RESEARCH NOTE

Study on Scramble Method for Speech Signal by Using Random-Bit Shift of Quantization

Zhi Zhu[1], Katsuhiko Yamamoto[1], Masashi Unoki[1] and Naofumi Aoki[2]

[1] School of Information Science, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan
[2] Graduate School of Information Science and Technology, Hokkaido University
Kita-14 Nishi-9, Kita-ku, Sapporo 060-0814, Japan
E-mail: {kyamamoto, zhuzhi, unoki}@jaist.ac.jp, aoki@ime.ist.hokudai.ac.jp

Abstract **Speech scrambling aims to eliminate intelligibility of original speech in order to preventing eavesdropping and copyright infringement. There is, however, a problem in that completely recovering scrambled speech into the original speech cannot be achieved with conventional methods. In this paper, we propose a speech scrambling method that uses the random-bit-shift of quantization bits with common keys. We evaluated the confidentiality and efficiency of the proposed method by using two objective measures, SER and PESQ. As a result we confirmed that speech signals can be scrambled into completely unintelligible sounds with the proposed method. Moreover, it is possible to restore a scrambled speech signal into the original one completely. In addition, we also confirmed that the scrambled speech signal could not be descrambled correctly with the wrong key.**

Keywords: speech scramble, random bit shift, PCM speech signal, secure voice

1. Introduction

Information protection and the confidentiality of digital content are becoming more important today. With the rapid development of multimedia and Internet technologies, digital speech content is being transmitted over networks more frequently. Therefore, reliable security for the storage and transmission of digital speech contents is required for many areas, such as for preventing eavesdropping and copyright infringement. The speech scrambling method is a method where the sender encrypts digital speech information and only the people who the sender allows can decrypt the scrambled speech signal to use it [1]. It is similar to but not a direct application of normal cryptographic techniques. The method is very important for preventing eavesdropping and copyright infringement.

Two representative speech scrambling methods are reverse scrambling in the frequency domain [2] and time-segment permutation in time-domain scrambling [3]. Many other speech-scrambling methods for speech coding, such as scrambler using variation in bit-reversal sensitivity [4] and one that focuses on the AMR codec [5], were proposed. They are effective at speech signal encryption. However, poor voice quality may result and there is a problem in that completely recovering a signal seems unachieved.

In this paper, we propose a speech scrambling method that uses the random-bit shift of quantization with a common key. Since the descrambling process is an inverse operation of the scrambling process. The scrambled speech signal can be completely recovered into the original signal with the proposed method.

2. Proposed Method

Figure 1 shows an overview of the proposed method. The main idea of the method is that a bit sequence of each sample in a PCM speech signal is subjected to a circular left shift of random bit position in order to completely

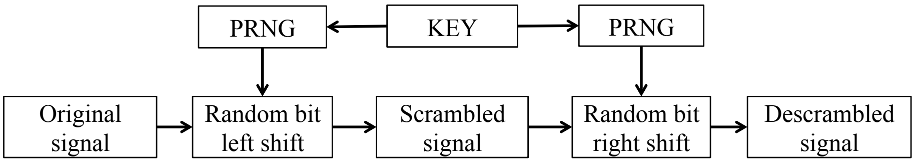


Fig. 1 Overview of proposed method

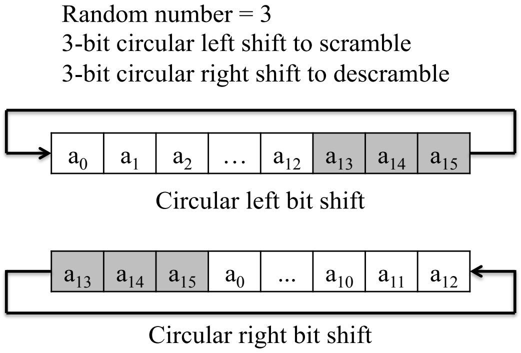

Random number = 3
3-bit circular left shift to scramble
3-bit circular right shift to descramble
a0 a1 a2 … a12 a13 a14 a15
Circular left bit shift
a13 a14 a15 a0 ... a10 a11 a12
Circular right bit shift

Fig. 2 Scrambling based on circular bit shift

scramble the original speech signal. A common key is used as the seed of pseudorandom number generator to produce a predictable sequence of numbers. Each number is the bit position that should be shifted. The common key could be any nonnegative integer. In addition, in order to be able to completely recover scrambled speech into the original speech, we carry out circular right shift of random bit position operation on each sample. The bit position is shifted equal to the value when it was scrambled by taking advantage of the common key.

2.1 Random number generator

The Mersenne twister is used as a random number generator [6], which is the default pseudorandom number generator (PRNG) in Matlab. Mersenne twister is the first PRNG to provide fast generation of high-quality pseudorandom integers that have the colossal period of 2^{19937}-1 iterations. The function “rng” in Matlab is used to control the seed of the PRNG (key for scrambler) so that the random number generator can produce a predictable sequence of numbers. In other words, if the key is identical, the sequence of random numbers will always be identical. The length of the sequence of random numbers is equal to the number of samples in a PCM speech signal. Each random number is the shift amount of the bit shift for each sample. Each random number is the shift amount of the bit shift for each sample.

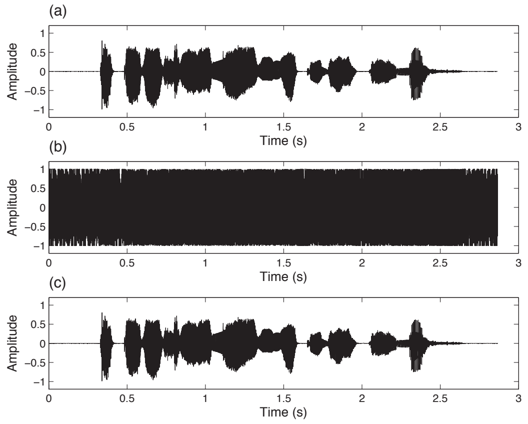


Fig. 3 Example of proposed method: (a) Original speech, (b) Scrambled speech, (c) Descrambled speech

2.2 Random-bit shift

The scrambling and descrambling are done by circular bit shift operation, as illustrated in Fig. 2. With circular bit shift, the value of the amplitude for a PCM speech signal will be changed. There is, for instance, the binary number (0100110)2 = (38)10. With a circular left bit shift of 3-bit positions, this number will be changed to (0110010)2 = (50)10. In the proposed method, the bit position of circular bit shift operation is based on the sequence of random numbers generated by the pseudorandom number generator. The steps for scrambling by using the proposed method are shown as follows.

(1) A PCM speech signal is expressed as a sequence of binary numbers:

$$x(m) = (a_0 a_1 a_2 \cdots a_{n-2} a_{n-1})_2 \qquad (1)$$

where m = 0, 1, …, M-1, and M is the number of samples of the signal; n is the quantization bit rate.

(2) Use an integer (common key) as the seed for the pseudorandom number generator to generate a sequence of uniform random integers *r(i)*. The random numbers should be drawn from the discrete uniform distribution on the interval [0, n-1].

(3) For sample *m*, the circular bit left shift *r(m)* bit operation is done on *x(m)*.

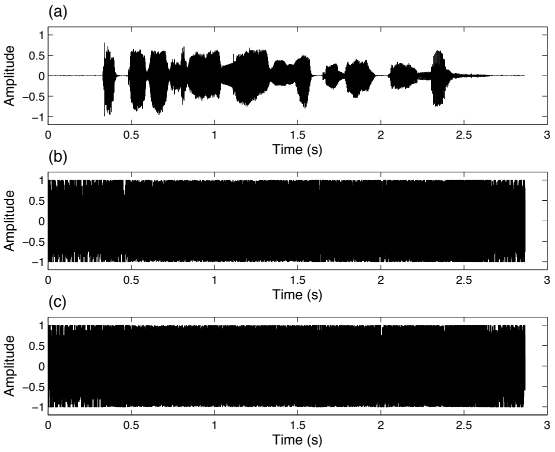


Fig. 4 Example of proposed method: (a) Original speech, (b) Scrambled speech, (c) Speech descrambled with wrong key

(4) Apply step (3) to each sample of the original speech signal.

The descrambling process involves changing the circular bit left shift to a right shift in step (3), and the other steps are all the same as in the scrambling process. This is an inverse operation of the scrambling process. If the key used to generating random integers in step (2) is correct, scrambled speech signal can be descrambled into the original one completely.

An example of the proposed method is shown in Fig. 3. An integer "5" is used as the common key here. Figures 3(a)-(c) show the original speech signal, scrambled speech signal, and the speech signal descrambled from scrambled speech. In Figs. 3(b) and (c), as a result, we can see that the original speech was scrambled into an unintelligible sound signal and the scrambled speech signal was recovered into the original speech signal completely.

3. Evaluations

The confidentiality and efficiency of the proposed method were evaluated by carrying out two objective measures, signal-to-error ratio (SER) in dB and perceptual evaluation of speech quality (PESQ) with an objective difference grade (ODG). SER shows the error ratio between the target signal and the original signal and is defined by the following equation.

$$SER = 10\log_{10}\left(\frac{\sum f_1^2}{\sum (f_1 - f_2)^2}\right) \text{ [dB]} \qquad (2)$$

Table 1 Results of evaluations
(mean and standard deviation)

	Scrambled signal	Descrambled signal	Descrambled by wrong key
SER [dB]	-12.1 (1.31)	∞ (0)	-12.1 (1.29)
PESQ [ODG]	0.62 (0.19)	4.5(0)	0.62 (0.19)

A smaller value of SER means a higher error between the target signal and the original signal. When the value is ∞ dB, the result indicates that the target signal is the same as the original signal. PESQ is a standard for objective voice quality testing. It provides scores in the range –0.5 to 4.5 to map the value of mean opinion score (MOS) from 1 (Bad) to 5 (Excellent). 20 stimuli from a database produced by Fujitsu Laboratory were used to evaluate the proposed method. The same sampling frequency of 22.05 kHz and 16-bit resolution were used for all stimuli.

Table 1 shows the results of the evaluation. For the scrambled speech signal, the mean value of SER was negative, and the mean value of PESQ was close to 0. As a result, we confirmed that the original speech signal had been completely scrambled by proposed method. In other word the original speech signal was converted into similar white noise, and the linguistic information of the original speech was broken completely. In addition, for the descrambled speech signal, the mean value of SER and PESQ were ∞ dB and 4.5 ODG. The results confirmed that the scrambled speech signal could be completely recovered into the original speech signal. The results were also confirmed with preliminary subjective tests.

In addition, a case of where the scrambled speech is descrambled with the wrong key is shown in Fig. 4(c). For the scrambled speech, the common key integer "5" was used as the seed of the random number generator. Figure 4(c) is speech incorrectly descrambled with the wrong key "6". From the waveform we can see that the scrambled speech could not be recovered into the original speech signal. The results of objective evaluation are also shown in Table 1. We can see there was not much change both in the SER and PESQ. This means that the speech descrambled with the wrong key is still unintelligible.

4. Discussion

As an additional consideration, we investigated the confidentiality of the proposed method. The random-bit shift is applied to each sample with the proposed method.

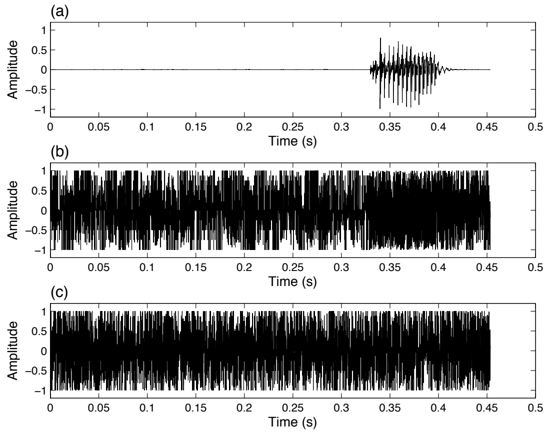


Fig. 5 Example of improved method: (a) Original speech signal, (b) Scrambled speech signal, (c) Scrambled speech signal with disrupting the order of samples

However, the order of the samples of the scrambled speech is the same as the original one. To improve confidentiality, we try to disrupt the order of samples at random in order to make descrambling by brute force attack more difficult. The amount of calculation needed for an attacker to break the whole scrambled speech signal of *n* seconds and *fs* kHz sampling frequency is $16^{n\times fs}$ for the proposed method. However, by disrupting the order of samples, an adversary has to complete $(n\times fs)!\times 16^{n\times fs}$ tries before they can recover the scrambled speech into the original speech. For a scrambled speech signal of 1 s and at a sampling frequency of 20 kHz, this number is $20000!\times 16^{20000}$. This security level appears to be sufficient.

Figures 5(a)-(c) show an example of the above-mentioned method. Figure 5(a) is a speech signal with one word. Figure 5(b) is the scrambled speech without the order of samples disrupted and Figure 5(c) is the scrambled speech with the order of samples disrupted. As a result, we can see that the effect of scrambling on the non-speech period was less than the speech period in Fig. 5(b). The reason is the value of samples in the non-speech period was very small; therefore, the effect of circular bit shift was not that much as well. It is easy to distinguish the speech period and non-speech period from the waveform. However, by disrupting the order of samples, the impact of the scrambling increased, and the scrambled speech was more like Gaussian white noise.

Table 2 Results of evaluations on improved method (mean and standard deviation)

	Scrambled signal	Descrambled signal
SER [dB]	-12.1 (1.31)	∞ (0)
PESQ [ODG]	0.46 (0.20)	4.5(0)

The results of objective evaluation of the improved method with the same stimuli are shown in Table 2. We can see that the mean value of SER did not change that much, whereas the mean value of PESQ decreased from 0.62 to 0.46. This means that the voice quality deteriorated further, and the original speech was scrambled more deeply. We also confirmed that the scrambled speech signal could still be recovered into the original one.

5. Conclusion

In this paper, we proposed a speech scrambling method that uses the random-bit shift of quantization bits with a common key. The results of evaluations showed that the original speech signal could be scrambled into an unintelligible audio signal. We also confirmed that the scrambled speech signal could be recovered into the original signal completely. We also discussed improving the confidentiality of the proposed method by disrupting the order of samples. Our future work will be on using the proposed method for not only PCM speech signals but also compressed speech signals like MPEG.

Acknowledgments

This work was supported by a Grant-in-Aid for Scientific Research (B) (No. 23300070) and the Fostering ICT Global Leader Program.

References

[1] N. S. Jayant: Analog scramblers for speech privacy, Computers & Security, Vol. 1, pp. 275-289, 1982.

[2] L. Lee, G. Chou and C. Chang: A new frequency domain speech scrambling system which does not require frame synchronization, IEEE Transactions on Communications, Vol. 32, No. 4, pp. 444-456, 1984.

[3] N. S. Jayant, B. McDermott, S. Christensen and A. Quinn: A comparision of four methods for analog speech privacy, IEEE Transcations on Communications, Vol. COM-29, No. 1, pp. 18-23, 1981.

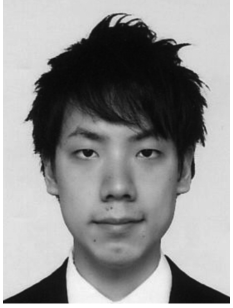
[4] A. Kataoka and S. Hayashi: A cryptic encoding method for G.729 using variation in bit-reversal sensitivity, IEICE Transactions (Japanese Edition), Vol. J87-D-II, No. 6, pp. 1224-1232, 2004.

[5] J. F. de Andrade, M. L. R. de Campos and J. A. Apolinaˊrio: Speech privacy for modern mobile communication systems, Proc. ICASSP2008, pp. 1777–1780, 2008.

[6] M. Matsumoto and T. Nishimura: Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator, ACM Transactions on Modeling and Computer Simulation (TOMACS), Vol. 8, No. 1, pp. 3-30, 1998.

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

Zhi Zhu received his B.E. degree in communication engineering from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2012. Now he is studying at the Master's course in the Japan Advanced Institute of Science and Technology (JAIST). He is interested in the secure voice and speech individuality.



Katsuhiko Yamamoto received his B.E. degree from Kobe City College of Technology in 2013. Now he is studying at the Master's course in Japan Advanced Institute of Science and Technology (JAIST). He is interested in auditory perception and models.

Masashi Unoki received his M.S. and Ph.D. degrees from the Japan Advanced Institute of Science and Technology in 1996 and 1999, respectively. His main research interests are in auditory motivated signal processing and the modeling of auditory systems. He was a Japan Society for the Promotion of Science (JSPS) research fellow from 1998 to 2001. He was associated with the ATR Human Information Processing Laboratories as a visiting researcher from 1999-2000, and he was a visiting research associate at the Centre for the Neural Basis of Hearing (CNBH) in the Department of Physiology at the University of Cambridge from 2000 to 2001. He has been with the Faculty of School of Information Science at JAIST since 2001 and is now an Associate Professor. He is a member of the Research Institute of Signal Processing (RISP), the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan, and the Acoustical Society of America (ASA). He is also a member of the Acoustical Society of Japan (ASJ), and the International Speech Communication Association (ISCA). Dr. Unoki received the Sato Prize from the ASJ in 1999 and 2010 for an Outstanding Paper and the Yamashita Taro "Young Researcher" Prize from the Yamashita Taro Research Foundation in 2005.

Naofumi Aoki received his Ph.D. degree from Hokkaido University in 2000. He is currently an Assistant Professor at the Graduate School of Information Science and Technology, Hokkaido University. He is engaged in the research on media information processing.

(Received April 24, 2014; revised September 3, 2014)