Study of Natural-Voice-Like Vibration Sound for Electrolarynx

Takafumi Hei^{a,*}, Yuto Tanaka^a, Mitsunori Mizumachi^a, Yoshihisa Nakatoh^a, Kenji Matsui^b

^aKyushu Institute of Technology, 1-1 Sensui-cho, Tobata-ku, Kitakyushu 804-8550, Japan

^bOsaka Institute of Technology, 5-16-1 Oomiya, Asahi-ku, Oosaka 535-0002, Japan

*Corresponding Author: n349430t@mail.kyutech.jp

Abstract

In this paper, we found a new vibration sound for an electrolarynx (EL) that sounds like natural voice. The difference of acoustic characteristics between natural speech and speech using an EL is stress from its laryngectomees using it. Also, this becomes a factor that hinders smooth communication for them.

Thus, we focus on the difference of the sound generated by an EL and by human being's real vocal cords. We created two pseudo vocal cords vibration waveforms that are Rosenberg wave and LPC residual wave. People can produce utterance by using a vibration loud speaker. We conducted listening experiments to figure out which sound is more similar to natural speech by using a variant of the Nakaya pairwise comparison of Scheffe. The results show that the speech by the LPC residual wave is closer to natural speech of human beings than EL.

Keywords: Laryngectomees, Electrolarynx, utterance support, sound source vibration.

1. Introduction

Voice is the primary device for communication. However, some people cannot produce natural sound because of handicaps.

Laryngectomy is a common technology for treatment of laryngeal cancer. As a result of the treatment, the patients' larynx have to be removed completely. Thus, laryngectomees cannot speak through sound source vibration. The ways for them to communicate with others are writing and facial expressions and these give them emotional distress and take a lot of effort to mutual understanding. Hence, alternative vocalizations are required⁽¹⁾.

The population of laryngectomees is estimated to be

Fig. 1. The route of the air flowed from the lung. Left figure shows the case of non-handicapped person. The right figure shows the case of the laryngectomees.

about less than 20,000 in Japan. A larynx cancer is common in elder people, and the number of them is increasing each year due to aging. Because an electrolarynx (EL) can be easily obtained, the demand of it is increasing as alternative vocalization among elder laryngectomees. However, speech using an EL is unnatural.

The difference of sound source vibration between an laryngectomees and non-handicapped person is focused in this study. We created two pseudo vocal cord vibration waveforms that are similar to the sound source actual vibration. After this operation, we output them by using a vibration loud speaker.

2. Electrolarynx sound

2.1 The present conditions of the laryngectomees

When a non-handicapped person utters, exhalation from lungs vibrate vocal cords and the vibration generate sounds. Then, providing a voice and a phoneme sound by controlling the articulatory organs such as the tongue,

thereby resonating the sound source vocal tract. However, in the case of laryngectomees, he doesn't have his own vocal cords, and his trachea and esophagus are completely separated. Furthermore, his breath goes through a hole called a tracheal stoma in his neck. Therefore, he cannot generate a sound source by himself. Thus alternative vocalization is required. The main alternative vocalizations for them are using an EL and esophageal speech.

Esophageal speech utterance is a method that uses burp. A sound source is generated by keeping air which taken from the mouth in the esophagus, and it vibrates the false glottis of the upper esophagus by discharging the air. By using this method, people can utter without using a special auxiliary instrument and the voice sounds like a natural voice than the alternative dictions. However, the amount of air that can be used at one utterance is small, generation time and the volume is limited. In addition, on that it takes more than six months to learn esophageal speech and it would use physical strength to utterance. For the people who have poor physical fitness like elders, who are high percentage among the Laryngectomees, an EL has an advantage because mastering how to use it is easier than mastering the esophageal speech. Hence, we focus on the method that uses EL in this study.

2.2 Electrolarynx utterance

EL is one of the most common support equipment which generates a sound source. It can generate a sound by pressing it to under of lower jaw. The sound source which is generated by vibration from a vibrator which is driven by a battery is transmitted to the vibration sound in a vocal tract. Thus, the user can move his mouth as non-handicapped person's utterance. This method has a larger volume than other comparison method, and it is easier to learn. Therefore, it is effective for elder people who find it hard to do esophageal speech. However, the sound of it is like a mechanical voice and far from natural. Also, the sound source is leaked around. Thus, using this method sometimes makes laryngectomees hard to do daily communication.

Thus, we thought that the difference between the vocal sound source of natural voice and source by an EL makes their communication hard.

3. Creating a pseudo vocal sound source

The EL vibration sound is similar to a pulse waveform, and it has less natural characteristics and the speaker's characteristics. Therefore, we created two pseudo vocal

(a) Left figure shows EL and the right figure shows the Vibration loud Speaker.

(b) Use example of Vibration loud Speaker.

Fig. 2. Instruments for use.

sound sources in imitation of a vocal sound source of non-handicapped people.

We used a vibration loud speaker for utterance instead of an EL. Fig2 (a) is the photograph of EL and vibration loud speaker. Vibration loud speaker is pressed against a throat as Fig2 (b) shows. We use vibration loud speaker that is VR3000, KOBATEL. This speaker applies a dynamic loudspeaker's structure and contains of a diamond Fulham (vibration version), a magnetic circuit, a spring coil and housing. A general speaker generates sound by shaking the vibration board in it. But a vibration loud speaker generates sound by shaking a contact surface of other objects and has no vibration board in it. Thus, we decided to use this speaker as a base instead of an EL.

3.1 Rosenberg wave

The Rosenberg wave is a transformed chopping wave⁽²⁾. It is proposed by Rosenberg and improved by Klatt. Because its waveform is similar to a natural vocal cords sound source, its used as a vocal sound source of voice synthesis.

3.2 LPC residual wave

The voice signal can be estimated that a fixed tuning filter (vocal tract information), which depends on the shape of a vocal tract, oral cavity and nasal cavity, convolutes with a sound source signals (vocal sound source) generated by a vibration of the vocal cords⁽³⁾. Thus, we created two

pseudo vocal sound sources by doing inverse filtering with the vocal tract properties which is demanded by linear predictive coding (LPC analysis) to original vocal sound. We call this waveform LPC residual wave.

4. Evaluation experiment of the electric larynx and the para-vocal cords sound source

We conducted listening experiments to find which sound is similar to natural speech and to use a variant of the Nakaya pairwise comparison of Scheffe for analysis of the results⁽⁴⁾.

4.1 Evaluation experiment

We conducted listening tests with three sound of EL, Rosenberg wave (RW) and LPC residual wave (LPC) to find which sound is similar to natural speech. We used a variant of the Nakaya pairwise comparison of Scheffe to analyze the results.

The subjects are ten university students who don't hear an EL sound routinely. The number of evaluation sounds is 15 that contain /a/, /i/, /u/, /e/, /o/ of the EL utterance, LPC residual wave utterance and Rosenberg wave utterance. The subjects heard two sounds that were selected in random and they chose the one similar to natural speech. The type of choices in this evaluation is five-grade (sound A is very good, sound A is good, same level, sound B is good, sound B is very good).

4.2 Experiment result

In this experiment, we used Nakaya variation of pair comparison method of Scheffe (five-grade). When sound B was at the same level as sound A, evaluation point is 0-point. Also, we aggregated to grant 2-points from -2-points in each grade. We show rating average of /a/ at Table 1.

In addition, Fig.3 (a) shows in the one-dimensional standard based on the count result. I found that speech by LPC residual wave is superior in every speech. Therefore, we performed a variance analysis and examined 99% confidence interval. Bar such as "T" of Fig3 (b) shows 99% confidence interval. There is no significance when bar steps over 0. Thus, /a/, /i/, and /u/ of EL and RW do not have significant differences. On the other hand, LPC is significantly different from other speech. Thus, there is a dominant impression difference in LPC residual wave statistically.

Table 1. Experiment result

Rating average of /a/.

	EL(A)	LPC(B)	RW(C)
EL(A)		0.61	-0.05
LPC(B)	-0.61		-0.47
RW(C)	0.05	0.47	
sum	-0.56	1.08	-0.52

(a) One-dimensional standard.

(b) 99% confidence interval.

Fig. 3. Experiment result.

5. Relationship of LPC residual wave sound and difference of phonological

We investigate whether LPC residual wave voice is affected by the difference in /a/, /i/, /u/, /e/, and /o/.

5.1 Evaluation experiment

We named the pseudo cords sound source that extracted from each vowel sound as LPC/a/: A, LPC/i/: B, LPC/u/: C, LPC/e/: E and LPC/o/:D. We evaluated them by using 25 sounds in total and tested them on a condition explained in 4.1. The subjects chose one that is similar to a

natural voice in every vowel sound.

5.2 Evaluation result

We used Nakaya variation of pair comparison method of Scheffe (five-grade) again. When sound B was at the same level as sound A, evaluation point is 0-point. We aggregated to grant 2-points from -2-points in each grade. We show rating average of /a/ at Table 2. In addition, Fig.4 shows the one-dimensional standard based on the count result. We found that there was little influence by phoneme characteristics in any place other than LPC/a/: A from Table.2.

In addition, utterance of E was judged as the most natural voice in tendency as a whole. Here, we confirmed 95% confidence interval of E. There is a significant difference that E is dominant by 65% in sounds and there is no significant difference that E is inferior level. On the other hand, there is no significant difference by 35% in

Table 2. Experiment result
Rating average of /a/ at LPC residual wave voice.

	A	B	C	D	E
A		-0.11	-0.12	-0.065	0.115
B	0.11		0.065	0.02	0.055
C	0.12	-0.065		0.08	0.035
D	0.065	-0.02	-0.08		0.035
E	-0.115	-0.055	-0.035	-0.035	
sum	0.18	-0.25	-0.17	0	0.24

Fig. 4. Experiment result
One-dimensional standard at LPC residual wave.

remaining sounds. According to the above-mentioned results, we found that there is a meaningful impression difference in E statistically.

6. Conclusion

In this paper, we examined the sound source vibration which is similar to a natural voice for the smooth communication of laryngectomees. We made two pseudo vocal cords sound source of LPC residual wave and Rosenberg wave, and uttered using a vibration speaker. Then, we evaluated these sounds and on EL sound using a pair comparison of Scheffe. According to the results, an LPC residual wave is similar to a natural voice than others. In addition, there was no influence except LPC/a/:A when we investigated relations of LPC residual wave sound and difference of phonological. We found that LPC/o/:E was the most dominant in LPC residual wave sound source when we examined which one is similar to a natural sound.

However, we intend to find the cause of deteriorating the EL sound and make the sound more similar to a natural sound. Because of the sound is not equal to the natural voice.

Acknowledgment

This work was partially supported by a Grant-in-Aid for Science Research (No.24500664) from the Japan Society for the Promotion of Science.

References

- (1) Hitomi Nagasaki and Michiko Nakamura, : "Characteristics of Communication Perceived by Patients with Laryngectomy", Journal of Yamanashi Nursing, Vol. 6, No. 2, pp. 11-16, 2008
- (2) Dennis H.Klatt and Laura C.Klatt : "Analysis, synthesis, and perception of voice quality variations among female and male talkers", J.Acoust, Soc.Am.87, pp.820-857, 1990
- (3) Sadaoki Furui, Shigeo Tsuji, Kunihiro Asada, Yoshinori Sakai, Masao Nakagawa and Yoishi Muraoka : "Onsei joho syori(Speech Information Processing)", MORIKITA PUBLISHING, pp.26-29, 1998
- (4) Sato Makoto : "Tokeiteki kannokensaho(Statistical Sensuality Laboratory Procedure)", Nikkagiren(JUSE), pp.231-270, 1985