ARCHIVES OF ACOUSTICS
33, 4 (Supplement), 183–188 (2008)

ESTIMATION OF THE VOCAL FOLDS VIBRATION FUNDAMENTAL FREQUENCY BY HIGER ORDER SPECTRUM

Wiesław WSZOŁEK, Maciej KŁACZYŃSKI

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

(*received June 15, 2008; accepted October 25, 2008*)

Research studies carried out by many authors prove that a lot of information concerning the phonation activity of speech organ can be gathered by accurate determination of parameters related to the fundamental tone F_0. It is also reckoned that the knowledge of acoustic parameters based on the measured time dependence of the source's F_0 parameter during the phonation process contains valuable information regarding larynx pathology, personal features as well as the physical and emotional condition of the speaker. At present many methods are applied for determination of the F_0 function. The present work discusses the feasibility and accuracy of fundamental frequency determination based on Higher-Order Spectra Analysis (HOSA).

Keywords: pitch detection, fundamental frequency, speech analysis, higher order spectral analysis.

1. Introduction

From the physical point of view the speech, or more precisely – the acoustic speech signal, presents an important and interesting research object. Multiple efforts are made in order to implement this optimal signal in communication systems with both human-human and human-machine interfaces as well as medical diagnosis systems.

An accurate determination of the time-function of vocal cords vibrations is extremely essential in the voice organ studies. The fundamental tone function F_0 can be estimated by internal measurements (e.g. optical methods) or external measurements (like acoustic or electrical methods). The optical methods include: stroboscopy, cinematography, videokymography (VKG), photoglottography (PGG), electrolaryngography (ELG) and two-point holographic interferometry. The acoustic methods include ultrasonography (USG), multi-dimension speech signal analysis and test evaluation of the voice acoustic pressure, while the electrical method is usually electroglottography (EGG) [3, 4, 8].

Most algorithms for determination of F_0, employing the acoustic speech signal, are based on the time-domain or frequency-domain analyses, which include the methods making use of the auto-correlation functions [3], cepstral analysis [7], zero-crossing analysis [3], the subharmonic-to-harmonic ratio analysis [9]. In the present paper, the attention has been focused on possible applications of higher-order spectral analysis to realization of the research task considered.

2. Higher-order statistical methods

An often encountered problem in the field of signal processing is the separation (as complete as possible) of the required signal from the noise, created in the process of signal transmission through a communication channel (e.g. for the case of F_0 function estimation for the vocal folds vibrations, it is necessary to separate the acoustic speech signal sent into the time-dependent generator-source signal and the pulse response of the voice channel).

The higher-order statistical methods, also known as cumulant methods, are related to the more popular concept of statistical moments. In the same way as the Fourier transform of autocorrelation function (power spectrum) is a useful analytical tool, the result of Fourier transform of cumulants, called polyspectrum, can be also useful. The moments and their spectra are more useful for analysis of deterministic signals, while the cumulants and their respective spectra are more suitable for analysis of random signals [5, 6].

Higher-order statistical moments are natural development (generalization) of the autocorrelation function, while cumulants are non-linear combinations of these moments. First order cumulant is the well-known average value:

$$C_{1x} = E\{x(t)\}. \tag{1}$$

Higher-order cumulants (second – autocorrelation, third, fourth) of the $x(n)$ process with zero average value are consecutively defined as [10]:

$$C_{2x}(k) = E\{x(n)x(n+k)\}, \tag{2}$$

$$C_{3x}(k,l) = E\{x(n)x(n+k)x(n+l)\}, \tag{3}$$

$$\begin{aligned} C_{4x}(k,l,m) = E\{x(n)x(n+k)x(n+m)\} - C_{2x}(k)C_{2x}(l-m) \\ - C_{2x}(l)C_{2x}(k-m) - C_{2x}(m)C_{2x}(k-l). \end{aligned} \tag{4}$$

Higher-order spectra (poly-spectra) are defined as Fourier transforms of the respective cumulants:

• power spectrum

$$S_{2x}(f) = \sum_{k=-\infty}^{\infty} C_{2x}(k)e^{-j2\pi fk}, \tag{5}$$

• bispectrum

$$S_{3x}(f) = \sum_{k,l=-\infty}^{\infty} C_{3x}(k,l) e^{-j2\pi(f_1 k + f_2 l)}, \tag{6}$$

• trispectrum

$$S_{4x}(f) = \sum_{k,l,m=-\infty}^{\infty} C_{4x}(k,l,m) e^{-j2\pi(f_1 k + f_2 l + f_3 m)}. \tag{7}$$

An essential property of the cumulants is the fact that their values are completely independent of all processes characterized by normal distribution. By applying the higher-order statistical methods to analysis of a useful signal, not characterized by a normal distribution and accompanied (disturbed) by a Gaussian noise, one effectively increases the signal-to-noise (S/N) ratio. Majority of real signals do not exhibit normal distribution (e.g. the signals generated by systems with nonlinear dynamics, including speech signal), while the measurement noise can be, to a high degree of accuracy, described as a colored process with a normal distribution. Therefore the utility of higher-order statistical methods is very essential in many practical applications. An additional feature distinguishing the cumulants and poly-spectra is the fact that they contain information concerning the amplitude and phase of a given process (e.g. harmonic fluctuations), while the correlation function and power spectrum contain only the information concerning signal amplitude.

3. Bispectral estimation of fundamental frequency

For the signal $x(n)$ and its Fourier transform $X(f)$, according to the bispectrum definition (6), it can be written as:

$$S_{3x}(f_x, f_y) = X(f_x)X(f_y)X(f_x + f_y). \tag{8}$$

On the other hand, the diagonal slice of the bispectrum is given by [13]:

$$S_{3x}(f) = X(f)X(f)X(2f). \tag{9}$$

From the bispectrum definition it follows that the operation enhances the fundamental frequency of the analyzed signal, by making use of the second harmonic [13]. The estimation of the fundamental frequency F_0, determined from the acoustic speech signal, comprises the specification of a local maximum in the amplitude spectrum (9). An additional criterion taken into account during the search for F_0, is the expected frequency range, e.g. 70–500 Hz for men and 160–960 Hz for women, according to the following formula:

$$\exists F_i = F_0 : \left[i \in \left\langle \frac{fs}{\mathrm{LF}}; \frac{fs}{\mathrm{HF}} \right\rangle \wedge \; F_i = \max(|S_{3x}(f)|) \right], \tag{10}$$

where f_s – sampling frequency for the speech signal [Hz], LF – 70 Hz for men, 160 Hz for women, HF – 500 Hz for men, 960 Hz for women.

Figure 1a, b present the diagonal slice of the bispectrum for $/a/$, $/i/$ vowels with prolonged phonation.

a)

b)

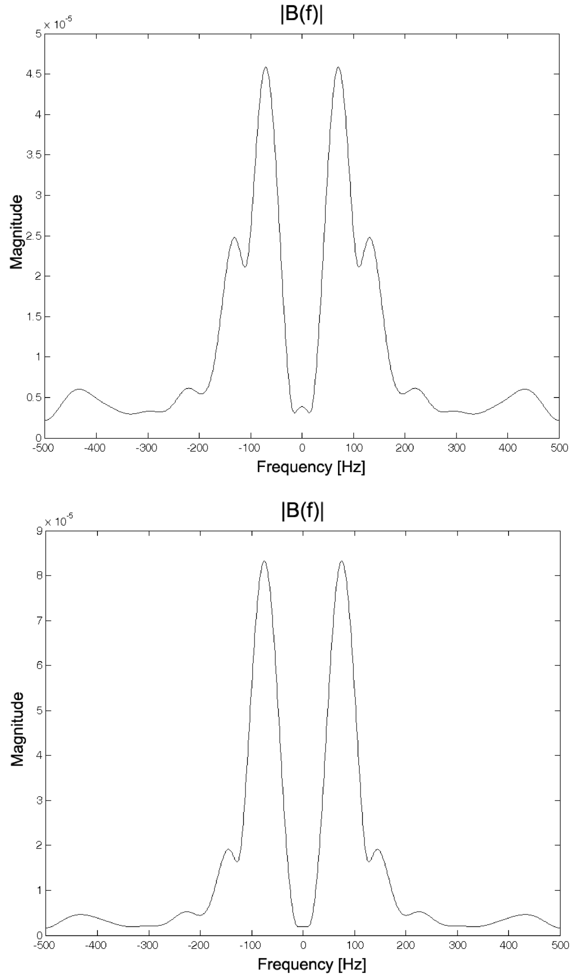

Fig. 1. a), b) Diagonal slice of bispectrum – the vowels $/a/$, $/i/$ with prolonged phonation.

4. Results and conclusions

The analysis has been applied to recordings of vowels with prolonged phonation (/a/, /e/, /i/, /u/) pronounced by a group of 22 persons (men), with correct but untrained pronunciation. The above-mentioned recordings have been already used by the authors in their previous studies and analyses [11, 12].

The algorithms carrying out the determination of fundamental tone based on the HOSA analysis have been implemented in the MATLAB environment. Sample results of such determination of F_0 are presented in Fig. 1a, b. In Table 1 detailed results are listed for F_0 determined for the $/a/$ vowel. Additionally, in Table 1 there are also results for F_0^*, determined by the EGG[1] method and example of reference results for relative error ΔF_0, determined according to formula (11), for the evaluation of fundamental tone frequency F_0, for the group of 22 persons examined.

$$\Delta F_0 = \frac{F_0(i) - F_0^*}{F_0^*} \cdot 100\%. \tag{11}$$

Table 1. Example of results for calculation of relative error of F_0 function for the $/a/$ vowel with prolonged phonation (22 persons, single utterance).

samle ID	1	2	3	4	5	6	7	8	9	10	11
$F_0(i)$ [Hz] bispectrum	118	108	107	105	116	117	110	116	98	114	145
F_0^* [Hz] EGG	120	108	104	101	120	123	106	120	92	113	140
ΔF_0 [Hz] bispectrum	1.7	0.0	2.9	4.0	3.3	4.9	3.8	3.3	6.5	0.9	3.6
sample ID	12	13	14	15	16	17	18	19	20	21	22
$F_0(i)$ [Hz] bispectrum	118	127	116	111	130	123	92	102	122	131	114
F_0^* [Hz] EGG	116	124	119	107	133	135	86	101	125	132	115
ΔF_0 [Hz] bispectrum	1.7	2.4	2.5	3.7	2.3	8.9	7.0	1.0	2.4	0.8	0.9

These preliminary results also show that higher-order spectral analysis can be reckoned as one of the available methods for determination of fundamental tone F_0. The power of speech signal is distributed to its fundamental frequency and harmonics. This makes the fundamental frequency component of power spectrum much weaker than pure sinusoidal signals. Furthermore, because the noise in a short duration sample is often not strictly white, some harmonic components of power spectrum may be higher than the fundamental frequency component. Estimation of the fundamental frequency based on the power spectrum is then a difficult task.

A valuable advantage of the HOSA method, presented in [13], is the fact that the method produces good results (increase of the S/N ratio) for the signals recorded with external interference e.g. noise. For the signal to noise distance less than 10 dB, the bispectrum function is two times more accurate (effective) than the power spectrum of the same signal. That fact may strongly support the application of this algorithm for the speech signal samples registered outside the specially, dedicated chambers or when inferior quality registering equipment has been used.

References

[1] Abeysekera S.S., *Multiple pitch estimation of poly-phonic audio signals in a frequency-lag domain using the bispectrum*, Circuits and Systems, ISCAS '04. Proceedings of the 2004 International Symposium, **3**, 469–472 (2004).

[1] The EGG method, described in [7, 11], is regarded as one of the more accurate methods used for determination of the F_0 function.

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

[3] HESS W., *Pitch Determination of Speech Signals*, Heidelberg, New York, Tokyo, Berlin 1983.

[4] MARASEK K., *Electroglottography description of voice quality*, Universität Stuttgard, 1997.

[5] NIKIAS C.L., MENDEL J.M., *Signal Processing with Higher-Order Spectra*, IEEE Signal Processing Magazine, 10–37 (1993).

[6] NIKIAS C.L., PETROPULU A.P., *Higher – Order Spectra Analysis*, New Jersey, PTR Prentice Hall, Englewood Cliffs, 1993.

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

[8] PAWŁOWSKI Z., *Phoniatric diagnostics of singing and speaking voice emission*, Kraków, Oficyna Wyd. "Impuls", 2005.

[9] SUN X., *Pitch determination and voice quality analysis using subharmonic to harmonic ratio*, Proc. of ICASSP2002, Orlando, Florida, **1**, 333–336 (2002).

[10] SWAMI A., MENDEL J.M., NIKIAS C.L., *Higher-Order Spectral Analysis Toolbox for use with Matlab*, Natick, The MathWorks Inc., 1995.

[11] WSZOŁEK W., KŁACZYŃSKI M., ENGEL. Z., *The acoustic and electroglottographic methods of determination the vocal folds vibration fundamental frequency*, Archives of Acoustics, **32**, 4 (Supplement), 143–150 (2007).

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

[13] XUDONG J., *Fundamental frequency estimation by higher order spectrum,* Acoustics, Speech and Signal Processing, IEEE International Conference, **1**, 253–256 (2000).