

Analysis of Polish Pathological Speech by Higher Order Spectrum

W. WSZOLEK AND M. KŁACZYŃSKI*

Faculty of Mechanical Engineering and Robotics, AGH University of Science and Technology
al. Mickiewicza 30, 30-059 Kraków, Poland

This paper presents a next, consecutive stage of the authors' research, concerning the search for additional signal processing methods, which could be used for objective detection and registration of pathological changes in the larynx and vocal tract area. This paper presents pathological speech analyzing by suitably directed higher order spectra analysis (HOSA).

PACS numbers: 43.70.Dn, 43.72.Ar, 43.60.-c

1. Introduction

The speech signal emitted by humans may be a source of useful diagnostic and prognostic information. An acoustic signal contains semantic and individual voice features. Except this, the signal may become, indirectly by some selected parameters, an additional source of information concerning the condition of vocal tract anatomy, as well as physiology and pathology (deformation) of human's internal organs. Research studies carried out by many authors prove, that significant amount of information concerning phonation activity of speech organ can be gathered by accurate determination of parameters related to the fundamental tone F0. It is also recognized that the knowledge of acoustic parameters based on the measured time dependence of the source's F0 parameter during the phonation process contains valuable information regarding larynx pathology, personal features as well as the physical and emotional condition of the speaker [1]. In case of the pathological speech signal analysis (most often resulting from chronic illnesses of vocal track) problem in the field of signals processing relates to separating essential information from noise. In some pathological voice samples, or in speech deformed on purpose, is a greater presence of nonstationary noise because of incorrect work of glottal, or else part of vocal track. This type of phenomenon can be detected and analyzed by suitably directed higher order spectra analysis (HOSA).

2. Higher-order statistical methods

The higher-order statistical methods, also known as cumulant methods, are related to a more popular concept of statistical moments. In the same way as the Fourier transform of autocorrelation function (power spectrum) makes a useful analytical tool, however results of the Fourier transform of cumulants, called polyspectrum, may also prove 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 [2].

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:

$$C_{1x} = E\{x(n)\}, \quad (2.1)$$

where: $E\{\cdot\}$ — denotes the ensemble expectation operator.

Higher-order cumulants (second — autocorrelation, third) of the $x(n)$ process with zero average value are consecutively defined as [3]:

$$C_{2x}(k) = E\{x(n)x(n+k)\}, \quad (2.2)$$

$$C_{3x}(k, l) = E\{x(n)x(n+k)x(n+l)\}. \quad (2.3)$$

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

$$\text{power spectrum —} \\ S_{2x}(f) = \sum_{k=-\infty}^{\infty} C_{2x}(k) e^{-j2\pi f k}, \quad (2.4)$$

bispectrum —

$$S_{3x}(f) = \sum_{k, l=-\infty}^{\infty} C_{3x}(k, l) e^{-j2\pi(f_1 k + f_2 l)} \quad (2.5)$$

Utility of higher-order statistical methods is essential for numerous practical applications. An important property of the cumulants is their values being 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 [4]. 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. An additional feature distinguishing the cumulants and poly-spectra is the fact that they contain information concerning amplitude and phase of

* corresponding author; e-mail: mklaczyn@agh.edu.pl

a given process (e.g., harmonic fluctuations, quadratic phase coupling [5]), while the correlation function and power spectrum contain only the information concerning signal amplitude [6]. The quadratic phase coupling is present in healthy voice in larger extend as opposed to deformed (pathological) speech. This fact is observed because the pathological voice is characterized by vocal tract which is more markedly linear than in healthy speech [7]. The observation of non-linear phenomenon in speech signal is possible by bicoherence function which is defined [3]:

$$B_{3x}(f_1, f_2) = \frac{S_{3x}(f_1, f_2)}{\sqrt{S_{2x}(f_1)S_{2x}(f_2)S_{2x}(f_1 + f_2)}}. \quad (2.6)$$

One-dimensional bicoherence index is defined as:

$$B_{1-\text{dim}}(f) = \sum_{f_i=0}^N B_{3x}(f, f_i). \quad (2.7)$$

3. Material

In Poland, diseases of the voice organ constitute over 20% of all cases of occupational diseases. In addition, it is estimated that half of the number of people complaining of dysphonia suffer from hypertrophy changes on their vocal folds, such as laryngeal polypus, Reinke's oedema, vocal cord nodules, laryngeal papillomatosis, coccus and contact ulcer. Nearly 60% of all malignant diseases localized in the area of the head and neck are laryngeal cancers [8].

During a study conducted in co-operation with the Chair and Clinic of Otolaryngology at Collegium Medicum, Jagiellonian University (CM UJ; Cracow, Poland) the applied test was arranged and proposed by a group of otolaryngologists and phoniatriest experts. It had been used in earlier research of authors [8, 9]. A recording of sound pressure time waveforms, $p(t)$, of a speech signal was made in the anechoic chamber of the Department of Mechanics and Vibroacoustics, AGH University of Science and Technology, and in a room for hearing examinations and audiometric measurements of the Chair and Clinic of Otolaryngology, CM UJ.

The equipment used for recording the patients' voice, consisted following parts: Microphone G.R.A.S type 40 AF, preamplifier Norsonic type 1201, Amplifier G.R.A.S type 12AA and professional digital magnetic recorder (PDR 1000; HHB Communications, the UK) for recordings in the frequency range from 20 to 20 000 Hz and a dynamic range exceeding 80 dB. The patients displayed disease changes in their vocal folds, glottis and larynx. The database of deformed speech signals comprised recordings of 96 patients treated in the Clinic. The correct speech of 128 persons, both male and female, with no pathology of voice, were recorded as a reference for standard Polish.

4. Results

Figures 1, 2 present the graphical representation of the bicoherence index (2.6) of correct speech sample data.

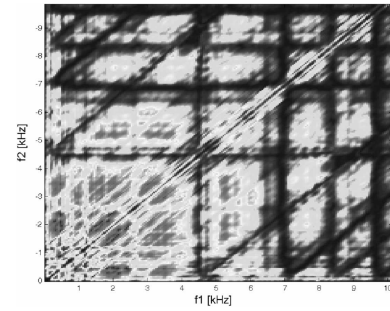


Fig. 1. Representation of the bicoherence. Correct speech. Vowel /a/ (e.g., /pat/) with prolonged phonation.

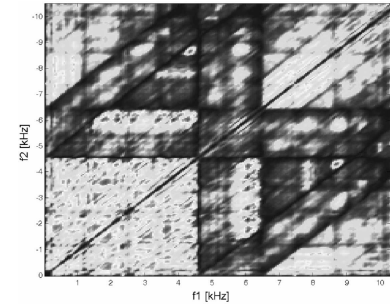


Fig. 2. Representation of the bicoherence. Correct speech. Vowel /a/ (e.g., /pat/) with prolonged phonation.

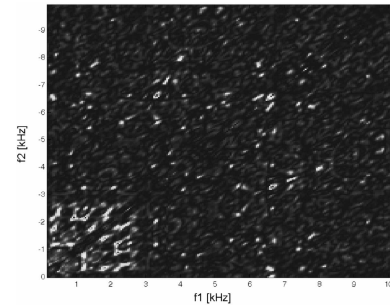


Fig. 3. Representation of the bicoherence. Deformed speech — laryngeal cancer. Vowel /a/ (e.g., /pat/) with prolonged phonation.

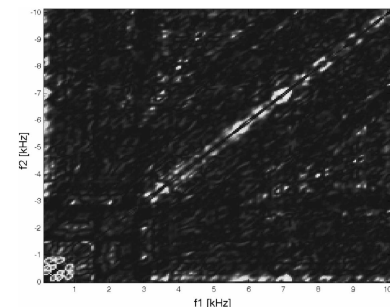


Fig. 4. Representation of the bicoherence. Deformed speech — laryngeal polypus. Vowel /a/ (e.g., /pat/) with prolonged phonation.

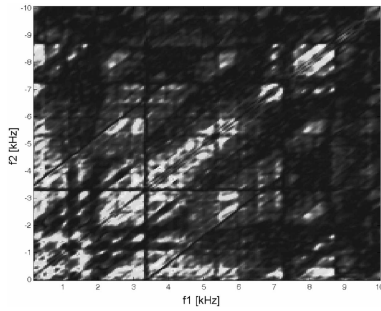


Fig. 5. Representation of the bicoherence. Deformed speech — chronic laryngitis. Vowel /a/ (e.g., /pat/) with prolonged phonation.

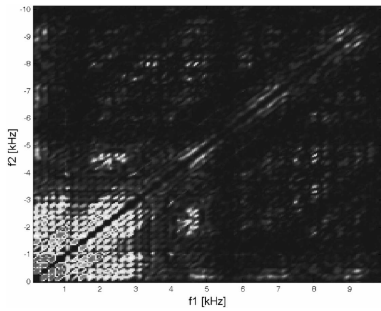


Fig. 6. Representation of the bicoherence. Deformed speech — chronic laryngitis. Vowel /a/ (e.g., /pat/) with prolonged phonation.

Figure 3–6 present the graphical representation of the bicoherence index (2.6) of pathological speech sample data.

According to formula (2.7) the one-dimensional bicoherence of healthy and pathological voice sample data is presented in Fig. 7.

5. Discussion

In correct speech there is a greater presence of quadratic phase coupling which is quantified by energy irregularity of the bicoherence index and the magnitude of

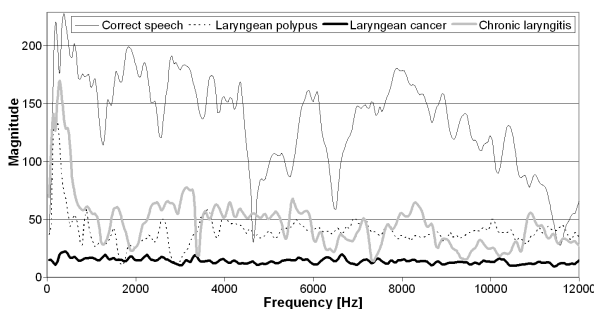


Fig. 7. Representation of one-dimensional bicoherence of voice sample data. Vowel /a/ (e.g., /pat/) with prolonged phonation.

the one-dimensional bicoherence index. The higher spectral components in correct speech (thin black line) are in opposing tendency to deformed speech. Greater number of quadratic phase coupling taking place in healthy speech. This fact could indicate that model of vocal track in normal speech is marked as more non-linear. It is possible to clearly separate such diseases as the most dangerous laryngeal cancer (bold black line) from laryngeal polypus (dotted line) and from chronic laryngitis (bold gray line), which medically and personally may be considered as a much less serious problem. This analysis, conducted among groups of patients, showed that speech pathology caused by various laryngeal diseases can be assessed using Higher Order Spectral Analysis based on acoustical signal. It is possible that this method (close to the other objective methods) would prove very useful for the automatic laryngeal pathology detection. Additionally, except the classification of the pathological speech, on the basis of received results it is possible to calculate the degree of speech deformation by using a distance metric (for example: Euclidean, Hamming, Camber, Bray-Curtis) [8, 9]. A quantitative estimation of similarity and the size of the differences of the deformation speech measure in the relation to the correct speech would have been executed by using the one-dimensional bicoherence of voice sample data, presented on the Fig. 7.

Acknowledgments

This study was supported by Foundations of Polish Science.

References

- [1] I.R. Titze, *Principles of voice production*, Prentice Hall, Englewood Cliffs, 1994, p. 301.
- [2] C.L. Nikias, J.M. Mendel, *Signal Processing Magazine IEEE* **10**, 10 (1993).
- [3] A. Swami, J.M. Mendel, C.L. Nikias, *Higher-Order Spectra Analysis Toolbox*, The Mathworks Inc., Natick 1995, p. 5.
- [4] X. Jiang, in: *IEEE International Conference Acoustics, Speech, and Signal Processing, 2000*, Vol. 1, p. 256.
- [5] Y. Kim, E. Powers, J. Edward, *Plasma Science IEEE Transactions* **7**, 123 (1979).
- [6] B. Boyanov, S. Hadjitodorov, T. Ivanov, *IEEE Electronics Letters* **27**, 2268 (1991).
- [7] J.B. Alonso, J. de Leon, I. Alonso, M.A. Ferrer, *EURASIP Journal of Applied Signal Processing: Nonlinear Signal and Image Processing* **4**, 279 (2001).
- [8] Z. Engel, M. Kłaczyński, W. Wszolek, *Int. J. Occupational Safety Ergonomics* **13**, 367 (2007).
- [9] W. Wszolek, M. Kłaczyński, *Arch. Acoust.* **30** (Supp.), 193 (2005).