for encryption.

## III. Proposed method

This section introduces the proposed hybridization of information hiding and encryption method. The main scheme consists of the emitter side and the receiver side, as illustrated in Fig. 1. The watermark signal is transformed using key $K_{A0}$ and embedded into the host signal to produce a watermarked signal, which is later encrypted using $K_{B0}$ to be sent through the communication channel on the emitter side. The watermarked signal is decrypted using key $K_{B1}$ and then decoded to obtain the secured watermark, which is later transformed using $K_{A1}$ to obtain the original watermark on the receiver side. The details of each side are as follows

### A. Emitter side

A detailed diagram of the emitter side in the proposed scheme is shown in Fig. 2. The left-hand side of Fig. 2 (a) shows the procedure for creating a secured watermark. The right-hand side (b) shows the procedure for inserting secured watermarks into the host signal to obtain the watermarked signal and encrypting the watermarked signal.

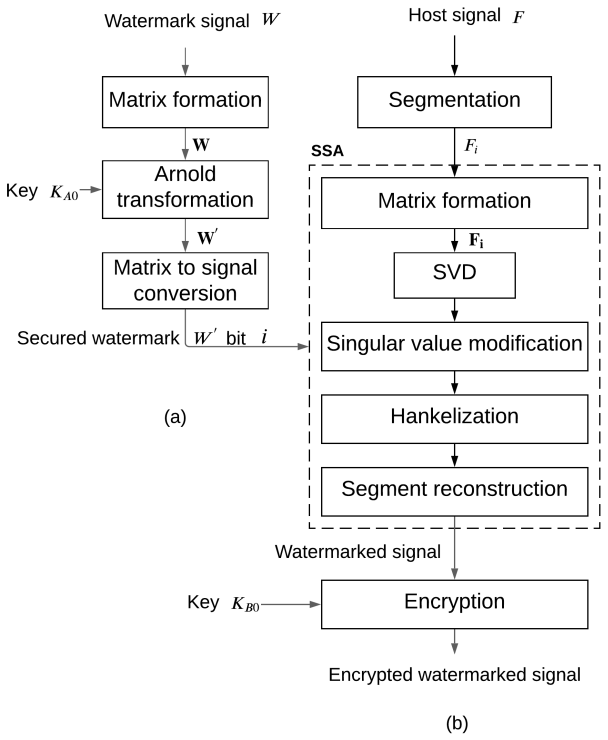


Fig. 2. Emitter side of proposed scheme.

Fig. 2 (a), the watermark signal $W$ is divided into vectors converted to $N$x$N$ matrix $\mathbf{W}$, and Arnold transformation alters the watermark matrix $\mathbf{W}$ to obtain altered watermark matrix $\mathbf{W}^{'}$ using key $K_{A0}$, where key $K_{A0}$ is a pre-defined number of transformation iterations. Next, the altered watermark matrix $\mathbf{W}^{'}$ is converted to a secured watermark signal $W^{'}$ to be embedded into the host speech signal. The confidentiality of the watermark signal can be strengthened as a result.

Fig. 2 (b), the secured watermark is embedded into the host signal to obtain the watermarked signal, and the encryption is applied in the last step. The steps are detailed as follows:

1) *Segmentation.* The host speech signal is segmented into frames of equal length $M$, where $M$ is the total number of samples in each frame.
2) *Matrix formation.* A signal $F$ of each frame is mapped to a trajectory matrix $\mathbf{F}$ of the size $L\times K$, where $F = [f_0\ f_1\ ...\ f_{M-1}]^{\mathrm{T}}$. The signal $F$ is mapped to matrix $\mathbf{F}$ by the following relation

$$\mathbf{F} = \begin{bmatrix} f_0 & f_1 & \cdots & f_{K-1} \\ f_1 & f_2 & \cdots & f_K \\ \vdots & \vdots & \ddots & \vdots \\ f_{L-1} & f_L & \cdots & f_{M-1} \end{bmatrix}. \quad (2)$$

where $L$ is a *window length*, and $2 \le L \le M$, and $K$ is $M-L+1$.

3) *Singular Value Modification*. A singular spectrum is modified on the basis of the secured watermark bit to be embedded. Given a singular spectrum $\{\sqrt{\lambda_0}, \sqrt{\lambda_1}, ..., \sqrt{\lambda_q}\}$, a specific part of this singular spectrum, which is $\{\sqrt{\lambda_p}, \sqrt{\lambda_{p+1}}, ..., \sqrt{\lambda_q}\}$, is modified on the basis of the secured watermark bit $w$ with

$$\sqrt{\lambda_i^*} = \begin{cases} \sqrt{\lambda_i} + \alpha_i(\sqrt{\lambda_p} - \sqrt{\lambda_i}), & \text{if } w = 1, \\ \sqrt{\lambda_i} \quad \text{(i.e., unchanged)}, & \text{if } w = 0, \end{cases} \quad (3)$$

where $\sqrt{\lambda_i^*}$ is the modified singular value for $i = p$ to $q$, $\sqrt{\lambda_p}$ is the largest singular value that is less than $\gamma\cdot\sqrt{\lambda_0}$, $\alpha_i$ is an embedding strength, as defined in [11]. Note that $\gamma$ is a pre-defined value to control the number of singular values to be modified, and $\gamma$ was set to 0.008 as same as prior SSA-based method [11].
4) *Hankelization*. A watermarked matrix $\mathbf{X}^*$ is computed as the product of $\mathbf{U\Sigma}^*\mathbf{V}^{\mathrm{T}}$ and then hankelized to obtain the signal $F^*$, which is the watermarked segment. The hankelization is the average of the anti-diagonal $i+j = k+1$, where $i$ and $j$ are the row index and the column index, respectively, of an element of $\mathbf{X}^*$, and $k$ ( for $k=0$ to $M-1$) is the index of element $F^*$.
5) *Segment Reconstruction*. The watermarked signal is finally produced by sequentially concatenating all watermarked segments.
6) *Encryption*. The watermarked signal from the previous step is transformed into an $R$x$R$ matrix and encrypted using key $K_{B0}$ to scramble its elements. The encrypted matrix is reshaped into one dimension resulting in an encrypted watermarked signal to be sent through the communication channel.

Note that the Arnold transformation was applied to secure a watermark signal and to encrypt the watermarked signal. However, the process was referred to as a *transfomation* when performed on a watermark signal, and *encryption* when performed on the encrypted watermarked signal. This is to clarify which signal is being transformed as each process differs slightly. For example, the matrix size $N$x$N$ of the watermark signal may differ from the matrix size $R$x$R$ of the encrypted watermarked signal due to the signals' size difference. The matrix only represents the square matrix, and its value $N$ and $R$ can be pre-defined. Since the matrix sizes differ, matrix construction and signal reconstruction differ as well.

### *B. Receiver side*

A detailed diagram of the receiver side in the proposed scheme is illustrated in Fig. 3. There are two main procedures on the receiver side, (a) extracting a secured watermark, and (b) retrieving the original watermark.

The left-hand side of Fig. 3 shows the five steps in (a). The first step is *Decryption*. The received signal is reshaped into $R$x$R$ matrix and is decrypted using key $K_{B1}$ to produce the watermarked signal. Note that key $K_{B1}$ on the receiver side matches $K_{B0}$ on the emitter side. The decrypted watermarked

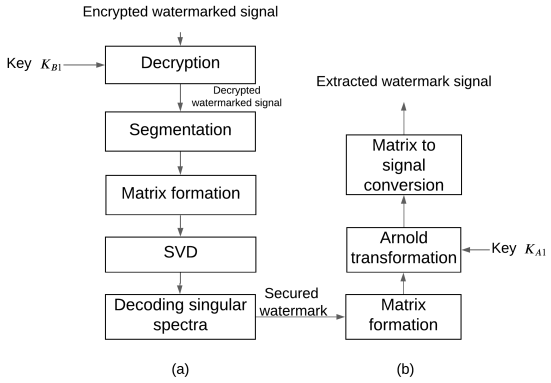


Fig. 3. Receiver side of proposed scheme.

signal is then passed through the next three steps, which are *Segmentation*, *Matrix formation*, and *SVD*, as is done on the emitter side. The last step is *Decoding the singular spectra*. The secured watermark bits are extracted by decoding the singular spectra, and how the spectra are decoded depends on how they are modified in the embedding process. The embedding rule in equation (3) results in the concave part on the singular spectra if embedding bit 1. Therefore, we can use the following condition to determine the secured watermark bit $w^*$.

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

where $l(i)$ is the corresponding values on the line connected $\sqrt{\lambda_p^*}$ and $\sqrt{\lambda_q^*}$, which is defined by

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

The right-hand side of Figure 3 (b) shows how the secured watermark bit $w^*$ is transformed to obtain the extracted watermark bit $\hat{w}^*$. The secured watermark signal is divided and converted to an $N$x$N$ matrix. The Arnold transformation transforms the secured watermark matrix using key $K_{A1}$ to recover the original watermark. Note that key $K_{A1}$ on the receiver side matches of $K_{A0}$ on the emitter side.

## IV. Evaluations and Results

In the experiment, twelve speech signals of Japanese sentences uttered by six men and six women from the ATR database (B set) were used [21]. The speech signals were one-channel with a 16-kHz sampling rate and 16-bit quantization. Since the proposed scheme is a hybridization of speech information hiding and encryption for double-layer security in speech communication, we evaluated the proposed scheme with respect to information hiding and encryption. We also evaluated the robustness of the entire system.

### A. Evaluation of information hiding

Our proposed scheme is based on adding secured watermarks to the host speech signal before encrypting the watermarked signal and extracting the original watermark signal in the decryption process. The purpose of information hiding is to enhance the doubled security achieved by an imperceptible watermark signal in a watermarked signal and secured watermark signals. As an example, if the algorithm for encryption is compromised so that the attackers can capture the contents of speech signals, they cannot perceive the watermarks hidden in the contents. Since the watermark bits were transformed before they were hidden in speech signals, the key is needed to discover the original watermark.

Three measurements were used to assess the imperceptibility of the watermark signal in a watermarked signal: the *log-spectral distance* (LSD), the *signal-to-distortion ratio* (SDR), and the *Perceptual Evaluation of Speech Quality* (PESQ). The LSD is the distance between the spectrum of the host speech and that of the watermarked signal (in dB), and defined as

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

where $P(\omega)$ and $P^*(\omega)$ are the spectra of the original signal and the watermarked signal, respectively. SDR is the power ratio between the signal and the distortion (in dB), defined as

$$\mathrm{SDR} = 10 \log \frac{\sum_n [A(n)]^2}{\sum_n \left[ A(n) - \hat{A}(n) \right]^2}, \tag{7}$$

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

The PESQ represents the sound-quality degradation of the watermarked signal compared with that of the host signal. The results principally model mean opinion scores (MOS) ranging from 1 (poor) to 5 (excellent).

The criteria for acceptable imperceptibility are as follows. The LSD should be less than 1 dB, SDR should be greater than 25 dB, and PESQ should be greater than 3. These criteria were set to comply with "information hiding and its criteria for evaluation" [22] and the prior criteria of compared methods [10], [11]. The proposed method was evaluated on these measures, the results of which are shown in Table I. The proposed method satisfies all three measures, which indicates that even though the speech contents could be heard, the hidden information was imperceptible. Additionally, the hybrid system's imperceptible properties are well-performing as in the pure SSA-based information hiding method of [10], and [11].

### B. Evaluation of encryption and decryption

The correlation coefficient and signal-to-noise ratio (SNR) were measured to evaluate encryption and decryption. The correlation coefficient measures the linear relationship between the original speech, the encryption speech, and the decrypted

speech, while SNR measures the noise content in the encrypted speech signal. The correlation coefficient between the original signal and the decrypted signal should be close to 1, which indicates no difference between the two signals, and the SNR should be high. On the other hand, the correlation coefficient between the original signal and the encrypted signal should be close to 0, which indicates the difference between the two, and the SNR should be small. Note that the original speech to be encrypted in this proposed method is the watermarked signal. Table II shows that the encryption and decryption performance of the proposed method was comparable to that of previously developed speech encryption methods [23] and [24].

### C. Robustness of proposed scheme

The robustness of the proposed scheme was evaluated by the sensitivity of the encryption algorithm to changing one or multiple keys and the watermark-extraction precision of the information hiding. The following were measured to assess the sensitivity to key changes: the number of sample change rates (NSCR), the correlation coefficient, and the bit error rate (BER). The NSCR is defined by

$$\text{NSCR} = \frac{1}{Len}\sum_{n=1}^{Len} D_n, \tag{8}$$

where $Len$ corresponds to the length of the speech signal, and $D_n$ is determined according to the rule

$$D_n = \begin{cases} 1, & \text{if } A_n \neq A_n{}', \\ 0, & \text{otherwise}, \end{cases} \tag{9}$$

where $A_n$ and $A_n{}'$ are the amplitudes of the speech encrypted with true keys and those of the encrypted speech with different keys, respectively.

The BER is defined as

$$\text{BER} = \frac{1}{B}\sum_{j=1}^{B} w(j) \oplus \hat{w}^*(j), \tag{10}$$

where $B$ is a total number of frames, $w(j)$ and $\hat{w}^*(j)$ are the embedded-watermark bits and the extracted-watermark bits, respectively.

BER was also used to represent the precision of watermark extraction in the proposed method. The NSCR and the correlation coefficient show the degree of variation between two encrypted speech signals when the keys are modified, and BER indicates the extraction precision when the keys were changed. Table III shows the measurements obtained when detecting the encrypted watermarked signal with a different key series (including the true key and wrong keys). If the true keys were applied, the ideal values for NSCR, correlation coefficient, and BER are 0%, 1, and 0%, respectively. The experimental results show that with the true keys, these three measurements are almost perfect values. The key changes slightly from the true keys to demonstrate the value when the wrong key was applied. The NCSR demonstrates that the two decrypted speech signals with slightly different keys hold different samples with near 100%, and the correlation coefficient is close to zero. The BER with the wrong keys is as high as 50%, while a BER with a true key is less than 1%. The measurement values indicate that the hidden information is secured. Note that the security of this hybridized method focuses mainly on the ability to access hidden information. The variations of keys for encryption will be tackled for remaining work to enhance the security of watermarked signals.

TABLE I
COMPARISON OF IMPERCEPTIBLE PROPERTIES BETWEEN PROPOSED AND OTHER METHODS

| Method | LSD (dB) | SDR (dB) | PESQ |
|---|---|---|---|
| Parameterized SSA-based method [10] | 0.65 | 31.58 | 3.70 |
| SSA-based method [11] | 0.69 | 30.96 | 3.64 |
| Proposed method | 0.65 | 31.56 | 3.70 |

TABLE II
COMPARISON OF CORRELATION COEFFICIENT AND SNR (IN DB) BETWEEN ORIGINAL SPEECH (ORI), ENCRYPTED SPEECH (ENC), AND DECRYPTED SPEECH (DEC) FOR THE PROPOSED METHOD AND OTHER ENCRYPTION METHODS. NOTE THAT NA IS NOT APPLICABLE DATA

| Method | Corr-coef (ori,enc) | SNR (ori,enc) | Corr-coef (ori,dec) | SNR (ori,dec) |
|---|---|---|---|---|
| Chaotic shift keying method [23] | 0.04 | NA | 0.99 | 123.57 |
| FFT with chaotic method [24] | 0.02 | NA | 0.99 | 33.52 |
| Proposed method | 0.10 | -2.52 | 0.99 | 31.74 |

TABLE III
KEY SENSITIVITY AND BER

| Keys ($K_{A0}, K_{A1}, K_{B0}, K_{B1}$) | Corr-coef (range [0,1]) | NSCR (%) | BER (%) |
|---|---|---|---|
| True key (5, 7, 11, 19) | 1 | 0 | 0.13 |
| True key (6, 6, 12, 18) | 1 | 0 | 0.07 |
| Wrong key(5, 6, 11, 18) | 0.0072 | 99.29 | 50.72 |
| Wrong key (6, 7, 11, 20) | 0.0068 | 99.28 | 49.51 |
| Wrong key (6, 9, 12, 19) | 0.0069 | 99.27 | 50.15 |
| Wrong key (6, 7, 12, 19) | 0.0073 | 99.33 | 50.12 |

## V. DISCUSSION

Let us consider the comparison of imperceptible properties between the proposed and other methods, as shown in Table I. The proposed method was compared with the pure SSA-based information hiding method of [10], and [11]. These two methods were chosen because both are SSA-based information hiding methods, and the dataset used in the experiment was the same dataset with the same embedding capacity. The experimental result showed that the hybridized system's imperceptible properties are well-performing as in the pure SSA-based information hiding method. This result indicates that applying Arnold transformation to secure hidden information does not degrade speech quality.

Let us consider Table II, which shows that the encryption and decryption performance of the proposed method was comparable to that of previously developed speech encryption methods [23] and [24]. These two methods were chosen

because the encryption was applied to speech signals. The correlation coefficient between original speech (ori), encrypted speech (enc), and decrypted speech (dec) for the proposed method and other encryption methods were similar, but the SNR (in dB) was slightly different. The reason is that the dataset used in the proposed method and the compared method were different. In this experiment, we compared the evaluation result with the existing papers. For a fair comparison, the dataset and experimental condition should be matched. We will tackle this problem in the future. However, from the results, the proposed method shows its promising effective results compared with the existing method.

## VI. Conclusions

We proposed a hybridized system for enhanced speech security. Arnold transformation was performed on watermark signals to create secured watermarks, which were then embedded into host speech using SSA-based information hiding, producing a watermarked signal. The watermarked signal was encrypted before being sent through the communication channel. The experimental results showed considerable differences between the correlation coefficient and SNR of the watermarked signal and those of the encrypted watermarked signal. The key sensitivity indicated that only authorized persons with the watermarked encryption key could access the speech contents. The imperceptible watermark and significant difference of BER with and without a watermark key indicated that access to the hidden information was limited. This hybridized system increased speech security and limited accessibility to the data at varying levels.

## VII. Acknowledgment

This work was supported by a grant from the SIIT-JAIST-NSTDA Dual Doctoral Degree Program, a Grant-in-Aid for Scientific Research (B) (No.17H01761), Fund for the Promotion of Joint International Research (Fostering Joint International Research (B))(20KK0233), and I-O DATA Foundation.

## References

[1] S. Latif, J. Qadir, A. Qayyum, M. Usama, and S. Younis, "Speech technology for healthcare: Opportunities, challenges, and state of the art," *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 342-356, 2020.

[2] M. I. Rahman, S. R. Fahim, S. S. Avro, Y. Sarker, S. K. Sarker, and T. Tahsin, "Voice-activated open-loop control of wireless home automation system for multi-functional devices," *IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE),* pp. 1–4, 2019.

[3] F. Salahdine and N. Kaabouch, "Social engineering attacks: a survey," *Future Internet,* vol. 11, no. 4, p. 89, 2019.

[4] O. G. Abood and S. K. Guirguis, "A survey on cryptography algorithms," *International Journal of Scientific and Research Publications,* vol. 8, no. 7, pp. 495–516, 2018.

[5] M. M. Rahman, T. K. Saha, and M. A.-A. Bhuiyan, "Implementation of rsa algorithm for speech data encryption and decryption," *International Journal of Computer Science and Network Security (IJCSNS),* vol. 12, no. 3, p. 74, 2012.

[6] N. Radha and M. Venkatesulu, "A chaotic block cipher for real-time multimedia," *Journal of Computer Science,* vol. 8, no. 6, p. 994, 2012.

[7] D. Slimani and F. Merazka, "Encryption of speech signal with multiple secret keys," *Procedia computer science,* vol. 128, pp. 79–88, 2018.

[8] E. Mossa, "Security enhancement for aes encrypted speech in communications," *International Journal of Speech Technology,* vol. 20, no. 1, pp. 163–169, 2017.

[9] M. Unoki and R. Miyauchi, "Detection of tampering in speech signals with inaudible watermarking technique," *Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing,* pp. 118–121, 2012.

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

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

[12] K. Galajit, J. Karnjana, M. Unoki, and P. Aimmanee, "Semi-fragile speech watermarking based on singular-spectrum analysis with cnn-based parameter estimation for tampering detection," *APSIPA Transactions on Signal and Information Processing,* vol. 8, 2019.

[13] S. Wang, W. Yuan, J. Wang, and M. Unoki, "Detection of speech tampering using sparse representations and spectral manipulations based information hiding," *Speech Communication,* vol. 112, pp. 1–14, 2019.

[14] Y. Lin and W. H. Abdulla, "A secure and robust audio watermarking scheme using multiple scrambling and adaptive synchronization," *6th International Conference on Information, Communications & Signal Processing,* pp. 1–5, 2007.

[15] R. Shelke and M. Nemade, "Audio encryption algorithm using modified elliptical curve cryptography and arnold transform for audio watermarking," *3rd International Conference for Convergence in Technology (I2CT),* pp. 1–4, 2018.

[16] Q. Wu and M. Wu, "Adaptive and blind audio watermarking algorithm based on chaotic encryption in hybrid domain," *Symmetry,* vol. 10, no. 7, p. 284, 2018.

[17] J. Karnjana, M. Unoki, P. Aimmanee, and C. Wutiwiwatchai, "An audio watermarking scheme based on singular-spectrum analysis," *International Workshop on Digital Watermarking,* pp. 145–159, 2014

[18] J. Bao and Q. Yang, "Period of the discrete arnold cat map and general cat map," *Nonlinear Dynamics,* vol. 70, no. 2, pp. 1365–1375, 2012.

[19] G. Gaspari, "The arnold cat map on prime lattices," *Physica D: Nonlinear Phenomena,* vol. 73, no. 4, pp. 352–372, 1994.

[20] P. Sankhe, S. Pimple, S. Singh, and A. Lahane, "An image cryptography using henon map and arnold cat map," *International Research Journal of Engineering and Technology (IRJET),* vol. 5, no. 4, 2018.

[21] "ATR Japanese Speech Database (set B) ATR-Promotions, Inc., Japan," http://www.atr-p.com/products/sdb.html, accessed: 2021-03-13.

[22] K. Iwamura, M. Kawamura, M. Kuribayashi, M. Iwata, H. Kang, S. Gohshi, and A. Nishimura, "Information hiding and its criteria for evaluation," *IEICE TRANSACTIONS on Information and Systems,* vol. 100, no. 1, pp. 2–12, 2017.

[23] P. Sathiyamurthi and S. Ramakrishnan, "Speech encryption using chaotic shift keying for secured speech communication," *EURASIP Journal on Audio, Speech, and Music Processing,* vol. 2017, no. 1, pp. 1–11, 2017.

[24] P. Sathiyamurthi and S. Ramakrishnan, "Speech encryption algorithm using fft and 3d-lorenz–logistic chaotic map," *Multimedia Tools and Applications,* pp. 1–19, 2020.