Acoustic Vocal Tract Model of One-year-old Children

Milan Vojnović, Ivana Bogavac, and Ljiljana Dobrijević

***Abstract* — The physical shape of vocal tract and its formant (resonant) frequencies are directly related. The study of this functional connectivity is essential in speech therapy practice with children. Most of the perceived children's speech anomalies can be explained on a physical level: malfunctioning movement of articulation organs. The current problem is that there is no enough data on the anatomical shape of children's vocal tract to create its acoustic model. Classical techniques for vocal tract shape imaging (X-ray, magnetic resonance, etc.) are not appropriate for children. One possibility is to start from the shape of the adult vocal tract and correct it based on anatomical, morphological and articulatory differences between children and adults. This paper presents a method for vocal tract shape estimation of the child aged one year. The initial shapes of the vocal tract refer to the Russian vowels spoken by an adult male. All the relevant anatomical and articulation parameters, that influence the formant frequencies, are analyzed. Finally, the hypothetical configurations of the children's vocal tract, for the five vowels, are presented.**

***Keywords* — formant frequencies, larynx height index, maximal vowel spaces, speech therapist, vocal tract shape, vowel.**

I. Introduction

EARLY detection of atypical pronunciation of certain sounds, as well as the speech, is very important. Experience from speech therapist work shows that the biggest success can be achieved in cases of early detection of voice anomalies. From birth, motor skills and speech are the main indicators of unimpeded growth of the child. The absence or improper development of the common phases of speech communication (cooing, babbling, the pronunciation of certain phonemes, imitations etc.) points to problems not only in the development of speech, but may be due to some untypicality in other areas of a child's development.

Together with speech analysis, the synthesis of speech is also used, i.e. the modelling of speech process. The modelling of speech process allows a deeper and comprehensive voice analysis. By changing the initial conditions of the vocal tract (VT) model, one can analyze the influence of various parameters on speech. In general, the modelling of speech process involves the analysis of sound propagation through the VT. The fundamental precondition for this modelling is the knowledge of the VT geometry, i.e. its acoustic model. The transfer characteristics of the VT and its physical shape are directly related. This fact is one of the arguments for the declaration of speech as a biometric parameter and is the basis of forensic speaker identification. When the pronunciation of vowels is analyzed, then the resonance of the VT transfer characteristic is called formant frequencies. In summary, the analysis of vowel formant frequencies indirectly analyzes the VT shape, i.e., the position of articulation organs.

In the field of engineering, modelling of the pronunciation of a certain phoneme is done by analogy. The analogy is based on the identity of wave equations, which describe a short homogeneous line and a short cylindrical tube. The first step in the process of modelling is to convert VT into an acoustic model. The shape of VT is approximated by short cylindrical tubes with different cross-sectional areas. Therefore, the only information that is required is a cross-sectional area of VT as a function of the distance from the glottis. The acoustic VT model is converted into an equivalent electrical model, which is further analyzed using the standard methods of electrical circuit theory [1] [2].

In the fifties and sixties of the last century, the VT shape was obtained using X-ray imaging. During the pronunciation of sustained vowels, VT was recorded laterally and frontally. Later on, VT cross-sectional area was determined from these images. Today, magnetic resonance is used for VT shapes reconstruction.

In the case of children's speech, the X-ray or magnetic resonance imaging cannot be used for VT shape estimation. In this case, shape estimation starts from the VT data for adults. Of course, there are certain rules and restrictions in mapping an adults' VT into a child's VT.

If the estimation of children's VT shape starts from the

Paper received March 13, 2014; revised September 2, 2014; accepted September 22, 2014. Date of publication November 15, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Irini Reljin.

This paper is a revised and expanded version of the paper presented at the 21th Telecommunications Forum TELFOR 2013.

This paper is created within the project „E-Speech therapist" number TR32032 and the project of basic research „Interdisciplinary Research of Verbal Communication Quality" number OI178027, which are partly funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

Milan Vojnović, Life activities advancement center - Inovation center, Gospodar Jovanova 35, Belgrade, Serbia (e-mail: vojnovicmilan@yahoo.com).

Ivana Bogavac, Institute of experimental phonetics and speech pathology, Gospodar Jovanova 35, Belgrade, Serbia (e-mail: ivbogavac@gmail.com).

Ljiljana Dobrijević, Institute of experimental phonetics and speech pathology, Gospodar Jovanova 35, Belgrade, Serbia (e-mail: lilijen@ymail.com).

shapes of adults, the following facts have to be taken into account:

- the length of the children's VT is smaller,
- the cross-sectional areas of the children's VT are smaller,
- articulations of children and adults are different and
- there are differences in the morphological structure of the VT of adults and children.

According to the available data from references [3], the average length of VT is: 7.1 cm for a newborn, 10.5 cm for a four-year-old child, 16 cm for an adult female and 17.3 cm for an adult male.

A smaller VT in children involves a shorter length and a lower volume, i.e. smaller cross-sectional areas. A children's VT cannot be approximated by a linear reduction of an adults' VT, because there are significant differences in the shape (different morphological structures). The differences in the VT shapes come from its non-linear growth, as well as from the different articulations of adults and children. It is a known fact [4] that the ratio of the pharyngeal and oral cavity length (LHI - larynx height index) is different in children and adults. This ratio is 0.5 in newborn, and 1.1 in adult males. This means that, during the growth of a child, the pharyngeal cavity increases more than the oral cavity. This fact means that the VT shape of adults should be "compressed" in the region of the pharyngeal cavity, when a children's VT shape is estimated.

From birth, a child begins to pronounce certain phonemes, or short syllables. Some of these phonemes sound like vowels, but their discrimination is quite problematic. The pronunciation of voiced phonemes is highly centralized which, in terms of articulation, means "unstressed" articulation. This means that the articulation organs move in a limited range. More specifically, the range of VT cross-sectional area change is smaller. The VT shape of one-year-old child is "smoother" and more like a uniform tube than in adults.

These are the most important facts that should be taken into consideration when children's VT shape is estimated.

II. Influence of Vocal Tract Shape on Formant Frequencies

Before the estimation of children's VT shape, it is necessary to examine more closely the parameters that have the most important influence on its transfer characteristics. The transfer characteristics of the VT are defined by resonant (formant) frequencies. Because of that, the impact of the VT shape changes on vowels formant frequencies will be analyzed.

The whole analysis will first be done for the case of an adult VT. Initial configurations refer to the Russian vowels spoken by an adult male [1]. In the procedure of transfer characteristics simulation, the VT was modelled with losses and with infinite impedances of the VT wall, glottis and subglottis system. The radiation impedance of the mouth is approximated by a radiation circular piston set in a spherical baffle [5]. This model of VT was used in all further simulations. Formant frequencies (resonant frequencies and transfer characteristics of VT) were calculated by the FFOR program [6], which is based on algorithms given in [7]. The PRAAT program [8] was used in the case of formant frequencies estimation from real speech signals.

The first case to be analyzed is linear, percentage scaling of VT length. The formant frequencies are estimated for the following cases:

- unchanged length of VT,
- VT was reduced by 12.5%,
- VT was reduced by 25%,
- VT was reduced by 37.5% and
- VT was reduced by 50%.

The results of this analysis show that the first three vowel formant frequencies increase linearly if the length of VT is reduced. If the length of VT is reduced by 50%, formant frequencies are increased by about 100%. Among all the parameters analyzed below, the VT length change has the greatest impact on formant frequencies.

The second analyzed case is VT cross-sectional area scaling. The cross-sectional areas were reduced by the same percentage values as in the previous case: 0%, 12.5%, 25%, 37.5% and 50%. There are no significant changes in formant frequencies when the cross-sectional area of VT is reduced by the same percentage value. If the cross-sectional area of VT is reduced by 50%, formant frequencies are increased by less than 1%. In this case, as in the previous one, there is a linear relationship between the percentage of scaling the cross-sectional VT area and vowel formant frequencies change. Somewhat larger changes of formant frequencies were obtained with cross-sectional area reduced over 50%. All these results confirm the well-known fact that formant frequencies do not vary significantly with the VT volume, but only with its shape. Decreasing or increasing the VT volume does not affect the formant frequencies, if the VT length is unchanged and VT has the same shape, i.e., relations between the cross-sectional areas are unchanged.

As indicated, formant frequencies are unchanged as long as the relative ratio of the VT cross-sectional area is unchanged. However, the case of non-uniform changes in the VT cross-sectional area is much more important. With these non-uniform changes, differences in the articulation of children and adults can be simulated. Certainly, in children aged up to one or two years the ability of articulation organs movement is less than in adults. The articulation of the one- year-old child is unstressed, i.e. "shallow". In the domain of the VT physical dimensions this means that the range of cross-sectional area changes is smaller in children. The shape of the children's VT looks like a uniform tube.

The gradually changing VT shape to a uniform tube simulates the "grade" of articulation, i.e. transforming a VT into a uniform cylindrical tube of the same length. Converting a VT into a uniform cylindrical tube is done for each vowel separately on the following principle: for cylindrical segments with the cross-sectional area smaller than 5 cm^2, the area was increased in five steps (a linear percentage increase). Conversely, for cylindrical segments with the cross-sectional area larger than 5 cm^2, the area was decreased in five linear percentage steps. As a final

result of this converting, a uniform cylindrical tube with a cross-sectional area of 5 cm^2 was obtained. The gradual transformation of VT shape to a uniform cylindrical tube leads to drastic changes in formant frequencies. The first three formant frequencies gravitate to the following frequencies: 480, 1440 and 2400 Hz. These frequencies correspond to quarter wave resonances for a tube of length 17.5 cm. The mean VT length for five Russian vowels is 17.6 cm. It was noticed that the percentage changes are more significant at lower formant frequencies (the first two formants) than at higher formants.

In the simulation of different articulations of adult man and child, lips protruding is also important. Lips protruding is less expressed for children than for adults. In practice, this means that the perturbation of VT length is smaller for children in comparison with adults.

Finally, the simulation of different larynx height indices [4] is performed. This index in fact shows the length ratio of the pharyngeal and oral cavity, and it is 0.5 for the newborn and 1.1 for an adult male. In the pronunciation of different vowels, the boundaries between oral and pharyngeal cavity are not always clearly defined, so this part of simulation will be partially simplified. It was assumed that the first two centimetres above the glottis were larynx. The following eight centimetres are the pharyngeal cavity and the remainder is the oral cavity. The simulation of LHI values of 0.5 is done by "compressing" the VT shape in the pharyngeal cavity region by a factor of 0.7 and "stretching" it by a factor of 1.4 in the region of the oral cavity. Fig. 1 shows how the VT of an adult male looks for two values of the larynx height index during the pronunciation of Russian vowel /*a*/.

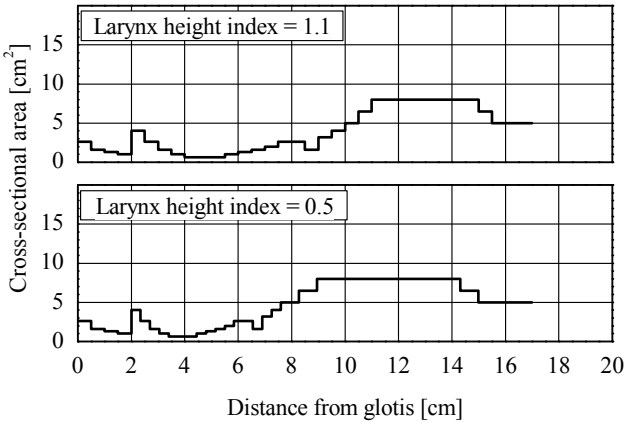

Fig. 1. VT shape of adult male in the case of pronunciation of vowel /*a*/ and two values of larynx height index.

For these two values of larynx height index, the second formants for the middle and back vowels (/*a*/, /*o*/ and /*u*/) have the biggest changes. Percentage changes of these formant frequencies are about 10%.

On the example of vowel formant frequencies analysis, it has been shown that the following four parameters are important for the estimation of the VT shape:

- length,
- cross-sectional areas,
- articulation and
- larynx height index.

If the adult VT is used as starting data, then these four parameters must be respected in the process of children's VT estimation.

III. Estimation of Children's Vocal Tract Shape

The starting point is the VT shape of an adult man for cases when he pronunces the Russian vowels. The first step in the estimation of a child's VT shape is length scaling. According to the data available in the literature [3], the average VT length of a one-year-old child is about 8 cm. To simplify estimation, a twice-shorter VT length in comparison to an adult male was chosen. This principle of shortening the VT length was applied for each vowel. The total length of VT can always be adjusted by changing the length of cylindrical segments. In this case, the length of the cylindrical segment is 0.25 cm.

The next step is to change the LHI value from 1.1 to the value of 0.5. This means that the pharyngeal cavity is compressed and the oral cavity is stretched (Fig. 1).

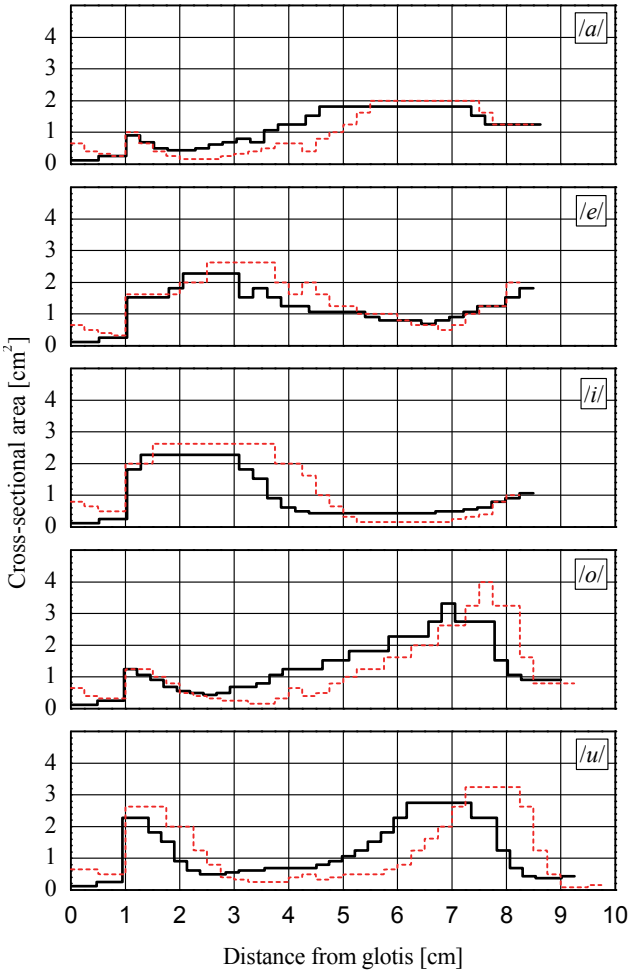

Fig. 2. Estimated VT shapes for one-year-old child (thick line) and scaled VT shapes for adult male (thin dashed line).

Children's VTs are shorter and have a smaller volume and cross-sectional area. It is taken that the cross-sectional area is four times smaller in comparison with an adult male. In addition, the larynx of one-year-old child was modelled with four 0.25 cm long cylindrical tubes with the following cross-sectional areas: 0.125, 0.125, 0.25 and

0.25 cm^2 [4]. Therefore, the same shape of the larynx is modelled in all five vowels.

In order to better simulate the centralized vowel pronunciation in children, the cross-sectional area was increased by 20% in cases where the surface was smaller than some referent (mean) value. Opposite, the cross-sectional area was reduced by 20% where the surface was greater than the referent values. The mean cross-sectional area with adult males is 5 cm^2 and 1.25 cm^2 (four times smaller) with one-year-old children. With this correction of cross-sectional area, the shape of VT is "smoother", i.e. it has become more like a uniform cylindrical tube.

With regard to different articulations in adults and one-year-old child, a correction of the VT length was made in the sense of simulating the smaller ability of lips protruding in children. The average length of an adult male's VT is about 17.5 cm. According to the criteria adopted in this paper, the average length of one-year-old child's VT is 8.75 cm (twice shorter). Limited lips protruding, in some way, involves equalizing the VT length in the case of vowel pronunciation. The following principle, of the equalization of VT length, was used: if the length of the VT, during the pronunciation of a vowel, is greater than 8.75 cm, then the VT length is reduced. On the other hand, if the length of VT is smaller than 8.75 cm, then this length is increased.

Fig. 2 shows the estimated VT shapes of one-year-old child, and the first three formant frequencies of these configurations VT are displayed in Table 1.

TABLE 1: FORMANT FREQUENCIES OF RUSSIAN VOWEL PRONOUNCED BY ADULT MALE AND SIMULATED FORMANT FREQUENCIES FOR ONE-YEAR-OLD CHILD.

Vowel formant	Adult male [Hz]	One year old child [Hz]	Frequency ratio
F_1 [*a*]	641,3	1244.9	1.941
F_2 [*a*]	1083.8	2928.5	**2.702**
F_3 [*a*]	2468.9	5035.6	2.040
F_1 [*e*]	419.4	911.5	2.173
F_2 [*e*]	1973.4	3746.6	1.899
F_3 [*e*]	2819.1	5770.3	2.047
F_1 [*i*]	226.9	689.7	**3.040**
F_2 [*i*]	2276.1	3816.5	1.677
F_3 [*i*]	3109.4	6161.7	1.982
F_1 [*o*]	504.2	1090.0	2.162
F_2 [*o*]	866.7	2438.5	**2.814**
F_3 [*o*]	2390.0	4669.1	1.954
F_1 [*u*]	236.8	784.8	**3.314**
F_2 [*u*]	599.8	1876.7	**3.129**
F_3 [*u*]	2383.0	5012.2	2.103

The VT shapes for an adult male (Fig. 2) are drawn with thin dashed lines, but with scaled length (reduced twice) and the cross-sectional area (reduced four times). This scaling is done in order to facilitate comparison of VT shapes. As it can be seen, there are differences caused by different LHI (pharyngeal cavity is compressed and mouth cavity is stretched) and differences in articulation (a smaller range of changes in the cross-sectional area and the total length of VT in the case of one-year-old child).

Data presented in Table 1 show that the ratio of formant frequencies of one-year-old child and adult males is about 2. However, there are exceptions, where the ratio is much higher:

- the first formants of vowels /*i*/ and /*u*/ and
- the second formants of vowels /*a*/, /*o*/ and /*u*/.

The frequency ratio of the second formant vowel /*i*/ is 1.68. This valus is lower than 2.

This fact shows that the estimation of vowel formant frequencies, spoken by children, cannot be made by applying the simple principle of formant frequency scaling.

IV. DISCUSSION

In previous sections, the changes in vowel formant frequencies were analysed for the four most important parameters: the VT length, VT volume, articulation and larynx height index.

Some of these parameters have a larger or smaller effect on the formant frequencies, but all of them have to be involved in the estimation of one-year-old child's VT shape. It is shown how much and in which way these parameters affect the formant frequencies of vowels spoken by an adult male. These four parameters are the basis for the estimation of the children's VT shape from the shape of the VT of adults. As a result of this whole analysis, the hypothetical VT shapes of one-year-old child were obtained (Fig. 2). The VT shapes correspond to the case of five vowels pronunciation.

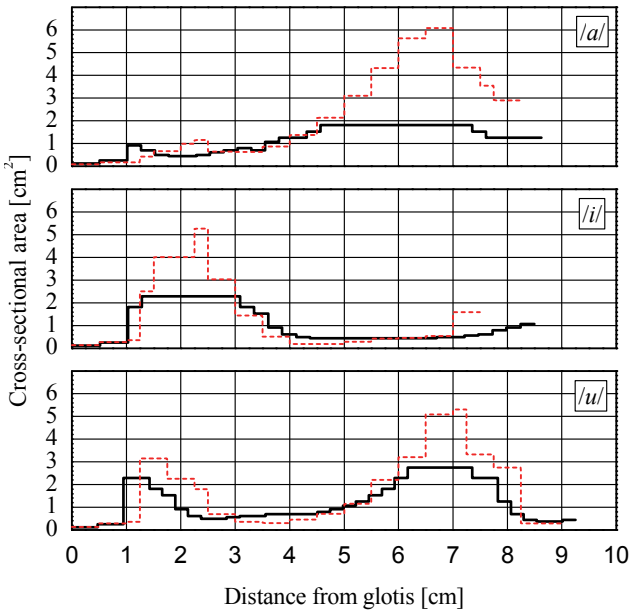

Fig. 3. Estimated VT shapes for one-year-old child (thick line) compared with the results in [4] (thin dashed line).

Shapes of the VT for the newborn [4] are slightly different from the results presented in this paper. In [4] only three VT configurations are presented: for vowels /*i*/, /*a*/ and /*u*/. Fig. 3 shows the VT configuration for these vowels (thin dashed line) together with VT configurations

resulting from this work (thick line).

The biggest differences between the estimated VT shapes are the values of cross-sectional area. They are considerably higher in the estimated configurations presented in [4]. There are significant differences in the VT length, for example vowel /*i*/. In principle, in [4] they have a shorter VT because it is a newborn's VT model. This is reasonable because it is a newborn's VT. However, the differences in cross-sectional areas are unreasonable because they should be smaller in the newborn than in a one-year-old child. It is interesting to note that the general shapes of VT are similar.

There are significant differences in the formant frequencies for VT configurations shown in Fig. 3. These differences are of the order of up to twenty percent, Table 2.

TABLE 2: SIMULATED FORMANT FREQUENCIES FOR ONE-YEAR-OLD CHILD (RESULTS FROM THIS PAPER) AND SIMULATED FORMANT FREQUENCIES FOR NEWBORN (RESULTS FROM [4]).

Vowel formant	One year old child [Hz]	Newborn [Hz]	Percentage difference [%]
F_1 [*a*]	1244.9	1542.2	23.9
F_2 [*a*]	2928.5	2843.5	-2.9
F_3 [*a*]	5035.6	5899.9	17.2
F_1 [*i*]	689.7	571.0	-17.2
F_2 [*i*]	3816.5	4752.0	24.5
F_3 [*i*]	6161.7	7804.9	26.7
F_1 [*u*]	784.8	671.9	-14.4
F_2 [*u*]	1876.7	1434.5	-23.6
F_3 [*u*]	5012.2	5859.3	16.9

In the analysis of children's speech at an early age (up to one year) the speech mechanisms are not fully developed, and the vowel pronunciation cannot be analyzed in a traditional manner. In these situations, the analysis of the formant frequencies spreading is used (maximal vowel spaces), irrespective of the fact which specific vowel is pronounced. With the VT configurations that have been proposed in [4], a significantly larger maximum vowel space is obtained, which does not correspond to reality. A larger space of vowel formant frequencies means greater freedom of articulation organs' movement.

Measurements of formant frequencies on children's real speech recordings should show which VT model is more accurate, i.e., which corresponds more to a realistic situation. The preliminary results of formant frequencies estimation of voiced phonemes, spoken by a one-year-old child, show good agreement with the simulated maximal vowel spaces [9]. The simulations of maximal vowel spaces are based on the one-year-old child's VT configurations shown in Fig. 2.

The results of real vowel formant frequencies, which are pronounced by the one-year-old-child, should be considered for possible correction of VT shape. In any case, the main corrections should be directed to the value of cross-sectional area and VT length.

V. CONCLUSION

The estimation of the VT shape of children aged one year is not an easy task, because the traditional imaging methods, such as X-rays and magnetic resonance, cannot be used. What remains is to perform estimation based on the data on the VT shape of adults. In doing so, the child's VT shape cannot not be obtained by simply scaling the VTof adults. Different anatomical, morphological, articulation, etc., parameters that affect speech must be taken into account

The paper has presented a procedure for children's VT shape estimation based on the data of adult VT shape. In order to obtain a one-year-old child's VT shape, in an adult VT shape we should: reduce the length, reduce the volume (reduce cross-sectional areas), reduce cross-sectional areas dynamics and correct a larynx height index.

Estimated VT configurations are the basis for determining the maximal vowel spaces, which can be used for the early detection of atypicality in children's speech development.

REFERENCES

[1] G. Fant, *Acoustic theory of speech production*, Mouton, The Hague. 1970.

[2] J. Flanagan, *Speech analysis, synthesis and perception*, Springer-Verlag, New York, 1972.

[3] L. Ménard, J. -L. Schwartz, L. -J. Boë, J. Aubin, "Articulatory-acoustic relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model", *Journal of Phonetics,* vol. 35, pp. 1-19, 2007.

[4] U. G. Goldstein, "An articulatory model for the vocal tract of the growing children", Thesis of Doctor of Science, MIT, Cambridge, Massachusetts, 1980.

[5] M. Vojnović, M. Mijić, "An improved model for the acoustic radiation impedance of the mouth based on an equivalent electrical network", *Applied Acoustics*, vol. 66, pp. 481-499, 2005.

[6] M. Vojnović, "Influence of the mask on acoustic and articulation speech characteristic", Thesis of Doctor of Science, School of Electrical Engineering, Belgrade, 2008 (In Serbian).

[7] P. Badin, G. Fant, "Notes on vocal tract computation", *STL-QPSR,* vol. 2-3/1984, pp. 53-108, 1984.

[8] P. Boersma, D. Weenink, "PRAAT: A system for doing phonetics by computer", *http://www.praat.org/*, 1992-2005.

[9] M. Vojnović, I. Bogavac, Lj. Dobrijević, "Vowel formant frequencies for one-year-old children", *XXI Telecommunications Forum TELFOR*, pp. 781-784, Belgrade, 2013 (In Serbian).