ARCHIVES OF ACOUSTICS
32, 4 (Supplement), 143–150 (2007)

THE ACOUSTIC AND ELECTROGLOTTOGRAPHIC METHODS OF DETERMINATION THE VOCAL FOLDS VIBRATION FUNDAMENTAL FREQUENCY

Wiesław WSZOŁEK, Maciej KŁACZYŃSKI, Zbigniew ENGEL

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

(*received July 15, 2007; accepted October 18, 2007*)

There are many methods of measuring fundamental frequency of vocal folds' vibrations. The mentioned function can be estimated in several ways, either through internal measurements (optical methods) or external measurements (the acoustic analysis of the voice, electric method). Electric methods (EGG), based on analysis of the electroglottographic signal are often applied. However acoustic methods or registration and processing the speech signals can be applied. The authors' research proves that maximum information on phonetic action of larynx can be assembled by delimitation of the fundamental frequency parameters. Thus the exact determination of pitch function makes a priority in the glottal area research. The spectrum character of speech wave is connected with the fundamental frequency (F_0) of human vocal folds vibration. As it is considered, F_0 of the source during voicing contains an abundance of information on the larynx pathology, individual trait, the emotional state and ethnographical origin of speaker. The literature demonstrates only limited research that conducted simultaneous measurement of fundamental frequency by the EGG and with the acoustics methods. The present paper presents results of such research, including simultaneous processing EGG and acoustic data. The analysis of the F_0 function exactitude and the usefulness of these methods were executed too.

Keywords: speech analysis, speech recognition, pitch determination, electroglottographic signal.

1. Introduction

An acoustic speech signal, defined as a variation of acoustic pressure in time, has a complex graph, being a reflection of its complex articulation process. On the parameters of the signal influence both its source (i.e. the vibrating vocal folds or sound caused by turbulent air flow through the narrowing of speech organs) and dynamical properties of the vocal channel, forming the structure of the signal. The signal can be described in the time domain as the convolution of time-dependent signal source $pg(t)$ and pulse-like answer of the voice channel $h(t)$ [1, 14]:

$$p(t) = \int_0^T h(t-\tau)\, p_g(\tau)\, \mathrm{d}\tau. \tag{1}$$

Thus, in the acoustic time-dependent speech signal, properties of the source and properties of the voice channel, forming the sound, are closely bound.

When a phoneme articulation takes place in the larynx (i.e. for voiced sounds), the shape of voice wave spectrum is related to the fundamental-tone frequency of the human vocal folds vibrations, F_0. A representative spectrum of the fundamental tone has the amplitude decreasing exponentially in such a way that the overtones are attenuated with the slope about – 12 dB/octave, yet the overtones with the frequencies thirty times larger than the fundamental one are still clearly visible (Fig. 1) [14, 16].

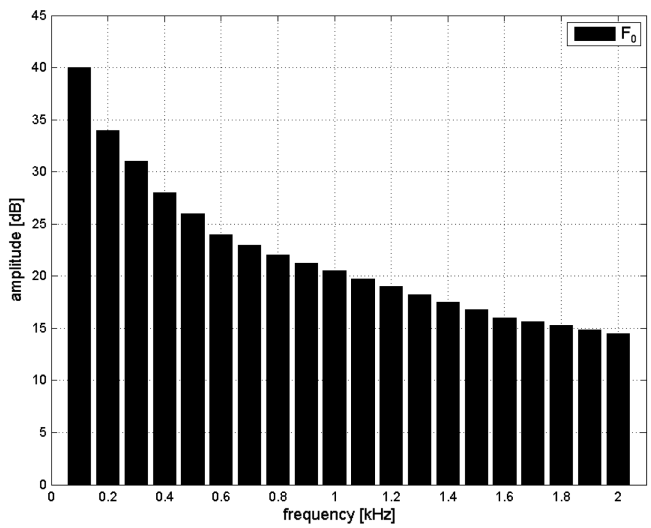

Fig. 1. Spectrum of the laryngeal tone.

Most of the researchers find that the shape of the larynx tone during the phonation contains information about the individual characteristics, emotional state, ethnographical origin and pathology of the vocal tract. Deformation of the vocal organ, related to the larynx dysfunction, manifests itself in the change of the vocal cords vibrations parameters, what influences on F_0 [4, 6, 11, 19]. Exact determination of the fundamental tone function becomes a priority in the voice generator research. However, the lack of the direct access to the signal F_0, generated by this source, makes the estimation of F_0 difficult.

There are several different techniques known and used to measure and determine the fundamental tone function of the human vocal folds vibrations, both internal measurements (optical methods) and external ones (acoustic analysis of voice, electrical method EGG) among them.

Literature review suggests that only few researches have been made so far in such a way that fundamental tone parameters have been measured by electrical (EEG) and acoustic method simultaneously. In this paper we present results of the research carried out in this way.

2. Methods of the larynx tone determination

A number of the methods used for the larynx tone determination can be found in the literature (G. Fant, J. Flanagan., J. Kacprowski, W. Jassem, R. Gubrynowicz, I. Titze and others). An accurate and frequently used method is the electrical method, EGG, based on the recording and analysis of the electroglottographic signal [3, 8, 9, 12]. From the other hand, the acoustic methods have been used more and more frequently, thanks to the professional methods of the acoustic signals recording and processing, being developed all the time. All these methods can be roughly divided in two groups:

- with the help of electroglottography,
- based on the acoustic speech signal.

2.1. Electroglottography

Electroglottography is a noninvasive method of the glottis electrical impedance measurement. The impedance is measured between two electrodes placed on the subject's skin on the larynx level. The electrical impedance changes during the vocal folds movements. Its value is smaller when the vocal folds are clenched and larger when they are open. The impedance change, with the vocal folds clenched, is equal to 1–2% of the total neck impedance in the place of examination. The sinusoidal signal, with the frequency usually in the range from 300 kHz up to 5 MHz, is sent with the help of the aforementioned electrodes and penetrates well the less-conductive skin layer. After demodulation, the signal is recorded and processed, and next stored in the computer memory as a time-depended function. The archived signal is the subject of preemphasis and of the subsequent analysis. Electroglottography does not determine the movements of each fold individually. However, it allows representing the phase of vocal folds closing and openings more accurately than other methods, especially in the vertical direction, what gives an opportunity to measure directly the time dependence of the fundamental tone and to determine its value.

The disadvantage of this method is the necessity of using of relatively expensive measurement equipment. If applied to children or patients after a larynx area surgery, they may feel very uncomfortable, making the procedure impossible to conduct.

The glottogram, i.e. the function of the electroglottographic signal vs. time, for the vowel $/a/$ with the prolonged phonation of the normal male voice is depicted in Fig. 2.

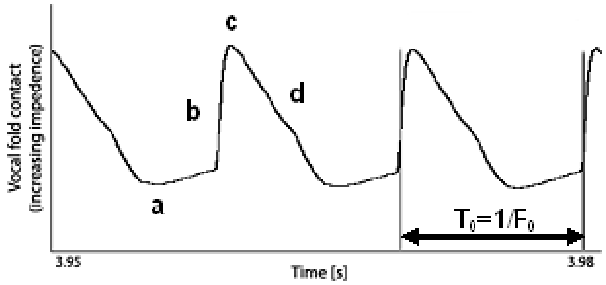

Fig. 2. Time dependence of the EGG signal – glottogram.

In addition, the period T_0 of the EGG signal and 4 phases of the folds movements, i.e. a) the complete closing phase, b) the process of opening, c) the full opening phase and d) the process of closing, are marked. The detailed description of this method is given in [9].

2.2. Acoustic methods

There are several dozen of algorithms (based on the acoustic signal) for determination of the fundamental tone F_0, which differ in accuracy. A substantial majority of them is based on the signal in the time domain or in the frequency domain. The method utilizing the properties of the autocorrelation function is described in [4]. The authors of the paper [19] presented (both the theoretical principles and the numerical applications of) three methods of F_0 estimation: ZCA – Zero Crossing Analysis method, CEP – Cepstrum method, and SHR Subharmonic-to-Harmonic Ratio method. The SHR method, described in the paper [13], was adapted and refined in [19], taking into consideration the analysis of the deformed speech signal of Polish tongue. These algorithms were employed in the research carried out in the frame of this project. Their detailed description can be found in Ref. [17]; thus we have omitted it here.

3. Research material and methodology

The research, aimed at finding the accuracy of the fundamental tone determination and checking the practicability of methods described in Sec. 2, was carried out on the group of 22 men, age 20 to 25, so-called standards of Polish tongue, without any pathologies that could affect the voice quality. The time-dependent acoustic speech signal and the EGG signal were recorded simultaneously in an anechoic chamber, at the Department of Mechanics and Vibroacoustic, AGH University of Science and Technology, Kraków, Poland. The diagram of the measurement setup is shown in Fig. 3.

The task of the group people being the subjects of examination was to read out the phonetic text slowly and without any intonation. They had to repeat three times: the vowels – */a/, /e/, /i/, /u/*; the vowels with the prolonged phonation – /a/, /e/, /i/, /u/; the words – */ala/, /as/, /ula/, /ela/, /igwa/* (i.e. 4 Polish names and Polish equivalent for

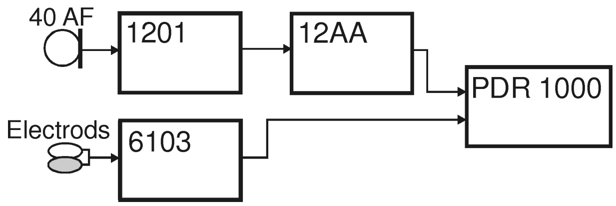

Fig. 3. The block diagram of the measurement setup, where: 40 AF – the measuring microphone, G.R.A.S; 12AA – the microphone preamplifier, G.R.A.S; 6103 – the electroglottograph (EGG), KAYELEMETRICS; PDR 1000 – digital magnetic recorder, HHB.

"needle") and the sentence – */dzis'/ /jest/ /wadna/ /pogoda/* (i.e. the Polish equivalent of the sentence "We have a good weather today").

4. Results and conclusions

To depict and compare the fundamental tone spectra, determined by the acoustic and the electroglottographic methods, the analysis in the frequency domain was made, using Short Term Fourier Transforms, STFTs. Before the frequency analysis, the collected data was subjected to the process of preemphasis with the band-pass FIR filter, with $f_{\text{low}} = 50$ Hz and $f_{\text{upp}} = 400$ Hz. (The FIR filters have a capability to give the linear phase-frequency characteristic.) Implementation of the algorithm and the calculations were made in the MATLAB (MathWorks Inc.) environment [15]. First, the evaluation of the spectrum shape was made. Employing STFTs, the dynamical spectrum $W(t, f)$ containing 56 lines with the $\Delta f = 10$ Hz width, made with the $\Delta t = 0.1$ s time quantum and the level quantum equal to $\Delta L = 0.2$ dB, was obtained. The subject of analysis was 4 vowels pronounced by each person ($88F_0$ spectra in total). The goal of the analysis was to determine the difference between the spectra obtained from the acoustic and the EGG signals. These vowels have a fundamental significance in the examination of the voice channel condition (especially of the glottis) because of their stationary-like time dependence. The examples of the over-time-averaged spectra of the vowels with the prolonged phonation, obtained from the EGG and the acoustic signals, are presented in Fig. 4 and Fig. 5.

From the analysis of the frequency spectra carried out for each investigated signal sample, it appears that differences (in shape and envelope) between the fundamental tone spectra determined from the acoustic signal and from the EGG signal are small.

The substantial differences, observed in the relative level (amplitude) of recorded signal, are related to the signal normalization process. For each group (acoustic signal sample, EGG signal sample), the averaged minimal value for all samples recorded in the given group was used as a reference level in the logarithmic scale.

In the second part of research, the comparison between the averaged values of F_0 obtained by the acoustic methods and the F_0 value determined with the help of EGG was made. From the comparison made, it appears that in the investigated group of per-

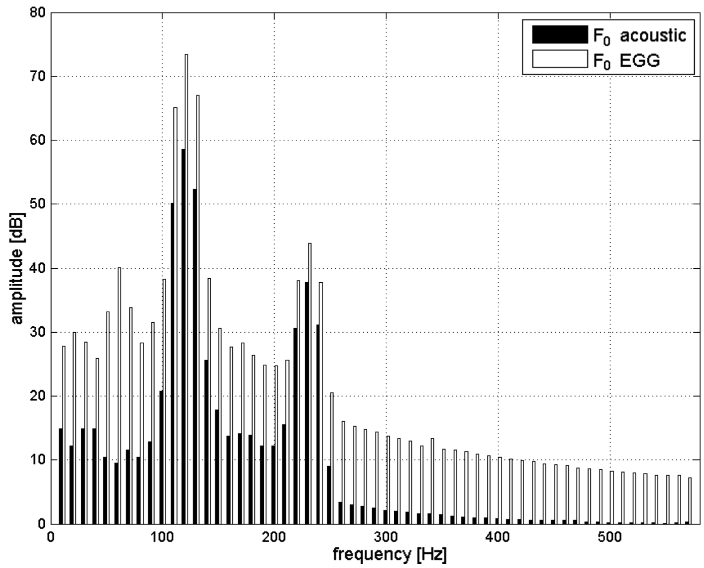

Fig. 4. Averaged spectrum of the vowel “a” with the prolonged phonation.

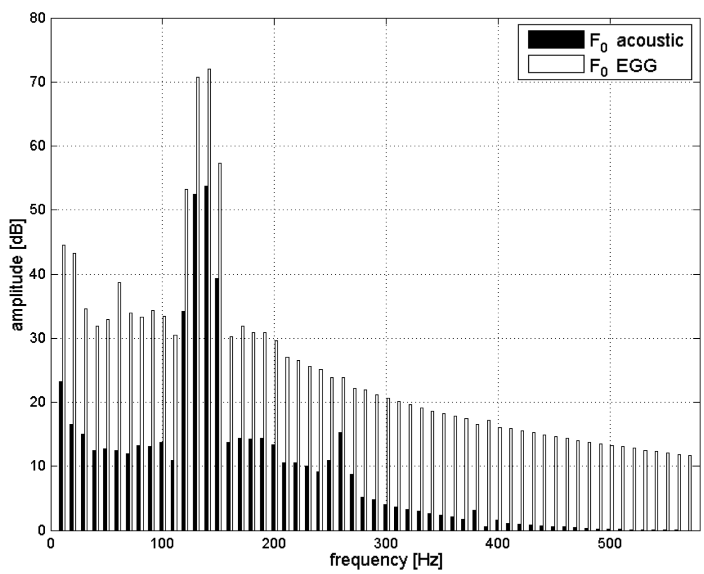

Fig. 5. Averaged spectrum of the vowel “i” with the prolonged phonation.

sons, for all analyzed vowels with the prolonged phonation, the mean squared error in the determination of fundamental tone with the help of the acoustic methods does not exceed 2 Hz for the “ZCA” algorithm, 1.5 Hz for the “cepstrum” algorithm and 1 Hz for the “SHR” algorithm. The detailed description of the aforementioned methods is presented in [16, 17].

From the preliminary researches made so far it is clear that the acoustic methods for the determination of the fundamental tone F_0 are effective and accurate.

They can be an alternative tool in the examination of the vowel quality and the glottis condition for the diagnostics and for the medical therapy, especially in the postoperative cases, when the using of EGG is difficult.

Acknowledgments

The study was co-financed within the research project of the Ministry of Science and Higher Education No N501 01731/1182.

References

[1] Basztura Cz., *Źródła, sygnały i obrazy akustyczne*, WKiŁ, Warszawa 1988.

[2] Deller J. R., Proakis J. G., Hasen J. H., *Disrete-Time Processing of Speech Signals*, Macmillan Publising Company, New York 1993.

[3] Dikshit P. S., Goshom E. L., Seaman R. L., *Differences in fundamental frequency of deaf speech using FFT and electroglottograph*, Biomedical Engineering Conference, Proceedings of the Twelfth Southern, pp. 111–113, New Orleans, LA, USA 1993.

[4] Engel Z., Modrzejewski M., Wszołek W., *Akustyczna ocena operacji krtani z wykorzystaniem parametrów tonu podstawowego*, Zeszyty Naukowe AGH Mechanika, **16**, 1, 67–79 (1997).

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

[6] Gubrynowicz R., *Metody akustyczne w diagnostyce narządu mowy*, [in:] Problemy Biocybernetyki i Inżynierii Biomedycznej. Tom 2: Biopomiary, L. Filipczyński, W. Torbicz [Eds.], pp. 161–181, WKiŁ, Warszawa, 1990.

[7] Hess W., *Pitch Determination of Speech Signals*, Springer-Verlag Berlin, Heidelberg, New York, Tokyo 1983.

[8] Kania R., Hartl D., Hans S., Maeda, S., Vaissiere J., Brasnu D., *Fundamental frequency histograms measured by electroglottography during speech: A Pilot Study for Standardization*, Journal of Voice, **20**, 1, 18–24 (2006).

[9] Marasek K., *Electroglottography description of voice quality*, Phonetic AIMS, Univesität Stuttgard, 1997.

[10] Noll A. M., *Cepstrum pitch determination*, JASA, **41**, 2, 293–309 (1967).

[11] Pawłowski Z, *Foniatryczna diagnostyka wykonawstwa emisji głosu śpiewaczego i mówionego*, Oficyna Wyd. „Impuls”, Kraków 2005.

[12] Rossato S., Audibert N., AUBERGÉ V., *Emotional voice measurement: a comparison of articulatory-EGG and acoustic-amplitude parameters*, Paper of International Conference Speech Prosody, pp. 91–95, Nara, Japan, 23-26 March 2004.

[13] Sun X., *Pitch determination and voice quality analysis using subharmonic to harmonic ratio*, Proc. of ICASSP2002, Orlando, Florida, 1, pp. 333–336, 2002.

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

[15] THE MATHWORKS INC. – *Signal Processing Toolbox User's Guide version 6.7*, pdf document, http://www.mathworks.com.

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

[17] TITZE I. R., *Workshop on acoustic voice analysis-summary statement*, Denver, Colo: National Center for Voice and Speech, 1995.

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

[19] WSZOŁEK W., KŁACZYŃSKI M. *Comparative study of the selected methods of laryngeal tone determination*, Archives of Acoustics, **31**, 4S, 219–226 (2006).