Coll. Antropol. **36** (2012) Suppl. 2: 111–114
Original scientific paper

Objective Assessment of Tracheoesophageal and Esophageal Speech Using Acoustic Analysis of Voice

Ljiljana Širić[1], Dario Šoš[1], Marinela Rosso[1] and Siniša Stevanović[2]

[1] »J. J. Strossmayer« University, Osijek University Hospital Center, Department of Otorhinolaryngology and Head and Neck Surgery, Osijek, Croatia
[2] Virovitica General Hospital, Department of Otorhinolaryngology, Virovitica, Croatia

A B S T R A C T

The aim of this study was to analyze the voice quality of alaryngeal tracheoesophageal and esophageal speech, and to determine which of them is more similar to laryngeal voice production, and thus more acceptable as a rehabilitation method of laryngectomized persons. Objective voice evaluation was performed on a sample of 20 totally laryngectomized subjects of both sexes, average age 61.3 years. Subjects were divided into two groups: 10 (50%) respondents with built tracheoesophageal prosthesis and 10 (50%) who acquired esophageal speech. Testing included 6 variables: 5 parameters of acoustic analysis of voice and one parameter of aerodynamic measurements. The obtained data was statistically analyzed by analysis of variance. Analysis of the data showed a statistically significant difference between the two groups in the terms of intensity, fundamental frequency and maximum phonation time of vowel at a significance level of 5% and confidence interval of 95%. A statistically significant difference was not found between the values of jitter, shimmer, and harmonic-to-noise ratio between tracheoesophageal and esophageal voice. There is no ideal method of rehabilitation and every one of them requires an individual approach to the patient, but the results shows the advantages of rehabilitation by means of installing voice prosthesis.

***Key words**: objective assessment, acoustic analysis, tracheoesophageal speech, esophageal speech, laryngectomy*

Introduction

Voice is a complex sound which has a fundamental tone and higher harmonic tones which are grouped into separate frequency groups, ie. formants. Basic features of the voice are height, volume, timbre and duration. Pitch is a perceptual phenomenon, and depends on the fundamental frequency. The strength of voice depends on the amplitude of oscillation and pressure, it is determined by the intensity and we perceive it subjectively as a volume. Timbre, or tone of the voice, makes each voice unique, and is the result of resonance[1–3]. Different structures, integrated with functions, involved in voice production, are estimated by various methods, procedures and techniques, and each provides different information, such as instrumental and behavioral analysis of the production, analysis of the products, the analysis of a person who produces the voice and process analysis[4–8]. Objective voice analysis includes acoustic and aerodynamic measurements. Acoustical measurements consist of determining the fundamental frequency, frequency range, jitter which means the frequency oscillations, intensity, intensity range, shimmer which means the intensity oscillations and the ratio of harmonic tones and noisy components. Aerodynamic measurements include determining phonation and friction times[3–8].

Table 1 shows the average values of acoustic parameters of laryngeal voice of men and women with no vocal pathology, chronological age ranging from 60– 69 years[5]. Total laryngectomy is a mutilitant surgery which leaves behind multiple serious consequences. Loss of voice, respectively loss of loud speech, which adversely affects the psychosocial life of the individual, changes the overall quality of life at all levels. Rehabilitation is a challenge to surgeons and speech therapists since 1873 when was made first laryngectomy by Billroth.

Postoperative voice-speech rehabilitation is individual and in the past was included two possible methods: learning esophageal speech and speaking with the help of

Received for publication October 15, 2011

TABLE 1
AVERAGE VALUES OF ACOUSTIC PARAMETERS OF LARYNGEAL VOICE

	F_0 (Hz)	I (dB)	JITTER (%)	SHIMMER (dB)	HNR (dB)	MPT (sec)
Male	112.20	68.60	0.52	0.39	20.26	24.56
Female	202.20	67.42	0.52	0.39	20.26	15.09

Fo – pitch in Hertz, I – intensity in decibels, JITTER – frequency oscillations in percentage, SHIMMER – intensity oscillations in decibels, HNR – harmonic-to-noise ratio in decibels, MPT – maximum phonation time in seconds

mechanical, digital aids. For the last 30 years tracheoesophageal speech became the most preferred method in voice restoration following total laryngectomy, but the other methods are still useful[9–13].

The aim of this study was to analyze the voice quality of alaryngeal tracheoesophageal and esophageal speech, and to determine which of them is more similar to laryngeal voice production, and thus more acceptable as a rehabilitation method of laryngectomized persons.

Subjects and Methods

Sample

The sample consisted of 20 totally laryngectomized persons of both sexes. The study involved 17 men (85%) and 3 women (15%). The average age of respondents was 61.3 years. All patients underwent total laryngectomy and primary closure of the pharynx with unilateral or bilateral neck dissection. Duration of the voice-speech rehabilitation was different and individually determined for each of the respondents. The study included only those respondents who were sufficiently rehabilitated, and whose speaking ability and reading skills were at a sufficient level to perform the test task. Examination of the medical records of each of the participants excluded the existence of a significant hearing loss or other similar obstructions, which could affect the control of prosodic elements of speech, and therefore the measurement results. Subjects were divided into two groups: 10 respondents (50%), which included 2 women and 8 men with built tracheoesophageal prosthesis and 10 respondents (50%), which included 1 woman and 9 men with learned esophageal speech.

Measuring instruments and variables

The average frequency and average intensity of speaking basic alaryngeal tones and their standard deviations were obtained by prolonged phonation of vowel /a/ and reading a paragraph of text adapted for fluent reading. The frequency of individual votes in the test material is proportional to their frequency in the standard spoken Croatian language[14]. Average alaryngeal speech tone in speaking laryngectomized subjects is described with six variables: five parameters of acoustic voice analysis (the fundamental frequency of speech alaryngeal tone, frequency oscillations of alaryngeal voice tone, voice intensity alaryngeal tone, intensity oscillations of alaryngeal tone, the ratio of harmonic tones and noisy components) and one parameter of aerodynamic measurements (maximum phonation time of vowel /a/).

The manner of performing tests

After selecting the laryngectomized subjects, according to the previously mentioned criteria, they began recording individual respondents. Recording was done in Osijek at Department of Otorhynolaryngology and Head and Neck Surgery in Osijek University Hospital Center. The room in which the recording is started is not acoustically isolated, but it is not exposed to external noise, and is considered adequate space to record voice samples. The sound input unit used was directional microphone that doesn't capture possible background noise. Recording is carried out without time constraints the individual duration of each participant was required to execute all the test tasks.

Methods of data processing

The recorded speech samples were processed in the computer program Cool Edit 2000. Further acoustic analysis was performed in Praat 4.3.21. computer program. Acoustic analysis included the calculation of the fundamental frequency in Hertz (Hz), average intensity in decibels (dB), the value of frequency oscillations in percentage (%), and the value of intensity oscillations in decibels (dB), then the ratio of harmonic tones and noisy components in decibels (dB) and the maximum time of phonation in seconds (sec). Statistical analysis was made in Statistical Package for Social Sciences (SPSS 13.0) computer program. Testing normality of distribution was done by One-sample Kolmogorov-Smirnov test. The obtained data was statistically analyzed by analysis of variance – the difference between the testing of small independent samples (Independent samples T-Test).

Results

Table 2 shows the average results of the minimum and maximum values of each variable in both groups of laryngectomized subjects. The minimum value of the fundamental voice frequency in first group shows deeper alaryngeal voice, but it is acceptable. The maximum value of fundamental frequency was found amongst female respondent which is distinctly pathological and does not correspond to a usual level of female voice. In second group the minimum value of the fundamental frequency is very low and the maximum value is acceptable. Average scores of intensity values are very different for both groups and shows that tracheoesophageal speech is louder than esophageal. The maximum value was found amongst male respondent and shows very loud tracheoesophageal speech. The values of jitter during prolonged

TABLE 2
AVERAGE RESULTS OF THE MINIMUM AND MAXIMUM VALUES OF TRACHEOESOPHAGEAL AND ESOPHAGEAL VOICE

Variable	TES		ES	
	MIN	MAX	MIN	MAX
Fo (Hz)	124.93	424.03	75.31	147.13
I (dB)	73.18	90.42	38.52	51.01
JITTER (%)	0.9	6.66	1.37	7.96
SHIMMER (dB)	0.2	1.99	0.66	1.93
HNR (dB)	1.56	11.24	2.31	4.97
MPT (s)	1.48	20.04	1.11	2.75

TES – tracheoesophageal speech (group 1), ES – esophageal speech (group 2), Fo – pitch in Hertz, I – intensity in decibels, JITTER – frequency oscillations in percentage, SHIMMER – intensity oscillations in decibels, HNR – harmonic-to-noise ratio in decibels, MPT – maximum phonation time in seconds, MIN – minimum value, MAX – maximum value

vowel phonation and speech should be less, as well as the values of shimmer. These values are high and correspond to a large amount of breathiness and roughness in the alaryngeal voice. The values of harmonic-to-noise ratio should be larger and in these subjects are small which means less tones and more noise in voice in both groups of respondents. Results of prolonged phonation are very different in both groups. Respondents with voice prosthesis have longer phonation than subjects who speaks esophageal. Table 3 shows the results of statistical analysis, with calculated arithmetic means and their standard deviations for each variable, then these F values were compared with those tabular F values according to the degrees of freedom and p value.

In testing differences between arithmetic means of first and second group of respondents for the variables that describe the fundamental voice frequency and intensity alaryngeal voice and speech, and maximum vowel phonation time, a statistically significant difference was found, in the confidence interval of 95%. Tracheoesophageal speech, respectively replacement speech by an internal voice prosthesis is significantly different from esophageal speech in the average fundamental frequency, intensity and duration of prolonged phonation.

Testing of differences between the first and second groups of respondents for the variables that describe the frequency and intensity oscillations during the speech, and the ratio of harmonic tones and noisy components in the speech, showed no statistically significant difference between the tested variables. Tracheoesophageal and esophageal replacement speech of laryngectomized subjects contains an equal amount of breathiness and roughness.

Discussion and Conclusion

Comparing the results between both groups of respondents can bring us to conclusion that tracheoesophageal speech has a higher frequency range of esophageal speech. Also, the values of the minimum and average fundamental frequency tracheoesophageal voice correspond to fundamental frequency values of laryngeal voice, and therefore this method is more appropriate speech replacement for women. Significantly larger volume, which is adequate average intensity of the laryngeal voice and suitable to communicate regardless of the circumstances in which laryngectomized person at the moment when they speak can be achieved by speaking using the built in voice prosthesis. On the other hand, esophageal speech is significantly reduced in the intensity and isn't suitable for communication when the background noise is present. The values of jitter and shimmer did not differ significantly in either of the groups, and differ from the values of jitter and shimmer laryngeal voice. Both modes of speech replacement sound very hoarse and wheezing, with tracheoesophageal speech having higher values for both tested variables. A higher value of jitter means more frequency oscillations, or subjectively perceived as a greater amount of roughness in a voice. A higher value of shimmer means more intensity oscillations, which is a subjective assessment of sound as breathiness. The values of harmonic-to-noise ratio in both groups were similar and differ from the average value of laryngeal voice. In tracheoesophageal voice, as well as esophageal, there is more noisy components than

TABLE 3
TESTING THE SIGNIFICANCE OF THE DIFFERENCE IN ACOUSTIC AND AERODYNAMIC PARAMETERS OF ALARYNGEAL VOICE BETWEEN TRACHEOESOPHAGEAL AND ESOPHAGEAL SPEECH

Variable	TES		ES		t	df	F	p
	MEAN	SD	MEAN	SD				
Fo	227.25	96.25	106.56	22.71	3.859	18	16.497	0.001
I	78.31	4.68	44.68	4.69	16.031	18	0.967	0.000
JITTER	3.94	1.91	5.64	1.92	–1.982	18	0.102	0.063
SHIMMER	1.39	0.68	1.37	0.37	0.090	18	8.565	0.929
HNR	4.28	3.83	3.77	0.84	0.408	18	12.385	0.688
MPT	6.92	5.44	1.76	0.58	2.983	18	9.245	0.008

TES – tracheoesophageal speech (group 1), ES – esophageal speech (group 2), Fo – pitch in Hertz, I – intensity in decibels, JITTER – frequency oscillations in percentage, SHIMMER – intensity oscillations in decibels, HNR – harmonic-to-noise ratio in decibels, MPT – maximum phonation time in seconds, MEAN – arithmetic mean, SD – standard deviation, t – t value, df – degree of freedom, F – F value, p – p value the components of harmonic tones. These values of the surveyed variables can significantly reduce speech comprehension of laryngectomized person. Average values of the maximum phonation time of tracheoesophageal and esophageal speech are similar, but comparing the individual results we notice the difference. The maximum duration of the prolonged phonation with voice prosthesis is significantly different from the maximum duration of phonation of respondents who use esophageal speech. This result is understandable considering that tracheoesophageal speakers use a natural reservoir of air for phonation, unlike esophageal speakers. The longer time of phonation allows speech fluency and overall quality of prosody and thus is more comprehensible.

These results largely confirm previous findings. There is no ideal method of rehabilitation and every one of them requires an individual approach to the patient, but the average value of the fundamental frequency and intensity alaryngeal voice, and the value of the maximum phonation time demonstrate the advantages of rehabilitation by means of installing voice prosthesis. Arias et al. (2000.) in a similar study found that the average voice fundamental frequency of esophageal and tracheoesophageal speech statistically significantly differs and that the voice of respondents with voice prosthesis is closer to laryngeal voice. Mušura et al. (2003.) examined 24 alaryngeal speakers, and the results also showed the advantage of surgical rehabilitation[15]. Salihović et al. (2001.) compared the acoustic parameters between tracheoesophageal and laryngeal voice, and got the results which show that they do not differ significantly in the average elementary and the lowest frequency, but significant differences were obtained in the highest frequency, frequency range, jitter and shimmer[16]. Further tests should include more variables that will be tested and should be conducted on a larger sample.

REFERENCES

1. HEĐEVER M, Osnove fiziološke i govorne akustike (Edukacijsko--rehabilitacijski fakultet, Zagreb, 2005). — 2. BOGERT BP, The Acoustics of Speech. In: TRAVIS LE, Handbook of Speech Pathology (Appleton--Century-Crofts. Inc., New York, 1957). — 3. BRACKET IP, Parameters of Voice Quality. In: TRAVIS LE, Handbook of Speech Pathology and Audiology (Appleton-Century-Crofts. Inc., New York, 1971). — 4. CVEJIĆ D, KOSANOVIĆ M, Fonijatrija I. dio–glas (Zavod za udžbenike i nastavna sredstva, Beograd, 1982). — 5. VEČERINA VOLIĆ S, KIRINIĆ PAPEŠ V, PAPEŠ Z, Fonijatrija, accessed 2011/08/11. Available from: URL: http://www.foni.mef.hr/Prirucnik. — 6. BRESTOVCI B, Logopedska dijagnostika (Edukacijsko–rehabilitacijski fakultet, Zagreb, 2006). — 7. BOONE DR, McFARLANE SC, The voice and Voice Therapy (Allyn and Bacon, Needham Heights MA, 2000). — 8. HORII Y, Journal of Speech and Hearing Research, 22 (1979) 5. — 9. PERKINS WH, Vocal Function: A Behavioral Analysis. In: TRAVIS LE, Handbook of Speech Pathology and Audiology (Appleton-Century-Crofts. Inc., New York, 1971). — 10. EMANUEL FW, Folia Phoniat., 25 (1973) 110. — 11. SANSONE D, Journal of Speech and Hearing Research, 13 (1970) 489. — 12. KROM G, Journal of Speech and Hearing Research, 36 (1993) 254. — 13. ŠTAJNER KATUŠIĆ S, Glas i govor nakon totalne laringektomije (vlast. naklada, Zagreb, 1998). — 14. ŽIVKOVIĆ IVANOVIĆ T, Logopedska rehabilitacija, accessed 2011/03/28. Available from: URL: http://www.larynx-hr.org/logopedska-rehabilitacija-o-link-84.html. — 15. ŠKARIĆ I, Govorni dijelovi III. dio. In: BABIĆ S, BROZOVIĆ D, MOGUŠ M, PAVEŠIĆ S, ŠKARIĆ I, TEŽAK S, Povijesni pregled, glasovi i oblici hrvatskog jezika (Globus, Zagreb, 1991). — 16. ARIAS MR, RAMON JL, CAMPOS M, CERVANTES J, Otorhinogol Head Neck Surg, 122 (2000) 743. — 17. SALIHOVIĆ N, AKŠAMIĆ N, BRKIĆ F, Glas i govor osoba s ugrađenom traheoezofagealnom protezom. In: Glas – zbornik radova (1. znanstveni skup s međunarodnim sudjelovanjem, Opatija, 2001).

Lj. Širić

»J. J. Strossmayer« University, Osijek University Hospital Center, Department of Otorhinolaryngology and Head and Neck Surgery, J. Huttlera 4, 31000 Osijek, Croatia
e-mail: ljsiric@gmail.com

OBJEKTIVNA PROCJENA TRAHEOEZOFAGEALNOG I EZOFAGEALNOG GOVORA POMOĆU AKUSTIČKE ANALIZE GLASA

SAŽETAK

Cilj ovog ispitivanja bio je analizirati kvalitetu alaringealnog glasa između treheoezofagealnog i ezofagealnog govora, te utvrditi koji je prema kvaliteti govorne produkcije sličniji laringealnom glasu, a time i prihvatljiviji kao rehabilitacijska metoda laringektomiranih osoba. Objektivna glasovna procjena provedena je na uzorku od 20 totalno laringektomiranih ispitanika oba spola, prosječne dobi 61,3 godine. Ispitanici su podijeljeni u dvije skupine: 10 (50%) ispitanika s ugrađenom traheoezofagealnom protezom i 10 (50%) s usvojenim ezofagealnim govorom. Ispitivanje je obuhvatilo 6 varijabli: 5 parametara akustičke analize glasa i 1 parametar aerodinamičkog mjerenja. Dobiveni podaci statistički su obrađeni analizom varijance. Analiza prikupljenih podataka pokazala je statistički značajnu razliku između dviju skupina u vrijednostima intenziteta i fundamentalne frekvencije, te maksimalnom vremenu fonacije vokala na razini značajnosti od 5% i u rasponu pouzdanosti od 95%. Statistički značajna razlika nije utvrđena u vrijednostima jittera, shimmera i omjera signal-šum između traheoezofagealnog i ezofagealnog glasa. Nijedna metoda rehabilitacije nije idealna i zahtjeva individualan pristup pacijentu, ali dobiveni rezultati pokazuju prednost rehabilitacije ugradnjom govorne proteze.