The Effects of the Active Hypoxia to the Speech Signal Inharmonicity

Zoran N. MILIVOJEVIĆ[1], Darko BRODIĆ[2], Marina Z. MILIVOJEVIĆ[3]

[1] College of Applied Technical Sciences Niš, Aleksandra Medvedeva 20, 18000 Niš, Serbia
[2] University of Belgrade, Technical Faculty in Bor, V. J. 12, 19210 Bor, Serbia
[3] University of Niš, Medical Faculty, Bulevar dr Zorana Đinđića 81, 18000 Niš, Serbia

zoran.milivojevic@vtsnis.edu.rs, dbrodic@tf.bor.ac.rs

Abstract. *When the people are climbing on the mountain, they are exposed to decreased oxygen concentration in the tissue, which is commonly called the active hypoxia. This paper addressed the problem of an acute hypoxia that affects the speech signal at the altitude up to 2500 m. For the experiment, the speech signal database that contains the articulation of vowels was recorded at different altitudes. This speech signal was processed by the originally developed algorithm, which extracted the fundamental frequency and the inharmonicity coefficient. Then, they were subjected to the analysis in order to derive the effects of the acute hypoxia. The results showed that the hypoxia level can be determined by the change of the inharmonicity coefficient. Accordingly, the degree of hypoxia can be estimated.*

Keywords

Acute hypoxia, inharmonicity coefficient, fundamental frequency, speech analysis, speech processing.

1. Introduction

Hypoxia is a condition of decreased concentration of oxygen in blood, cells and tissues of an organism [1]. The brain is the most sensitive organ, so the condition of hypoxia causes disturbances of mental activities (disturbances of memory, sight and speech, etc.). Among other factors which can cause hypoxia there is inhaling the air with lowered oxygen concentration. This condition can occur in tunnels, during diving, during flights with planes, and during stay on high mountains [2]-[4]. There are: (i) acute, and (ii) chronic hypoxia conditions. Under acute hypoxia, we understand the situation when someone whose natural environment is at low altitude, is exposed the stay on high mountains. In the period of adaptation we talk about the acute hypoxia. Under the chronic hypoxia, we understand the condition of the lowered amount of oxygen in people whose natural environment is on high mountains. Refs. [3], [4] showed that the fundamental frequency of the speech signal changes due to hypoxia.

According to [5], the processing of the fundamental frequency as well as its relation to the other frequencies in the spectrum can designate the emotional state of the speaker (happiness, sadness, anger, anxiety, boredom, disgust, and neutral) and the stress [6]-[9]. Ref. [10] shows the results of digital signal processing made by filtering of dissonant spectral ranges. This approach is incorporated in order to improve quality of real speech in different ambient noise. The process included the fundamental frequency estimation, which was the base for determination of dissonant interval limit. According to that, the spectral components from defined intervals have been eliminated by filtering. The dissonant ranges known as Devils intervals are defined by the theory of music [11]. Refs. [12], [13] show that along with the fundamental frequency change, hypoxia has repercussions on the amplitude speech characteristic. This way, it changes the energy of the spectral components that belongs to the dissonant intervals.

In this paper, the influence of the acute hypoxia to the real speech inharmonicity is described. Till now, this problem hasn't been completely studied. The basic idea is connected to the research of the inharmonicity in the stringed musical instruments [14]. Although, the inharmonicity linked with the oscillations of stretched strings was well known, the effects of the musical instruments inharmonicity weren't investigated. First important results that analyzed the piano strings inharmonicity were published in [15], [16]. Up to now, the inharmonicity of different musical instruments like piano and guitar has been made [17]-[23]. These works show that due to the finite elasticity of the string the higher harmonics deviate from the position given by the integer multiple of the fundamental frequency. This deviation is modeled by the coefficient β. Hence, the higher value of β initiates the higher level of instruments inharmonicity.

Hypoxia effects on the fundamental frequency and the energy of the spectral components [12], [13]. Furthermore, it changes the frequency of the partials which leads to the inharmonicity. To determine the degree of the inharmonicity, the inharmonicity coefficient has been calculated for the persons under the effect of hypoxia. This is made by the originally proposed algorithm. Then, the real speech signals from database are processed by the algorithm. This database consists of the voices from different speakers, which live at the small altitude. These speakers pronounce the following vowels: A, E, I, O, and U at the different altitude, which are recorded and further archived in the WAV format files. From the obtained results the relation between the degree of hypoxia and inharmonicity coefficients is established.

The organization of the paper is the following one. In Section 2 hypoxia is described. In Section 3 the inharmonicity coefficient is determined according to the proposed algorithm. In Section 4 testing results and comparative analysis are presented. In Section 5 conclusions are made.

2. Hypoxia

Hypoxia is a condition of insufficient concentration of oxygen in blood, cells and tissues, which causes functional disorder of organs, the nervous system and cells. Due to insufficient concentration of oxygen cells die off, tissues decay or the function of many organs is disturbed: brain, lungs, heart, blood vessels, liver, and kidneys. Accordingly, the hypoxia can affect some organs or the whole organism. The brain is the most sensitive to lowering and insufficiency of oxygen. Hence, studying of this condition is mandatory. The causes of hypoxia can be multiplied in following cases: deficiency of oxygen in the atmosphere (staying on high mountains, during incidents in the mines in underground pits, in aviation, cosmic flights, underwater activities etc.), lung diseases, disorders of the breathing center, diseases of the blood vessels, increased need of tissues for oxygen during extreme work of muscles which usually happen to sportsmen and physical workers.

The effect of hypoxia on the organism are: (i) lowering of mental activities (indicated through short memory, forgetfulness, slow thinking, sleepiness, euphoria, headache, nausea, sight and speech disturbances and finally jerks, convulsions and coma), (ii) lowering of the working capability of muscles (manifested in slow walk, feeling of powerlessness, weak and slow reflexes, bad coordination of motor movements, bad accommodation of eyes), and (iii) depression of the respiratory center (losing consciousness, coma and death). Furthermore, the hypoxia affects the complete state of an organism and the human speech apparatus as well. Some investigations have been carried out concerning the effect of hypoxia on the speech as a consequence of changing the altitude [3], [5]. The concentration of oxygen in the atmosphere air is approximately 21% when the pressure is 100 kPa at the altitude. With higher altitude, the air density decreases as well as the quantity of oxygen. With higher altitude anyone gets less oxygen, which brings the hypoxia reinforcement. It has been demonstrated that a considerable effect of hypoxia appears even on 1600 m altitude. Accordingly, it is called the reaction threshold. However, the highest altitude where human settlements were formed is at 5500 m altitude [4]. This height is considered to be the outmost limit of human adaptability power. A healthy man/woman, who is not accustomed to this height, can preserve his full working ability and state of full consciousness for at most 30 minutes. Still, anyone can adapt to this height and to stay there for a long time. In contrast, it has been shown that it is not possible to survive for a long time above 6700 m. Consequently, no one can live for as long as 10 minutes at that altitude without any adequate equipment.

The characteristics of the speech signal influenced by the acute hypoxia have been analyzed in [3], [12], [13]. The results of the fundamental frequency analysis in two test groups have been presented, by those who live naturally at 400 m altitude and by those who live at 1600 m altitude. The analyses have been made in order to investigate the effect of acute and chronic hypoxia. The effect of the acute hypoxia has been analyzed due to measurements of the fundamental frequency in the tested persons who live at 400 m altitude, and those who live at 1600 m altitude. Furthermore, the chronic hypoxia has been analyzed on the tested persons who live at 1600 m altitude. The examinations have been performed in cases of separately uttered vowels A, E, I, O and U. It was stated that it comes to the slight increase of the fundamental frequency because of the acute hypoxia. Consequently, the effect of the chronic hypoxia led to considerable increase of the fundamental frequency, i.e. from 140 to 170 Hz. However, the effect of hypoxia to inharmonicity of the speech signal has not been analyzed in the literature so far. Hence, in this paper the results of the analysis of inharmonicity as a function of hypoxia, i.e. the elevation has been presented. In order to determine the inharmonicity limits of the speech signal, the musicological definition of inharmonicity coefficient and the realized relation to the spectrum of the speech signal will be described below.

3. Inharmonicity of the Partials

Sound source generates sound with the fundamental frequency, along with the tone aliquots. Different number of aliquots as well as theirs different relative intensity in sound determines the color of tone [11]. The aliquots are called partials. With the respect to the fundamental frequency, the partials can be with frequencies that corresponds to: (i) integer multiples f_0 (harmonics), and (ii) fractional multiples of f_0 (inharmonic) [22].

3.1 Inharmonicity Coefficient

The partials are integer multiples of fundamental frequency f_0. It is defined as:

$$f_{ph} = p \cdot f_0, \quad p = 1,2,... \tag{1}$$

where f_0 is the fundamental frequency, p is the order of the partials, and f_{ph} is the frequency of the harmonic partial. Frequency displacement of the partials from the harmonic position represents inharmonicity of the tone. It is defined through inharmonicity coefficient β. The frequencies of the harmonic and inharmonic partials are:

$$f_p = p \cdot f_0 \sqrt{1+\beta \cdot p^2} = f_{ph} \sqrt{1+\beta \cdot p^2}, \quad p = 1,2,... \tag{2}$$

Inharmonicity coefficient β depends on the type of string material that produces a tone. It is given as:

$$\beta = \frac{\pi^3 \cdot Q \cdot d^4}{64 \cdot l^2 \cdot F}, \tag{3}$$

where Q is the Young's coefficient of the material elasticity that the string is made of, d is the radius of the string, l is the length of the string, and F is the straining force.

The speech signal originates by the vocal cords vibration with the interaction of the whole vocal tract [24]. The pronouncement of the vowels (in contrast to the consonants) generates the acoustic signal which contains fundamental frequency and the high number of partials. Hence, the analysis of the amplitude characteristic of the speech signal can elaborate the effects of harmony or inharmonicity. As an example, the effects of the partial frequency displacement due to inharmonicity for the vowel A is shown in Figs. 1-3. Fig. 1 shows the speech signal in real time domain.

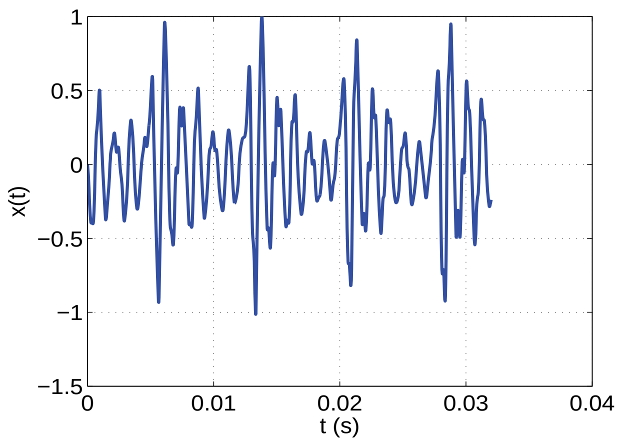

Fig. 1. Speech signal (vowel A).

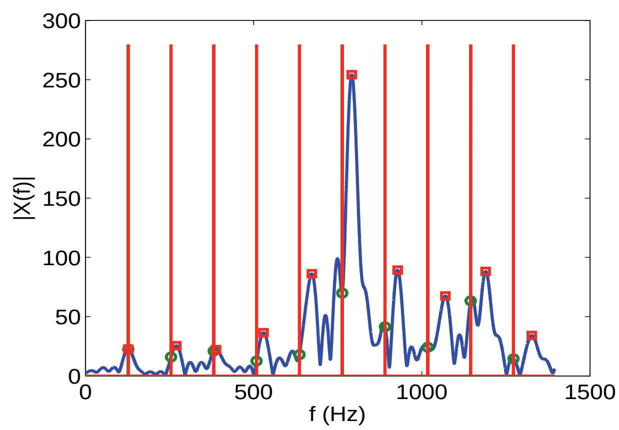

Fig. 2. Amplitude characteristics (vowel A). Vertical red lines represent the positions of the theoretical harmonic partials (10 partials).

Fig. 2 shows the amplitude characteristic of the speech signal. Vertical red lines and sign 'o' represent the positions of expected frequency of partial (harmonics), while sign '□' represents the real position of the partial (inharmonic). It can be noticed that the frequencies of partials are displaced with respect to the frequencies of harmonics (see Fig. 3 for reference). Furthermore, the difference is even more displaced for the highest order of partials. As a consequence, the speech signal is inharmonic. The degree of inharmonicity can be determined by the estimation of inharmonicity coefficient.

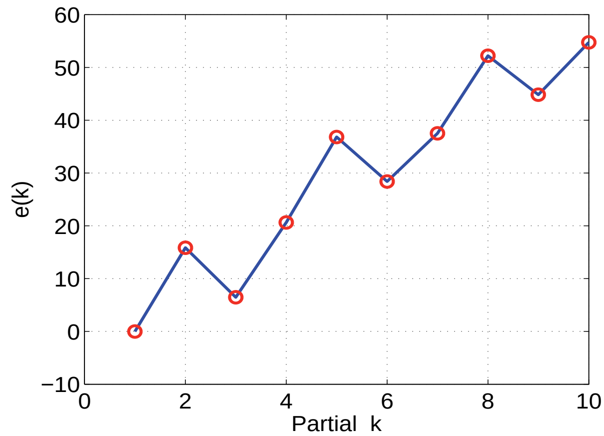

Fig. 3. The frequency difference between harmonic and inharmonicity components (vowel A).

3.2 Algorithm for the Estimation of Inharmonicity Coefficient

Many algorithms have been proposed for the determination of the inharmonicity coefficient β. Ref. [23] describes the iterative algorithm in which the estimation accuracy is increased by analyzing more partials as well as by positioning window according to the previously calculated values. Furthermore, the coefficient determination is based on the third degree polynomial, which approximates the error value. According to the polynomial coefficient, the inharmonicity coefficient is calculated. However, this algorithm is complex and computer time intensive. Due to the use of the high number of partials (up to 50), its application has been limited to the real speech signal processing. Ref. [18] described the algorithm for the determination of the inharmonicity coefficient β based on the interpolation function. Ref. [21] proposed the algorithm for the calculation of the inharmonicity coefficient β according to the frequency of two partials and without the use of the fundamental frequency. However, the proposed algorithm showed inconsistent results under test. It means that the smallest change in the estimation of the partial frequency will lead to a considerable change of the inharmonicity coefficient β.

The algorithm for the estimation of the fundamental frequency mean value and inharmonicity coefficients for the vowels of the speech signal is as follows:

Input: The speech signal **x**, the length of samples L, the length of frame N_F, the frame overlapping N_0.

Output: Fundamental frequency f_0, inharmonicity coefficient β.

Step 1: Speech signal $x(n)$, $n = 1,...,L$ is divided into frames. The frame length is given as N_F, while overlapping N_0 is enforced. The total number of frames is:

$$B = \lceil (L - N_F)/(N_F - N_0) + 1 \rceil \tag{4}$$

where $\{\mathbf{x}_b, b = 1,...,B\}$.

Step 2: It estimates the fundamental frequency f_0 and inharmonicity coefficient β. This step is given in pseudo-code and realized as follows:

FOR $b = 1$ **TO** B

a) Calculation of the frame amplitude characteristics:

$$\mathbf{X}_b = abs(DFT(\mathbf{x}_b, NDFT)). \tag{5}$$

b) Frequency partial determination by picking the maximum.

b.i Determination of the fundamental frequency f_0:

b.i.1 Picking the maximum and creation of the subsequence:

$$X_b^0 = \{x_b(k_m^0 - K),...,x_b(k_m^0),...,x_b(k_m^0 + K)\},$$

where the length is $2K + 1$.

b.i.2 Estimation f_0 by Parametric Cubic Convolution:

$$f_0^b = PCC(X_b^0, \text{PCC\_kernel}) \tag{6}$$

where $2K + 1$ represents the length of the convolution kernel.

b.ii Determination of the partial frequency f_p:

FOR $p = 1{:}N_P$

b.ii.1 Picking the maximum and creation of the subsequence

$$X_b^p = \{x_b(k_m^p - K),...,x_b(k_m^p),...,x_b(k_m^p + K)\},$$

b.ii.2 Estimation partials by Parametric Cubic Convolution

$$f_p^b = PCC(X_b^p, \text{PCC\_kernel}). \tag{7}$$

b.ii.3 Determination of the frequency partial displacement

$$e_p^b = f_p^b - p \cdot f_0^b \tag{8}$$

END (*p*)

c) Estimation β_b:

β_b=0

$$MSE\_new = \frac{1}{N_P}\sum_{p=1}^{N_P}\left(f_p^b - p \cdot f_0^b\right)^2. \tag{9}$$

REPEAT

$$MSE\_old = MSE\_new$$

$$\beta_b = \beta_b + \beta_s, \tag{10}$$

$$f_p^b = p \cdot f_0^b \sqrt{1 + \beta_b p^2}, \quad p = 1,..., N_P, \tag{11}$$

$$MSE\_new = \frac{1}{N_P}\sum_{p=1}^{N_P}\left(f_p^b - p \cdot f_0^b\right)^2, \tag{12}$$

UNTIL $MSE\_old - MSE\_new \geq 0$

END (*b*)

Step 3: It calculates mean value of the fundamental frequency and the inharmonicity coefficient as follows:

$$F_0 = \frac{1}{B}\sum_{b=1}^{B} f_0^b, \tag{13}$$

$$\beta = \frac{1}{B}\sum_{b=1}^{B} \beta_b. \tag{14}$$

4. Results and Discussion

4.1 The Real Speech Database

For testing the efficiency of the proposed algorithm (for estimation of inharmonicity coefficients), the real speech signal database has been created. Consequently, a test group, which consists of persons who lived at 200 m altitude, was composed. This group was made of 5 persons. They were males from 18 to 50 years old. The measurements were performed at different altitudes, i.e. at 200, 800, 1400, 1800, 2200, and 2600 m altitude. Every tested person articulated the following vowels: A, E, I, O and U. It has been done three times on each height with a pause of 5 minutes between the utterances. The speech signal was recorded and stored in the form of wav files. This way, the speech signal database (f_S = 44.1 kHz) was formed.

4.2 The Speech Database

The speech signal from database was subjected to the algorithm with the following parameters: T = 32 ms, N_F = 1411, overlapping $N_0 = N_F/4$, $NDFT$ = 2048, PCC_kernel: Keys, the length of the kernel $2K + 1 = 3$.

The obtained results are shown in Figs 4-7. Fig. 4 shows the trajectories of the fundamental frequency for different persons at different heights.

Fig. 5 shows the mean of the fundamental frequency (for vowels) as a function of the elevation.

Fig. 6 shows the mean of inharmonicity coefficients for vowels A, E, I, O and U.

Fig. 7 shows the mean of inharmonicity coefficients for the vowels.

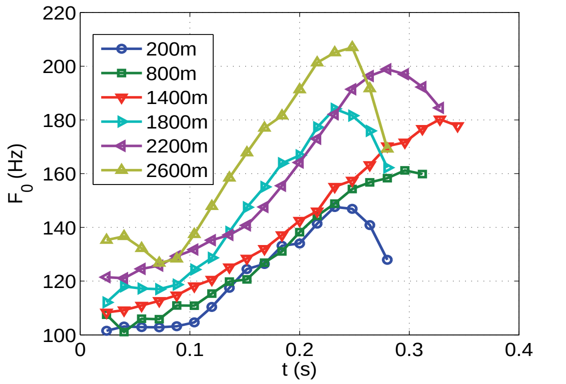

Fig. 4. Trajectories of the fundamental frequency for 200, 800, 1400, 1800, 2200 and 2600 m altitude (vowel A).

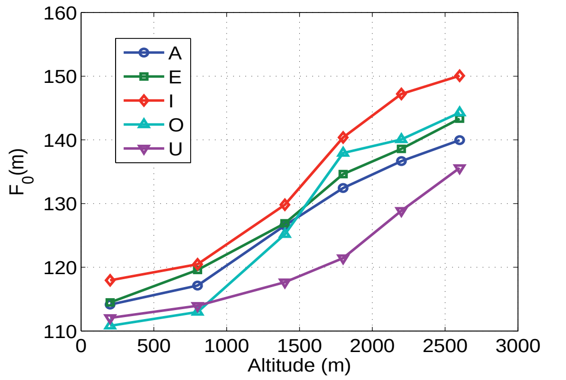

Fig. 5. Mean of the fundamental frequency for vowels as a function of the elevation.

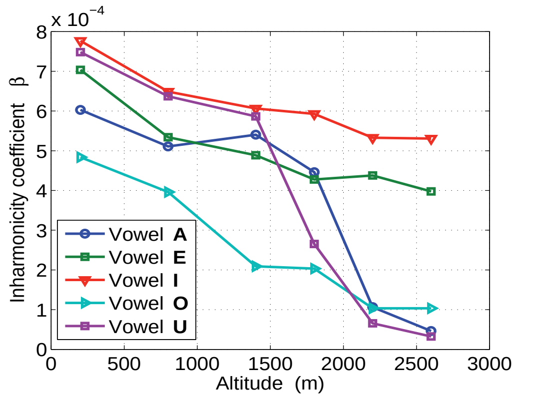

Fig. 6. Inharmonicity coefficient as a function of the elevation.

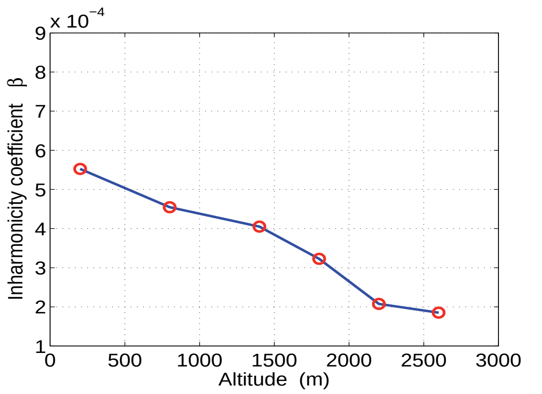

Fig. 7. Mean of the inharmonicity coefficient for the vowels.

4.3 Results and Discussion

According to the analysis of the results presented in Figs 4-7 and the results from [3] and [13] the following can be concluded: (i) with the increase of the altitude, the fundamental frequency increases, and (ii) with the increase of the altitude the value of the inharmonicity coefficient β decreases. These conclusions are obtained due to the following facts:

- In relation to 200 m altitude, the fundamental frequency increases by 2.58% at 800 m altitude, 9.95% at 1400 m altitude, 17.1% at 1800 m altitude, 21.42% at 2200 m altitude and 25.25% at 2600 m altitude.
- In relation to 200 m altitude, the inharmonicity coefficient β reduces by 17.74% at 800 m altitude, 26.67% at 1400 m altitude, 41.61% at 1800 m altitude 62.42% at 2200 m altitude and 66.49% at 2600 m altitude.

The analysis of inharmonicity coefficient β showed that it is possible to estimate the degree of hypoxia. This can be valuable in many circumstances. For instance, during the flight on great heights the pilot's mask can fail or the plane can be exposed to the decompression. In such situations it is possible to discover the signs of hypoxia by the processing the pilot's conversation with the flight control.

5. Conclusion

This paper addressed the problem of the acute hypoxia. In order to determine the degree of hypoxia the speech signal has been processed by the originally developed algorithm which is based on the analysis of the fundamental frequency and the inharmonicity coefficient. For the experiment, the speech signal database that contains the articulation of vowels was recorded at different altitudes. The analysis showed the dependence of the hypoxia from the inharmonicity coefficient. Hence, this coefficient can be used as a parameter for determination of the hypoxia degree.

References

[1] LIERE, V., EDWARD, J., CLIFFORD, J. *Hypoxia*. The University of Chicago and London, 1963.

[2] RICHALET, J., KACIMI, R., ANTEZANA, A. The control of chronotropic function in hypobaric hypoxia. *International Journal of Sports Medicine*, 1992, vol. 13, p. S22-S24.

[3] OBRENOVIC, J. Fundamental frequency of the voice in relation to hypoxia as a stressor. *Facta Universitatis, Series: Philosophy, Sociology and Psychology*, 2002, vol. 2, p. 683-689.

[4] OBRENOVIC, J., NESIC, M. Kognitivna vrednost govornog signala u uslovima akutne hipoksije (Cognitive values of speech signal under the conditions of acute hypoxia.). *Godišnjak za psihologiju,* 2003, vol. 2, p. 127-137. (In Serbo-Croatian).

[5] YANG, B., LUGGER, M. Emotion recognition from speech signals using new harmony features. *Signal Processing*, 2010, vol. 90, p. 1415-1423.

[6] MURRAY, I., BABER, C., SOUTH, A. Towards a definition and working model of stress and its effects on speech. *Speech Communications*, 1996, vol. 20, p. 3-12.

[7] RUIZ, R., ABSIL, E., HARMEGNIES, B., LEGROS, C., POCH, D. Time and spectrum related variabilities in stressed speech under laboratory and real conditions. *Speech Communications*, 1996, vol. 20, p. 111-129.

[8] DELIC, V., BOJANIC, M., GNJATOVIC, M., SECUJSKI, M., JOVICIC, S. T. Discrimination capability of prosodic and spectral features for emotional speech recognition. *Elektronika Ir Elektrotechnika*, 2012, vol. 18, no. 9, p. 51-54.

[9] SIGMUND, M. Influence of psychological stress on formant structure of vowels. *Elektronika Ir Elektrotechnika*, 2012, vol. 18, no. 10, p. 45-48.

[10] KANG, S. Dissonant frequency filtering technique for improving perceptual quality of noisy speech and husky voice. *Signal Processing*, 2004, vol. 84, p. 431-433.

[11] MILIVOJEVIĆ, Z., BALANESKOVIĆ, D. Enhancement of the perceptive quality of the noisy speech signal by using of DFF-FBC algorithm. *Facta Universitatis, Series: Electronics and Energetics,* 2009, vol. 22, p. 379-392.

[12] MILIVOJEVIĆ, Z., MILIVOJEVIĆ, M., BRODIĆ D., MILIVOJEVIĆ, D. The acute hypoxia indication by the dissonant intervals of the speech signals. *International Journal of Physical Sciences*, 2012, vol. 7, p. 334-339.

[13] MILIVOJEVIĆ, Z., MILIVOJEVIĆ M., BRODIĆ, D. The effects of the acute hypoxia to the fundamental frequency of the speech signal. *Advances in Electrical and Computer Engineering*, 2012, vol. 12, p. 57-60.

[14] ROSSING, T. *The Science of String Instruments*. New York: Springer, 2010.

[15] SCHUCK, L., YOUNG, R. Observations on the vibrations of piano strings. *Journal of the Acoustical Society of America*, 1943, vol. 15, p. 1-11.

[16] YOUNG, R. Inharmonicity of plain wire piano strings. *Journal of the Acoustical Society of America*, 1952 ,vol. 24, p. 267-273.

[17] GALEMBO, A., ASKENFELT, A. Measuring inharmonicity through pitch extraction. *Journal STL-QPSR*, 1994, vol. 35, p. 135-144.

[18] GALEMBO, A., ASKENFELT, A. Signal representation and estimation of spectral parameters by inharmonic comb filters with application to the piano. *IEEE Transaction on Speech and Audio Processing*, 1999, vol. 7, p. 197-203.

[19] JÄRVELÄINEN, H., KARJALAINEN, M. Perceptibility of inharmonicity in the acoustic guitar. *Acta Acustica United with Acustica*, 2006, vol. 92, p. 842-847.

[20] ORTIZ-BERENGUER, L., CASAJUS-QUIROS, J., TORRES-GUIJARRO, M., BERACOECHEA, J., ARANDA, J. Simple modeling of piano inharmonicity due to soundboard impedance. In *Proceeding of 120th AES Convention*, 2006, vol. 66, p. 1-6.

[21] HODGKINSON, M., WANG, J., TIMONEY, J., LAZZARINI, V. Handling inharmonic series with median-adjustive trajectories. In *12th Int. Conference on Digital Audio Effects DAFx-09*. Como (Italy), 2009.

[22] LIXIN, S., JUNXING, Z., LINGBIN, P. Piano fundamental frequency estimation algorithm based on weighted least square method. In *Proceedings of the International Forum of Information Technology and Application* (*IFITA*). Chengdu (China), 2009, vol. 2, p. 155-158.

[23] BARBANCHO, I., TARDON, L., SAMMARTINO, S., BARBANCHO, A. Inharmonicity-based method for the automatic generation of guitar tablature. *IEEE Transaction on Audio, Speech, and Language Processing*, 2012, vol. 20, p. 1875-1868.

[24] RABINER, L., SCHAFER, R. *Digital Processing of Speech Signals*. Englewood Cliffs: Prentice-Hall, 1978.

About Authors…

Zoran MILIVOJEVIĆ was born in Svrljig, Serbia, in 1959. He received his B.E.E., M.E.E., and Ph.D. degrees in Electrical Engineering from the Faculty of Electronic Engineering, University of Niš in 1984, 1994 and 2002, respectively. He is a Professor at the Technical College in Niš. His primary research interests are digital signal processing: algorithms and applications. He is the author and coauthor of over 180 journal papers (over 15 in journals indexed by Thomson SCI/SCIE JCR) and conference papers.

Darko BRODIĆ was born in Konjic, Bosna and Herzegovina, Yugoslavia, in 1963. He received his B.E.E. and M.E.E. from the Faculty of Electrical Engineering, University of Sarajevo in 1987 and 1990, respectively, as well as Ph.D. from the Faculty of Electrical Engineering, University of Banja Luka in 2011. Currently, he is an Assistant Professor at the University of Belgrade, Technical Faculty in Bor. His research interests include image processing, pattern recognition and algorithms. He is the author and coauthor of over 80 journal papers (over 20 in journals indexed by Thomson SCI/SCIE JCR) and conference papers as well as few books and the textbook.

Marina MILIVOJEVIĆ was born in Niš, Serbia, in 1982. She received M.D. from the Medical Faculty, University of Niš in 2008. Her research interest includes the effects of hypoxia. She is the author of 10 manuscripts.