Intensity and fundamental frequency control in tracheoesophageal voice

Controllo dell'intensità e della frequenza fondamentale nella voce tracheoesofagea

A. SCHINDLER, A. CANALE[1], A.L. CAVALOT[2], R. ALBERA[2], P. CAPACCIO, F. OTTAVIANI, O. SCHINDLER[1]
Otorhinolaryngology Clinic IV, University of Milan; [1] Audiology-Phoniatrics Clinic, University of Turin;
[2] Otorhinolaryngology Unit, University of Turin, Italy

Key words

Total laryngectomy • Tracheo-oesophageal voice • Intensity • Fundamental frequency

Parole chiave

Laringectomia totale • Voce tracheo-esofagea • Intensità • Frequenza fondamentale

Summary

Tracheo-oesophageal voice prostheses are currently widely used following total laryngectomy. Data on maximum phonation time and spectrum have been studied by various Authors and are well known. On the contrary, intensity and fundamental frequency control have received little attention. Intensity and fundamental frequency play an important role in the prosodic aspects of speech. Fundamental frequency variations have been studied in tone language speakers, but the ability to voluntarily change intensity and fundamental frequency remain to be fully investigated. Aim of the present study was to analyse the ability of tracheo-oesophageal voice users to change intensity and fundamental frequency. A total of 12 male subjects who underwent total laryngectomy, in whom a tracheo-oesophageal prosthesis had been inserted, were considered. Maximum phonation time was calculated. Each subject was asked to utter an /a/ as loud as possible and an /a/ as soft as possible. Each subject was then asked to utter an /a/ at comfortable pitch and then at an interval of a fifth. Intensity as well as fundamental frequency variations were compared using Wilcoxon signed rank test. Correlation between maximum phonation time and variation in intensity and in fundamental frequency as well as between the two latter variables was calculated using Spearman's rank correlation coefficient. Mean maximum phonation time was 8 (± 3.8) sec. Mean energy was 50 (± 4.8) dB SPL for soft phonation and 68 (± 4.7) dB SPL for loud phonation. The difference observed was statistically significant ($p < 0.02$). Mean fundamental frequency values were 106 (± 14) Hz and 135 (± 34) Hz at the interval of a fifth. The difference observed was statistically significant ($p < 0.02$). Tracheo-oesophageal voice users were able to change intensity and fundamental frequency, but their control was rather poor. Variations in intensity, as well as fundamental frequency, did not show any correlation with maximum phonation time, and were not correlated with each other. In conclusion, the tracheo-oesophageal voice allows small fundamental frequency variations, but their control appears difficult. On the contrary, intensity variations appear larger and control somewhat easier.

Riassunto

Le protesi fonatorie dopo laringectomia totale sono largamente utilizzate al giorno d'oggi. I dati sul massimo tempo di fonazione e sullo spettro sono stati studiati da diversi Autori e sono ben noti. Al contrario il controllo dell'intensità e della frequenza fondamentale (Fo) non hanno ricevuto molta attenzione. L'intensità e la Fo hanno un ruolo importante negli aspetti prosodici della parola. Le variazioni della Fo sono state studiate nella lingua a toni, ma l'abilità di cambiare volontariamente intensità e Fo sono stati finora raramente oggetto di indagine. Lo scopo del lavoro è analizzare l'abilità di chi utilizza la voce tracheo-esofagea nel cambiare intensità e Fo. Dodici soggetti di sesso maschile sottoposti a laringectomia totale, in cui è stata inserita una valvola fonatoria, sono stati presi in considerazione. È stato calcolato il massimo tempo di fonazione (MPT). A ogni soggetto è stato chiesto di produrre una /a/ quanto più forte possibile e una /a/ quanto più piano possibile. Ad ogni soggetto è stato successivamente chiesto di produrre una /a/ ad altezza comoda e ad un intervallo di una quinta. Le variazioni di intensità e di Fo sono state confrontate con il test dei ranghi di Wilcoxon. La correlazione fra MPT e le variazioni in intensità e in Fo così come fra queste due ultime variabili è stata calcolata attraverso il coefficiente dei ranghi di Spearman. Il MPT medio era di 8 (± 3,8) sec. L'intensità media era di 50 (± 4,8) dB SPL per la fonazione a basso volume e di 68 (± 4,7) dB SPL per la fonazione ad alto volume. La differenza osservata era statisticamente significativa ($p < 0,02$). I valori medi della Fo media erano di 106 (± 14) Hz e 135 (± 34) Hz all'intervallo di una quinta. La differenza osservata era statisticamente significativa ($p < 0,02$). I soggetti con voce tracheoesofagea erano in grado di cambiare intensità e Fo, ma il loro controllo era abbastanza ristretto. Le variazioni in intensità così come in Fo non hanno mostrato alcuna correlazione con il MPT, né fra di loro. In conclusione, la voce tracheoesofagea consente piccole variazioni in Fo, ma il loro controllo appare difficoltoso. Al contrario le variazioni in intensità appaiono di grado maggiore e il controllo più semplice.

Introduction

There are three major methods of communication used by patients after undergoing total laryngectomy: artificial larynx, oesophageal speech and tracheo-oesophageal (TE) speech. Oesophageal speech was the mainstay of alaryngeal communication until the early 1980s, but a large number of people are unable to use this technique adequately [1]. In TE speech, a small device made of silicone is inserted into a surgically created opening between the trachea and oesophagus to allow air from the lungs to reach the pharyngo-oesophageal segment. The TE puncture technique for voice restoration after total laryngectomy provides an excellent rehabilitative option for the laryngeal patient; since its introduction in 1979, it has become widely accepted [2]. Maximum phonation time (MPT) in the TE speaker has been studied by various Authors [3-6]: data ranges from 6.5 sec to 13 sec. Also the spectrographic characteristics of TE speech have been studied extensively [5-10]; a more stable fundamental frequency (Fo) and more clearly defined harmonics compared to oesophageal voice users have been shown. Data on the characteristics of neoglottic vibration have been obtained starting from stroboscopic images and source waveform analysis [10-12]. The characteristics of the signals appeared highly variable; the normal excitation source could be described as a complex, quasi periodic, triangular shaped signal.

Intensity and Fo control of speech are used not only in singing but in everyday communication for the prosodic aspect of verbal communication. Prosody provides the melodic contour and rhythm of speech and facilitates in decoding syntactic and lexical meaning as well as the emotional content of the spoken language. It helps speakers convey intended meanings and it helps listeners understand them. Prosody consists of variation between two parameters that are defined perceptually as pitch and amplitude, acoustically as Fo and intensity. Deference, formality, humour and irony are just a few examples of the attitudes conveyed through prosody [13]. Therefore, prosody plays a key role in the pragmatic competence of verbal communication. It has been stated that pragmatic competence is at least as important as language competence in verbal communication; in fact, it is estimated that more than half of what we say is not literally what we mean [14].

Prosodic functions in TE speakers have not been studied so far. The aim of the study is to analyse the ability of TE voice users to voluntarily control the intensity and Fo of their voices.

Material and methods

A total of 12 male subjects, ranging in age from 63 to 81 years (mean 70), took part in the study. All subjects were provided with a voice prosthesis and had additional voice therapy. Age of subjects, year of laryngectomy surgery and of TE puncture are shown in Table I. The criteria for selection for the study were: tracheo-oesophageal fistulization and fitting of a Blom-Singer prosthesis dating back at least one year and no evidence of disease. Spectrographic analysis showed that TE voices may differ enormously one from another: voices with a clear harmonic structure exist, together with voices with evident formant structure or with an almost chaotic structure [10]. Voice showing a clear harmonic structure was another inclusion criterion.

MPT was calculated for each subject. The MPT was determined by measuring the longest sustained /a/ in

Table I. Age, year of laryngectomy surgery, year of tracheo-oesophageal puncture and type of voice prosthesis in study population.

Subject	Year of birth	Year of surgery	Year of TE puncture
1	1921	1973	1998
2	1937	1994	1996
3	1935	1999	2000
4	1940	1999	2000
5	1925	1992	1995
6	1939	1994	1995
7	1925	1994	1998
8	1939	1992	1995
9	1928	1994	1996
10	1936	1997	1998
11	1931	1995	1995
12	1937	1994	1998

three productions on the basis of the oscillogram signal [15]. Each subject was then asked to utter an /a/ as loud as possible and an /a/ as soft as possible; finally all subjects had to utter an /a/ first at a comfortable pitch, then at an interval of one fifth. Acoustic analyses were performed with the Computerized Speech Lab (CSL) programme with a 4300 B external module from Kay Elemetrics Corporation. All voices were recorded using a professional electrodynamic microphone AKG D-90 S positioned approximately 30 cm from the mouth and slightly below the chin to reduce the effect of airflow. All speech samples were recorded with a sample frequency of 20,000 Hz and directly stored in the computer. A spectrography of each production at FFT-1024 points ranging between 0 and 2 kHz was performed. Mean energy in dB SPL was determined for the loudest and softest /a/ production. The vibration rate of the (pseudo-) voice source – Fo – based on the spectrogram, was determined for the /a/ production at a comfortable pitch level and at the interval of one fifth. The study was carried out according to the Declaration of Helsinki. Intensity as well as Fo variations were compared using Wilcoxon signed rank test. Finally, correlations between MPT and variations in intensity and in Fo as well as between the two latter variables were calculated through Spearman's rank correlation coefficient.

Results

Data on MPT, intensity of soft and loud phonation as well as on Fo at a comfortable pitch and at an interval of one fifth are shown in Table II. Mean MPT was 8 (± 3.8) sec. Mean energy was 50 (± 4.8) dB SPL for soft phonation and 68 (± 4.7) dB SPL for loud phonation. The difference observed was statistically significant ($p < 0.02$). Spectrographic analysis showed that pitch control is possible even if restricted (Fig. 1). Mean Fo values were 106 (± 14) Hz and 135 (± 34) Hz at the interval of one fifth. The difference observed was statistically significant ($p < 0.02$). Even if an evident difference was observed in all subjects between soft and loud phonation, Fo variations were restricted to < 10 Hz, in three subjects.
Correlation between MPT and intensity ($r = - 0.21$; $p = 2.0$) or Fo variation ($r = 0.20$; $p = 2.0$) did not reach statistical significance. No correlation was found between intensity and Fo variations ($r = -0.06$; $p = 2.0$).

Discussion and conclusion

Control of two vocal parameters – intensity and Fo – have been investigated in 12 TE voice speakers. So far, little attention has been focused on intensity and Fo control in the TE voice. Our subjects were able to change the Fo of their voice, even if they managed to reach, in the experimental conditions, an interval of about one third. Since the Fo variations, in the laryngeal speaking voice of an average adult speaker cover an interval of about one fifth, the difference measured seems to allow an almost normal prosody. Max et al. [18] observed a wide range between the highest and lowest voice by, confirming that Fo control is possible for TE voice users. Fo control in alaryngeal speakers has been studied in tone language speaking people [16 17]. In tone languages, where Fo characteristics are important to semantic distinction, word

Table II. Data on MPT, soft and loud intensity, Fo at comfortable pitch (Fo low) and at an interval of one fifth (Fo high) from the twelve subjects studied.

Subject	MPT (sec)	Soft intensity	Loud intensity	Low Fo	High Fo
1	3.39	51	63	100	134
2	3.75	50	72	136	204
3	4.35	55	68	104	138
4	4.15	50	75	102	136
5	7	49	74	130	135
6	7.28	45	71	114	118
7	8.63	49	64	88	112
8	8.93	54	68	98	104
9	9.98	45	64	90	124
10	12.27	62	71	95	105
11	12.8	49	62	108	110
12	14.66	46	63	112	204

MPT is expressed in seconds, intensity in dB SPL, Fo in Hz.

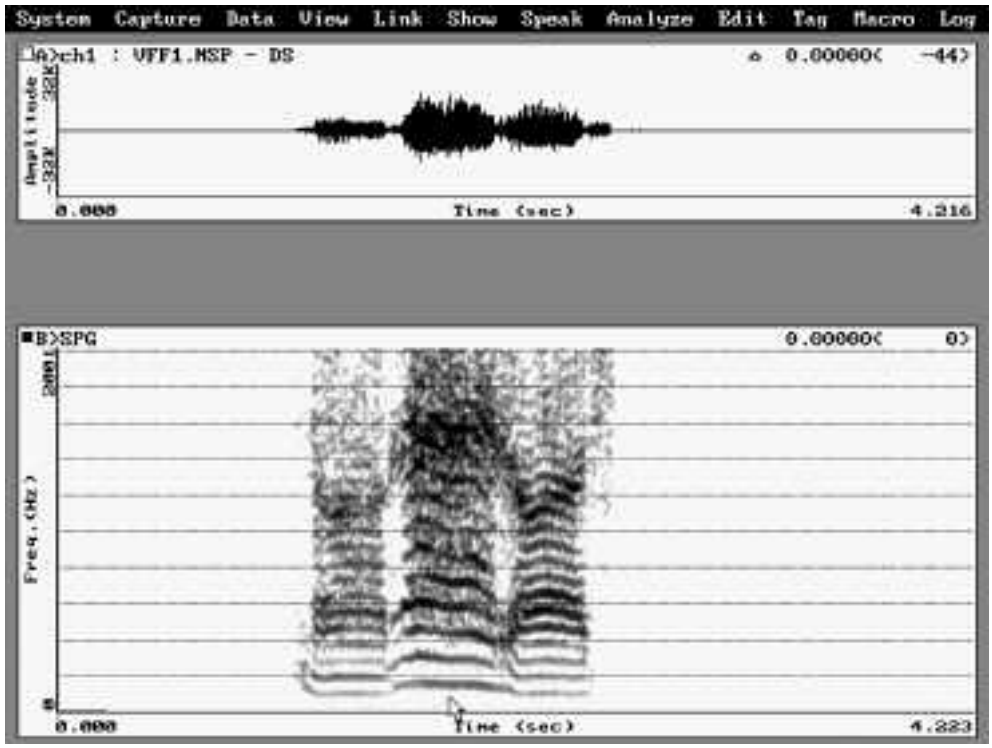

Fig. 1. Spectrography of the /a/ production at comfortable pitch level, at the interval of one fifth, and again at comfortable pitch level. Sample frequency is 20,000 Hz. Spectrography is at FFT-1024 points, ranging between 0 and 2 kHz. A clear variation of the first harmonic and of the whole spectrum is visible. A: oscillogram; B: spectrography.

meaning is distinguished only by pitch level patterns. Fo contours appeared to be the acoustic cues associated with the signal that mainly contributed to the perception of meaning, allowing satisfactory speech in tone language. Fo variations in alaryngeal voice of tone language speakers represent an important example of the role of Fo control in everyday communication for TE voice speakers.

MPT, in our subjects, is in keeping with reports in the literature [3-6]. No correlation was found between MPT and Fo variations, suggesting that other variables may influence Fo control.

The difference observed in intensity was about 20 dB SPL, which is not very different from the variation measured in laryngeal speakers in a reading test. Max et al. [18] found similar results, confirming that a satisfactory intensity control is possible for TE voice speakers. Even for this parameter, no correlation was found with MPT. This finding is somewhat surprising, since both MPT and intensity are clearly related to respiratory control.

Deschler et al. [19] demonstrated that increased sound pressure level is a significant factor for the increase in Fo noted in alaryngeal speakers, even though changes in sound pressure level account for only a small portion of the modulation in pitch achieved in TE speakers. In our patients, no correlation was found between intensity and Fo variation, suggesting that control on intensity does not influence control on Fo. Moreover, Fo control was rather poor in some of the subjects; an external variable seems, therefore, to influence this ability.

The data obtained in the present investigation show that TE voice users are able to voluntarily change Fo, but their ability is restricted to a small interval. On the contrary, TE voice users showed a wider range in intensity control. The variables influencing the control in Fo and intensity of TE voice have not yet been clearly established. In particular, voice therapy does not seem to play a role [18]. Omori et al. [20] demonstrated anatomic alterations in the pharyngo-oesophageal segment of TE speakers during phonation. The contractions of the pharyngeal musculature surrounding the pharyngo-oesophageal segment served to narrow the segment and were simultaneous with voice production. Although some component of this muscle activity may be reflex mediated and not under volontary control, the ability to actively control contrac-

tion, at the level of the pharyngo-oesophageal segment, would provide alaryngeal speakers with a method of actively modulating Fo and intensity. Whether this ability could be trained is not yet known; to our knowledge, pitch and intensity control constitute only a small part of voice therapy, in TE voice speakers [21].

The subjects included in this study were "star patients", since a clear harmonic structure on sound spectrography was one of the inclusion criteria; therefore, the data observed can not be generalized to the general population of TE voice speakers, until a larger study is carried out. Control on the two main parameters of speech prosody appeared to be only slightly worse than that found in laryngeal speakers, thus confirming the positive value of TE speech in subjects who have undergone total laryngectomy.

References

[1] Simpson CB, Postma GN, Stone RE, Ossof RH. *Speech outcomes after laryngeal cancer management*. Otolaryngol Clin North Am 1997;30:189-205.

[2] Singer MI, Blom ED. *An endoscopic technique for restoration of voice after laryngectomy*. Ann Otol Rhinol Laryngol 1980;89:529-33.

[3] Robbins J, Fisher HB, Blom EC, Singer MI. *A comparative acoustic study of normal, esophageal, and tracheoesophageal speech production*. J Speech Hear Disord 1984;49:202-10.

[4] Max L, Steurs W, De Bruyn W. *Vocal capacities in esophageal and tracheoesophageal speakers*. Laryngoscope 1996;106:93-6.

[5] Van As CJ, Hilgers FJM, Verdonck-de Leeuw IM, Koopmans-van Beinum FJ. *Acoustical analysis and perceptual evaluation of tracheoesophageal prosthetic voice*. J Voice 1998;12:239-48.

[6] Capaccio P, Schindler A, Tassone G, Ottaviani F. *Comparative study of esophageal and tracheo-esophageal voice by digital spectrographic analysis*. Acta Phon Lat 2001;23:322-6.

[7] Loebe LP, Gall V, Pfau EM. *Comparison of phoniatric parameters in oesophageal and voice prosthetic speech.* Rev Laryngol 1987;108:125-7.

[8] Debruyne F, Delaere P, Wouters J, Uwents P. *Acoustic analysis of tracheo-oesophageal versus oesophageal speech*. J Laryngol Otol 1994;108:325-8.

[9] Bertino G, Bellomo A, Miani C, Ferrero F, Staffieri A. *Spectrographic differences between tracheal-esophageal and esophageal voice*. Folia Phoniatr Logop 1996;48:255-61.

[10] Piroddi P, Ottaviani F, Amorelli V, Cantarella G. *Instrumental evaluation of tracheoesophageal voice*. Otolaringol 1990;40:201-5.

[11] Omori K, Kojima H. *Neoglottic vibration in tracheoesophageal shunt phonation*. Eur Arch Otorhinolaryngol 1999;256:501-5.

[12] Qi Y, Weinberg B. *Characteristics of voicing source waveforms produced by esophageal and tracheoesophageal speakers*. J Speech Hear Res 1995;38:536-48.

[13] Myers PS. *Right hemisphere damage – disorders of communication and cognition*. San Diego: Singular Publishing Group 1999. p. 136-8.

[14] Paradis M. *The other side of language: pragmatic competence*. J Neurolinguist 1998;11:1-10.

[15] Hirano M. *Examination of phonatory ability*. In: Hirano M. *Clinical examination of voice*. New York: Springer-Verlag; 1981.

[16] Wong SHW, Cheung CCH, Yuen APW, Ho WKH, Wei WI. *Assessment of tracheoesophageal speech in a tonal language*. Arch Otolaryngol Head Neck Surg 1997;123:88-92.

[17] Ng ML, Gilbert HR, Lerman JW. *Fundamental frequency, intensity and vowel duration characteristics related to perception of cantonese alaryngeal speech*. Folia Phoniatr Logop 2001;53:36-47.

[18] Max L, Steurs W, de Bruyn W. *Vocal capacities in esophageal and tracheoesophageal speakers*. Laryngoscope 1996;106:93-6.

[19] Deschler DG, Doherty ET, Reed CG, Singer MI. *Effects of sound pressure levels on fundamental frequency in tracheoesophageal speakers*. Otolaryngol Head Neck Surg 1999;121:23-6.

[20] Omori K, Koikma H, Nonmura M. *Mechanism of tracheoesophageal shunt phonation*. Arch Otolaryngol Head Neck Surg 1994;120:648-52.

[21] Roth FP, Worthington CK. *Treatment resource manual for speech-language pathology. Second Edition*. Dekmar: Singular; 2001.

■ Received: April 3, 2004
Accepted: May 17, 2005

■ Address for correspondence: Prof. O. Schindler, Università di Torino, S.C.U. Audiologia-Foniatria, Azienda Ospedaliera "S. Giovanni Battista", Via Genova 3, 10126 Torino, Italy - E-mail: oskar.schindler@unito.it