Study on the Reproducton of Vocal Fluctuations using Box-Muller method for Artificial Synthesis

Tomohisa OKAWA[1], Takaaki KONDO[1], Shun KADOI[1], Kyouhei KAMIYAMA[1], and Hiroyuki KAMATA[2]

[1]Graduate School of Science and Technology, Meiji University, Japan
[2]School of Science and Technology, Meiji University, Japan

1 Introduction

The human voice contains the fluctuations of fundamental frequency and human speech waveform. According to a study, these fluctuations are producing the naturalness of human voice.[1] Today, text-to-speech synthesis has become mainstream in speech synthesis. This technology, however, requires a large amount of data to make high-quality speech synthesis. For this reason, it is impossible to create a synthesized voice of many people. In this study, we aim to create a synthesized speech by focusing on fluctuation of the vocal cords. By using the Box-Muller method, we try to reproduce the fluctuations of the vocal cords. In the Box-Muller method, the average and variance values of the Gaussian function is used. In this paper, we extract the fundamental frequency and frequency variance value from the human voice. We expect that this result will be an indicator of reproduction of vocal fluctuations by the Box-Muller method.

2 Fluctuation of vocal cords

At first sight, the human speech waveform seems to be a series of similar shape. However, this form is changing slightly for each shape. Cycle width is also changing as well. The difference in width of the period is called the fluctuation of the vocal cords.[2]

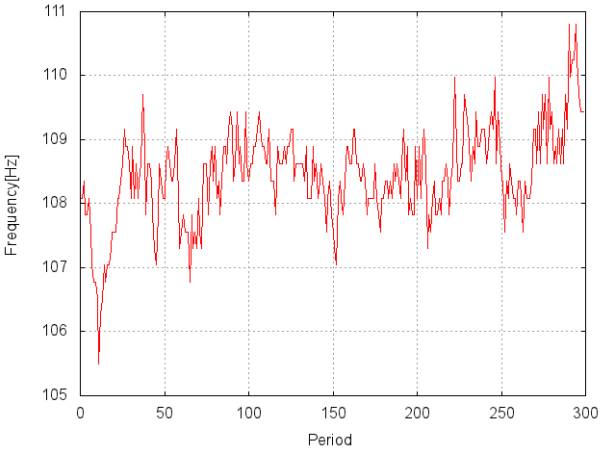

Fig. 1. The vocal fluctuations

3 Derivation of the fluctuation of the vocal cords

In this chapter, we describe the method of reproducing the fluctuation of vocal cords. In this study, we try to reproduce the fluctuation of vocal cords by using the Box-Mueller method. This method uses the statistical distribution of the fluctuation of vocal cords of human voice. Therefore, we calculate the fundamental frequency and variance value in order to the statistical distribution. Fundamental frequency f is calculate from the average frequency of the waveform of period 300, the variance value σ^2 is calculated by the following equation.

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(f_i - f)^2 \tag{1}$$

Here, f_i is the frequency of each cycle. Using this result, we generate a random number according to the normal distribution created by Box-Muller method. How to generate random numbers Z_1,Z_2 by Box-Mueller method is shown below.

$$Z_1 = \sigma\sqrt{-2 \times \ln x}\cos(2\pi y) \tag{2}$$
$$Z_2 = \sigma\sqrt{-2 \times \ln x}\sin(2\pi y) \tag{3}$$

Here, x and y are mutually independent random variables, which are subject to the uniform distribution on (0,1).

4 Result

We show the result of relationship between the fundamental frequency and variance value from a human voice. According to Fig. 2, the variance value is somewhat high when the fundamental frequency is low. and the variance value become lower with increasing fundamental frequency. Considering the relationship between fundamental frequency and variance value, we made a synthesis of the fluctuation of vocal cords by using the Box-Muller method.

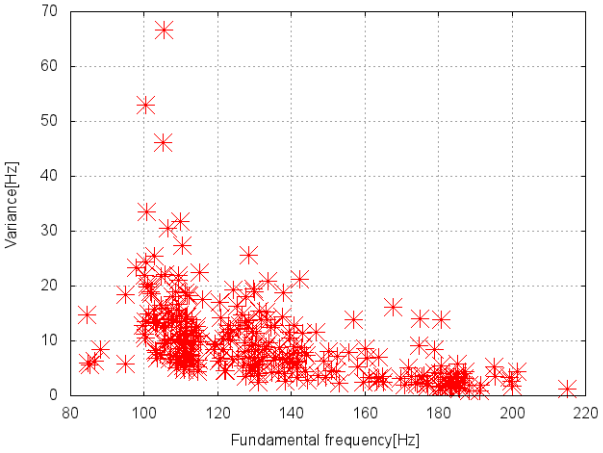

Fig. 2. Relationship between the fundamental frequency and variance value

Next, we show the result of the synthesis of the fluctuation of vocal cords by using Box-Muller method.

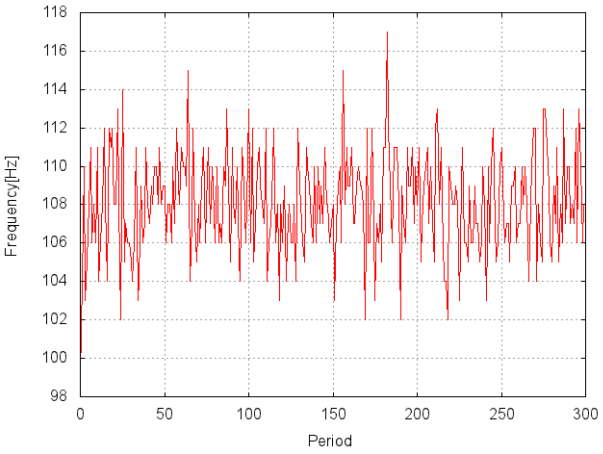

Fig. 3. The synthesized fluctuations(fundamental frequency =108[Hz], variance value = 2,71[Hz])

5 Conclusion

In this study, in order to reconstruct the speech synthesis with highly naturalness, we focused on the fluctuations of vocal cords in a real voice signal /a/. In this result using the Box-Muller method, some different were observed between the reproduced frequency and the frequency of the real voice. Major difference is that real voice has small fluctuations in adjacent frequency, however the fluctuations of the reproduced frequency sometimes becomes large. By this result, in order to reproduce the frequency of the fluctuations of vocal cords, it is necessary to consider the frequency of the previous cycle, not only to generate a normal random.

In future, we plan to a comparison of real speech and synthetic speech with this result.

References

[1] Naohumi AOKI, Tohru IFUKUBE, Frequency Characteristics of Amplitude and Period Fluctuations in Sustained Vowels and Their Psychoacoustic Effects, IEICE A, 82, 649–657, 1999

[2] Yuuki NANIWA, Takaaki KONDO, Kyouhei KAMIYAMA, Hiroyuki KAMATA, Study on the Artificial Synthesis of Human Voice using Radial Basis Function Networks, IEICE technical report NC, 110, 199–204, 2011