ARCHIVES OF ACOUSTICS
32, 4 (Supplement), 151–158 (2007)

ACOUSTIC ANALYSIS OF ESOPHAGEAL SPEECH IN PATIENTS AFTER TOTAL LARYNGECTOMY

Wiesław WSZOŁEK(1), Maciej MODRZEJEWSKI(2), Monika PRZYSIĘŻNY(2)

(1) AGH University of Science and Technology
Department of Mechanics and Vibroacoustics
Al. Mickiewicza 30, 30-059 Kraków, Poland
e-mail: wwszolek@agh.edu.pl

(2) Jagiellonian University
Collegium Medicum
Św. Anny 12, 31-008 Kraków, Poland
e-mail: m.mod@poczta.fm, bprzysiezny@rdn.com.pl

(*received July 15, 2007; accepted October 23, 2007*)

Laryngeal cancer is one of the most frequent types of malignant tumor in male patients and it is located at position 5 among all the malignant tumors. The primary method of laryngeal carcinoma treatment for the cases with advanced cancer stages is the surgical treatment. In such a treatment the most frequent version is the total extraction of the organ. An effective postoperative therapy leads to the extension of the patients life-time, but the price is a serious disability, considerably impairing the patients communication with other persons. Therefore there is a problem of creation of a substitute voice and speech for those patients. The creation mechanism of the esophageal speech is the formation of vibration generator in the upper part of oesophagus. The substitute air container is formed by the oesophagus itself and the air vibrations in the upper part of oesophagus, generated during the anti-peristaltic motion of the air removed from oesophagus, create the primary pitch. The modulation of the fundamental pitch is obtained by the unchanged articulation organs and resonance cavities of the body. In the present paper an acoustic study of esophageal speech is described. An attempt has been made to determine the acoustic signal parameters, which can be useful in evaluation of esophageal speech quality.

Keywords: speech analysis, speech recognition, pitch determination, pathological speech.

1. Introduction

Professional methods of signal processing and analysis, in particular the acoustic methods, offer several possibilities of speech signal quality evaluation enabling multilateral speech analysis as well as visualization of the results, including their variability during the speech generation process [1, 14]. However direct analysis of such variability is very complex and requires great experience, particularly in cases when pathological speech is analyzed. In the present paper an attempt has been made to analyze the voice of patients after total larynx extraction (laryngectomy). After total extraction of the voice organ the patients develop a distressing psychological problem related to the voice loss and direct consequence of that fact, i.e. the loss of social communication abilities. Therefore the task of creation of a substitute voice and speech abilities, crucial for their social and professional adaptation, is a problem of primary importance. For patients after total laryngectomy the process of communication with other persons may take place by means of:

- formation of a substitute voice and speech (esophageal voice, pharyngeal voice),
- formation of substitute voice and speech by creation of voice fistulas, with application of proper prosthetic elements (so-called fistula voice),
- communication using electronic devices – so-called "artificial larynx" speech aid,
- communication using oropharyngeal "pseudo-whisper",
- communication using facial expression, gestures, writing, character tables – non-verbal communication,
- communication by simultaneous use of several ways – combined method [8, 10].

For such patients the best method of restoring interpersonal communication seems to be the substitute speech, in particular the esophageal speech. The review of the subject literature shows that actually in Europe the most preferred method for formation of a substitute speech is the esophageal speech. On the other hand in USA about 30% of patients uses the fistula speech with inserted valve voice prosthesis [6].

2. Experimental material and research methodology

The acoustic examinations have been carried out for 33 patients after total laryngaectomy resulting from laryngeal cancer, for whom various degree of esophageal speech formation (very good, good, sufficient) has been found. The study group consisted of 93.9% of men (31 persons) and 6.1% of women (2 persons). The average age of the group was 61 years and the age of individual patients was contained between 45 and 80 years. The majority of patients – 26 persons – managed to learn the esophageal speech in 1–3 months, for 3 patients the required rehabilitation time exceeded 6 months and 4 persons managed to master the esophageal speech by themselves. Identical study has been carried out for a group of 35 normal persons, with full hearing abilities, in the age between 40 and 60 years, and this group has been regarded as a reference group. The registration of acoustic signal of the esophageal voice has been carried out in the anechoic chamber of the Department of Mechanics and Vibroacoustics at AGH-University of Science and Technology in Kraków. The block diagram of the measurement setup has been shown in Fig. 1.

The presented setup enabled registration of acoustic signals in the 20 Hz–20 kHz band, with a dynamic range not less than 80 dB.

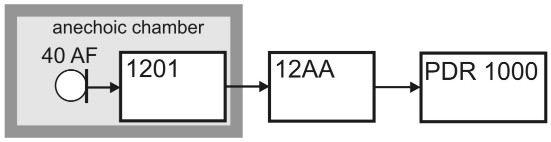

Fig. 1. Block diagram of the measurement setup, where 40 AF – measurement microphone by G.R.A.S, 12AA – microphone preamplifier by G.R.A.S, PDR 1000 – digital magnetic registration device by HHB.

Both the patients and the persons from the reference group pronounced the same text (three times), consisting of: vowels (/a/, /u/, /e/, /i/), words containing vowels (/ala/, /aS/, /ula/, /ela/, /igwa/), and a test sentence "/dzis'/ /jest/ /wadna/ /pogoda/" (the weather is nice today) [9]. For the case of vowels list the patients have pronounced three short utterances: one just as in physiological speech, and the other ones in a prolonged manner. Such a registered material in the form of time samples of acoustic signals has been treated as the experimental material, on which the further analyses have been based. The product obtained after the preliminary processing of the registered signal took the form of dynamic spectra $W(i,j)$, digitized in time, frequency and amplitude [3]. In order to standardize the procedure and to make the results comparable the same signal processing scheme has been used in the whole study. Thus the dynamic spectra $W(i,j)$ have been calculated with amplitude resolution of $\Delta L = 0.2$ dB, homogeneous frequency quantization of $\Delta f = 125$ Hz in the band from $f_d = 125$ Hz to $f_g = 12$ kHz and time quantization (sampling) of $\Delta t = 9$ ms [13, 14]. Figure 2 presents averaged spectrum of the $/u/$ vowel, for the correct speech and esophageal speech with prolonged phonation.

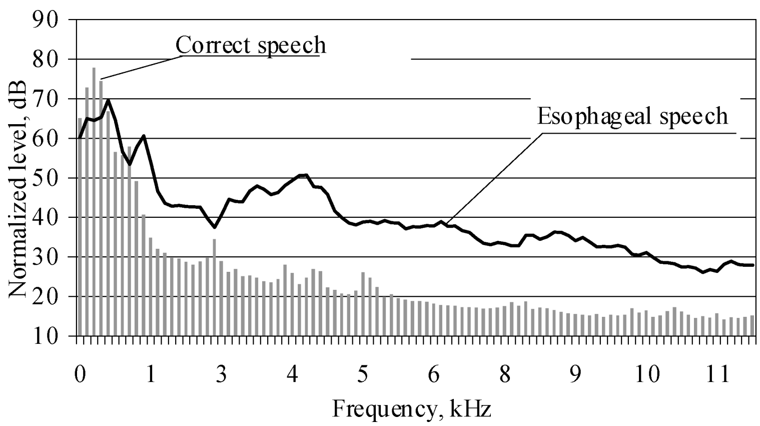

Fig. 2. Typical spectrogram of averaged /u/ phoneme for correct and esophageal speech.

In further analysis of the results the Multi-Dimensional Voice Program (MDVP) Model 5105 by KAYELEMETRICS has been also used. The MDVP is a new computer system for acoustic analysis of pathological voice signals and screening of laryngeal diseases [7].

3. Selected results of the study

In the present section selected results of the study will be described. Figures 3, 4 present averaged spectra of the $/a/$ vowel for the correct and esophageal speech in the band up to 12 kHz.

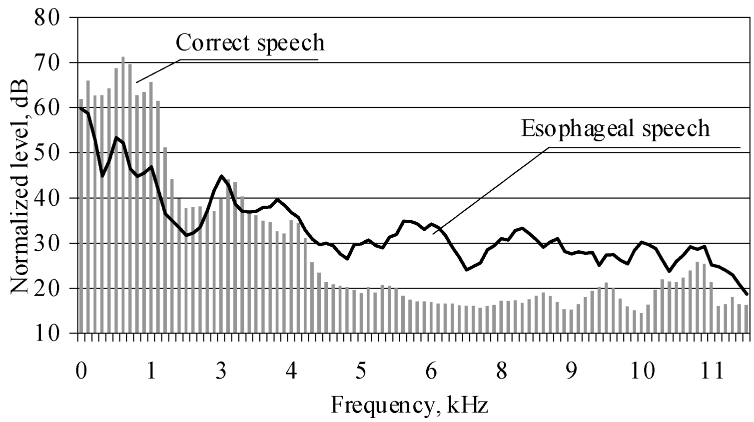

Fig. 3. Spectrogram of averaged /a/ phoneme for correct and oesophagus speech, patient No. 1.

One can determine the level differences in individual frequency bands between the correct and esophageal speech.

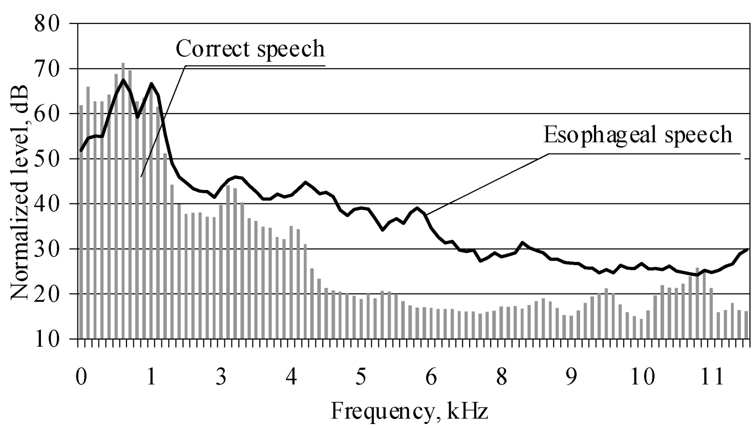

Fig. 4. Spectrogram of averaged /a/ phoneme for correct and oesophagus speech, patient No. 10.

On the other hand Fig. 5 presents averaged spectra of the $/u/$ vowel in esophageal speech for two different patients, after a narrow-band analysis with $\Delta f = 10$ Hz, applied in order to bring out the detailed changes of the signal in the formant frequency bands.

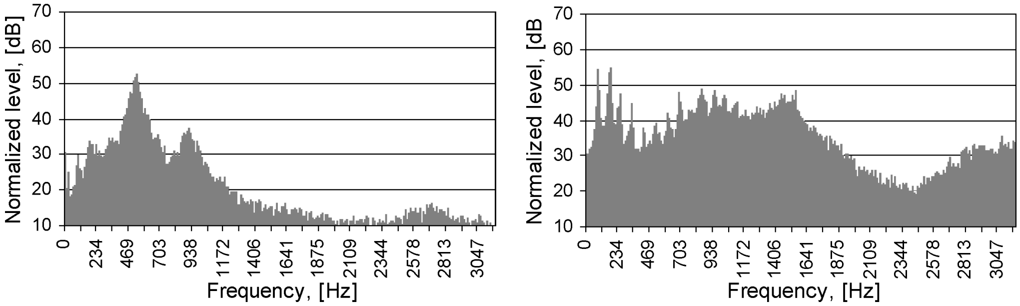

Fig. 5. Spectrogram of averaged /u/ phoneme, $\Delta f = 10$ Hz.

Table 1 contains the description of selected parameters (obtainable from the MDVP program), which have been applied for the esophageal speech analysis.

Table 1. Selected acoustic parameters applied for the esophageal speech analysis.

Parameter (abbr.)	Parameter description
SD F_0	Standard deviation of the fundamental frequency.
NF_0	Measure of the fundamental frequency changeability. $\mathrm{NF}_0 = \frac{F_{0\,\max} - F_{0\,\min}}{F_{0m}}$, $F_{0\,\max}$ – maximal fundamental frequency, $F_{0\,\min}$ – minimal fundamental frequency, F_{0m} – mean fundamental frequency
Pitch Perturbation Quotient (jitter)	$\mathrm{PPQ} = \frac{N \sum_{i=2}^{N-1} \left\lvert \frac{P_{i-1} + P_i + P_{i+1}}{3} - P_i \right\rvert}{(N-2) \sum_{i=1}^{N} P_i}$, P_i – pitch of the i-th glottal cycle, N – number of cycles contained in the considered segment of the voice signal, This method for calculation reduces the influence of slow amplitude and pitch variations (tremors) due to non-pathological factors (physical overtension).
Amplitude Perturbation Quotient (shimmer)	$\mathrm{APQ} = \frac{N \sum_{i=2}^{N-1} \left\lvert \frac{A_{i-1} + A_i + A_{i+1}}{3} - A_i \right\rvert}{(N-2) \sum_{i=1}^{N} A_i}$, A_i – amplitude of the i-th glottal cycle, N – number of cycles (as above).

Table 2 contains the values of parameters calculated for selected patients (esophageal speech) and a representative person from the reference group.

Table 2. Calculated results for selected acoustic parameters

Patient	F_0 [Hz]	NF_0	SDF_0 [Hz]	PPQ [%]	APQ [%]
1	93.0	0.7	16.9	9.0	69.1
2	513.2	1.0	118.6	6.0	30.8
3	83.8	0.3	4.8	1.3	7.9
4	247.6	2.2	159.4	6.5	21.0
5	440.0	0.8	51.5	5.8	23.8
6	84.8	1.0	8.6	2.0	9.5
7	298.8	1.0	253.7	17.4	20.4
8	72.8	0.1	2.4	1.8	11.6
9	293.4	2.0	207.5	6.9	22.1
10	89.7	1.0	12.4	4.9	20.8
standard	524.5	0.4	59.9	8.8	70.0

In Figs. 6, 7 it has been shown how the parameters of acoustic speech signal (shown in Table 2) distinguish between the quality of the patients' speech and the correct speech.

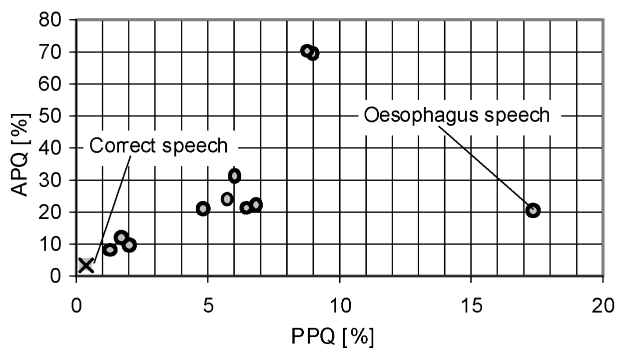

Fig. 6. Esophageal speech on the SDF_0-NF_0 scale.

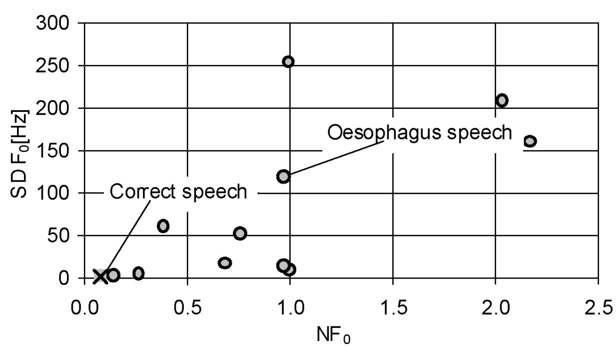

Fig. 7. Esophageal speech on the APQ-PPQ scale.

4. Conclusions

The paper presents the results of a preliminary study concerning esophageal voice and speech in patients after total laryngaectomy resulting from laryngeal cancer, with various levels of the esophageal speech formation: very good, good and sufficient. It should be stressed that the study is one the first esophageal speech studies carried out in Poland. On basis of that study a new research methodology has been proposed, oriented towards making the voice quality evaluation fully objective. This methodology allows the evaluation of the fundamental frequency pitch, variation range of that frequency, variation of the fundamental frequency in time, acoustic description of the phonation time and acoustic parameters of the vowels. The results of the preliminary studies indicate that in the long run they can useful in the voice diagnosis and therapy for patients after total laryngeactomy.

The presented results are collected from a study that has been only begun. In further stages the authors will focus their attention on the verification of feature space with respect to the diagnostic applications. In future the presentation of results from such esophageal speech studies should also be modified in order to improve their suitability for the rehabilitation process.

References

[1] BASZTURA CZ., *Źródła, sygnały i obrazy akustyczne*, Wyd. Komunikacji i Łączności, Warszawa 1988.

[2] ENGEL Z., MODRZEJEWSKI M., WSZOŁEK W., *Akustyczna ocena operacji krtani z wykorzystaniem parametrów tonu podstawowego*, Zeszyty Naukowe AGH, Mechanika, **16**, 1, 67–79 (1997).

[3] DELLER J. R., PROAKIS J. G., HASEN J. H., *Disrete-Time Processing of Speech Signals*, Macmillan Publising Company, New York 1993.

[4] GROCHOLEWSKI S., *Statystyczne podstawy systemu ARM dla języka polskiego*, Wydawnictwo Politechniki Poznańskiej, Poznań 2001.

[5] GUBRYNOWICZ R., *Metody akustyczne w diagnostyce narządu mowy*, [in:] Problemy Biocybernetyki i Inżynierii Biomedycznej, Tom 2: Biopomiary, FILIPCZYŃSKI L., TORBICZ W. [Eds.], 161–181, WKiŁ, Warszawa 1990.

[6] HAGEN R., *Stimmrehabilitation nach totaler Laryngektomie in der Bundesrepublik Deutschland*, HNO., 105, 487 (1990).

[7] HADJITODOROV S., MITEV P., *A computer system for acoustic analysis of pathological voices and laryngeal diseases screening*, Medical Engineering & Physics, **24**, 419–429 (2002).

[8] JANCZEWSKI G., OSUCH–WÓJCIKIEWICZ E., *Rak krtani i gardła dolnego*, a-medica press, Bielsko Biała 2002.

[9] MODRZEJEWSKI M., *Skuteczność chirurgicznego leczenia chorych na raka krtani piętra głośni*, Rozprawa habilitacyjna, Collegium Medium UJ, Kraków 1996.

[10] PRUSZEWICZ A., *Foniatria kliniczna*, PZWL, Warszawa 1992.

[11] TADEUSIEWICZ R., *Sygnał mowy*, WKiŁ, Warszawa 1988.

[12] TITZE I. R., *Principles of Voice Production*, Prentice-Hall Inc., Englewood Cliffs, New Jersey 1994.

[13] WSZOŁEK W., *Selected methods of pathological speech signal analysis*, Archives of Acoustics, **31**, 4, 413–430 (2006).

[14] WSZOŁEK W., KŁACZYŃSKI M., *Acoustics methods of voice estimation after surgical treatment of the vocal tract*, Archives of Acoustics, **30**, 4S, 193–197, 2005.