Evaluation of the Voice Quality of Tracheoesophageal and Esophageal Speech

Koichi OMORI, Kazuhiko SHOJI, Shuji FUJITA*, Hideyuki FUKUSHIMA, and Hisayoshi KOJIMA

INTRODUCTIOM

Esophageal speech is a good method for speech rehabilitation after total laryngectomy. However, there are remarkable differences among individuals during its learning process. Tracheoesophageal speech with voice prosthesis, first reported by Singer et al.[1], has been widely spread as a good method for speech rehabilitation. Using this method, good speech can be easily and surely acquired. What differences in the voice characteristics are there between these two speaking methods? Although tracheoesophageal speech is considered to be superior in duration and intensity of the voice to esophageal speech, there are few objective studies on the voice quality. In this study, we compared the voice quality of tracheoesophageal speech with esophageal speech by the acoustic and perceptual methods which we previously reported.[2,3]

SUBJECTS

The subjects consisted of 27 esophageal speakers (ES group; 25 males, 2 females) and 12 tracheoesophageal speakers using voice prosthesis (TE group; all males). Their ages ranged from 45 to 79 years in the ES group and from 49 to 71 years in the TE group (average age: 63 years old in both groups). The period of usage with each speaking method ranged from 5 to 221 months (81 months on the average) in the ES group, and from 6 to 24 months (14 months on the average) in the TE group. (Table. 1)

Table 1 Subjects

	ES group (n=27)	TE group (n=12)
Sex	Male 25 Female 2	Male 12
Age	45–79 (ave.63)	49–71 (ave. 63)
Period of speaking (m)	5–221 (ave. 81)	6–24 (ave. 14)

Koichi OMORI (大森孝一), Kazuhiko SHOJI (庄司和彦), Hideyuki FUKUSHIMA (福島英行), Hisayoshi KOJIMA (児嶋久剛): Dept. of Otolaryngology, Faculty of Medicine, Kyoto Univeristy
Shuji FUJITA (藤田修治): *Dept. of Otolaryngology, Osaka Red Cross Hospital

METHODS

The sustained vowel /a/ pronounced under easy phonation was recorded, and converted into digital signals with an antialiasing filter (5 KHz, 80 dB/Oct) and an AD converter (10 KHz, 12 bit). Three acoustic parameters and perceptual impression were examined as follows:

(1) Maximum phonation time (MPT)

The duration of the voice was measured visually on the CRT display of the personal computer and confirmed auditorily because the phonation time in the ES group was too short to measure manually.

(2) Fluctuation of intensity

Effective sound pressure (Pe) is defined as:

$$Pe = \sqrt{\frac{1}{N}\sum_{K=1}^{N} Ak^2} \qquad Ak\text{: amplitude} \qquad -[1]$$

$$N = 500 \text{ (50 msec)}$$

Flucturation of sound intensity (Ft) can be derived as:

$$Ft = 20 \log \frac{Pt}{Po} (dB) \qquad -[2]$$

Po: effective sound pressure at voice onset

Pt: effective sound pressure at time t

Ft was plotted every 20 msec from voice onset to 500 msec and the range of fluctuation ($F_{max} - F_{min}$) was used as the parameter to represent the fluctuation of intensity.[3] Fig. 1 shows a sample of intensity fluctuation. The time course of intensity was plotted.

(3) Extraction of the fundamental frequency

Spectral analysis by Fast Fourier Transform (204.8 msec, Hanning window) was performed and the fundamental frequency was extracted from the interval of its harmonics.[3]

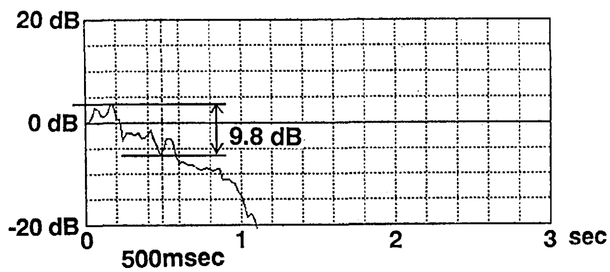

Fig. 1. The time course of intensity.
The range of flucturation was 9.8 dB in this case.

(4) Perceptual evaluation

The vowel sound /a/ reproduced from onset to 500 msec was used for perceptual evaluation. Initially, we ranked the perceptual impression of esophageal voices using the paired comparison method— scored by 3 otolaryngologists and 2 speech therapists. We then ranked the tracheoesophageal voices using the rankging of the esophageal voices as a standard.

RESULTS

The average 'maximum phonation time' was 1.2 sec in the ES group and 14.8 sec in the TE group. There was a significant difference between the two groups.

Fluctuation of intensity ranged stably within 11 dB in the TE group. On the other hand, the intensity fluctuated widely in some cases of the ES group (Fig. 2). However, there were no significant differences between the two groups. (Wilcoxon's test).

We divided the subjects into two groups depending on whether or not the fundamental frequency and its harmonics could be extracted. Fig. 3(A) shows an example with the fundamental frequency clearly detected, i.e. with periodical vibration. Fig. 3(B) shows a case where the fundmental frequency was not detected. The fundamental frequency could be extracted in 11 of 12 cases in the TE group, and in 18 of 27 cases in the ES group (Table. 2). There were no significant differences between the two groups (λ-square test). In the cases with the fundamental frequency successfully extracted, frequencies ranged from 78 Hz to 249 Hz (mean: 129 Hz) in the ES group, and from 73 Hz to 244 Hz (mean: 113 Hz) in the TE group.

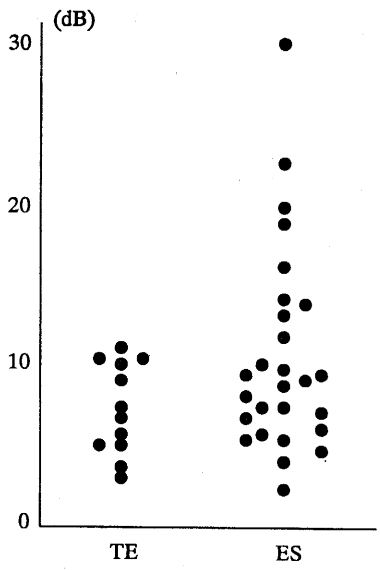

Fig. 2. Distribution of intensity fluctuation.

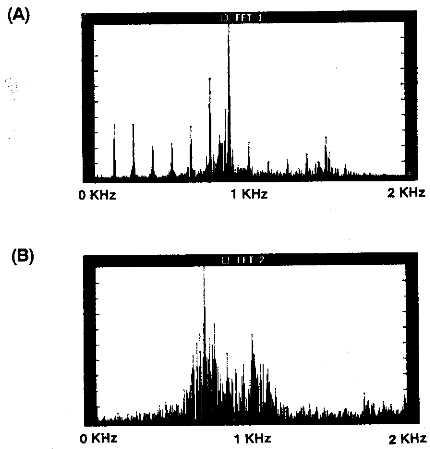

Fig. 3. Spectral analysis.

Table 2 Extraction of the fundamental frequency

	Extracted	Not extracted
TE group (cases)	11	1
ES group (cases)	18	9

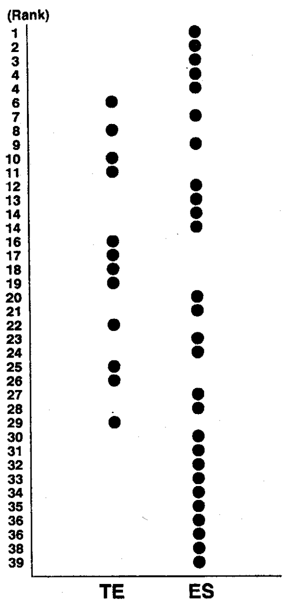

Fig. 4. Perceptual evaluation ranking.

There were no significant differences in the ranking of perceptual evaluation between the ES and the TE groups (Wilcoxon's test). Despite this, more lower ranked voices were found in the ES group. (Fig. 4)

Fig. 5 shows the relationship between fluctuation of intensity and perceptual

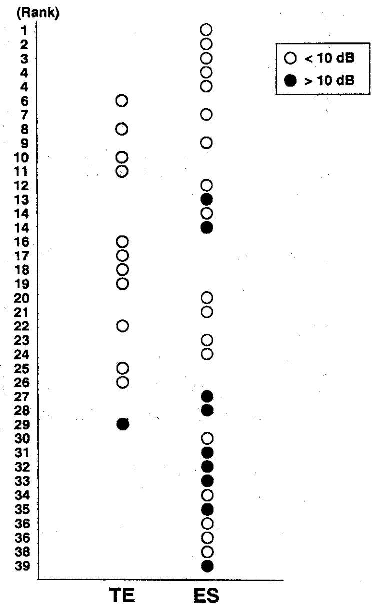

Fig. 5. Perceptual evaluation and fluctuation of intensity. Open circles represent less than 10 dB of intensity fluctuation and closed circles represent more than 10 dB.

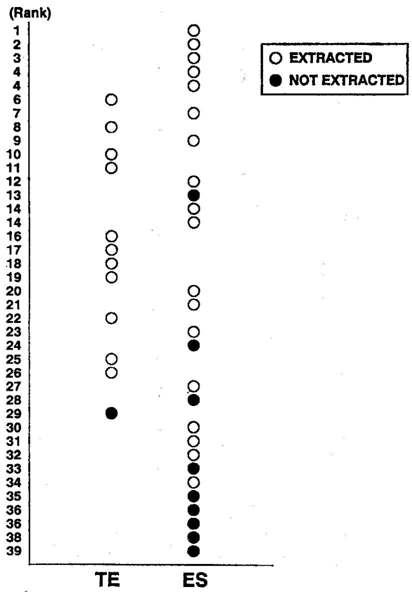

Fig. 6. Perceptual evaluation and extraction of fundamental frequency.
Open circles indicate cases with fundamental frequency extracted, and closed circles indicate cases not extracted.

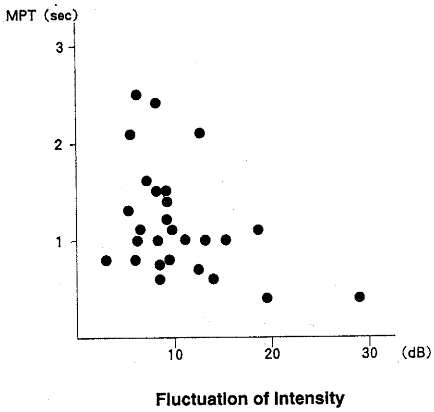

Fig. 7. Fluctuation of intensity and MPT (esophageal voice).

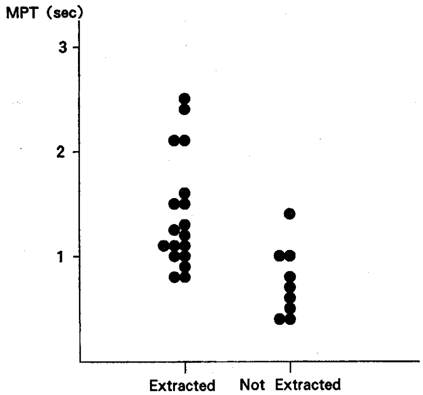

Fig. 8. Extraction of the fundamental frequency and MPT (esophageal voice).

ranking in both groups. Lower ranked cases had a tendency to show a larger fluctuation of intensity; these cases were found mostly in the ES group. Fig. 6 shows the relationship between perceptual ranking and periodicity of vibration in both groups. Lower ranked voices without periodical vibration were found in the ES group.

Fig. 7 shows the relationship between maximum phonation time and fluctuation of intensity in the ES group. Some cases with a short maximum phonation time of less than 1.5 seconds had a large intensity fluctuation. Maximum phonation time and the existence of periodical vibration in the ES group were plotted in Fig. 8. Periodical vibration was not observed in cases with a short maximum phonation time of less than 1.5 seconds.

Discussion

In esophageal speech, a small amount of air injected into the esophagus drives the mucosal membrane of the pharyngoesophageal (PE) segment. The skill of this speech depends mainly on air injection. Although esophageal speech is considered to be the best method for laryngectomized patients, some patients give it up because of the difficulty of air intake. In tracheoesophageal speech, sufficient expiratory air flow drives the mucosal membrane of the PE segment and causes good conversational ability. These two speaking methods have the same kind of vibrating portion, i.e. the PE segment, but a different driving force.

Voice anaysis shows that the maximum phonation time was significantly longer in the TE group. The distribution of the voices with large fluctuations of intensity, non-periodical vibrations and lower perceptual ranks were localized in the ES group, although the differences between the two groups were not significant. The fact that cases of lower perceptual ranking that tended to have larger fluctuation and no periodical vibration were found more often in the ES group indicates that lower ranked esophageal voices seem to have poor vibrating status. In these cases, the insufficient driving force had a bad effect on the vibrating portion. Some cases of esophageal voice with a short maximum phonation time had large intensity fluctuation and no periodical vibration. These results suggest that short maximum phonation time, i.e. insufficient driving force, adversely affected on the vibrating status of the esophageal voice.

Several authors[4-7] reported that tracheoesophageal speech was superior to esophageal speech acoustically and perceptually and Sedory et al[8]. reported that there were no significant differences in listener preference between tracheoesophageal speech and excellent esophageal speech. Our investigations also suggest that the voice quality of the TE group seemed to be superior when compared to the lower ranked cases of the ES group, even though there were no significant differences in the voice quality between the two groups on the whole. Therefore, laryngectomized patients using tracheoesophageal speech could be able to acquire good voice quality comparable to that of skillful esophageal speakers.

References

1. Singer, MI and Blom, ED: An endoscopic technique for restoration of voice after laryngectomy. Ann Otol Rhinol Laryngol 89: 529–533, 1980.
2. Omori, K, Shoji, K, Fukushima, H, Kojima, H: Auditory impression of esophageal voice. Practica Otologica (Kyoto) Vol 82: 967–970, 1989.
3. Omori, K, Shoji, K, Fujita, S, Fukushima, H, Kojima, H: Objective evaluation of esophageal voice quality. Practica Otologica (Kyoto) Vol 83: 1423–1427, 1990.
4. Robbins, J, Fisher, HB, Blom, EC, Singer, MI: A comparative acoustic study of normal esophageal and tracheoesophageal speech production. J Speech Hearing Disorders Vol 49: 202–210, 1984.
5. Schutte, HK, Bors, EFM, de Boer, GHA, Nieboer GLJ, Annyas, AA.: Evaluation of speech with and without a "Groningen Type" voice button. Speech Restoration Via Voice Prostheses (ed by Herrmann IF). pp 135–138, Springer-Verlag, Berlin, Heiderberg, 1986.
6. Williams, SE and Watson, JB: Speaking proficiency variations according to method of alaryngeal voicing. Laryngoscope Vol 97: 737–739, 1987.
7. Nieboer, GLJ, de Graaf, T, Schutte, HK: Esophageal voice quality judgements by means of semantic differential. J Phonetics Vol 16: 417–436, 1988.
8. Sedory, SE, Hamlet, SL, Connor, NP: Comparisons of perceptual and acoustic characteristics of tracheoesophageal and excellent esophageal speech. J Speech Hearing Disorders Vol 54: 209–214, 1989.