computer visualization of speech signal,
electroglotography, laryngograph

Jolanta ZIELIŃSKA*

ESTIMATION OF POSSIBILITIES CONNECTED WITH USAGE OF ELECTROGLOTOGRAPHY METHOD IN SPEECH SIGNAL ANALYSIS

The research presented in this paper deals with the speech signal with use of elecroglotography method analysis issue. This is an instrumental analysis, so the device called Laryngograph is presented, as a practical application. In this paper capabilities of this device are estimated. The very interesting fact is that the visualization of the speech signal obtained using Laryngograph allows to detect its acoustically and phonetically most important features, and presenting them in a graphical form. The analysis process performed using a computer and the specified computer attachment is easier, faster and ensures higher quality than other methods. Computer voice recording enables not only visualization but also objective assessment and its repetitiveness. In the context of presented questions, practical capabilities of integrated system for speech examination - Speech Studio are discussed.

1. INTRODUCTION

The speech analysis is a very complicated process because it concerns highly complex signal. This is the main reason it is a subject of interest of specialists from many scientific disciplines, so varying from linguistics [4] to computer science [8]. Acoustic phonetics concern the technical and physical aspect of the speech signal. It analyzes the speech signal, as a sound realization of a language code, with use of properly extended acoustic system. The research in this scope [5, 6, 2] presents classical measurement methods and their results that include mainly the descriptions of spectral polish sounds features. The very important aspect of high-tech phonetic research is an instrumental analysis of speech signal with use of such devices as: x-ray machine, electromyograph (EMG), spectrograph, minograph, Laryngograph. Its goal is speech signal visualization, which makes possible to detect its most important features and present them in a graphical form. The analysis process performed with use of computer is easier, faster and ensures higher quality than other methods. Computer voice recording enables not only visualization but also objective assessment and its repetitiveness. These are the main advantages of electroglotopraphy method. It enables vocal cords examination in a way which is independent of generated sound, through the measurement of physical displacements of vocal folds and registration of larynx vibrations. It is possible because the majority of human speech sounds is generated by stimulation of vocal cords by air streams, pulsating with the frequency, which is determined by vocal folds vibrations.

2. ELECTROGLOTOGRAPHY METHOD – GENERAL CHARACTERISTIC

Electrtoglotography is based on electrical impedance measurements between vocal folds during speaking process vibrations. It is a function of their mutual position and changes during larynx vibration. When vocal folds are cramped, the value of this function is smaller, than in case when they are opened. The monitoring of vocal folds movements during normal speaking process uses this phenomenon. However there are some disadvantages of devices, which use it directly. Individual regulation for each patient is required and some distortions can appear, as the electrodes, which are used for impedance measurement, can translocate on the patient throat. The technique, which presents the improved form of electroglotography and eliminates mentioned disadvantages is called laryngographical and was practically introduced in Laryngograph. The elimination of measurement errors was accomplished by:

- introduction of electrodes with additional surrounding conductive rings, applying screened connections and using tri-pole net, which output is dependent on electrical conditions on the patient's throat, only in the electrodes direct neighbourhood
- designing the circuit, which is dedicated to process the electrodes output and automatically compensates various impedancies on the speaker's throat in such a way, that output is dependent only on fast vocal folds vibrations.

The electrical power diffused on the patient's throat is about 20 mW (with frequency 1 MHz) and the top limit of the frequency is about 5 kHz. In case of normal male voice, the ratio of the final output signal to the noise is about 40 dB. For small children relative noise is much bigger, however successful results can be obtained also for new-born children.

The elegctroglotography signal obtained, which represents the functioning of vocal folds, establishes the basis for Fx basic frequency determination (presented in Fig. 1). Fx frequency is very important in phonetic research.

* Pedagogical University of Cracow, ul. Pochorążych 2, 30-084 Kraków

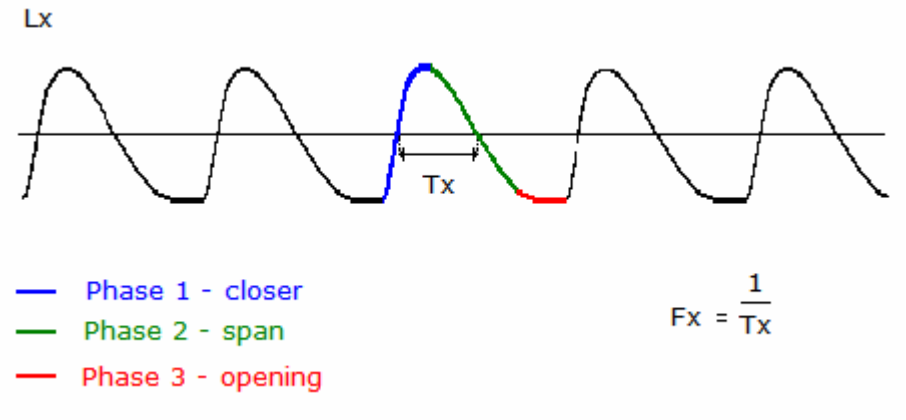

Fig.1 Sample curve of Lx signal, determination of basic Fx frequency [9]

The initial signal of the Lx signal the visualizes vocal folds and is the base for determination of basic frequency Fx. Analysis of the signal cycle Lx helps to separate its three phases. The first phase is a quick signal growth, corresponding to the quick closing of the vocal folds. The second phase corresponds to the slightly slower signal fall and is connected with their opening. The third phase, which is a flat waveform, corresponds to the state, when the vocal folds are open. Fx frequency is calculated while measuring of each functioning cycle length in the middle of the Lx signal amplitude value (in practice, the measurement of time intervals between transitions of the signal through that level is made). The Fx value is the reverse of Lx signal cycle length.

3. PRACTICAL USAGE OF ELECTROGLOTOGRAPHY METHOD – USING COMPUTER ATTACHMENTS LARYNGOGRAPH PROCESSOR PCLX

Computer attachment Laryngograph Processor PCLX operation is based the electroglotography method. It is connected to computer by interface card, installed directly on IBM computer bus, which constitutes coupler with Laryngograph and software. This card contains special system, which processes signals and makes it possible to visualize them on the computer screen and remember just in-time. PCLX software called Speech Studio allows signals activation, visualization, on-line analysis and remembering them on the hard drive for later analysis and recovery.

The most important element of the speech signal analysis is the way the signal is sampled and recorded by computer. Computer attachment Laryngograph Processor PCLX possessed two input signal detectors. These are coopery electrodes placed on the patient's throat with use of regulative band and microphone, which can be placed stably on the head by means of special bow. Necessary condition of acquiring correct research result is the way that electrodes are placed on the patient's throat. It depends on the age of person: if the person is younger, electrodes should be placed higher, assuming central location between 4-rd and 7-th neck vertebra in case of adults. This fact comes out directly of the physiological structure of the larynx [7, 3]. Fast growth of larynx takes place in case of every child below 5. The second period of growth falls on puberty age. The larynx structure is changing, it is growing in horizontal area, especially in case of boys, and it is associated with voice mutation. This process takes place when children are 13 -15 years old, but sometimes lasts longer, till 19. In this period the neck gets longer and the larynx moves down. At that age it is important to pay special attention to physiologically correct placement of laryngograph electrodes. The second decoder (microphone) should be placed about 2,5cm from the patient's mouth.

The prototype of presented device was introduced for the first time at the Faculty of Phonetics University College in London [1]. Nowadays it is produced in Laryngograph Ltd. Company. During several years it was put into practice in such domains as:

- speech therapy, in order to improve speech process,
- clinical examination of nose and throat, to diagnose work of vocal folds after surgical procedures,
- improvement of singers and announcers voice,
- examination relations between speech and work of vocal cords,
- teaching foreign languages.

4. ANALYSIS OF SPEECH SIGNAL USING COMPUTER ATTACHMENT LARYNGOGRAPH PROCESSOR PCLX

During voice generation, three basic phases take place and they can be fully monitored and analyzed by computer software called Speech Studio. These three phases are: generation of stimulated vibrations, its modification in vocal resonator and radiation [5, 6]. The signal which is generated in larynx during voiced sound pronunciation, by vocal cords in the role of stimulation generator, is a periodic flow, characterized by the basic frequency Fx, which was described in details earlier

(Fig. 1). In the narrowness of the voice channel the signal is modulated. The adjustment of voice route (articulators configuration) is characterized by formants, which are defined basing on average amplitude-frequency spectrum (presented in Fig. 3).

Fig. 2 presents the main screen of integrated system for voice examination Speech Studio, which has the possibility of speech signal analysis that was described previously. During work with this program there are two toolbars available (top and side one), which allow to perform several operations, starting with signal activation, its remembering, recovering, analysis, and configuring the amount and type of visible flows and scales.

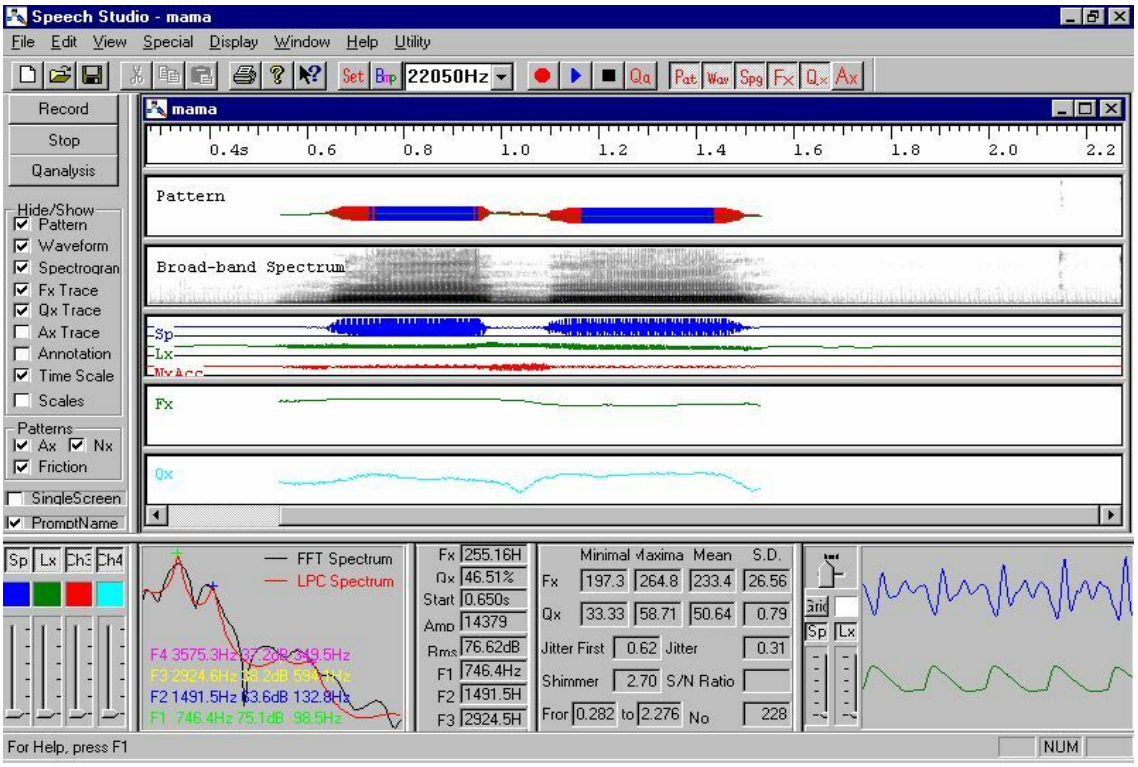

Fig.2 Integrated system for voice examination Speech Studio [9]

The fundamental part of the screen is seven flows, which characterize the speech signal. Two of the most important flows present changes of amplitude over time and the work of vocal folds. These charts can be enlarged in the bottom right window and allow to immediate, preliminary diagnose of voice pathology. Signals amplification can be regulated by attachment knobs or sliders placed in the bottom left corner of the screen. In the bottom part of the screen there are also two other windows. In the first one, average amplitude-frequency spectrum is displayed. In the second, there are values, which characterize the speech profile of the patient. In this profile there following parameters are defined: minimum, maximum, average value, standard deviation for the basic Fx frequency and occlusion factor Qx and also other parameters describing disturbances of the frequency (Jitter) and amplitude (Shimmer). The flow of the occlusion factor Qx depends directly on the Lx, and is presented as a percentage value. It determinates the percentage usage of time, when vocal folds are opened, that is the time of occlusion and opening during statement pronunciation (the sum of phase 1 and phase 2 durations in Fig. 1).

During this type of research, related to the speech signal analysis on the computer screen there are presented many flows simultaneously, characterizing the speech signal. On base of those flows we can define both the numerical parameters describing the voice during appropriate analyses, and qualitative conclusions. This program allows to reduce the number of presented charts to the selected.

The obtained result depends not only on vocal folds movements, but also on the size of the larynx and on the weight of vibrating muscles. The signal, which is marked by Sp, represents changes of the acoustic wave pressure. This is an oscillograme and basing on it, some initial information about the structural and functional pathological changes of the larynx can be obtained.

When the signal is more jagged and irregular, then physiological changes within the diagnosed organ are bigger. During the analysis of presented changes of vocal wave amplitude, it is possible to determine its basic acoustical features, sonority and nasality. The equality of random and specified elements and the connection between individual and semantic aspects is clearly reviled in it [8].

It is the main reason, why time flows cannot be used in automatic speech recognition (ASM), however they are widely used in acoustical features of speech research, especially in the obtained signal amplitude outline. It is a widely examined issue, which is described in many works concerning phonetics [5, 6, 2].

The second chart presented in Fig. 2 is called spectrograph. It is a broad-band device, which is used in acoustical evaluation of normal and pathological phonation. So it determines one of the phonetic parameters of speech at the basic level.

It allows also to differentiate opposition: sonority – soundlessness and evaluating prosodic elements of pronunciation, especially intonation. The flow presents changes of amplitude in time and its level, so the signal amplitude is specified by the degree of darkness.

Such representation is widely used not only in phonetic research, but also in ASM. It creates foundations of value of the basic frequency definition and formants specification, which is very important in this kind of research. Its connection with pathological functioning of the larynx, intonation, functioning of vocal cords and correct articulation was deeply analyzed and widely described in scientific literature.

The next spectrum, which is displayed on the computer screen (Fig. 2) is described by Sp parameter, and allows to visualize the time based pronunciation flows, and to define the degree of its intensity and estimate volume of sound. What is more, the proper analysis of the basic Fx frequency gives information about all elements of correct pronunciation. These are: dynamic breathing, economic management during breathing, work of vocal folds, pitch of the voice, prosodic elements of the statement, such as: speed, rhythm, accent and additional presentation of sounds resonance degree (if there is no resonance, there is also no image on the screen). In case of analyzing correct articulation, mainly the shape of the larynx – mouth vocal channel, where the shape is modulated, the amplitude-frequency spectrum of the statement is very useful (Fig. 3).

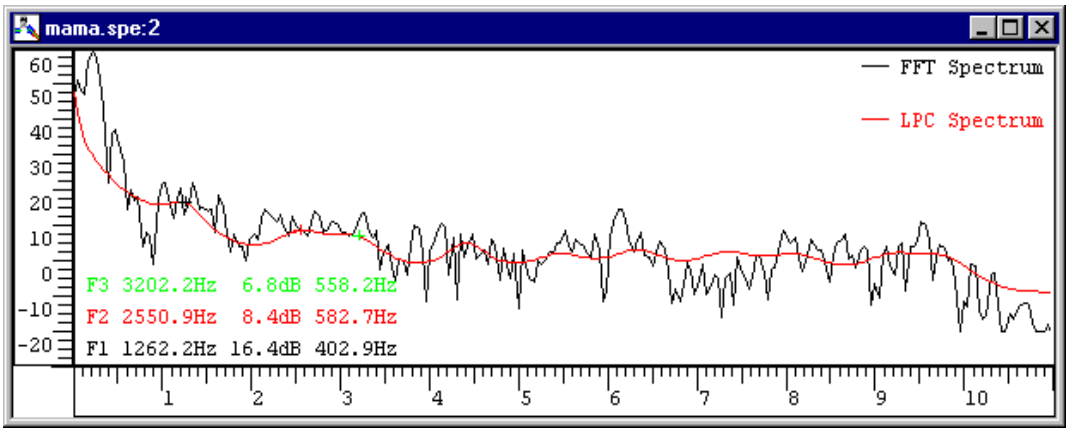

Fig.3 The average amplitude-frequency spectrum, „mama" statement, Speech Studio program [9]

The presented speech signal spectrum is obtained with appropriative computer procedures, which are based on FFT algorithm (Fast Fourier Transformation) and LPC (linear prediction algorithm). It allows to determine formants in the selected period of time, and the temporary spectrum at the specific point of time. For each formant, its maximum frequency, amplitude and frequency band is presented.

5. COMPUTER ANALYSIS OF RESEARCH MATERIAL

There are various possibilities of research material analysis, starting with statistical calculations, parametrical calculations, and graphical processing. In order to obtain more precise signal view, we can select any part of it and enlarge, so that even the smallest pathological change can be detected. This software enables also measurements of time intervals between individual points of the signal. It can be performed by placing two cursors on the screen, thus the time interval between them can be measured.

The Fx signal analysis is based on evaluation of 23 histograms, including CFx histogram, which presents the vocal cords work irregularity (Fig. 4).

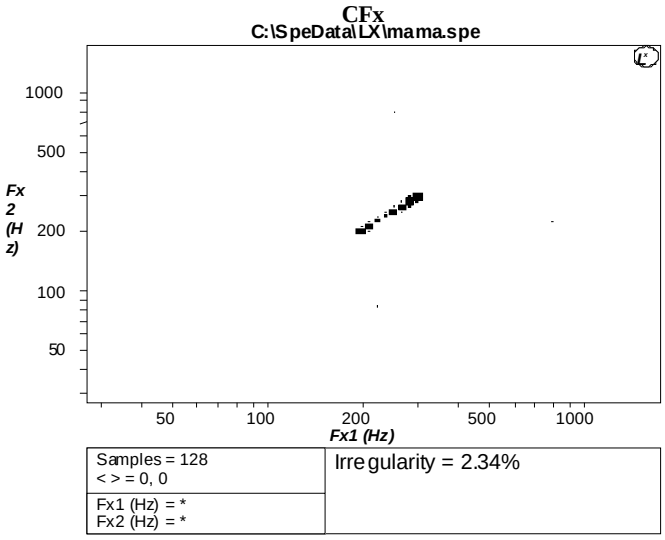

Fig.4 Sample CFx histogram, which presents the vocal cords work irregularity [9]

Computer analysis of speech signal allows to obtain the voice profile of the patient in form of acoustic parameters (Fig. 5).

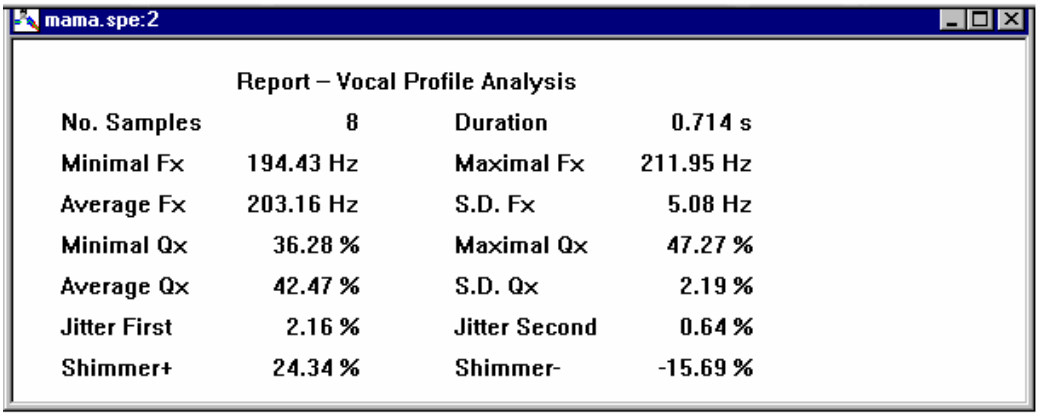

Fig.5 Sample voice profile for „mama" statement, Speech Studio program [9]

The voice profile contains the average, minimum and maximum value of basic Fx frequency. It also allows to determine the oscillations frequency, the standard deviation, the factor of frequency disturbance (Jitter), the percentage value of the average occlusion factor Qx, its minimum and maximum value and standard deviation. Additionally, as the last information, the value of voice amplitude disturbance factor (Shimmer) is presented.

Hardware and software capacities, presented in this paper, were used during empirical research. The main goal of that research was the comparison of acoustic parameters of voice in group of 88 deaf children (with bilateral, receiving hearing injury) and group of 25 healthy children. Both groups were between 7 and 13 years old. The obtained results are presented in table 1.

Table 1: Central values of acoustic parameters connected with basic frequency Fx, phonation irregularity and amplitude disturbance

	Fx avg.	Fx min.	Fx max.	S. D. Fx	Jitter	Shimmer	CFx Irregularity
Deaf children	294,22Hz	110,59Hz	424,15Hz	33,17Hz	8,59%	8,45%	21,09%
Healthy children	248,93Hz	228,58Hz	424,15Hz	5,07Hz	0,97%	4,22%	1,75%

Table 2: Central values of acoustic parameters connected with occlusion factor Qx

	Qx avg.	Qx min.	Qx max.	S. D. Qx
Deaf children	44,67%	14,21%	74,34%	1,75%
Healthy children	44,63%	30,72%	57,74%	0,60%

The research showed significant differences between both groups. The Fx basic frequency in case of deaf children phonation is higher (294,22Hz) than in case of healthy children (248,93Hz). That is why central values are higher in case of deaf children: standard deviation S.D. Fx – 33,17Hz, very extensive boundary frequencies values (110,59Hz do 424,15Hz) and high indicator of jitter - 8,59%. Similarly, the values of acoustic parameters related with Fx frequency in case of healthy children were much lower, just as amplitude disturbance indicator - shimmer – 4,22%. In case of voice of deaf children, the amplitude disturbance indicator was twice as higher - 8,45% and also the CFx irregularity indicator, connected with work of vocal cords, was much higher. The occlusion factor was similar in both groups, about 45%, so in case of deaf children there are no physiological changes in vocal cord structure.

The obtained results of acoustic research showed, that there are very significant acoustic changes both in case of voice and phonation of deaf children, and initially confirmed hiperfunctional dysphonia in this group. The lack of physiological changes can be a proof, that therapeutic activities should be performed as fast as it is possible, otherwise their vocal cords can be permanently damaged.

6. CONCLUSIONS

The usage of electroglotography method enables to deeply analyze the speech signal, which was stored in the computer memory. For example, it can be a general spectral analysis or detailed analysis based on their formants. By means of basic Fx frequency visualization on the computer screen, it is possible to estimate the voice pitch as well as to obtain the full acoustic voice profile, using instrumental analysis. Both speech signal and methods of its examination are very complex. The scope of electrotography usage depends mainly on the designated research goal and competence of the person performing the examination. The undisputed fact is that presented opportunities are really extensive and, what is the most important fact, comparatively easy to obtain and interpretate.

BIBLIOGRAPHY

[1] ABBERTON E., HOWARD D., FOURCN A., Laryngograph assessment of normal voice: a tutorial, Clinical Linguistic and Phonetics 3, pp. 243–259, Informa Healthcare, 1996.
[2] BASZTURA C., Źródła, sygnały i obrazy akustyczne, Wyd. Komun. i Łączn.,Warszawa, 1988.
[3] EBERHARDT G., Zaburzenia głosu u dzieci w wieku rozwojowym, [in:] Zaburzenia głosu – badanie – diagnozowanie – metody usprawniania, red. H. Mierzejewska, M., Przybysz-Piwkowa, Wyd. DIG, Warszawa, 1998.
[4] GRABIAS S., Logopedyczna klasyfikacja zaburzeń mowy, „Audiofonologia" t. VI, pp. 7–22, wyd. Polski Komitet Audiofoniologii, 1994.
[5] PRUSZEWICZ A., Foniatria kliniczna, PZWL, Warszawa, 1992.
[6] TADEUSIEWICZ R., Sygnał mowy, WKiL, Warszawa, 1992.
[7] ZIELIŃSKA J., Diagnoza i terapia sprawności ortofonicznej dzieci z uszkodzeniem słuchu wspomagane techniką komputerową, Wyd. Naukowe AP, Kraków, 2004.