Hindawi
Security and Communication Networks
Volume 2017, Article ID 3847092, 12 pages
https://doi.org/10.1155/2017/3847092

Research Article

Parameterization of LSB in Self-Recovery Speech Watermarking Framework in Big Data Mining

Shuo Li,[1] Zhanjie Song,[2] Wenhuan Lu,[3] Daniel Sun,[4] and Jianguo Wei[3]

[1]*School of Electrical and Information Engineering, Tianjin University, Tianjin, China*
[2]*School of Mathematics, Tianjin University, Tianjin, China*
[3]*School of Computer Software, Tianjin University, Tianjin, China*
[4]*Commonwealth Scientific and Industrial Research Organization, Campbell, ACT, Australia*

Correspondence should be addressed to Wenhuan Lu; wenhuan@tju.edu.cn

Received 18 August 2017; Accepted 9 October 2017; Published 12 November 2017

Academic Editor: Lianyong Qi

The privacy is a major concern in big data mining approach. In this paper, we propose a novel self-recovery speech watermarking framework with consideration of trustable communication in big data mining. In the framework, the watermark is the compressed version of the original speech. The watermark is embedded into the least significant bit (LSB) layers. At the receiver end, the watermark is used to detect the tampered area and recover the tampered speech. To fit the complexity of the scenes in big data infrastructures, the LSB is treated as a parameter. This work discusses the relationship between LSB and other parameters in terms of explicit mathematical formulations. Once the LSB layer has been chosen, the best choices of other parameters are then deduced using the exclusive method. Additionally, we observed that six LSB layers are the limit for watermark embedding when the total bit layers equaled sixteen. Experimental results indicated that when the LSB layers changed from six to three, the imperceptibility of watermark increased, while the quality of the recovered signal decreased accordingly. This result was a trade-off and different LSB layers should be chosen according to different application conditions in big data infrastructures.

1. Introduction

In recent years, the rapid development of Internet and mobile phones has resulted in thousands of exploded data. Even though it is convenient to get information, it is possible for digital data to be replaced with fake information, potentially by an adversary, or even lost as a result of poor communication conditions. Therefore, the question of how to best guarantee data integrity and recover the tampered data has become an important problem in big data mining infrastructures [1–3]. Watermarks, defined as the art of embedding secret message into the original signal, are effective ways to solve this problem [4, 5].

The self-recovery watermarking techniques are firstly popular in the image domain [6–8] and the pioneer study on image watermarks dates back to the last century [9]. There are a variety of different methods for watermark embedding and data recovery, such as discrete cosine transform [10, 11], multiple watermarks [12], and source-channel coding [13].

With an increasing amount of audio and speech data, the security and privacy of speech become an urgent problem. While the self-recovery methods are less explored in the speech domain, because human audio systems are more sensitive than human visual systems, it is necessary to design more accurate schemes to recover tampered speech data. Traditional research has focused on detecting the tampered area but not further recovering the tampered speech [14], which limits the application. A fragile segment-based watermarking scheme for speech detection and recovery is proposed in [15]. The algorithm can both detect the tampered area and recover the lost data, but there are two shortcomings: the tampering coincidence problem and the watermark data waste problem. To solve the two shortcomings at the same time, a novel method using reference-sharing mechanism [16] is proposed in [17]. Moreover, Reed-Solomon codes are fully utilized to design an effective speech self-recovery scheme [18]. In addition, many works focus on the various approaches to the watermark embedding based on its intrinsic characteristics,

such as synthesized echoes [19, 20], spread spectrum techniques [21, 22], and patchwork watermarking methods [23, 24].

To conform the complicated scenes in big data mining infrastructures, this paper discusses the influence of the LSB layers used for watermark embedding. Our work is based on a speech self-recovery framework proposed in [17]. In [17], six LSB layers are used for watermark embedding by experience. In this paper, the LSB layer is treated as a parameter and the relationship with other parameters is discussed. By exploring the quantitative relationship between LSB layers, the maximum quantized bits, and the hash bits, the best choices of other parameters are then deduced by the exclusive method when the LSB layers change. We also observed that three to six LSB layers should be chosen for watermark embedding when the total bit layers equal sixteen. When fewer than six LSB layers are used, the imperceptibility of the watermark and the quality of the recovered signal change in opposite directions. Different LSB layers should be chosen according to different big data infrastructures. Moreover, when we enhance the tampered rate, fewer reserved areas could provide efficient reference bits, which may cause worse quality of the recovered signal.

There are three contributions in this paper: First, once LSB layer is fixed, the best choices of other parameters are deduced using the exclusive method. Second, there is a finding that six LSB layers are the limit for watermark embedding, which has been verified through experiments. Third, in conclusion, the trade-off between the imperceptibility of the watermarked speech signal and the quality of the recovered speech signal is discussed; different LSB layers should be chosen to balance it in different big data infrastructures.

The remainder of the paper is organized as follows. The framework for the speech watermark embedding and tampered speech recovery is introduced in Section 2, which also covers the parameterization of LSB and the relationship with other parameters. Experimental results are presented in Section 3. Section 4 discusses how to choose LSB layers according to different big data infrastructures and introduces various aspects of the proposed scheme. Section 5 includes concluding remarks.

2. The Speech Self-Recovery Framework

This entire speech watermarking scheme can be divided into two sections: the watermark embedding procedure and the tampered area recovery procedure. The parameterization of LSB is also mentioned in the section. The details are as follows.

2.1. Watermark Embedding Procedure. Assume an original 16-bit 8 kHz speech signal has N samples. In the algorithm, a frame consisted of 64 neighbor samples, so there are totally $\lceil N/64 \rceil$ frames. If N is not the multiple of 64, add several zeros at the end of the signal until N can be divided by 64. These frames are then permuted randomly according to a secret key, which is known to both sides with consideration of privacy. A frame group consists of 16 neighbor frames in the random permutation. The total number of frame groups is $\lceil N/1024 \rceil$. The embedding and the recovery procedure are both carried out in one frame group.

Out of 16 bits, x LSB layers are dedicated to watermark embedding, while the remaining $16 - x$ most significant bit (MSB) layers are unchanged during the entire procedure. The watermark consists of two parts: reference bits and hash bits, which will be introduced below.

In each frame of a frame group, the amplitude of the original signal is divided by 16 to obtain the compressed information, which is a 64-dimensional vector:

$$v_j = \left[C_{j_1}, C_{j_2}, \ldots, C_{j_{64}}\right], \quad j = 1, 2, \ldots, 16, \tag{1}$$

where C_{j_i} ($i = 1, 2, \ldots, 64$, $j = 1, 2, \ldots, 16$) is the compressed information and j is the index of each frame. The vectors are then randomly permuted according to a secret key to form a vector whose dimension is 1024:

$$V = \left[v_{k_1}, v_{k_2}, \ldots, v_{k_{16}}\right], \tag{2}$$

where different subscripts are used to indicate the random permutation of frames. Next, calculate 368 reference values in each frame group in the following linear manner:

$$\left[r(1), r(2), \ldots, r(368)\right]^T = A \cdot V, \tag{3}$$

where A is a random matrix sized 368×1024 and the Euclidean norm of each row is 1. To generate A, the first step is to produce a matrix A_0 sized 368×1024 whose elements are derived from an independent identical distributed Gaussian distribution with zero mean. Then the elements of matrix A can be obtained as follows:

$$A(i, j) = \frac{A_0(i, j)}{\sqrt{\sum_{t=1}^{1024} \left[A_0(i, t)\right]^2}}, \quad 1 \le i \le 368, \ 1 \le j \le 1024, \tag{4}$$

where $A(i, j)$ and $A_0(i, j)$ are the elements of A and A_0, respectively. According to the central limit theorem, the reference values approximately follow Gaussian distributions with zero mean. There are 368 reference values and 16 frames in each frame group, so each frame carries 23 reference values randomly. The reference-sharing mechanism is used here to avoid both the coincidence problem and the watermark data waste problem effectively. As long as the 16 frames in a group have not been all tampered, it is possible to achieve high recovered quality.

To meet the storage constraint, the float reference values should be changed into integers. So the next step is to quantize the reference values:

$$\widehat{r} = \begin{cases} R_{\max} - 1, & \text{if } r \ge f_{R_{\max}} \\ t, & \text{if } f_t \le r \le f_{t+1} \\ -t - 1, & \text{if } -f_{t+1} \le r \le -f_t, \\ -R_{\max}, & \text{if } r \le -f_{R_{\max}}, \end{cases} \quad t = 0, 1, 2, \ldots, R_{\max}, \tag{5}$$

where

$$f_t = 1500 \times \frac{t}{R_{\max}}, \quad t = 0, 1, 2, \ldots, R_{\max}. \tag{6}$$

Each reference value is converted into an integer within $[-R_{\max}, R_{\max}]$ and can be represented by a bits (the maximum quantized bits):

$$2 \times R_{\max} = 2^a. \tag{7}$$

Thus, there are totally $23 \times a$ reference bits in one frame group.

For each frame, the index of the frame is represented by 64 bits, which are called position bits. There are $64 \times (16 - x)$ bits in MSB layers in a frame, which are called MSB bits. Then 64 position bits, $64 \times (16 - x)$ MSB bits, and $23 \times a$ reference bits are put into a hash function to produce y hash bits. To guarantee the privacy, y label bits are randomly generated and the exclusive-or results between hash bits and label bits are calculated as check bits:

$$c_i(j) = h_i(j) \oplus l(j), \quad i = 1, 2, \ldots, \left\lceil \frac{N}{64} \right\rceil, \ j = 1, 2, \ldots, y, \tag{8}$$

where $h_i(1), h_i(2), \ldots, h_i(y)$ are hash bits of the ith frame, $l(1), l(2), \ldots, l(y)$ are label bits which are the same in each frame, and $c_i(1), c_i(2), \ldots, c_i(y)$ are check bits of the ith frame, respectively.

In each frame, $23 \times a$ reference bits and y check bits are embedded into x LSB layers of the frame as watermark, which are used for detecting and recovering the tampered speech at the receiver end. The $16 - x$ MSB layers remain unchanged to ensure the invisibility. The watermark embedding procedure is shown in Figure 1.

2.2. Tampered Area Recovery Procedure. After receiving a speech signal that may have been tampered, the first step is to divide the received signal into several frames and frame groups according to the secret key, which is known to both sides.

For each frame, the $64 \times (16 - x)$ MSB bits, the 64 position bits, and the $23 \times a$ reference bits are extracted from the received speech. They are put into the same hash function to obtain y hash bits. The label bits are calculated by exclusive-or operator:

$$l_i(j) = h_i(j) \oplus c_i(j), \quad i = 1, 2, \ldots, \left\lceil \frac{N}{64} \right\rceil, \ j = 1, 2, \ldots, y, \tag{9}$$

where $h_i(1), h_i(2), \ldots, h_i(y)$ are hash bits that are calculated at the receiver end, $c_i(1), c_i(2), \ldots, c_i(y)$ are check bits which are extracted from the received signal, and $l_i(1), l_i(2), \ldots, l_i(y)$ are label bits of the ith frame, respectively.

Due to the property of the exclusive-or operator, the label bits of each frame should be the same if there has been no tampering at all. If a frame has been tampered, the label bits of the frame are different. Even though the receiver does not know the label bits specifically, the tampered area can also be detected by comparing the label bits of each frame. In addition, based on the property of hash function, the probability of a tampered frame being falsely judged as reserved is 2^{-y}, which is extremely low when y is large enough. This means that it is virtually impossible for false detection to occur. The reference values can be correctly extracted to recover the tampered speech.

After detecting which frames have been tampered, the next step is to recover the tampered content. In a frame group, if 16 frames are all reserved, the speech recovery is needless. If 16 frames have all been tampered, the recovery fails. Otherwise, assuming that there are z $(1 \le z \le 15)$ tampered frames in a frame group, only the reference values in $16 - z$ reserved frames can be used:

$$\left[r(\alpha_1), r(\alpha_2), \ldots, r(\alpha_M)\right]^T = A^{(M)} \cdot V, \tag{10}$$

where $r(\alpha_1), r(\alpha_2), \ldots, r(\alpha_M)$ are the reference values embedded into the reserved frames and $A^{(M)}$ is a matrix with rows taken from A corresponding to reserved frames. So (10) can also be rewritten as

$$\left[r(\alpha_1), r(\alpha_2), \ldots, r(\alpha_M)\right]^T = A^{(M,R)} \cdot V_R + A^{(M,T)} \cdot V_T, \tag{11}$$

where V_R and V_T are compressed information of the reserved frames and the tampered frames, respectively, and $A^{(M,R)}$ and $A^{(M,T)}$ are matrices whose columns are those in $A^{(M)}$ corresponding to V_R and V_T, respectively.

Note that the reference values extracted from LSB layers are quantized, which could not be directly used for recovery. The reference values that have not been quantized can be estimated by

$$r \in \begin{cases} \left[f_{R_{\max}}, +\infty\right), & \text{if } \hat{r} = R_{\max} - 1 \\ \left[f_{\hat{r}}, f_{\hat{r}+1}\right), & \text{if } 0 \le \hat{r} \le R_{\max} - 2 \\ \left[-f_{-\hat{r}}, -f_{-\hat{r}-1}\right), & \text{if } -\left(R_{\max} - 1\right) \le \hat{r} \le -1 \\ \left(-\infty, -f_{R_{\max}}\right), & \text{if } \hat{r} = -\mathrm{R}_{\max}. \end{cases} \tag{12}$$

Let

$$r' = \begin{cases} \dfrac{\left(f_{\hat{r}} + f_{\hat{r}+1}\right)}{2}, & 0 \le \hat{r} \le R_{\max} - 1 \\ \dfrac{\left(-f_{-\hat{r}} - f_{-\hat{r}-1}\right)}{2}, & -R_{\max} \le \hat{r} \le -1, \end{cases} \tag{13}$$

where $R_{\max}$ and f are the same as that in the embedding procedure. In other words, r' is the estimate of r at the receiver end, which is a median value of the corresponding interval. The estimate results in errors are related to the choices of $R_{\max}$. The larger the value of $R_{\max}$ is, the thinner the interval is, resulting in fewer errors. So (11) can be rewritten as

$$\left[r'(\alpha_1), r'(\alpha_2), \ldots, r'(\alpha_M)\right]^T - A^{(M,R)} \cdot V_R = A^{(M,T)} \cdot V_T. \tag{14}$$

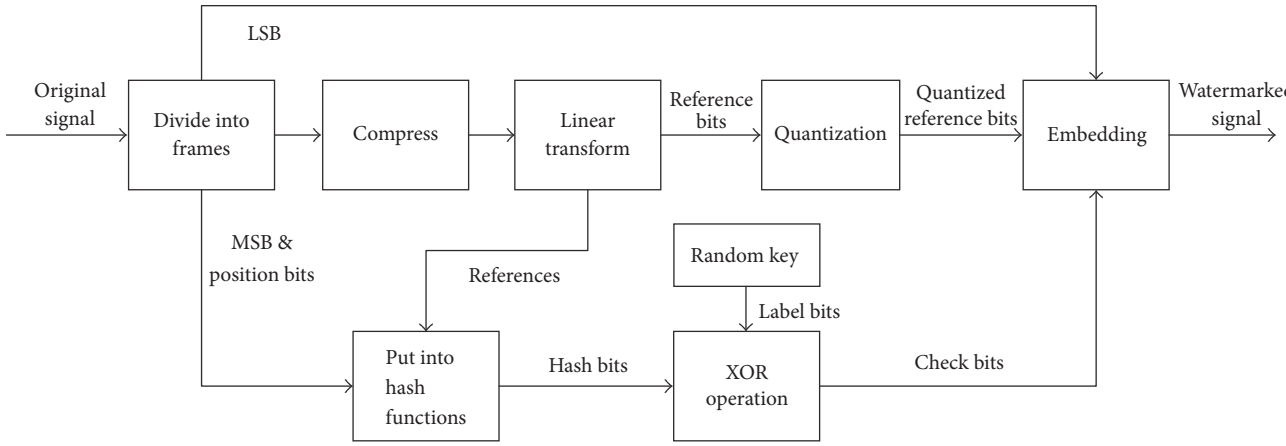

Figure 1: Sketch of watermark embedding procedure.

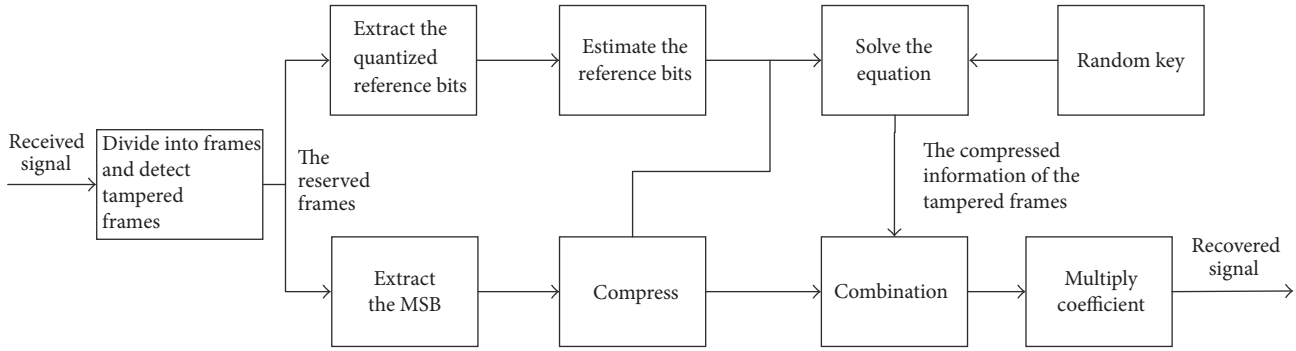

Figure 2: Sketch of speech recovery procedure.

In (14), $A^{(M,R)}$ and $A^{(M,T)}$ are already known to the receiver according to the secret key, which is known to both sides; V_R can be calculated in the reserved frames. Moreover, $r'(\alpha_1), r'(\alpha_2), \ldots, r'(\alpha_M)$ can be estimated according to (13). In other words, only V_T is unknown and can be obtained by solving (14) at the receiver end. The compressed sensing and compositive reconstruction can be used to solve (14).

Next, V_T and V_R are combined to obtain a vector V' whose dimension is 1024, which is the recovered compressed information of the original speech signal. Finally, the amplitude of the original signal can be obtained by multiplying 16. The entire procedure of recovery is showed in Figure 2.

2.3. Parameterization of LSB in the Framework. The best choices of parameters corresponding to different LSB layers are deduced in this subsection. From the perspective of quantity, the watermark consists of reference bits and hash bits, so the total bits of LSB are equivalent to the sum of reference bits and hash bits:

$$23 \times a + y = 64 \times x. \tag{15}$$

To ensure the imperceptibility of the watermark, the LSB layers x should be less than 8, which is the half of all the bit layers. There are three variables in (15) and all their choices are shown in Table 1, in which the values before and after "/" stand for the corresponding values of each variable.

The exclusive method is used to detect the best choices of each parameter.

Firstly, at the receiver end, the reference values are estimated using (13), which generally causes errors that relate to the values of $R_{\max}$: the larger the value of $R_{\max}$ is, the thinner the segmentation is, which indicates a better estimate. In addition, the relationship between $R_{\max}$ and a follows (7). Thus, the larger the value of a is, the larger the value of $R_{\max}$ is, which indicates fewer errors. In other words, the value of a should be large enough to ensure small error, so the conditions of $a = 1, 2, 3, 4, 5$ are exclusive.

Secondly, the hash bits are used for detecting the tampered area. Because of the property of the hash function, the probability of a tampered frame being falsely judged as reserved is 2^{-y} when y hash bits are used. Thus, the value of y should be large enough to reduce the false judged probability. In conclusion, the conditions of $y = 3, 6, 8, 11, 13$ are exclusive; the corresponding falsely judged probability is too high for our framework.

Finally, when the LSB layer is fixed, the larger the value of a is, the larger the value of $R_{\max}$ is, which indicates fewer errors. So smaller values of a should be exclusive in this step. Consequently, the conditions of $a = 6, 9, 12, 14, 15, 17, 20$ are

Table 1: All choices of parameters.

LSB layer	Total bits	Reference bits	Hash bits
x	$64 \times x$	a	y
8	512	22/21/20	6/29/52
7	448	19/18/17	11/34/57
6	384	16/15/14	16/39/62
5	320	13/12	21/44
4	256	11/10/9	3/26/49
3	192	8/7/6	8/31/54
2	128	5/4/3	13/36/59
1	64	2/1	18/41

Table 2: Best choices of parameter.

LSB layer	Total bits	Reference bits	Quantized result	Hash bits
x	$64 \times x$	a	$[-R_{\max}, R_{\max} - 1]$	y
8	512	21	[−1048576, 1048575]	29
7	448	18	[−131072, 131071]	34
6	384	16	[−32768, 32767]	16
5	320	13	[−4096, 4095]	21
4	256	10	[−512, −511]	26
3	192	7	[−64, 63]	31

exclusive. The remaining choices of a are the best choices when the LSB layers are 8, 7, 6, 5, 4, 3, respectively.

The theoretical results of choosing the parameters are shown in Table 2. Once the LSB layers are chosen, the other parameters can be obtained by looking up this table.

3. Experimental Results

Both objective and subjective experiments were carried out in this section. In the experiments, the above theoretical conclusions of the best choices of parameters were used. The values of the signal to noise ratio (SNR) of the watermarked speech signal and the recovered speech signal were calculated to deduce several useful conclusions. Moreover, the waveform of the original signal, the watermarked signal, and the recovered signal was shown in this section. In the subjective experiments, the listening tests were carried out to effectively verify the invisibility of the watermarked speech signal.

3.1. Objective Experimental Results. A 16-bit 8 kHz [17, 18] sampled speech signal with the length of 5 seconds was chosen as a sample in our experiments. The above theoretical results regarding the choices of maximum quantized bits and hash bits were used in the corresponding experiments. When the LSB layer was 8, 7, 6, 5, 4, 3, the values of SNR of the watermarked signal and the recovered signal were calculated, respectively. Due to the randomness of the algorithm, all experiments were carried out ten times and the means of the results were shown in Table 3. Additionally, the SNR of the watermarked signal and the recovered signal was shown in Figure 3.

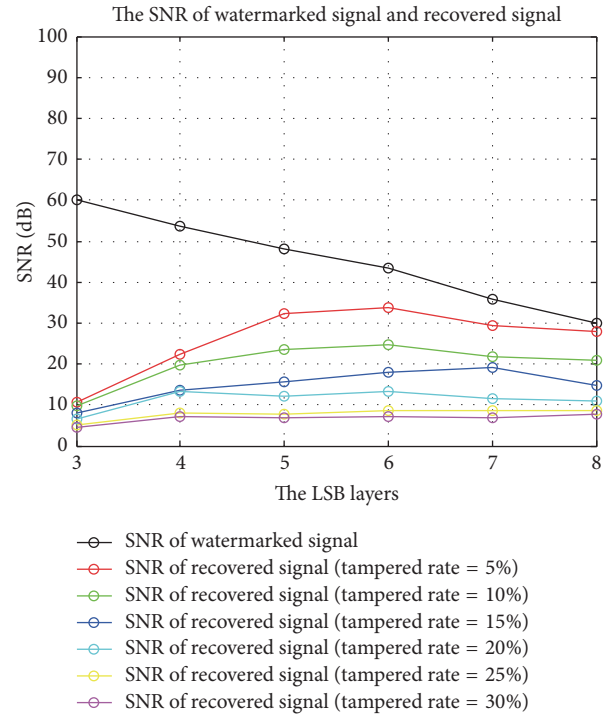

Figure 3: SNR of watermarked and recovered signal.

Several conclusions can be observed from Table 3 and Figure 3. If more LSB layers are dedicated for watermark embedding, the values of SNR of the watermarked speech signal decrease because of more changes in the original one. Moreover, as more LSB layers are used, the values of SNR of

Table 3: The SNR of watermarked and recovered signal.

SNR (dB)	LSB = 8	LSB = 7	LSB = 6	LSB = 5	LSB = 4	LSB = 3
Watermark	29.9407	35.6731	43.4935	48.1025	53.6022	59.9497
Tampered rate (%)						
5	27.9201	29.3410	33.7328	32.1921	22.5146	10.6607
10	20.9200	21.8606	24.6679	23.6509	19.6991	9.6582
15	14.6268	19.1903	17.9543	15.5811	13.6571	8.0125
20	10.9552	11.5015	13.1581	12.0458	13.2171	6.4483
25	8.6504	8.4837	8.6717	7.8364	8.0430	5.1675
30	7.6440	6.7256	7.0462	6.9904	7.1732	4.6055
35	4.4040	4.5564	4.5645	5.0154	4.4516	3.4517
40	3.2233	3.1574	3.0861	3.1238	3.2030	2.5935
45	2.4254	2.6394	2.4341	2.4978	2.5310	2.0356
50	2.1907	2.2027	2.3016	2.2885	2.2553	1.9224

recovered speech signal are not monotonous. If fewer than six LSB layers are used, the values of SNR of the recovered speech signal increase when the number of the LSB layers increases. This is because more bits are used for expressing the reference values, which leads to less error in estimate (13). If more than six LSB layers are used, the values of SNR of the recovered speech signal decrease when the number of LSB layers increases. This is because fewer MSB layers remain unchanged in the procedure, which leads to larger changes in the original signal. If more than six LSB layers are used, both the values of SNR of the watermarked and the recovered speech decrease as the number of LSB layers increases. In summary, the highest value of SNR of the recovered speech signal is achieved at six. According to this conclusion, fewer than six LSB layers should be chosen.

Our results are compared with that in [17], in which the six LSB layers are used. This condition is included in our framework. The comparison is shown in Figure 4. Though when the five to three LSB layers are used for watermark embedding the values of SNR of the recovered speech are a little lower than that in [17], the values of SNR of the watermarked speech are much higher. In other words, we extend the results in [17] by using different LSB layers for watermark embedding.

In addition, when the tampered rate increases, the values of SNR of the recovered speech decrease. This is because fewer reference values in the reserved area can be used for recovery. Through listening tests, when the value of SNR is larger than 7.2, the recovered signal is understandable, though there are big differences in naturalness. When the value of SNR is smaller than 7.2, the recovered speech signal is incomprehensible; these conditions are treated as failures. In other words, when the tampered rate is smaller than 30%, our framework could recover the tampered speech signal successfully.

To further verify the above discussion regarding the quality of the watermarked speech and the recovered speech, the first 20% of the signal was set as mute. The waveform of

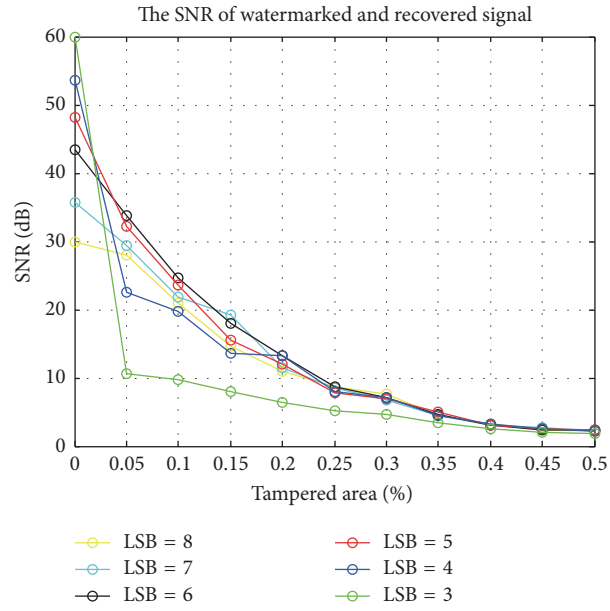

Figure 4: SNR of watermarked and recovered signal versus tampered area with different LSB layers.

the original signal is shown in Figure 5(a), and the waveform of the watermarked signal and the recovered signal with LSB layers of 8, 7, 6, 5, 4, and 3 is shown in Figures 5(b), 5(c), 5(d), 5(e), 5(f), and 5(g), respectively. Moreover, the spectrograms of the corresponding waveform are shown in the Appendix. Figure 5 indicates that when the LSB layers change from eight to three, the watermarked speech signal is more similar to the original one. Furthermore, the quality of the recovered speech signal firstly improved and then reduced. If six LSB layers are used, the quality of the recovered speech is highest.

3.2. Subjective Experimental Results. Subjective listening texts were carried out in order to test the perception of the watermarked speech signal. In these subjective listening

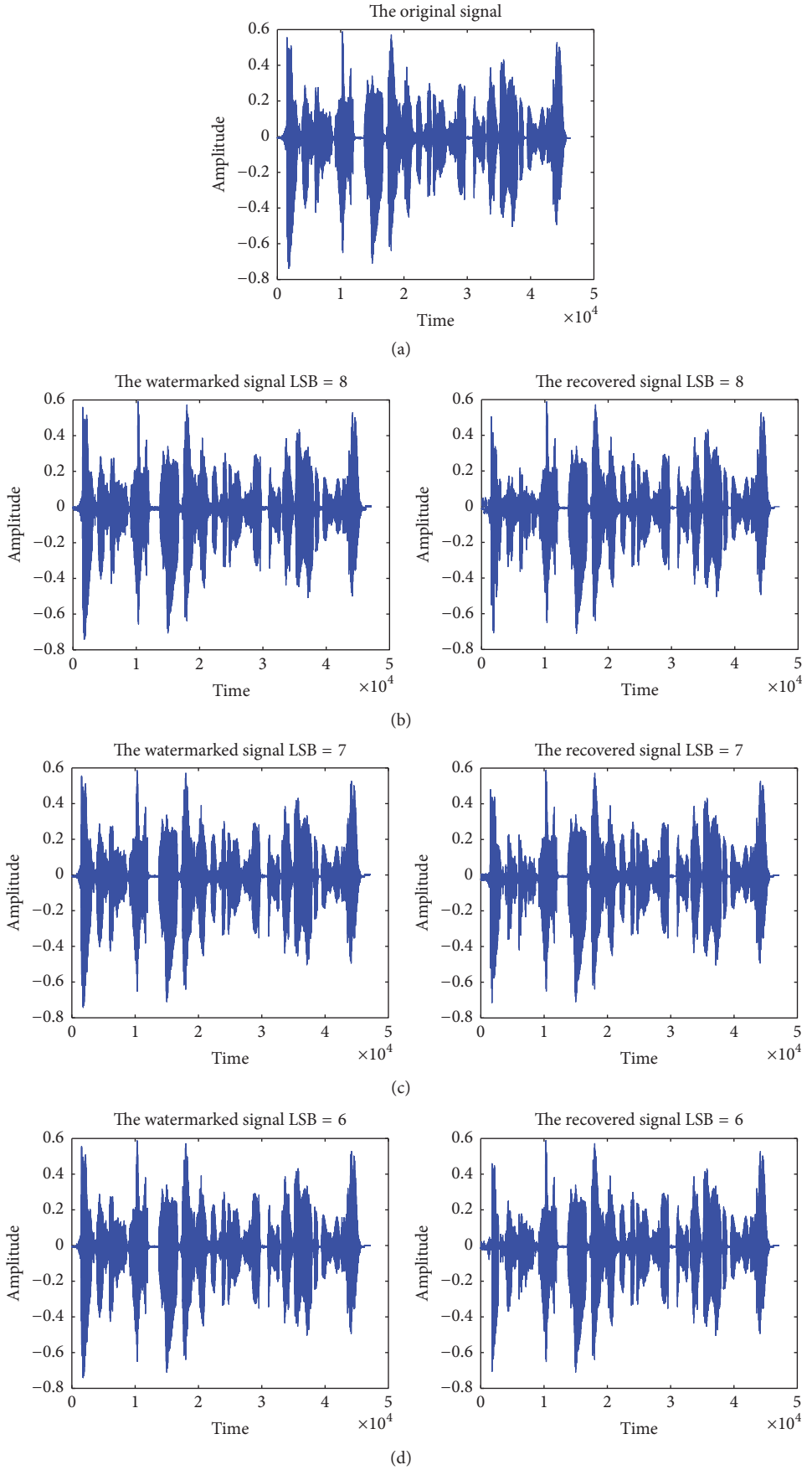

Figure 5: Continued.

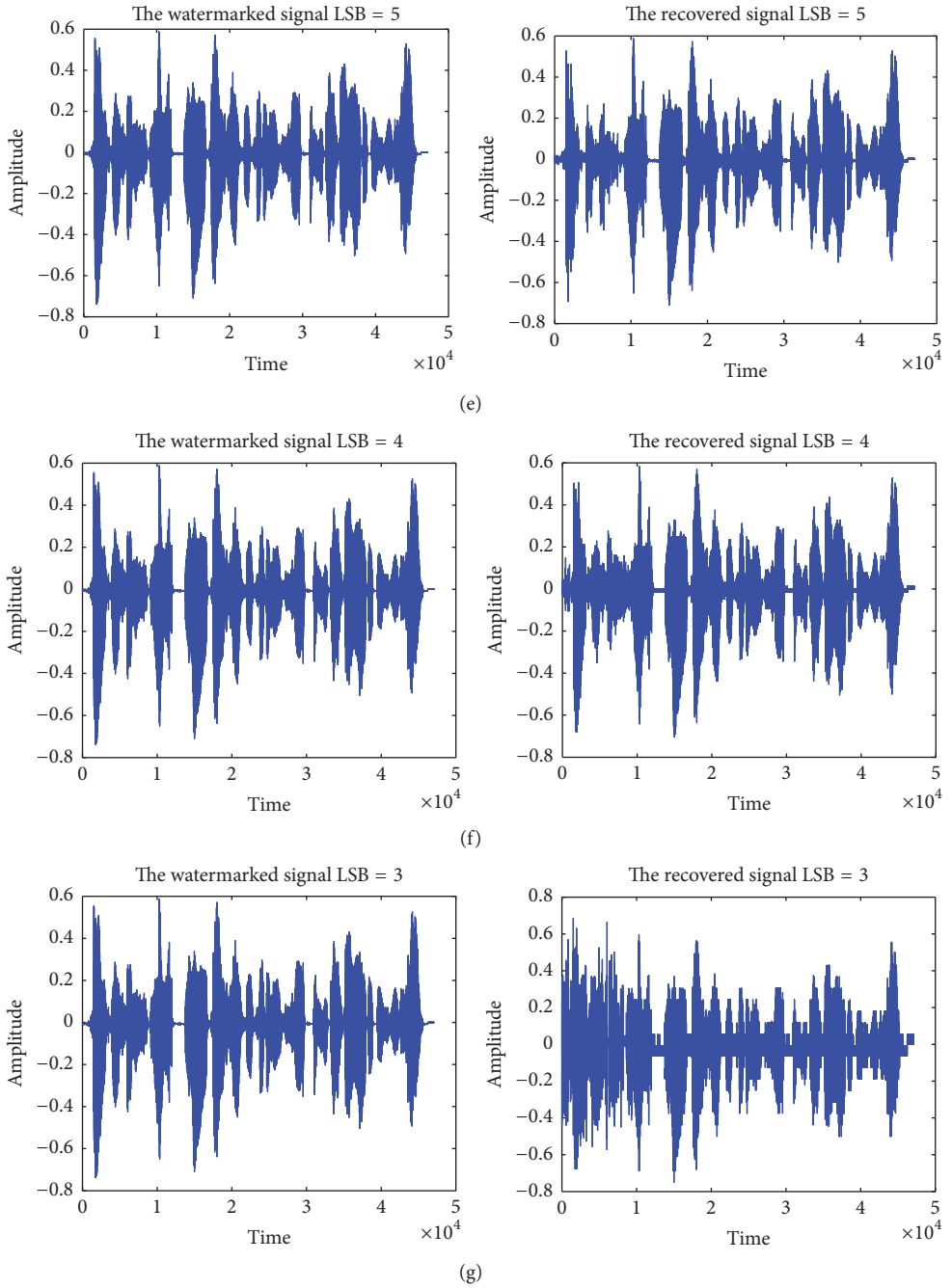

Figure 5: The waveform of the original, watermarked, and recovered signal. (a) The original signal. (b) The watermarked and recovered signal when LSB = 8. (c) The watermarked and recovered signal when LSB = 7. (d) The watermarked and recovered signal when LSB = 6. (e) The watermarked and recovered signal when LSB = 5. (f) The watermarked and recovered signal when LSB = 4. (g) The watermarked and recovered signal when LSB = 3.

tests, five sentences were randomly chosen from the CASIA-863 speech synthesis database. Ten 22- to 25-year-old participants, all with normal hearing ability, were trained to effectively evaluate the watermarked speech quality.

The subjective difference grade (SDG) is one of the most widely used subjective methods for evaluating the quality of a watermarked speech signal. The SDG ranges from 5.0 to 1.0 (from imperceptible to very annoying, as shown in Table 4). In subjective listening tests, the original and the watermarked speech signals were given to ten participants; they could classify the differences according to Table 4. The average SDG scores of five sentences and ten participants for watermarked speech signals were shown in Table 5. From the test results, we observed that the mean opinion score (MOS) ranged from 4.9 to 5.0 for all watermarked signals, indicating that all watermarked speech signals were almost

Table 4: The standard of subjective evaluation.

SDG	Description	Quality
5.0	Imperceptible	Excellent
4.0	Imperceptible but not annoying	Good
3.0	Slightly annoying	Fair
2.0	Annoying	Poor
1.0	Very annoying	Bad

Table 5: The evaluation of watermarked signal.

Watermark embedding	SNR (dB)	MOS	ABX (%)
LSB = 8	29.9407	4.91	53.7
LSB = 7	35.6731	4.93	46.2
LSB = 6	43.4935	4.97	54.8
LSB = 5	48.1025	4.95	51.5
LSB = 4	53.6022	4.99	50.4
LSB = 3	59.9497	5.00	49.8

imperceptible. As fewer LSB layers were used, the MOS is enhanced. This indicated that the fewer numbers of LSB layers resulted in higher quality of the watermarked speech signal.

We also employed the ABX method, another subjective quality assessment technique, to evaluate the quality of the watermarked speech signals. In tests, the original speech signal A and the watermarked signal B were presented to ten participants. A third speech signal X, which was either A or B, was presented to the participants in random order. The participants were asked to identify what X was. The correction percentage was used to evaluate whether the watermarked speech signal was perceptible or not. If the result is 50%, which was the probability of the random guess, it is suggested that the differences between the original signal and the watermarked one were imperceptible. The evaluation results were shown in Table 5. The results showed that the correct percentage ranged from 46% to 55%. This indicated that the watermarked speech was almost imperceptible. With fewer LSB layers, the ABX test result approached 50% generally, indicating that fewer LSB layers resulted in higher quality of the watermarked speech signal.

4. Discussion

If fewer than six LSB layers are used, when the number of LSB layers increases, the imperceptibility of the watermark and the quality of the recovered signal change in opposite directions. For the watermarked signal, using more LSB layers results in worse watermarked signal quality. Though there is no obvious difference in the MOS and ABX subjective listening results, the values of SNR of different watermarked signals are obvious. For the recovered signal, more LSB layers result in improved recovered quality. This can be seen in higher values of SNR and the waveform of the recovered speech signal.

Once the number of LSB layers reaches six, both the imperceptibility of the watermark and the quality of the recovered signal decline as more LSB layers are used. Consequently using more than six LSB layers is not recommended for watermark embedding in the framework when the framework is used in big data scenes.

In conclusion, six to three LSB layers are recommended in big data mining infrastructures. The specific choices of parameters depend on the requirement in real applications. Because of the complexity in big data applications, it is important to choose different LSB layers for watermark embedding and tune the trade-off between the imperceptibility of the watermark and the quality of the recovered speech. Concretely, if the watermark's imperceptibility is highly required, while the recovered speech's quality is less required, fewer LSB layers are recommended; otherwise, more LSB layers should be chosen. Additionally, while the tampered rate increases, less reserved areas could provide efficient reference bits, causing lower values of SNR of the recovered speech signal. Therefore, when tampering may not occur or when the tampered rate is estimated as low, such as in a stable communicating environment, fewer LSB layers are suitable. When the communicating environment is terrible and the tampered rate is estimated as high, a larger number of LSB layers should be chosen. Once the LSB layers are chosen, other parameters are also determined as shown in Table 2. In other words, choosing different LSB layers makes it possible to extend the applications of the self-recovery speech watermarking framework to fit the complicated scenes in big data infrastructures.

As mentioned before, only the erasure of the watermarked speech was carried out in our experiments. In fact, our proposed framework is suitable for any other falsifications. This is because hash bits are used for detecting the tampered areas. Once a frame is treated as a tampered one, its label bits are different from others. Then the reference values of the tampered frame are useless in the following steps of the recovery procedure.

In the above experiments, continuous tampering was carried out. It should be noted that discrete tampering is

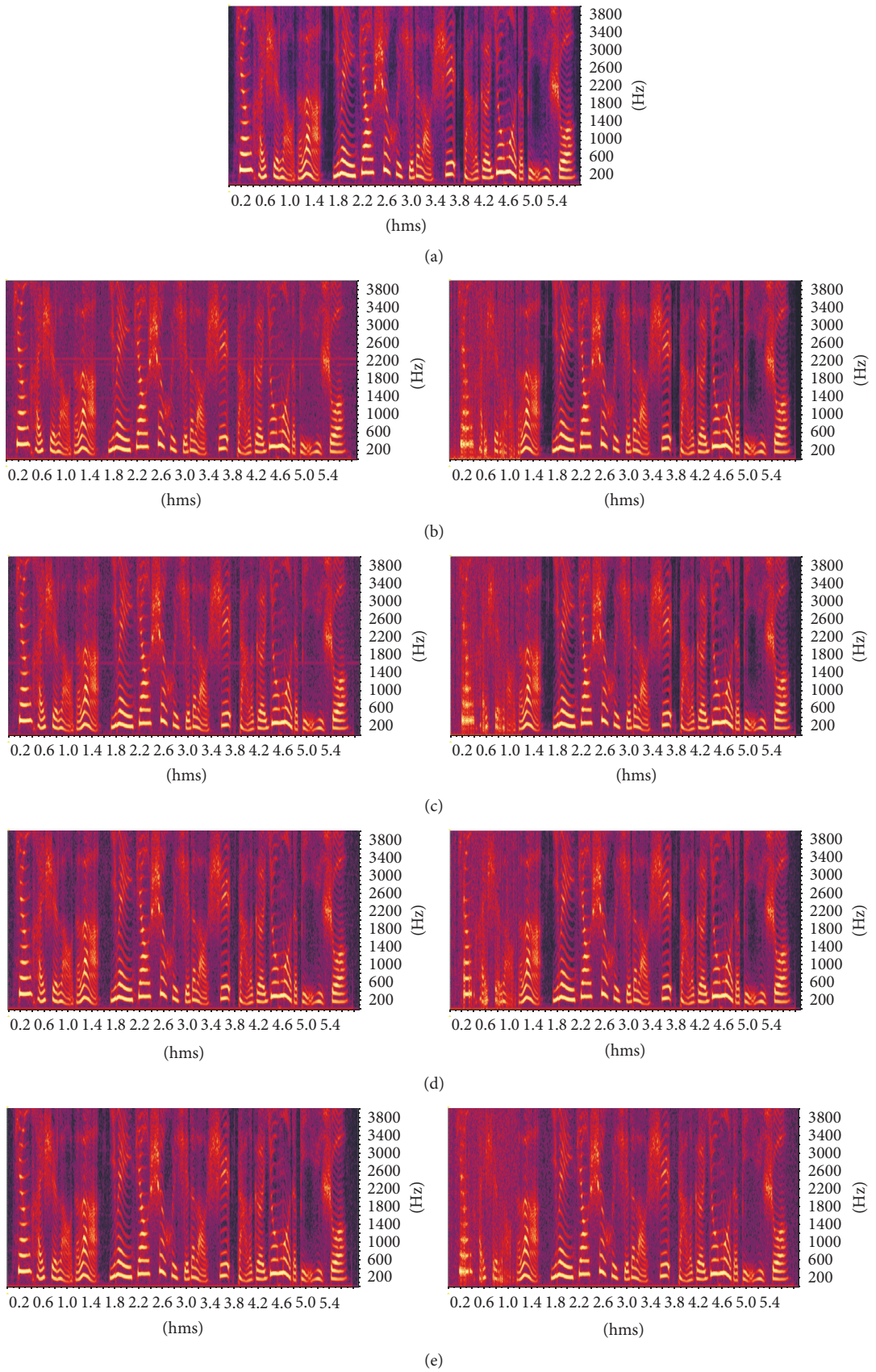

Figure 6: Continued.

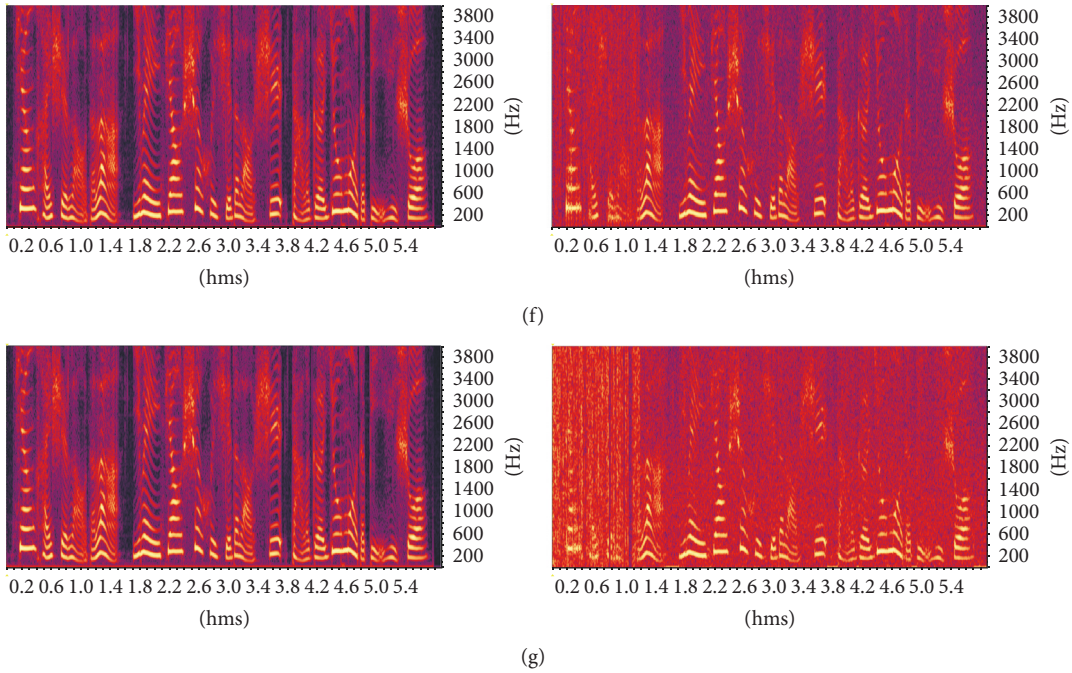

Figure 6: The spectrograms of the original, watermarked, and recovered signal. (a) The original signal. (b) The watermarked and recovered signal when LSB = 8. (c) The watermarked and recovered signal when LSB = 7. (d) The watermarked and recovered signal when LSB = 6. (e) The watermarked and recovered signal when LSB = 5. (f) The watermarked and recovered signal when LSB = 4. (g) The watermarked and recovered signal when LSB = 3.

equivalent to continuous tampering, as long as the total lengths of isolated tampering are equal to the single length of continuous tampering. This is because random permutations and hash bits are used in the algorithm. In other words, continuous tampering is an assumption closer to the real big data applications.

Moreover, the sampled rate in above experiments is 8 KHz, but it can be fully proved through extended experiments that our results are not sensitive to the sampled rate. In other words, we can arbitrarily change the sampled rate to meet the requirement.

5. Conclusions

In this paper, based on the speech self-recovery framework, the LSB layer is treated as a parameter; the relationship between the LSB layer and other parameters is expressed through an explicit mathematical equation; the best choices of parameters are given using the exclusive method. The proposed method can conform the complexity of the big data infrastructures. The six to three LSB layers are recommended for different real applications; the specific choices depend on the requirement and the communicating environment. The trade-off between the imperceptibility of the watermark and the quality of the recovered signal is also discussed in this paper; different LSB layers are chosen to balance this. In the future, we will explore more effective methods to enhance the imperceptibility of the watermarked speech and the quality of the recovered signal at the same time.

Appendix

The spectrograms of the corresponding waveform are shown here to further verify the conclusions drawn in this paper. When the LSB layers changed from eight to three, the spectrograms of watermarked speech signal were increasingly imperceptible to the original signal, indicating that the watermark was more imperceptible. In addition, the quality of the spectrograms for the recovered speech signal first improved and then declined. When six LSB layers were dedicated for watermark embedding in the procedure, the spectrogram of the recovered signal was most similar to the original signal. The theoretical results in our paper were fully verified through Figure 6.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Authors' Contributions

Shuo Li performed the experiments and wrote the paper, Wenhuan Lu and Jianguo Wei proposed the framework, and Zhanjie Song and Daniel Sun revised the paper.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (no. 61379014) and the Natural Science

Foundation of Tianjin (no. 16JCYBJC15900). Thanks are due to Letpub for helping the authors to improve the English presentation.

References

[1] D. Puthal, S. Nepal, R. Ranjan, and J. Chen, "A Synchronized Shared Key Generation Method for Maintaining End-to-End Security of Big Data Streams," in *Proceedings of the Hawaii International Conference on System Sciences*, pp. 6011–6020, Jan 2017.

[2] X. Zhang, C. Liu, S. Nepal, C. Yang, W. Dou, and J. Chen, "A hybrid approach for scalable sub-tree anonymization over big data using MapReduce on cloud," *Journal of Computer and System Sciences*, vol. 80, no. 5, pp. 1008–1020, 2014.

[3] C. Yang, C. Liu, X. Zhang, S. Nepal, and J. Chen, "A time efficient approach for detecting errors in big sensor data on cloud," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 2, pp. 329–339, 2015.

[4] C. De Vleeschouwer, J.-F. Delaigle, and B. Macq, "Invisibility and application functionalities in perceptual watermarking - An overview," *Proceedings of the IEEE*, vol. 90, no. 1, pp. 64–77, 2002.

[5] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Information hiding—a survey," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1062–1078, 1999.

[6] P. Korus and A. Dziech, "Efficient method for content reconstruction with self-embedding," *IEEE Transactions on Image Processing*, vol. 22, no. 3, pp. 1134–1147, 2013.

[7] P. Korus and A. Dziech, "Adaptive self-embedding scheme with controlled reconstruction performance," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 2, pp. 169–181, 2014.

[8] S. Sarreshtedari and M. A. Akhaee, "Source-channel coding approach to generate tamper-proof images," in *Proceedings of the 2014 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2014*, pp. 7435–7439, ita, May 2014.

[9] J. Fridrich and M. Goljan, "Images with self-correcting capabilities," in *Proceedings of the 6th International Conference on Image Processing (ICIP'99)*, pp. 792–796, Kobe, Japan, Oct 1999.

[10] X. Zhang, Z. Qian, Y. Ren, and G. Feng, "Watermarking with flexible self-recovery quality based on compressive sensing and compositive reconstruction," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 4, pp. 1223–1232, 2011.

[11] X. Zhang, S. Wang, and G. Feng, "Fragile watermarking scheme with extensive content restoration capability," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 5703, pp. 268–278, 2009.

[12] R. Chamlawi, A. Khan, and I. Usman, "Authentication and recovery of images using multiple watermarks," *Computers and Electrical Engineering*, vol. 36, no. 3, pp. 578–584, 2010.

[13] S. Sarreshtedari and M. A. Akhaee, "A source-channel coding approach to digital image protection and self-recovery," *IEEE Transactions on Image Processing*, vol. 24, no. 7, pp. 2266–2277, 2015.

[14] D. Kundur and D. Hatzinakos, "Digital watermarking for telltale tamper proofing and authentication," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1167–1180, 1999.

[15] F. Chen, H. He, and H. Wang, "A fragile watermarking scheme for audio detection and recovery," in *Proceedings of the 1st International Congress on Image and Signal Processing, CISP 2008*, pp. 135–138, China, May 2008.

[16] X. Zhang, S. Wang, Z. Qian, and G. Feng, "Reference sharing mechanism for watermark self-embedding," *IEEE Transactions on Image Processing*, vol. 20, no. 2, pp. 485–495, 2011.

[17] J. Li, W. Lu, C. Zhang, J. Wei, X. Cao, and J. Dang, "A study on detection and recovery of speech signal tampering," in *Proceedings of the Joint 15th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, 10th IEEE International Conference on Big Data Science and Engineering and 14th IEEE International Symposium on Parallel and Distributed Processing with Applications, IEEE TrustCom/BigDataSE/ISPA 2016*, pp. 678–682, China, August 2016.

[18] S. Sarreshtedari, M. A. Akhaee, and A. Abbasfar, "A Watermarking Method for Digital Speech Self-Recovery," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 23, no. 11, pp. 1917–1925, 2015.

[19] Y. Xiang, I. Natgunanathan, D. Peng, W. Zhou, and S. Yu, "A dual-channel time-spread echo method for audio watermarking," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 383–392, 2012.

[20] Y. Xiang, D. Peng, I. Natgunanathan, and W. Zhou, "Effective pseudonoise sequence and decoding function for imperceptibility and robustness enhancement in time-spread echo-based audio watermarking," *IEEE Transactions on Multimedia*, vol. 13, no. 1, pp. 2–13, 2011.

[21] A. Valizadeh and Z. J. Wang, "Correlation-and-bit-aware spread spectrum embedding for data hiding," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 2, pp. 267–282, 2011.

[22] X. Zhang and Z. J. Wang, "Correlation-and-bit-aware multiplicative spread spectrum embedding for data hiding," in *Proceedings of the 2013 5th IEEE International Workshop on Information Forensics and Security, WIFS 2013*, pp. 186–190, China, November 2013.

[23] I. Natgunanathan, Y. Xiang, Y. Rong, W. Zhou, and S. Guo, "Robust patchwork-based embedding and decoding scheme for digital audio watermarking," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, no. 8, pp. 2232–2239, 2012.

[24] Y. Xiang, I. Natgunanathan, S. Guo, W. Zhou, and S. Nahavandi, "Patchwork-based audio watermarking method robust to desynchronization attacks," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 22, no. 9, pp. 1413–1423, 2014.