A preliminary analysis of esophageal speech signals

Olga PUSTOVALOVA*
Moldova State University
pustovalova.olga@gmail.com

***Abstract* — The ESOTEK project by the University of Deusto (Spain) is aimed to provide computer assistance to patients who are learning esophageal speech. Esophageal speech is an alternative way of speech production that can be used by the persons who have lost their normal voice after the surgical procedure of removing the larynx. For analysis of that type of speech, digital signal processing techniques can be applied. In this paper, an analysis of esophageal speech properties is presented.**

***Index Terms* — assistive technologies, digital signal processing, esophageal speech, speech processing.**

I. INTRODUCTION

Laryngectomy is a surgical operation of removing the larynx (including vocal cords), usually because of cancer of the larynx at an advanced stage [1]. As the vocal cords are removed, patient loses its ability to communicate through use of speech. Restoration of voice after laryngectomy can be done in several ways, such as use of prostheses or an artificial larynx. However, esophageal speech remains the most natural way of voice restoration.

Esophageal speech is the generation of speech using esophagus. As a result of special exercises, a neoglottis is formed in the esophagus as well as an ability to use it. Instead of vocal fords' vibration, vibration of esophagus is involved in the speech production. Obtained speech acoustically differs from healthy speech in characteristics such as level of noise or pitch.

A typical result of comparison of speech fragments of length (sample rate * 3) is presented on the figure 1.

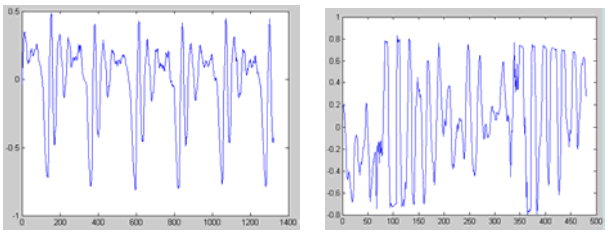

Figure 1. Waveforms of healthy and esophageal speech samples

As figure 1 illustrates, esophageal voice has lower frequency and less harmonic structure. A preliminary analysis of esophageal speech samples is presented in this paper.

II. MATERIALS AND METHODS

In order to compare healthy and esophageal speech, samples of the Spanish sound /a/ were selected from the database created by the group PAS of the University of Deusto.

* This work was supported by the European Commission doctoral exchange grant (141193-EM-1-2008-1-ES-ERASMUS-ECW-L06).

For the illustrations in the present paper, two samples of healthy voices and two of esophageal ones was used. For pitch measurement, five samples of each speech type were used. Sample rate was 44100 Hz for healthy voices and 16000 Hz for esophageal ones.

Typical pitch values for healthy speech are 80-210Hz for an adult male speaker and 150-320Hz for a female speaker [2]. For esophageal voice, this range is 50-100Hz, with an average of 60Hz [3] and with no difference between male and female voices. For esophageal speech, higher amount of noise is typical.

Signal processing toolbox [4], implementation of the YIN pitch measurement method [5], GammaTone Filter Kit for Matlab [6], and Praat tool [7] were used to analyze the signal.

III. RESULTS AND DISCUSSION

For periodogram analysis, for a signal *x* with sample rate *fs* a standard periodogram was plotted in Matlab, and the number of FFT points used to calculate the PSD estimate was equal to 512: periodogram(x,[],512,fs). As presented on Figure 2, for esophageal speech, power is spread less uniformly along the frequencies.

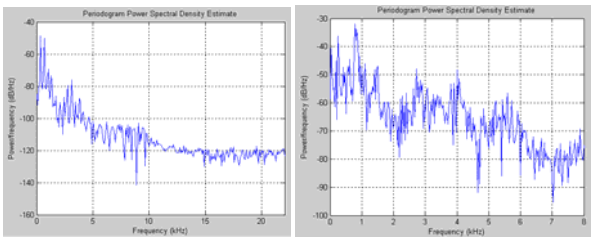

Figure 2. Periodograms of healthy and esophageal speech samples

Spectrograms obtained in Praat illustrate that for oesophageal voice, a less harmonic structure is typical, and the harmonics-to-noise ratio is higher than in case of normal voices.

In case of oesophageal voice, power of signal is concentrated on low frequencies, as analysis based on gammatone-based spectrogram shows.

Results of pitch measurement of healthy and esophageal voice are presented in the table 1. Default parameters of the YIN algorithm are minimal expected frequency of 30Hz (min_{f0} = 30) and maximal expected frequency of 4000Hz (max_{f0} = 4000). In pitch measurement of healthy speech, default expected pitch range was used. For esophageal speech, measurement was done for two other pitch ranges.

TABLE I. COMPARISON OF ESOPHAGEAL AND HEALTHY PITCH VALUES CALCULATED BY THE YIN METHOD

#	Esophageal speech			Healthy speech
	30-4000Hz	**20-120Hz**	**25-150Hz**	
1	63.74	28.27	28.22	198.13
2	798.36*	61.63	61.63	237.74
3	52.18	20.00	52.20	114.97
4	739.71*	67.45	67.42	203.03
5	933.12*	47.79	47.76	111.11

In the YIN algorithm, window size is calculated as the closest integer number that is greater or equal with sampling rate divided by minimal expected frequency. Results of pitch measurement clearly depend on window size selection.

As shown in the table 1, YIN algorithm failed for several samples of esophageal speech (values marked by asterisk (*)) and therefore needs to be revised in order to avoid errors caused by esophageal speech properties.

ACKNOWLEDGMENTS

This study was carried out as a part of research internship at the University of Deusto (Spain). The author would like to thank Begoña García Zapirain and Ibon Ruiz Oleagordia from the University of Deusto for their assistance, and Gheorghe Căpăţână and Nicolae Objelean from the Moldova State University for their support.

REFERENCES

[1] Edels, Yvonne. Laryngectomy: Diagnosis to Rehabilitation. Croom Helm, 1983, p. 16

[2] Michael Ashby, John A. Maidment. Chapter 10. Suprasegmentals. In: Introducing phonetic science. Cambridge University Press, 2005, p. 154.

[3] Дворниченко В.В. Проблема реабилитации голосовой функции у ларингэктомированных больных. Современная онкология, 4, 3 (2002), 115-118. // Dvornichenko V.V. Problem of rehabilitation of the voice function of the laringectomized. Sovremennaya onkologiyam 4-3(2002). [in Russian]

[4] Signal processing toolbox for Matlab® [computer program]. http://www.mathworks.com/products/signal/ - accessed in September 2008.

[5] Alain de Cheveigné, Hideki Kawahara. YIN, a fundamental frequency estimator for speech and music. Journal of Acoustic Society of America, 111(4), April 2002, pp. 1917-1930.

[6] Nick Clark. GammaTone Filter Kit for Matlab, version 2.0 [computer program]. Updated 20 September 2007.

[7] Praat, a system for doing phonetics by computer. Glot International 5(9/10), 2001, pp. 341-345.