Singing System by Electric-Larynx Using LPC Residual Wave

Takashi Himeno*, Yuto Tanaka†, Mitsunori Mizumachi‡, Yoshihisa Nakatoh§, Kenji Matsui¶

*Kyushu Institute of Technology, 1-1 Sensui-cho, Tobata-ku, Kitakyusyu 804-8550, Japan
†Osaka Institute of Technology, 5-16-1 Omiya, Asahi-ku, Osaka 535-0002, Japan

*Corresponding Author: o349426t@mail.kyutech.jp

Abstract

In this paper, we propose a singing system to transmit the pseudo-vocal sound source to the vocal tract. The purpose is the quality improvement of singing voice when using an Electric-Larynx (EL), which is an alternative vocal apparatus of laryngectomee. Thus, we focus on the LPC residual wave, but it has a problem that the signal-power of the vowels /i/ and /u/ are largely reduced. Therefore we tried to improve the quality of sound by adding a signal-power compensation.

We carried out listening experiments for the four sound sources "LPC residual wave", "LPC residual wave (signal-power Compensation)", "Rosenberg wave" and "EL". As a result, power-compensated LPC residual wave had the highest naturalness. However, the spectrum of the vowels /i/ differed greatly from the singing voice of the normal utterance.

Keywords: Singing System, Electric-Larynx, Laryngectomy, LPC residual wave.

1. Introduction

Songs are one of the major forms of entertainment, so a lot of people have fun with it. However, some people cannot produce a natural sound because of handicaps. Laryngectomy is a common technology for treatment of laryngeal cancer. As a result of the treatment, the patient’s larynx has to be removed completely. Thus, laryngectomees cannot emit voices through sound source vibration.

The population of laryngectomees is estimated to be more than 30,000 in Japan. Because Electric-Larynx (EL) can be easily obtained, the demand of it is increasing as an alternative vocalization among elderly laryngectomees(1). Also, EL is attached to a sing function called the Song-mode. However, the voice of Song-mode is unnatural.

Fig. 1. The route of the air flowing from the lungs. The left figure shows the case of non-handicapped people. The right figure shows the case of the laryngectomees.

2. The difference in the singing voice between a human voice and EL Song-mode

In this chapter, we describe the difference between the singing voice with a human voice and EL. Singing voice using an EL is a mechanical, naturalness is very low. Figure 2 is a graph of a singing voice pitch changes with human voice and EL. From Figure 2, while the human singing voice is smooth transition, singing voice using EL becomes choppy at the turn of the sound. To solve these problems, we propose a system described in the next chapter.

Fig. 2 Change of the two singing voice pitch
3. Singing system using a pseudo-vocal sound source

3.1 System summary

Figure 3 is a singing system using a pseudo-vocal sound source. First, one must create a pseudo-vocal sound source. Next, one must convert it to a vibration sound. Finally, the source is outputted from the vibration speaker, and singing is pressed against the throat. The system can reproduce a fine representation that cannot be reproduced in EL.

3.2 The generation of pseudo-vocal sound source

We have created two types of LPC residual waves and Rosenberg waves as a pseudo-vocal sound source. First, the LPC residual wave refers to a sound source which can be extracted by linear predictive coding of the speech. We extract the LPC residual wave from singing voice in the normal utterance. Songs to be used this time is "Twinkle Twinkle Little Star" in the nursery rhyme, because this song is inherent in EL Song-mode. In addition, we used the voice analysis software "Praat" to extract. Figure 4(a) is a waveform of LPC residual wave. Next, Rosenberg wave is used as a vocal sound source in speech synthesis \(^3\)\(^4\). It is created by using the equation (1). We set to \(\tau_1 = 0.95\), \(\tau_2 = 0.90\). This time, we have created a Rosenberg wave along to the music of "Twinkle Twinkle Little Star". Figure 3(b) is a waveform of Rosenberg wave.

\[
f(t) = \begin{cases} 
3 \left( \frac{t}{\tau_1} \right)^2 - 2 \left( \frac{t}{\tau_1} \right)^3 & (0 \leq t \leq \tau_1) \\
1 - \left( \frac{t-\tau_1}{\tau_2} \right)^2 & (\tau_1 \leq t \leq \tau_1 + \tau_2) 
\end{cases}
\]

3.3 Signal-power compensation to LPC residual wave

In LPC residual wave, the signal-power of the vowel /i/ and /u/ is reduced to 10-20 [dB]. Therefore, we have created an power-compensated LPC residual wave that was added to the signal-power increasing width of 15 [dB] to the signal-power reduction terms. Figure 5 is the signal-power waveform before the signal-power compensation and after it.
4. Evaluation experiment

4.1 Evaluation methods

In this experiment, we use a singing voice using a pseudo-vocal sound source and a normal human voice. The singing person speaks well using the EL. We output the pseudo-vocal sound source through a vibration speaker, and sing by pressing it against the throat. Figure 6 is the photograph of the EL and vibration speaker. We use a VR3000, KOBATEL vibration speaker. The recording location is a sound-proof room, and we set up a non-directional microphone 15 centimeters from the mouth. To evaluate the sound quality, we conducted listening tests with five sounds of “A: LPC residual wave”, “B: Power-compensated LPC residual wave”, “C: Rosenberg wave”, “D: EL-song mode”, and “E: A human voice”. We used a variant of the Nakaya pair-wise comparison of Scheffe to analyze the results(5). Also, we used four phrases: (1) kirakirahikaru, (2) osoranohoshiyo, (3) mabatakishiteha, (4) minnawomiteru from Twinkle Twinkle Little Star, and the subjects consisted of ten healthy individuals. The subjects heard two sounds that were selected at random and they chose the one most similar to natural speech. The type of choices in this evaluation consisted of five grades (-2: sound A was very good, -1: sound A was good, 0: same level, 1: sound B was good, 2: sound B was very good). Figure 7 is a screen of listening experiments.

4.2 The results of evaluation experiment

Figure 8 show the experimental result. Figure 5(a)’s the horizontal axis represents the evaluation value. It shows that the evaluation value is closer to -2, and that there is a high naturalness. (b) represents the degree of preference of

Fig. 6 The left figure shows EL and the right figure shows the vibration speaker.

Fig. 7 A screen of listening experiments.
each of singing voice with respect to the Power-compensated LPC residuals (95% confidence interval). The higher the degree of preference in the positive direction it represents a higher nature of the power compensated LPC residuals. From Fig. 8, we found that “Power-compensated LPC residual wave” is the most natural singing voice that uses the pseudo-vocal sound source. Therefore, we found that the sound quality is improved by the signal-power compensation. There is a large difference in naturalness between singing voices using “Power-compensated LPC residual wave” and “A human voice”. Thus, we focused on the spectrum of the vowel and we did a comparative analysis. We cutout vowel contained in the lyrics, and calculated the spectrums by FFT. Figure 9 shows two spectrums. In “A human voice”, the vowel /i/ has two valleys at 500-2000 [Hz] and 4000-5000 [Hz]. However, “Power-compensated LPC residual wave” does not have it. Therefore, we think that the singing voice’s using “Power-compensated LPC residual wave” pronunciation is not performed correctly.

5. Conclusions

In this paper, we verified whether perform sound quality improvement by adding a power compensated to the LPC residual wave. We made three pseudo-vocal sound source of LPC residual wave, Power-compensated LPC residual wave and Rosenberg wave, and uttered using a vibration speaker. Then, we evaluated variety of singing voice using a pair comparison of Scheffe. According to the results, we found that we can improve the sound quality by adding the signal-power compensation to the LPC residual wave. However, it has no unevenness compared to the human voice at spectrum of vowel /i/. From now on, we will consider compensating the method of spectrum in order to achieve a higher naturalness. Also, we want to increase the number of subjects of listening experiments.

Acknowledgment

This work was partially supported by a Grant-in-Aid for Science Research (No.15K01487) from the Japan Society for the Promotion of Science.

References

(2) Takashi Himeno, YutoTnaka, and Yoshihiisa Nakatoh: ”Denkishikizinkokotoh wo motiitautagoonsoitsu kaizen”,The Institute of Industrial Applications Engineers 2014