Acoustic Characteristics Analysis on the Tracheoesophageal Speech

Gang Lv (Corresponding author)

School of Electronic Information, Soochow University, Suzhou 215006, China

Tel: 86-512-6219-0898   E-mail: lvgang@suda.edu.cn

Abstract

Tracheoesophageal (TE) speech has been the current preferred treatment for speech rehabilitation after total laryngectomy, but due to the lack of objective evaluation criteria, it is tough to operate a qualitative analysis on the effects of postoperative speech rehabilitation. The present paper advanced an objective evaluation system involving acoustic characteristics parameters such as fundamental frequency, formants, jitter, shimmer, harmonics to noise ratio(HNR), the maximum phonation time(MPT), sound intensity and so on. The result showed that compared to subjective evaluation system, the application of objective evaluation system evaluated the effects of speech rehabilitation after total laryngectomy more objectively and precisely.

Keywords: Tracheoesophageal speech, Acoustic characteristics, Objective evaluation

1. Introduction

Tracheoesophageal speech is the voice resulted from the facts that through the establishment of channels between the trachea and esophagus, air exhaled by lungs entered the esophagus by way of fistula, vibrated esophageal mucosa, and transformed into sounds by way of sympathetic chord cavity. Tracheoesophageal speech has been the current preferred treatment for speech rehabilitation after total laryngectomy (Culton, 1998, PP. 458-463). In order to evaluate the effects of postoperative speech rehabilitation, parameters such as fluency, intelligibility, accuracy, hearing distance and so on were applied to assay TE speech in the past, the results were subjective with the characteristics such as hoarse voice, rough sound quality and poor understandability (Most, 2000, PP. 165-181; Van As, 1998, PP. 239-248), but due to the unified standards of subjective assessments, in addition to the subjective instability of conners, experimental test results were often unable to exert accurate comparative assessment. For example, TE speech usually was classified into three levels according to the hoarse degree, viz. good, middle, poor, but as for as which degree was good or poor was concerned, different judgers might have different subjective standards, and they might have different results on the same data. Consequently, the establishment of a set of objective evaluation system for assaying the effects of speech rehabilitation after total laryngectomy was absolutely necessary.

2. Methods

2.1 Research object

TE sound observation group consisted of 17 patients with the treatment of speech rehabilitation after total laryngectomy. Among them, there were 9 males and 8 females, with the average age of 64.7. The control group was composed of 20 healthy people with normal vocal cord function after laryngoscopy among which there were 10 males and 10 females with the average age of 63.3.

2.2 Research process

Two groups were tested in the environment with the yawp less than 45dB, taking comfortable sitting position, and sent out sustained and stable vowel "a" and "i". The sound source signal was recorded into the computer by means of desktop microphone, with the sampling frequency of 8kHz, pre-emphasis coefficient of 0.975, Hamming window of 256 samples per frame, frame shift of 1/4 frame length. Acoustic parameters of sample data such as fundamental frequency, formants, jitter, shimmer, harmonics to noise ratio, the maximum phonation time, sound intensity and so on were adopted by the software matlab7.0, and the compared results of the parameters were processed by Excel database. Finally, by means of SPSS11.0 statistical software, the difference significances between group data were compared by t Test (small samples).

3. Results

3.1 Fundamental frequency and formants

The fundamental frequency and formants of vowel "a" were listed in Table 1 in detail, while that of vowel "i" in Table 2. Among them, the fundamental frequency fluctuation scope of normal group ranged as follows: vowel "a", from 113.41 to 283.47Hz; vowel "i", from 122.84 to 310.58Hz. The fundamental frequency fluctuation scope of TE group ranged as follows: vowel "a", from 78.72 to 203.34Hz; vowel "i", from 81.97 to 212.44Hz. The formant fluctuation scope of normal group ranged as follows: vowel "a", F1: from793.34 to 1124.51Hz, F2: from 1206.14 to 1642.38Hz, F3: from 2243.54 to 3087.08Hz; vowel "i", F1: from 254.43 to 385.17Hz, F2: from 2376.37 to 2754.55Hz, F3: 2817.76 to 3341.18Hz. The formants fluctuation scope of TE group ranged as follows: vowel "a", F1: from 632.28 to 1128.44Hz, F2: from 1130.11 to 1672.53Hz, F3: from 2217.33 to 3163.27Hz; vowel "i", F1: from 226.14 to 391.85Hz, F2: from 2243.14 to 2672.64Hz, F3: from 2693.14 to 3276.71Hz.

As seen from Table 1 and 2, there were differences in vowel "a" and "i" fundamental frequencies between normal and TE groups (P<0.05), while there were no significant differences in two vowels formants (P>0.05).

3.2 Jitter, shimmer and harmonics to noise ratio

The jitter, shimmer and harmonics to noise ratio of vowel "a" and "i" were respectively listed in Table 3 and 4 in detail. Among them, the jitter fluctuation scope of normal group ranged as follows: vowel "a", from 0.17% to 3.46%; vowel "i", from 0.13% to 3.76%. The jitter fluctuation scope of TE group ranged as follows: vowel "a", from 1.12% to 8.76%; vowel "i", from 1.45% to 7.68%. The shimmer fluctuation scope of normal group ranged as follows: vowel "a", 0.86dB to 4.11 dB; vowel "i", from 0.91dB to 4.23dB. The shimmer fluctuation scope of TE group ranged as follows: vowel "a", 2.84dB to 9.31dB; vowel "i", from 2.36dB to 9.47dB. The harmonics to noise ratio fluctuation scope of normal group ranged as follows: vowel "a", from 18.33dB to 26.62dB; vowel "i", from 20.47dB to 29.56dB. The harmonics to noise ratio fluctuation scope of TE group ranged as follows: vowel "a", from 8.14dB to 12.46dB; vowel "i", from 9.23dB to 14.74dB.

As seen from Table 3 and 4, there were significance differences in vowel "a" and "i" jitter, shimmer and harmonics to noise ratios between normal and TE groups (P<0.05).

3.3 The maximum phonation time and sound intensity

The maximum phonation time and sound intensity of vowel "a" and "i" were respectively listed in Figure 1and 2 in detail. MPT fluctuation scope of normal group ranged as follows: vowel "a", from 15s to 23s; vowel "i", from 12s to 24s. MPT fluctuation scope of TE group ranged as follows: vowel "a", from 4s to 14s; vowel "i", from 5s to 13s. Sound intensity fluctuation scope of normal group ranged as follows: vowel "a", from 71dB to 93dB; vowel "i", from 68dB to 86dB. Sound intensity fluctuation scope of TE group ranged as follows: vowel "a", from 67dB to 84dB; vowel "i", from 70dB to 84dB. Small sample t Test results showed that MPT of normal group was higher than that of TE group, and there was no significant difference in sound intensity between two groups.

4. Discussions

Fundamental frequency was the lowest natural frequency of sound source periodical vibration, and reflected the regularity of glottal vibration. Since the patients with total laryngectomy used pharyngoesophageal mucosa as new "glottis" to send out TE speech, the new "glottis" was bloated with great quality, and the vibration frequency was lower than that of normal people. The observations proved that. Moreover, Xiao et al (2004, PP. 530-535) had checked 20 cases of patients with TE speech by video laryngoscope and found that there was kind of regularity in the vibration of pharyngoesophageal mucosa, which was the reason why there was no significant difference between TE and normal speech (p = 0.47).

Formants were the resonance generated by glottis and sound source after enunciation. Hou et al (2002, PP. 16-18) speculated that formants frequency were associated with tongue position, among which the first formant frequency was negatively correlated with the superioinferior position of tongue, and the second formant frequency was correlated with anteroposterior of tongue. As the surgery was to restore the normal airflow trend of lung at most, and failed to change the organs and tissues, such as tongue, maxilla, and lip and so on, the distribution of TE speech formants should be consistent with normal speech. Jin et al (2001, PP. 291-294) have assayed the formants frequency and energy in 26 cases of tracheoesophageal speech and 32 cases of normal speech, and results indicated that there was no significant difference among distribution groups of formant frequency and energy, which conformed to the present study.

Jitter referred to the change rate of fundamental frequency, shimmer referred to the change rate of sonic wave

amplitude, and they were all applied as short-term indices to reflect the stability of pronunciation system. As for normal speech, vocal cords regularly open, closed and vibrated by the coordination function of air force and nerve muscle of larynx, and acoustic characteristics of speech was relatively stable. As for TE speechs of the patients, due to the fleshy new "glottis" and scar tissue imposed by the post-operative fistula, the closure degree of "glottis" was poor, muscular control was weak, and thus there were obvious frequency perturbation and amplitude changes. After all these were reflected to the subjective evaluation system, rough tone of TE speech was resultant.

The harmonics to noise ratio was the ratios of periodic signal generated by vocal cord to irregular signal such as noises produced by vocal cord and sound channel, and was applied as an objective acoustic characteristics indicators to suggest the hoarse degree of speech. In general, higher harmonics to noise ratio showed that speech was clearer, and vice versa. As the patients with total laryngectomy lost normal vocal cord vibration, the harmonics to noise ration was significantly higher than that of normal speech.

The maximum phonation time was the longest pronunciation time of patients after deep breath as possible as they could, and was applied as acoustic characteristics parameters to suggest pronunciation capacity. The results showed that TE speech had stronger sound intensity and longer pronunciation time, which was primarily, attributed to the fact that aerodynamic force that produced TE speech came from the lung after diversion, and could engender enough vibration source air to promote esophageal mucosa to constitute new "glottis" and cause vibration sound.

Taken together, we concluded as follows: As for TE speech, the fundamental frequency was low, formant frequency and energy distribution were close to normal speech, and there were typical jitter and shimmer with high harmonics to noise ratio. These acoustic characteristics parameters were relative to the conclusions from subjective evaluation system, namely that as for TE speech, the tone was tough, sound intensity was strong, and it could pronounce continuously. Different from subjective evaluation system, these objective acoustic characteristics parameters could be treated with quantitative analysis. The application of objective evaluation system could evaluate the effects of speech rehabilitation of different patients after total laryngectomy and the same patient during different periods, and these objective evaluation results would be of important guiding significance for the practice of clinical vocal rehabilitation.

References

Culton G.L., Gerwin J.M. (1998). Current trends in laryngeetomy rehibilitation: A survey of speeeh-language Pathologists. *Otolaryngology-Head and Neck Surgery*, No.118 (4):458-463.

Hou L.Z., Han D.M., Xu W., Zhang.L. (2002). Study on Vowel Selection in Voice Acoustic Measurements. *Journal of Audiology and Speech Pathology*, No.10:16-18.

Jin G.W., Chen J., Li S.C. (2001). Acoustic analyses of vocal frequency spectrum after pronounce reconstruction. *Railway Medical Journal*, No.29 (5): 291-294.

Most T., Tobin Y., Mimran R.C. (2000). Acoustic and perceptual characteristics of esophageal and tracheoesophageal speech production. *Journal of Communication Disorders*, No.33 (2):165-181.

Van As C.J., Hilgers F.J., Verdonck-de Leeuw I.M., Koopmans-van Beinum F.J. (1998). Acoustical analysis and perceptual evaluation of tracheoesophageal prosthetic voice. *Journal of Voice*, No.12 (2): 239-248.

Xiao S.F., Li X.M., Shang Y.D., Li Q., He Y.F. (2004). Phonation assessment of voice restoration after laryngectomy. *Journal of clinical otorhinolaryngology*, No.18 (9):530-535.

Table 1. The comparison of vowel "a" fundamental frequency and formant between normal and TE groups

	Sample number	F0(Hz)		F1(Hz)		F2(Hz)		F3(Hz)	
		ϖ	σ	ϖ	σ	ϖ	σ	ϖ	σ
Normal group	20	176.32	50.47	950.63	130.11	1414.17	147.52	2712.36	274.52
TE groups	17	133.52	51.34	863.78	200.58	1384.62	217.38	2637.38	300.55
P value		<0.05		>0.05		>0.05		>0.05	

Table 2. The comparison of vowel "i" fundamental frequency and formant between normal and TE groups

	Sample number	F0(Hz)		F1(Hz)		F2(Hz)		F3(Hz)	
		ϖ	σ	ϖ	σ	ϖ	σ	ϖ	σ
Normal group	20	192.81	46.19	313.45	47.19	2569.74	131.42	3087.68	186.56
TE groups	17	142.63	53.54	297.87	63.18	2436.52	172.67	2917.17	196.30
P value		<0.05		>0.05		>0.05		>0.05	

Table 3. The comparison of vowel "a" jitter, shimmer and harmonics to noise ratio between normal and TE groups

	Sample number	Jitter (%)		Shimmer(dB)		HNR(dB)	
		ϖ	σ	ϖ	σ	ϖ	σ
Normal group	20	1.77	1.43	2.42	1.48	22.36	2.84
TE groups	17	4.96	3.32	5.96	3.05	10.81	1.54
P value		<0.05		<0.05		<0.05	

Table 4. The comparison of vowel "i" jitter, shimmer and harmonics to noise ratio between normal and TE groups

	Sample number	Jitter (%)		Shimmer(dB)		HNR(dB)	
		ϖ	σ	ϖ	σ	ϖ	σ
Normal group	20	1.32	1.47	2.41	1.33	24.38	3.92
TE groups	17	4.43	2.74	5.88	3.14	11.77	1.42
P value		<0.05		<0.05		<0.05	
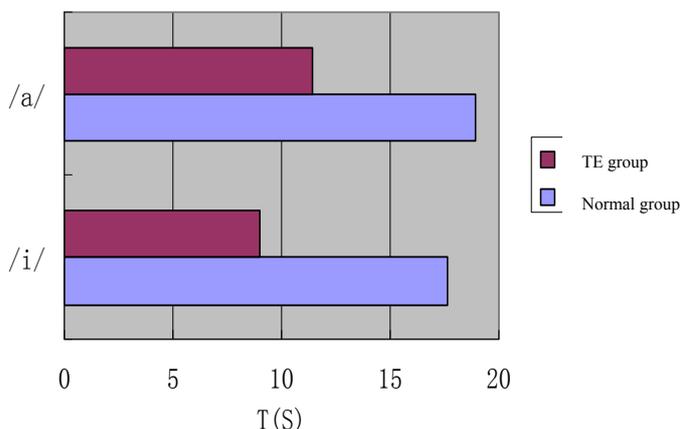
Figure 1. The comparison of MPT between normal group and TE group
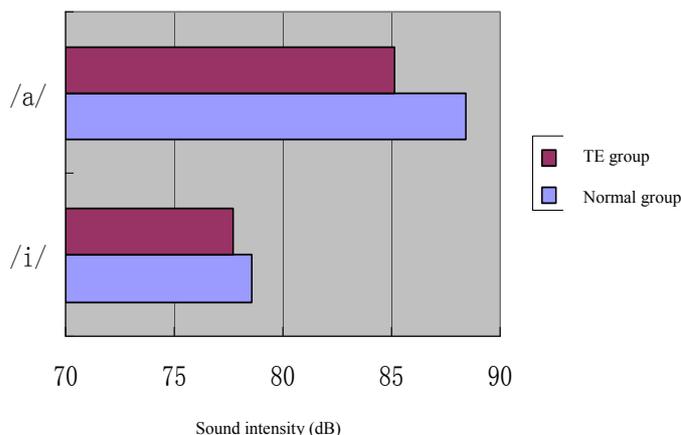
Figure 2. The comparison of sound intensity between normal group and TE group