elliptical model, vocal tract,
PARCOR, linear prediction,
disfluency, stuttering

Adam KOBUS[1], Wiesława KUNISZYK-JÓŹKOWIAK, Elżbieta SMOŁKA,
Waldemar SUSZYŃSKI, Ireneusz CODELLO

A NEW ELLIPTICAL MODEL OF THE VOCAL TRACT

In this paper a new model of the vocal tract is proposed. It is based on elliptical cylinders. It uses the vocal tract model based on PARCOR coefficients and midsaggital measurements of the voice tube. PARCOR coefficients were obtained from linear prediction coefficients which had been obtained by Levinson-Durbin method. Midsaggital lengths, understood as the height of a real vocal tract, were taken from X-Ray pictures, and they were averaged from the vocal tracts of a few people, who uttered the same vowels. The paper bases on Polish vowels: a,e,o,u,i,y.

1. INTRODUCTION

The voice signal is a signal which is produced by an air stream moving through the voice tract. The tract consists of the *lungs*, which bring the stream of air, the *bronchial tubes* and *trachea*, which lead the stream to the *larynx*, where the vocal cords tremble. These cords are the source of sound for the voiced parts of speech. Next, the sound is modulated in the resonance cavities, which consists of the *tongue*, *palate*, *teeth* and *lips*. A significant role is also played by the movements of the *jaw* and *cheeks*. The resonance cavities together are knowns as the resonator. In the case of the nasal sounds, the oral cavity becomes the shunt and, thanks to the appropriate arrangement of the uvula, a sound wave is emitted by the *nasal cavity* and *nostrils*. The opened lips and/or nostril constitute the last part of the vocal tract, which acts as a load that has its impedance [17].

After passing through the vocal cords, the air flow causes produces the sound named the *basic* or *laryngeal tone*. It is characterized by an extensive spectrum. In the case of the voiced noise sounds, the generated noise appears along with the basic tone. In the case of the unvoiced noise sounds, the noise totally substitutes the basic tone. The output spectrum of the sound is made by ovelapping of the laryngeal tone and/or noise characteristic of the vocal tract, where the resonances are the maxima of frequency characterizations named formants, which produces a spectrum in a shape which is dependent on the configuration of the speech organs during the expression of each sound. Each sound has a different configuration, so it allows the identification of a given sound [8,11,19].

In order to generate a sound, being a fragment of speech, it is necessary to determine the basic tone and/or noise related to it and to obtain the tract parameters which correspond to it.

The aim of modelling the vocal tract is to find an analogy between measurable acoustic attributes of the voice signal, in this case linear prediction coefficients [1,9,10], and the configuration of the vocal tract in the voice production process. That analogy can be helpful in the detection of incorrectness in the setting of the articulators in the case of malformed speech [15,18].

Hitherto prevailing cylinder model with constant wheel cross-section did not correspond to real cross-section of the vocal tract. Thus, basing on hitherto experiences, a new model is proposed. It is composed of cylinders with elliptical cross-sections. Real sections of the vocal tract were measured by means of X-Ray photos and, due to the fact, it cannot constitute the basis for speech disorders examinations [5].

In the diagnostics of pathological speech not only is it possible to establish that a speech disorder occurs, but also to find its relation to the current configuration of the vocal tract. The authors detects speech disfluencies in the speech of people who stutter with the use of linear prediction coefficients and neural networks [6,13,14,15]. The authors plan to continue these researches. Presently neural networks

[1] Institute of Computer Science, Marie Curie-Skłodowska University, Poland,
email: kobus.adam@gmail.com.

find their application in a number of domains of studies [2,3,4,16]. They will constitute an element of the speech disfluency detection, and it is planned to connect acoustic information with the information obtained from the model.

2. THEORETICAL BASICS

The basis for the discussion of the vocal tract modelling is the observation that there is a similarity between the vocal tract and a series of cylinders with variable cross-sections. Passing through the series of cylinders, the voice wave undergoes partial reflection on the cylinder connections and interferes with the forthcoming wave.

On the edge of two cylinders, the forthcoming wave and reflected wave are shown in Fig. 1 [12].

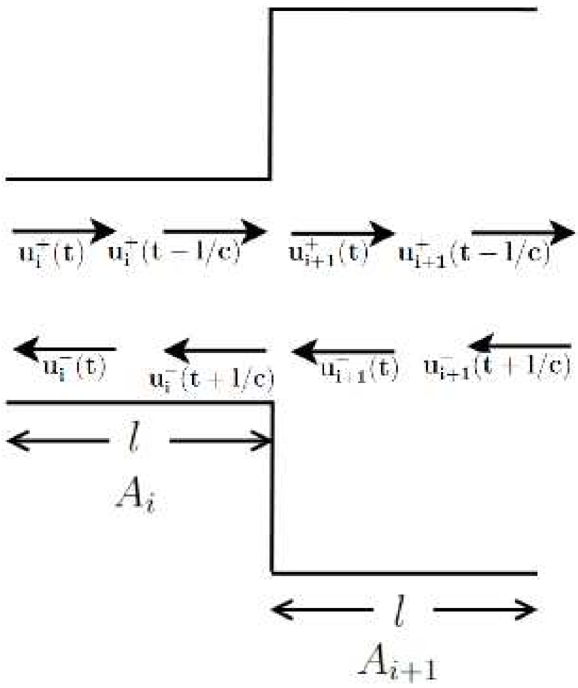

Fig. 1. A diagram of the connection of two lossless vocal tract tubes.

The transmission of sound on the edge of two cylinders can be described by the following equations [12]:

2.1.1. THE FORWARD AND BACKWARD TRANSMISSION OF SOUND

$$\begin{aligned} u_{i+1}^+(t) &= (1+r_i)(u_i^+(t-l/c)) + r_i(u_{i+1}^-(t)) \\ u_i^-(t+l/c) &= -r_i(u_i^+(t-l/c)) + (1-r_i)u_{i+1}^-(t) \end{aligned} \tag{1}$$

where u_i^+ and u_i^- are interpreted as travelling waves in the forward and backward direction from the larynx respectively, c is the velocity of sound, l is the length of the cylinder segment of the vocal tract, t is the time and the sound wave reflect coefficient r_i is described by the following equation [12]:

2.1.2. THE SOUND WAVE REFLECT COEFFICIENT

$$r_i = \frac{A_{i+1} - A_i}{A_{i+1} + A_i} \tag{2}$$

where A_i is the i'th value of cross-sectional area of the cylinder segment of the voice tube.

The vocal tract can be presented in the form of a digital filter. The transfer function of such an (exemplary) filter in the Z-transform notation is described by the following equation [12]:

2.1.3. THE TRANSFER FUNCTION

$$H(z) = \frac{\prod_{k=1}^{N} (1 + r_i) z^{-N/2}}{D(z)} \tag{3}$$

where N is the number of the cylinder segments, $D(z)$ is a polynomial recursion [12]:

2.1.4. THE DENOMINATOR OF THE TRANSFER FUNCTION

$$\begin{aligned} &D_0(z) = 1 \\ &D_i(z) = D_{i-1}(z) + r_i z^{-i} D_{i-1}(z^{-1}) \\ &D(z) = D_N(z) \end{aligned} \tag{4}$$

The reflection coefficients in the i'th pipe described by the equation 2 allow for the connection of its circuitry with the vocal tract cross-sections.

Based on that and that the system function of the i'th order inverse filter is described by the following equation [12]:

2.1.5. THE SYSTEM FUNCTION

$$A(z) = 1 - \sum_{i=1}^{p} \alpha_i z^{-i}, \tag{5}$$

where α_i are linear prediction coefficients, the system function is obtained as a result of the linear prediction analysis and can be obtained by the following recursion [12]:

2.1.6. THE SYSTEM FUNCTION AS RECURSION

$$\begin{aligned} &A^{(0)}(z) = 1 \\ &A^{(i)}(z) = A^{(i-1)}(z) - k_i z^{-i} A^{(i-1)}(z^{-1}), \\ &A(z) = A^{(p)}(z) \end{aligned} \tag{6}$$

where k_i are called PARCOR coefficients and their relation to the reflection coefficients is $r_i = -k_i$.

Thus, the partial correlation coefficient (PARCOR*) is defined and expressed by [12]:

* Partial Correlation

2.1.7. THE PARCOR COEFFICIENTS DEFINITION

$$A_{i+1} = \left(\frac{1-k_i}{1+k_i} \right) A_i, 1 \le i \le p, \tag{7}$$

where A_i are the fields of adjacent sections of the lossless tube model of the vocal tract shown in Fig. 2[12] and k_i are PARCOR coefficients.

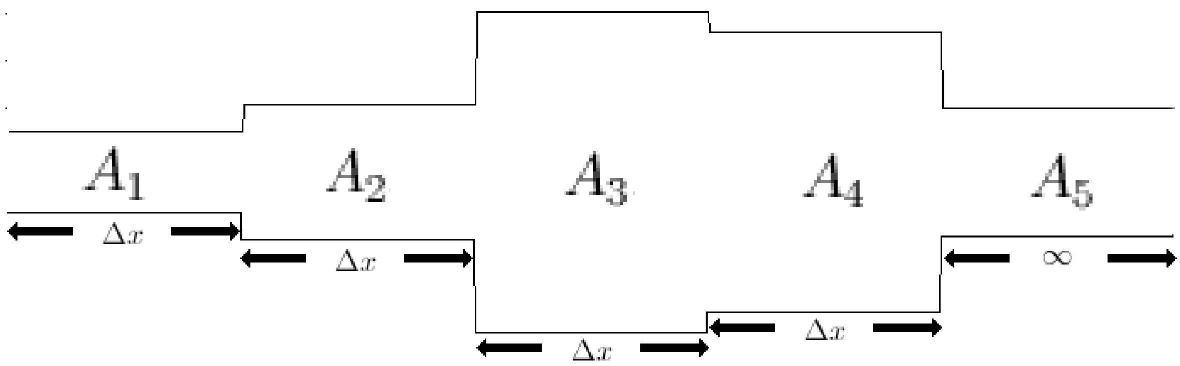

Fig. 2. Infinite lossless tube model of the vocal tract.

Each section's diameter is calculated by the equation of the base of the cylinder which is a section of such a model.

3. THE CYLINDRICAL MODEL WITH ELLIPTICAL CROSS-SECTIONS

3.1. THE IDEA OF THE ELLIPTICAL MODEL

At any given point the vocal tract cross-section is more similar to an ellipse than a circle, so in the present paper the model build of cylinders with elliptical cross-sections is proposed. For such a model, the field of i 'th segment's cross-section for such model is expressed by the following equation:

5.1.1. THE FIELD OF i 'TH SEGMENT'S CROSS-SECTION

$$A_i = \pi a_i b_i, \tag{8}$$

where a_i and b_i are the axes of an ellipse.

3.2. THE SOURCE OF THE HEIGTHS OF THE VOCAL TRACT

The axis b_i was calculated on the basis of the vocal tract cross-sections during the articulation of the Polish vowels: a, e, i, o, u, y. These heights the vocal tract were measured by means of X-Ray photos [5].

3.3. THE SOURCE OF THE FIELDS OF THE VOCAL TRACT CROSS-SECTIONS

The successive segment's cross-sections A_i where calculated with the use of the linear prediction analysis of the audio files which contained the utterances of each vowel. The Levinson-Durbin method was applied. It allowed the researchers to directly determine the PARCOR coefficients. A model composed of cylinders with an elliptical base was proposed.

3.4. NORMALIZATION OF THE FIELDS OF THE VOCAL TRACT CROSS-SECTIONS

The cross-sections A_i obtained from PARCOR coefficients were normalized with the assumption that the last cross-section at lips A_p is equal 1.

5.4.1. THE NORMALIZED FIELD OF p 'TH SEGMENT'S CROSS-SECTION

$$A'_p = \pi a'_p b'_p = 1, \tag{9}$$

Thus, each i 'th cross-section is equal $A'_i = A_i / A_p$.

Next, a_p and b_p measurements were made on the basis of labiograms[5] and cross-sections were normalized. Since $A'_p = \pi a_p \omega b_p \omega = 1$ it is true that:

5.4.2. THE NORMALIZATION COEFFICIENT

$$\omega = \frac{1}{\sqrt{\pi a_p b_p}}, \tag{10}$$

thus $a'_i = a_i \omega$.

3.5. THE SOURCE OF THE WIDTHS OF THE VOCAL TRACT

The successive widths of the vocal tract were calculated by the following equation:

5.5.1. THE SUCCESSIVE WIDTHS OF THE VOCAL TRACT

$$a'_i = \frac{A'_i}{\pi b'_i}. \tag{11}$$

The b_i values were obtained from the measurements of the real height of the vocal tract and fields A_i were obtained from the vocal tract model based on PARCOR coefficients.

4. RESEARCH RESULTS

In the experiment, model vowels a, e, i, o, u, y uttered by two women and one man were analyzed.

Calculations of PARCOR coefficients by means of the Levinson-Durbin method were made for audio files in the *wave* format for prediction orders of 15, 30 and 80 on a 512-sample window multiplied by Hanna window function. Based on that, the vocal tract model was built on it. It is composed of 15, 30 and 80 segments respectively for each of 15, 30 and 80 coefficients.

Real model measurements were done by: drawing a path of real cross-section's centers, determining successive steps at 0.17 cm intervals along it and after that measuring the widths of the vocal tract at each determined point of the path. This way, circa 80 width measurements for each real vocal tract were obtained. Such measurement made it possible to minimize the differences in the lengths of the recurved vocal tract for various walls. In Figure 3[5] the cross-section of the vocal tract during the articulation of the “a” vowel was showed.

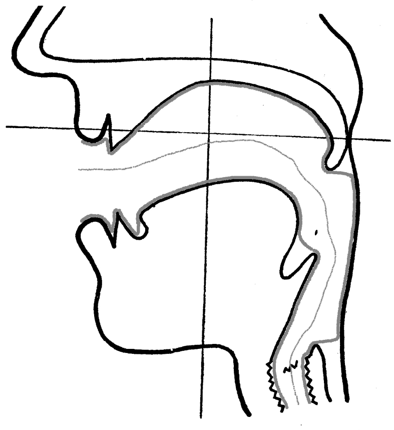

Fig. 3. The cross-section of the vocal tract for the “a” vowel.

In order to normalize the results of the measurements to the level of the fields obtained from PARCOR coefficients, the relations of height to width of lips parting on labiograms for vowels were taken. In Figure 4 the labiogram during the articulation of the “u” vowel was showed.

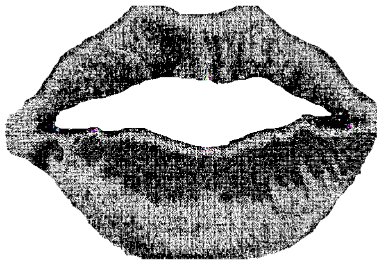

Fig. 4. The labiogram for the “u” vowel.

In literature, model settings of lips during the articulation of single vowels are known. In the elaborated model it was a cross-section of the last segment. On the basis of exemplary labiograms [5] for the utterances of the Polish vowels: a, e, i, o, u, y, a comparison was made of cross-section axes relations which were obtained based on the elaborated model and the ones measured based on the pictures of lips.

Table 1. The relations of height to width of lips parting on labiograms for vowels.

Vow	Height/width obtained from the
a	0.5043
e	0.4314
i	0.1792
o	0.5000
u	0.2353
y	0.2760

In successive figures, relative fields, heights and widths of the vocal tract of a single speaker for respective segments for the “a” vowel are presented. The numbers of the successive segments from larynx to lips are marked on the horizontal axis.

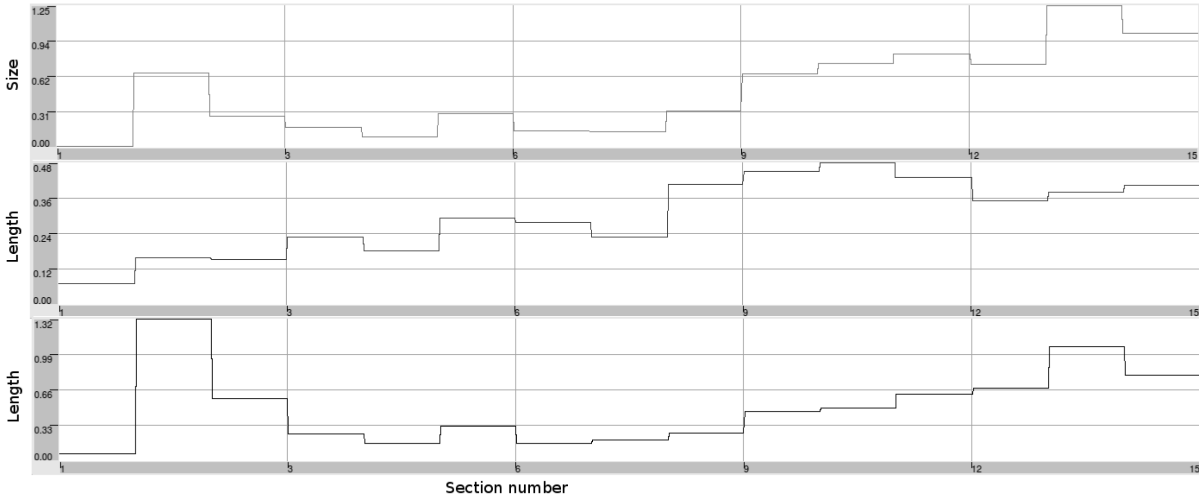

Fig. 5. The field (top graph), height (middle one) and width of the semimajor axis (lower graph) of the vocal tract cross-section for the “a” vowel for 15 PARCOR coefficients.

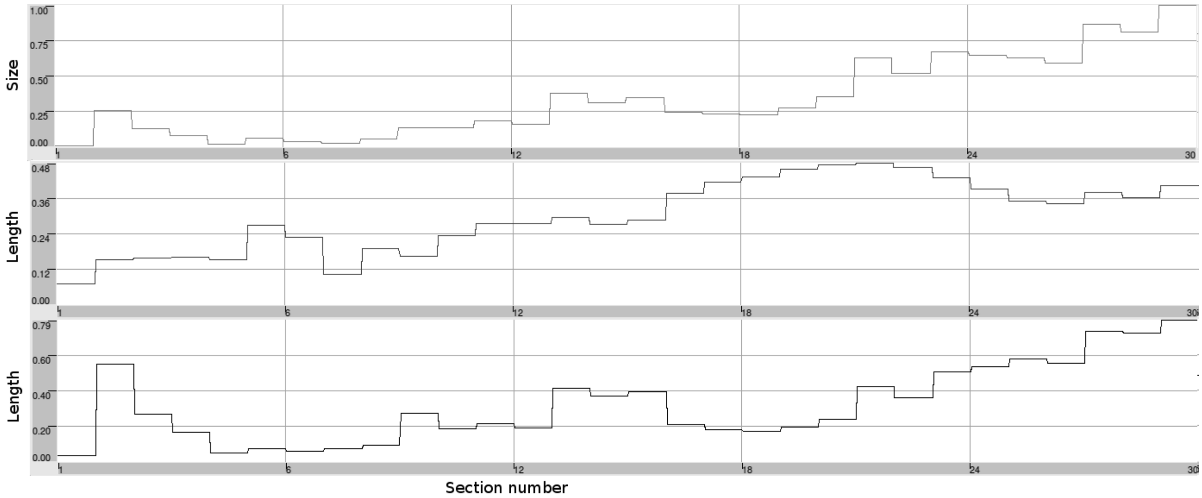

Fig. 6. The field (top graph), height (middle one) and width of the semimajor axis (lower graph) of the vocal tract cross-section for the “a” vowel for 30 PARCOR coefficients.

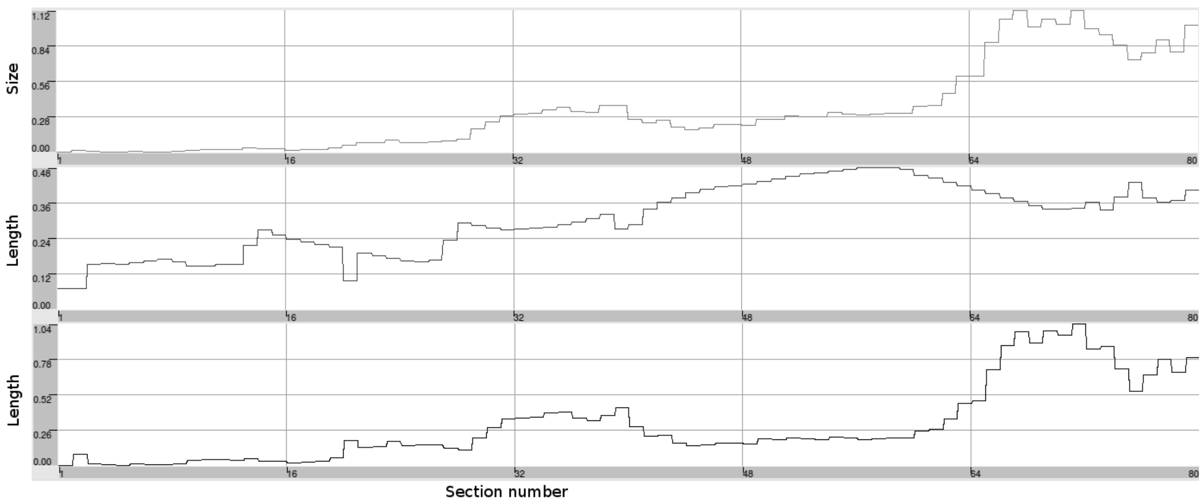

Fig. 7. The field (top graph), height (middle one) and width of the semimajor axis (lower graph) of the vocal tract cross-section for the “a” vowel for 80 PARCOR coefficients.

Graphs presented above make possible to see the tightenings and the widenings of the vocal tract in two surfaces: horizontal and vertical. Also they shows the obtained areas of the successive cross-sections. The comparison of two lower graphs in each figure, with taking into consideration the lengths of these axes located on the left side, allows to draw the model of the vocal tract.

In Figure 8 the example of the cylindrical model of the vocal tract with an elliptical base was showed.

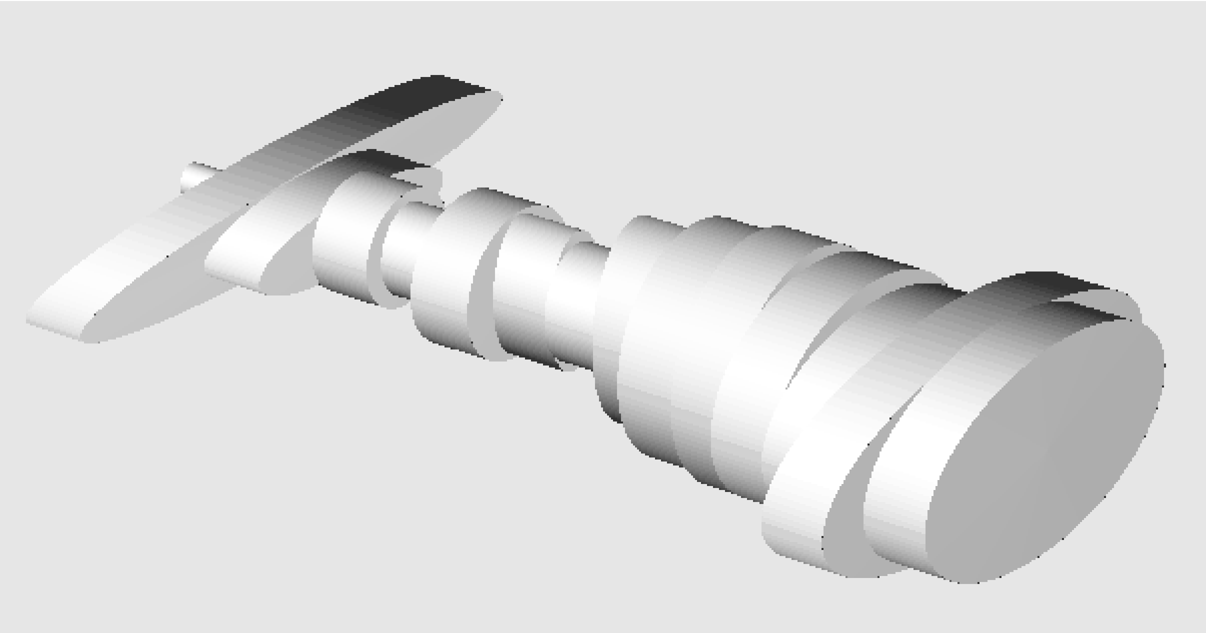

Fig. 8. The cylindrical model of the vocal tract with an elliptical base for the “a” vowel.

5. CONCLUSION

Taking the real shape of the vocal tract into consideration, it can be noted that it is more similar to the tract with an elliptic cross-section rather than circular. The simplification of the vocal tract model to the form of a set of cylinders with a circular base is appropriate, but the elliptic model, closer to a real one, is more precise.

When the model of the vocal tract is analyzed, it is possible to determine the characteristic points of the vocal tract, which causes the changes in the formant frequencies. Speech disfluency is characterized by sudden changes in these frequencies, which is caused by sudden changes in the configuration of the vocal tract[7].

When looking at the elliptic cross-sections of sounds, it can be observed that the expected changes in the tongue presentation occur, such as the similarity of the “o” vowel tract to the tract with cylinders with a circular base, or the decreasing proportion of vocal tract height to its length for the vowels: “e” and “i”. The tightening of the tract for the “i” vowel is particularly noticeable.

It is expected that future research will make it possible to determine the key points of the vocal tract configuration for that model, which will allow better detection of speech disfluencies for people who stutter.

6. ACKNOWLEDGEMENT

The scientific work co-financed from the means of the European Social Fund and national budget within the framework of the Human Capital Operational Programme. Measure 4.1. “Strengthening and development of didactic potential of universities and increasing the number of graduates from faculties of key importance for knowledge-based economy”. Submeasure 4.1.1: “Strengthening and development of didactic potential of universities". Project "Programmatic and structural teaching system reform on Faculty of Mathematics, Physics and Computer Science".

The authors thank Natalia Fedan for the language corrections.

BIBLIOGRAPHY

[1] CODELLO I., KUNISZYK-JÓŹKOWIAK W., Digital signals analysis with the LPC method, Annales UMCS Informatica, Vol. 5, Lublin, 2006, pp. 315n.

[2] IZWORSKI A., TADEUSIEWICZ R., WSZOŁEK W., Artificial intelligence methods in diagnostics of the pathological speech signals, Lecture Notes in Computer Science, Vol. 3215, 2004, DOI: 10.1007/978-3-540-30134-9 99, pp. 740–748.

[3] IZWORSKI A., TADEUSIEWICZ R., System for intelligent processing and recognition of auditory brainstem response (ABR) signals, Lecture Notes in Computer Science, Vol. 2690, 2003, DOI: 10.1007/978-3-540-45080-1 67, pp. 482-489.

[4] HORZYK A., TADEUSIEWICZ R., Self-optimizing neural networks, Lecture Notes in Computer Science, Vol. 3173, 2004, DOI: 10.1007/978-3-540-28647-9 26, pp. 150-155.

[5] KARCZMARCZUK B., Wymowa polska z ćwiczeniami, Wydawnictwo Polonia, Lublin, 1986, (in Polish).

[6] KOBUS A., KUNISZYK-JÓŹKOWIAK W., SMOŁKA E., CODELLO I., Speech nonfluency detection and classification based on linear prediction coefficients and neural networks, Journal of Medical Informatics & Technologies, Vol. 15, ISSN 1642-6037, 2010, pp. 135–144.

[7] KUNISZYK-JÓŹKOWIAK W., SUSZYŃSKI W., WARCHOŁ J., Spektrograficzna analiza niepłynności mowy, Logopedia 25, Lublin, 1998, pp. 37-47.

[8] LYONS R., Wprowadzenie do cyfrowego przetwarzania sygnałów, WKŁ, Warszawa, 2003, (in Polish).

[9] MAKHOUL J., Stable and efficient lattice methods for linear prediction, Proc. Of IEEE, 1970.

[10] MARKEL J., GRAY A., Linear Prediction of Speech, Springer Verlag, Berlin, 1976.

[11] PROKSA R., Visualization of stages of determining cepstral factors in speech recognition, Journal of Medical Informatics & Technologies, Vol. 13, ISSN 1642-6037, 2009, pp. 121–128.

[12] RABIBER L., SCHAFER R, Digital Processing of Speech Signals, Prentice Hall, New Jersey, 1978, pp. 396-461.

[13] SMOŁKA E., KUNISZYK-JÓŹKOWIAK W., DZIEŃKOWSKI M., SUSZYŃSKI W., ŚWIETLICKI M., Rozpoznawanie samogłosek w izolacji i mowie ciągłej z wykorzystaniem perceptronu wielowarstwowego, Biomedical Engineering, Kraków, 2005, pp. 143-148, (in Polish).

[14] SMOŁKA E., KUNISZYK-JÓŹKOWIAK W., SUSZYŃSKI W., DZIEŃKOWSKI M., SZCZUROWSKA I., Speech nonfluency recognition in two stages of Kohonen networks, Structures Waves - Human Health (ed. R. Panuszka), Vol. XIII, No. 2, Kraków, 2004, pp. 139-142.

[15] SZCZUROWSKA I., KUNISZYK-JÓŹKOWIAK W., SMOŁKA E., Speech nonfluency detection using Kohonen networks, Neural Computing & Applications, London, 2009.

[16] TADEUSIEWICZ R., About usefulness of neural networks in electrical engineering problems, Electrical Review (Przeglad Elektrotechniczny), Vol. 85, Nr. 2, ISSN 0033-2097, 2009, pp. 200-211.

[17] TADEUSIEWICZ R., Sieci neuronowe, Akademicka Oficyna Wydawnicza RM, Warszawa, 1993, (in Polish).

[18] WIŚNIEWSKI M., KUNISZYK-JÓŹKOWIAK W., SMOŁKA E., SUSZYŃSKI W., Improved Approach to automatic detection of speech disorders based on the hidden Markov Models Approach, Journal of Medical Informatics & Technologies, Vol. 15, ISSN 1642-6037, 2010, pp. 145–152.

[19] ZIELIŃSKI A., Teoria przetwarzania sygnałów, Wydział EAIiE AGH, Kraków, 2000 (in Polish).