Journal of Signal Processing, Vol.18, No.4, pp.205-208, July 2014

SELECTED PAPER AT NCSP’14

Study on Semi-scramble Method for Speech Signals Based on Phonemic Restoration

Katsuhiko Yamamoto[1], Zhi Zhu[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
E-mail: {kyamamoto, zhuzhi, unoki}@jaist.ac.jp

[2] Graduate School of Information Science and
Technology, Hokkaido University
Kita–ku, Kita–14 Nishi–9, Sapporo 060–0814, Japan
E-mail: aoki@ime.ist.hokudai.ac.jp

Abstract

Speech scrambling methods are widely used for copyright protection and encrypting digital speech signals in order to guarantee the confidentiality of the original signals. They are very important methods for preventing eavesdropping and unauthorized copying. However, it seems to be impossible to completely recover a scrambled speech signal into the original signal. Moreover, nobody can comprehend the partial speech content from speech signals scrambled with these methods. In this paper, we propose a semi-scramble method for speech signals based on phonemic restoration. By using a speech scrambling method based on the random-bit shift of quantization bits, speech signals are converted to scrambled signals in partial intervals. We evaluated the confidentiality and efficiency of the proposed method by using two objective measures, signal-to-error ratio (SER) and perceptual evaluation of speech quality (PESQ). As a result, we confirmed that the proposed method can play a role in copyright protection for an original signal and recover a semi-scrambled speech signal into the original one. Finally, we indicated that the acoustic characteristics of signal semi-scrambled with the proposed method enable the listener to understand the speech information.

1. Introduction

Recently, encryption and information protection for digital data have become more important. Speech scrambling methods are widely used for copyright protection and encrypting digital speech signals in order to guarantee the confidentiality of an original signal [1]. These methods are very important for preventing eavesdropping and unauthorized copying. There are, for example, typical methods such as speech codecs and reverse scrambling in the frequency domain. However, it seems to be impossible to completely recover a scrambled speech signal into the original signal. Moreover, nobody can make out the partial speech content from speech signals scrambled with these methods.

A semi-scramble speech method is one way to encrypt content to the extent that people can understand most of it. Although semi-scrambled data cannot be encrypted completely, it plays a role in copyright protection. In previous studies, there have been typical methods such as semi-scrambling a digital speech signal through coding by changing MDCT coefficients of the original signal [2], and reversing the order of the samples in the time domain. However, the same as with the previous scramble methods, it seems to be impossible to completely recover a semi-scrambled speech signal into the original signal.

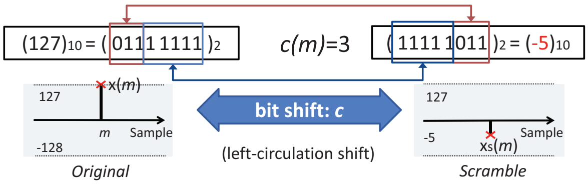

Figure 1: Example of scrambling with speech scrambling method based on random-bit shift of quantization (quantization bit length: $n = 8$, shifting frequency: $c(m) = 3$)

This study aims to achieve a semi-scramble method that plays a role in copyright protection for original speech signals and recovers a semi-scrambled speech signal into the original one completely. In this paper, we propose a semi-scramble method for speech signals based on phonemic restoration. Here, a speech scramble method based on the random-bit shift of quantization bits [3] is used to solve the problem.

2. Key Techniques

2.1 Speech scramble method based on random-bit shift of quantization bits

Figure 1 shows an example of scrambling by using the speech scramble method based on random-bit shift of quantization bits that the author proposed [3]. In this example, the numerical representation is binary or decimal, and the quantization bit rate is $n = 8$. When the value $x(m)$ of the original speech signals is represented as the decimal form of $(127)_{10}$, it is represented as the binary form of $(01111111)_2$ by the quantization bits. This scrambling method cyclically shifts the arrangement of the bits. The shift amount of the bit shift is based on the random number sequence c generated from a seed number as the public key. When the value of the shift amount is $c(m) = 3$, the value of the $x(m)$ is cyclically

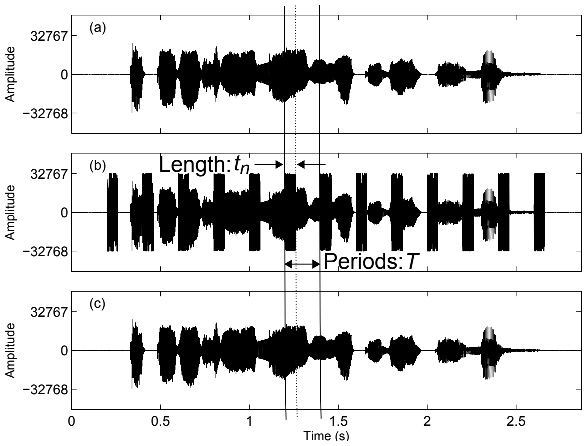

Figure 2: Examples of semi-scrambling with the proposed method: (a) Original speech, (b) Semi-scrambled speech, and (c) Descrambled speech

shifted 3 bits to the left. As a result, the value of the $x(m)$ is transformed to $x_s(m) = (11111011)_2$, and it is represented as $(-5)_{10}$ by considering two's complement. These procedures are applied to all samples of the original speech signal and convert it into a scrambled speech signal. Moreover, if the seed number is correct, the scrambled speech signal can be completely descrambled to the original one by applying the inverse process.

2.2 Phonemic restoration

When the partial intervals of a speech signal are replaced by gaps of silence, the speech's intelligibility is drastically reduced. However, when the gaps are filled with white noise that is louder than the recorded voice, the utterance sounds more natural and continuous. This phenomenon is referred to as the phonemic restoration effect. This effect occurs not only with speech sounds but also any sounds such as pure tones. These perceptual restorations are collectively called the "auditory continuity effect" [4].

When the effect of phonemic restoration occurs, the necessary conditions for the effect of auditory continuity are satisfied. Apart from that, co-articulation (temporal distribution of information for a phonetic segment) helps the listener to increase the intelligibility [5]. In addition, it is revealed that by using the semantic information from the context, speech continuity is kept from being interrupted any longer.

3. Proposed Method

We propose a semi-scramble method for speech signals based on phonemic restoration. The speech scrambling method based on the random-bit shift of quantization bits converts an original speech sounds into an unintelligible ones. By using the scrambling method on partial intervals of the original speech signal, speech information is partially encrypted. In the proposed method, the partial scrambling is repeated in the intervals of the semi-scrambled sound. However, the listener can restore each of the encrypted pieces of information of the semi-scrambled signal perceptually by phonemic restoration. Therefore, the proposed method is able to semi-scramble the information of speech sounds.

Figure 2 shows an example of semi-scrambling with the proposed method. In this example, the quantization bit length is $n = 16$, and the scrambling method is applied to the speech signal at two parameters, the length $t_n = 60$ ms and the period $T = 200$ ms of the scrambling. Figure 2(a) is the waveform of the original speech. The semi-scramble method is applied to the original speech signal, and the semi-scrambled speech signal that has the intervals of noise is converted, as shown in Fig. 2(b). The descrambled speech signal (Fig. 2(c)) is converted from the semi-scrambled speech signal. If the semi-scrambled speech is converted with the correct seed number of the random sequence number, the descrambled speech corresponds to the original speech completely.

4. Evaluations

4.1 Semi-scrambled speech signal

Two objective tests, signal-to-error ratio (SER) and perceptual evaluation of speech quality (PESQ) [6] of semi-scrambled and descrambled speech, were carried out to evaluate the confidentiality and efficiency of the proposed method. SER is represented as the error ratio of the target signal $x_1(m)$ to the standard signal $x_2(m)$ and is defined by the following equation,

$$\mathrm{SER} = 10\log_{10}\frac{\sum_{m=0}^{\mathrm{M}-1} x_1^2(m)}{\sum_{m=0}^{\mathrm{M}-1}(x_1(m) - x_2(m))^2} \quad \text{(dB)}$$

where $m = 0, 1, \ldots, \mathrm{M}-1$, and M is the total number of samples of signal. When the value of SER is positive, the error between the target and standard signal is small. Moreover, when the value is ∞ dB, the result indicates that the target signal is completely the same as the standard signal. When the value of SER is negative, the error between the target and standard signal is high. PESQ shows an objective difference grade (ODG) represented as a value from -0.5 to 4.5. These values correspond to mean opinion score (MOS) represented as a value from 1 (Bad; Very annoying) to 5 (Excellent; Imperceptible). 20 stimuli were selected from the Japanese multi-emotion single speaker Fujitsu database produced and recorded by Fujitsu Laboratory [7]. The sampling frequency was 22.05 kHz, the quantization bit rate was 16 bit, and the length of the stimuli was about 3 sec.

In this study, to evaluate the variations of SER and PESQ of the proposed method numerically, the stimuli, the semi-scrambled and descrambled speech signals, were generated by varying two parameters, the length t_n and the period T of the scrambling (Fig. 2). The periods of the scramble Ts were set to $100, 200, 300, 400,$ and 500 ms. The length t_n was

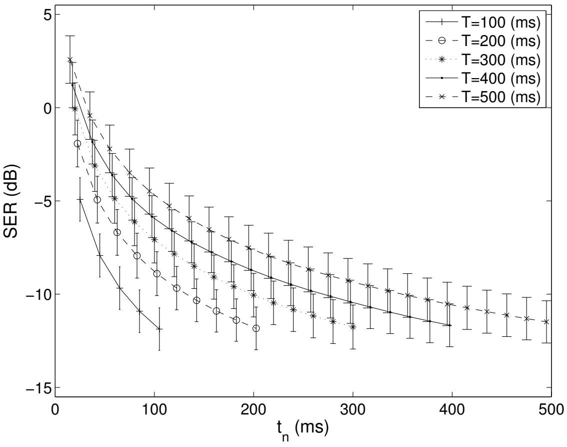

Figure 3: SER of semi-scrambled speech

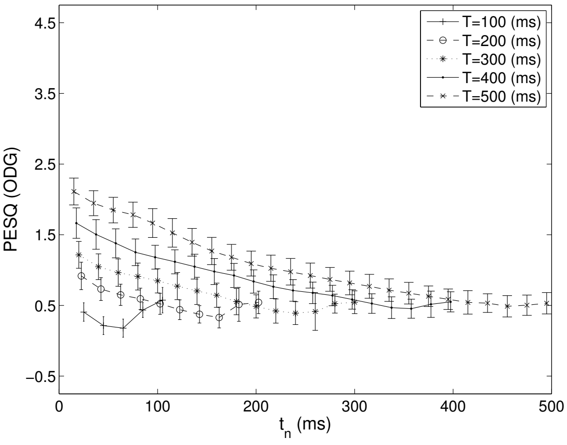

Figure 4: PESQ of semi-scrambled speech

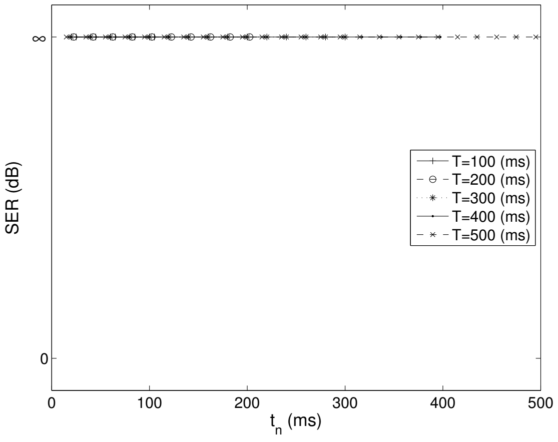

Figure 5: SER of descrambled speech

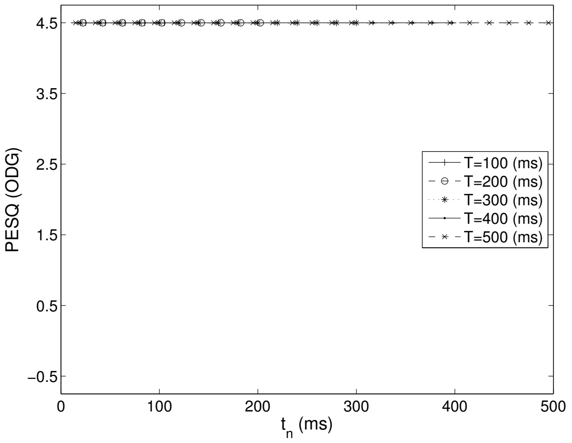

Figure 6: PESQ of descrambled speech

increased from 20 ms to T in steps of 20 ms. When the length t_n is 0 ms, the scramble process is not applied. Therefore, the semi-scrambled speech is the same as the original one. In addition, when the length t_n equals the period T, scrambled speech encrypted in all sections is made by our scrambling method [3].

Figures 3 and 4 show the results of SER and PESQ for semi-scrambled speech. Each of the values is represented as an average, and each of the error bars is represented as a standard deviation of the result for each parameter. When t_n was 0 ms, SER was ∞ dB and PESQ was 4.5 ODG in the entire period of the scramble T. As the length t_n became longer, SER and PESQ reduced gradually. When t_n reached T, the SER and PESQ of the semi-scrambled speech were the same as that of the scrambled speech [3]. These results could confirm that the proposed method could scramble the information of the original signal in partial intervals. and control the degree of the scrambling.

Figures 5 and 6 show the results of SER and PESQ for descrambled speech. For every length T and the period t_n, SER was ∞ dB, and PESQ was 4.5 ODG. These results confirmed that the proposed method could recover the semi-scrambled speech signal into the original one.

4.2 Semi-scrambled audio signal

An objective test, the perceptual evaluation of audio quality (PEAQ) [8], was carried out to investigate applying the proposed method to an audio signal. PEAQ shows ODG represented as a value from -4 (Bad; Very annoying) to 0 (Excellent; Imperceptible). 100 stimuli were selected from the Music Genre Database in the RWC Music Database [9]. These stimuli were processed to the monaural signals of the left channel in advance, and the length was set to 10 s (interval from 60 to 70 s in the original audio). As with the evaluation of the semi-scrambled speech signal, the stimuli of the semi-scrambled audio signal were generated by varying two parameters, the length t_n and the period T of the scrambling. The period of the scramble T was set to $100, 200, 300, 400$, and 500 ms. The length t_n was increased from 50 ms to T in steps of 50 ms.

Figure 7 shows the results of PEAQ for the semi-scrambled

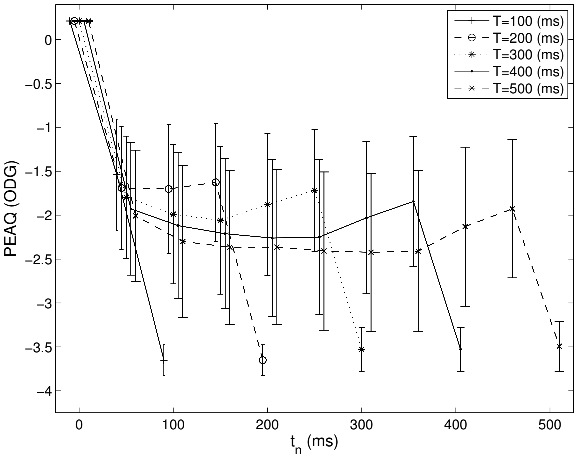

Figure 7: PEAQ of semi-scrambled audio

audio. Each of the values is represented as an average, and each of the error bars is represented as a standard deviation of the result for each parameter. For the parameter used for the semi-scrambled speech sound, the values of PEAQ were distributed from -1.5 to -2.5. These results could confirm that the proposed method deteriorated the sound quality of the audio signal.

5. Discussion

As an additional consideration, we investigated the acoustic characteristics of the scrambled signal in semi-scrambled speech. In this study, we hypothesized that the scrambled signal is white noise, and then analyzed the histogram of the amplitude, the spectrogram, and autocorrelation [10]. The results showed that each sample of the scrambled signal had a normal distribution with zero mean and a specific variance, flat spectrum over the range of frequencies, and its value at one time was uncorrelated with any other time. Therefore we confirmed that the scrambled signal with the proposed method was a white Gaussian noise. The characteristics satisfied the necessary conditions for the auditory continuity effect [5].

Thus, phonemic restoration enables the listener to comprehend the speech information even though the sound quality of the semi-scrambled speech is deteriorated. Moreover, when the lengths of the scrambling intervals are sufficiently short, the semi-scrambled speech can be perceived as the original one, as shown in our preliminary subjective tests. In addition, the continuity effect enables the listener to listen to music more naturally and continuously even though the sound quality of the semi-scrambled audio is deteriorated. However, these results indicate that the ideal length of the scrambling for the semi-scrambled audio, which does not have linguistic information, becomes shorter than that of the scrambling for semi-scrambled speech.

In a future study, to reveal the relationship between scrambling length and the ease of hearing semi-scrambled speech and audio, we will consider a subjective evaluation of the proposed method through hearing tests.

6. Conclusion

We proposed a semi-scramble method for speech signals based on phonemic restoration. Two objective measures, SER and PESQ, were used to evaluate the confidentiality and efficiency of the proposed method. These results confirmed that the proposed method plays a role in copyright protection for an original signal and recovers a semi-scrambled speech signal into the original one. An objective measurement, PEAQ, was used to investigate applying the proposed method to audio signals. These results confirmed that the proposed method deteriorated the sound quality of the audio signal. In addition, we also confirmed that the signal scrambled with the proposed method is a white Gaussian noise. Finally, we indicated that the acoustic characteristic of the signal semi-scrambled with the proposed method enables the listener to comprehend speech and music information.

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] Y. Nishihara, N. Iwakami, H. Fujii and K. Kushima: A research on semi-disclosure method of audio-music contents, IPSJ SIG Notes, Vol. 97, No. 28, pp. 31–36, 1997 (in Japanese).

[3] Z. Zhu, K. Yamamoto, M. Unoki and N. Aoki: Study on scramble method for speech signal based by using random-bit shift of quantization, Proc. NCSP'14, pp. 109–112, 2014.

[4] R. M. Warren: Auditory Perception: A New Analysis and Synthesis, Cambridge University Press, Cambridge, 1999.

[5] M. Kashino: Phonemic restoration: The brain creates missing speech sounds, Acoustical Science and Technology, Vol. 27, No. 6, pp. 318–321, 2006.

[6] H. Yi and C. L. Philipos: Evaluation of objective measures for speech enhancement, Proc. Interspeech2006, pp. 1447–1450, Pittsburgh, Pennsylvania, 2006.

[7] R. Elbarougy and M. Akagi: Improving speech emotion dimensions estimation using a three-layer model for human perception, Acoustical Science and Technology, Vol. 35, No. 2, pp. 86–98, 2014.

[8] P. Kabal: An examination and interpretation of ITU-R BS.1387: Perceptual evaluation of audio quality, TSP Lab. Technical Report, Dept. Electrical & Computer Engineering, McGill University, pp. 1–89, 2002.

[9] M. Goto: Development of the RWC music database, Proc. the 18th International Congress on Acoustics (ICA 2004), pp. I–553–556, 2004.

[10] K. Yamamoto, Z. Zhu, M. Unoki and N. Aoki: Study on semi-scramble method for speech signals based on phonemic restoration, IEICE Technical Report, Vol. 113, No. 290, pp. 59–64, 2013 (in Japanese).