

MAXIMAL VOWEL SPACE METOD IN ANALYSIS OF VOWELS IN PRELINGUAL SPEECH PHASE¹

Milan VOJNOVIĆ* , Ivana BOGAVAC**, Ljiljana JELIČIĆ DOBRIJEVIĆ**2

*Life Activities Advancement Centre**

Institute for Experimental Phonetics and Speech Pathology, Belgrade

*Life Activities Advancement Centre***

The main problems in the analysis of vowels which occur in prelingual speech phase are centralization of utterance and unknown dimension of vocal tract. Most researches in this field are based on the analysis of maximal vowel space (MVS) because discrimination of vowels is very difficult in this early period. MVS analysis includes the estimation of vocal tract (VT) physical dimensions. The aim of this research was to estimate and define changes in vowel pronunciation during prelingual speech phase. The analysis and voice recording were performed in a two month old child until he turned one. The recording was performed in 42 sessions, on average 4 sessions every month. Sound segments that look like vowel pronunciation were extracted from the recordings and were used for the formant frequencies estimation by PRAAT software. The Burg method was used for formant frequency estimation. Research results showed that MVS can be used in diagnostic procedure from a child's earliest age. MVS analysis is appropriate for a child's earliest age as a child needs to pronounce individual phonemes, and does not need to respond to speech stimuli. These results need to be confirmed on a larger sample when extended analysis should define criteria for discrimination of typical and atypical formant frequencies.

Key words: *prelingual phase, vocal tract, maximal vowel space, vowel formant frequencies*

1 This work has derived from the projects „Interdisciplinary research of verbal communication quality“ No 178027, & E-Speech therapist No 32032, (2011-2014), whose implementation is financed by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

2 E-mail: lilijen@gmail.com

INTRODUCTION

Intact organic basis for language development is necessary for a child during the period of language acquisition. That refers to intact auditory perception and discrimination, visual perception, motor control of speech organs, normal intelligence, and ability to focus and maintain attention, as well as intact social interaction and stimulation. Language is acquired by stimulation imitation and practice. A child acquires basic elements of language depending on its physical and psychical abilities and needs.

Between six and nine month of age, in the prelingual period, cackles become voices. They appear in syllables groups and they are more similar to the voices of adults. However, they do not have phoneme function; they represent conditional audio-articulation practice. When a child uses the first word, these voices become mother tongue voices – phonemes that are basis for the words with meaning (Kostić et al., 1995).

A child first acquires vowels, stop consonants and nasal voices. Vowels (/a/, /e/, /i/, /o/ and /u/) of a small child stay undefined for a long time. Open and then closed vowels are first central vowels developed. Qualities of voices are constantly changing during development. Sovilj (2002) gave very precise explanation of phases in the prelingual period as a result of research in our speaking countries. The first phase of cuing appears during the second half of a child's first month. When the physiological needs are satisfied, the child is in the homeostasis and it starts to express satisfaction with singing nuclei of the future phonemes – vowels /a/, /e/, /u/ (at the beginning they are very open), /a/ middle, /e/ front /u/ back vowel and regarding tongue position /a/ is low /e/ is middle and /u/ is high. The type of vowels and order of their appearance when compared with tongue movements show that these movements are based on the act of swallowing, suckling, opening and closing the mouth. The second phase of cuing at the beginning of the second month is vocalization of certain vowels. It proceed with continuous production of neutral voice, as well as plenty variable phonics (without articulatory movement – screaming and shrieking), with high intensity and frequency of vocal vibration. In this phase, cuing is an expression of children's feelings, emotions and need for communication. It can be accepted as a medium for communication with other people. The

third phase is at the beginning of the third month. During interaction with the environment a child develops more precise forms of vowels /a/, /e/ and /u/ using resonant area. The characteristics of cuing in this period are: appearance of diphthongs, shortening of vowels duration in sequence, their linkage in a certain way – three different vowels, where the first and the second are short, and third is long /e-a-u/; – three vowels, where the first and the third are the same with the difference that the first and the second are short and the third is long /e-a-e/.

It is interesting to perceive the possibility of the earliest phase when atypical speech can be detected. In that phase there is not classical speech or communication. A child reacts to sound, visual and other stimuli from the birth, which can be one form of communication, but not in a classic way. This period might be defined as „prelingual“. Experience of working with children at the youngest age (6-7 months), as well as the use of diagnostic procedures with a goal to monitor babies from normal and risk pregnancies (Dobrijević, 2013), have shown that atypical speech and language development can be detected in the early developmental period.

During prelingual period a child pronounces different sounds that seem like vowels. Discrimination of voices which are pronounced is unreliable and therefore should be avoided (Schwartz et al., 2005). It is better to analyze formant frequency changing area of voice sounds instead which should define variation field of geometrical VT shape, and thus the degree of speech apparatus movement. Therefore, the analysis of formant frequency changing area is performed, and not the qualitative analysis of pronounced vowels. This formant frequency changing volume shows changing of VT shape volume, and thus the volume of speech apparatus movement.

Maximal vowel space (MVS) is a method of global voice analysis known from before (Boë et al., 1989; Ménard et al., 2001). The first three formant frequencies are usually analyzed via F_1 - F_2 and F_2 - F_3 charts. Perceptual vowels discrimination is not primary in MVS analysis, only their formant frequencies. It is possible to obtain estimated MVS of person's real speech. On the other hand, if VT geometrical shape is known, then his acoustical model can be made, simulating sound spreading through that acoustical structure, and at the end, getting simulated (theoretical) MVS. Comparing of these two MVS can be useful for the discrimination of atypical speech during the development of speech mechanism. This is the basic idea of

this paper: to perceive the possibility of using MVS as diagnostic method during prelingual stage of a child's development.

The first problem is insufficient amount of information about anatomical and morphological VT structure at one year old children. Classical methods for generating VT acoustical model (Fant, 1970; Flanagan, 1972) include VT X-ray imaging during the pronunciation of steady vowels. Recently, X-ray imaging has been replaced with MRI (Story et al., 1996; Soquet et al., 2002). These methods are not appropriate for a one year old child, so another one must be found for estimating VT shape. The second problem is the MVS simulation based on known VT acoustical model itself. MVS means resonant frequencies estimation of all possible VT configurations. If accrue acoustical models of VT are used, then the number of possible configuration is enormous and all of them cannot be treated. A solution for this problem is to choose some part of possible configuration which is going to be representative for all the others. Configurations that are not achievable need to be rejected from the procedure of MVS estimation.

One of the ideas is to transform VT shape of an adult to a one year old child, considering anatomical, morphologic, articulation and other differences between an adult and a child. In engineering practice, theory of analogy is used in modeling pronunciation of a certain phoneme. Acoustic model of VT is converting into an equivalent electrical model. Further analyses are performed using standard methods of the theory of electrical circuits (Fant, 1970; Flanagan, 1972).

Acoustic model of VT is in the form of short cylindrical tubes (cascade connected), with defined cross-sectional area. Therefore, the only parameter that must be known in order to realize the modeling is the VT cross-sectional area in the function of distance from the glottis. Figure 4 shows acoustical and equivalent electrical four-tuber VT model.

General problems of generating VT acoustical model of a one year old child are consider first. One procedure for estimating VT shape is proposed. As a final result acoustical model is shown, thus dependence of the VT cross-section in the function of distance from the glottis is given. The problem of MVS simulation for a grown male is done first, and after this process is elaborated and defined, it was applied to a one year old child. At the end, the real MVS was estimated for one child, whose voice was

being recorded from two to twelve months of age. This paper considers in detail the emergence and establishment of vowels during prelingual period as a hallmark of speech and language development.

METHOD

Participants

One infant and his mother participated voluntarily in the recording. They are Serbian. The child is a boy and he was born and raised in Belgrade, Serbia.

Procedure

The child's pronunciation was recorded in the participant's home, i.e. in the living room without any kind of sound absorption treatment. H4n Handy Recorder ZOOM Corporation was used with built in stereo capacitor microphones arranged in an XY pickup pattern. The recorder was held by a mother or placed in a microphone stand during recording. Because of the child's movement, the distance between the recorder and the speaker varied. The recording was done during daily life activities, with no particular tasks, during the ten month period (from November 2011. to August 2012). The recording was done in 42 sessions, 4 recordings each month on average.

The child's voice was recorded in a stereo file with sampling frequency of 44100 Hz. No forms of speech signal improvements were used, such as automatic volume control, noise reduction, intelligibility enhancement etc.

In the pre-processing phase, all signals were cleaned from unwanted noises, converted to mono type signal and re-sampled to the frequency of 22050 Hz. In the next phase, recordings were extracted where the child pronounces vowels, or voices that sound like vowels. For these extracted segments formant frequencies were estimated with PRAAT software (Boersma, Weenink, 1992-2005). Burg method was used for formant frequency estimation (Anderson, 1978). This method of formant frequencies estimation is a classical procedure in speech signal analysis. Accuracy of

formant frequency estimation depends on many factors and it is not always ideal. Because of that, after software estimation, crosscheck with wideband spectrogram was done in order to obtain the highest possible confidence in the estimation of formant frequencies. All estimated formant frequencies that did not match wideband spectrogram and that were in the wrong order were rejected. This additional correction of formant frequency estimation needed to be done, because estimation error increases if fundamental frequency is high (van der Stelt et al., 2005).

Vocal tract shape in a one year old child

If we want to estimate the VT shape of children based on data for adults, we should keep in mind the following facts: the length of children VT is smaller, the cross-sectional areas of VT are lower in children, articulation of children and adults differs, and VT morphology of adults and children is different.

According to the literature (Ménard et al., 2007), 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. VT length for a one year old child is 8 cm, but it is suitable to take the half of the VT length of an adult. Smaller VT in children implies a smaller length and lower volume, i.e. smaller cross-sectional area. Comparing to an adult, VT cross-sectional areas of a one year old child are four times smaller. A linear scale of adults' VT cannot approximate the children's VT, because there are important differences in the shape (different morphological structure). It is well known (Goldstein, 1980) that the length ratio of the pharyngeal and oral cavity (LHI – Larynx Height Index) differs in children and adults. This ratio is 0.5 in a newborn and 1.1 for adult men. This means that during a child's growth, the pharyngeal cavity increases more than oral cavity. This fact means that the shape of VT in adults has to be „compressed“ in the region of the pharyngeal cavity, if one wants to estimate the shape of VT in a child. Pronunciation of vowels in a one year old child is significantly centralized and very similar in perceptual domain. Discrimination of spoken vowels is quite complicated (Schwartz et al., 2005). In terms of physical VT shape, this means that the dynamic of cross-section area changes will be smaller and the shape of VT will look like a uniform tube.

All of these listed factors must be taken into account when modeling VT in children in order to obtain sufficient precision of formant frequencies variation space. The significance of every single factor is analyzed through formant frequency changing. VT shapes simulation of an adult has been done to show the degree of formant frequencies changing. The whole analysis has been done for an adult.

During the stimulation, we used the VT model with losses in which the impedances of the VT wall, glottis and sub glottis system are infinite. Radiation impedance is approximated by radiation circular piston set in a spherical baffle (Vojnović et al., 2005). Starting from Fant's vowels (Fant, 1970), the formant frequencies are estimated for the following cases:

- Unchanged length of VT, than VT length was reduced by 12%, 25%, 37.5% and 50%.
- Unchanged VT cross-sectional area, than the cross-sectional area of each cylindrical segment (tube) is reduced by 12.5%, 25%, 37.5% and 50%.

Formant frequencies were calculated by program FFOR (Vojnović, 2008), which is based on algorithms given in (Badin et al., 1984). This model of VT was used in all subsequent simulations.

Formant frequencies do not vary significantly from the VT volume, but only in its shape (Vojnović, 2013a). However, insignificant influence of VT volume on formant frequencies should not be neglected, because there are some other problems related to MVS simulation.

In children under the age of one the shape of VT is like a uniform tube and the mobility of articulation organs is limited. Because of that, there are no significant differences in the VT shape when a child pronounces different vowels. In the domain of VT physical dimensions, the range of cross-sectional area change is smaller in children.

The „grade“ of articulation is simulated by the gradually changed VT shape to a uniform tube; transforming VT to a uniform cylindrical tube of the same length. This transformation 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 tube length of 17.5 cm. The mean VT length for five Russian vowels is 17.6 cm. The percentage changes are more

significant at lower formant frequencies (the first two formants) than the higher formants (Vojnović, 2013a).

In the simulation of different articulations of an adult man and a child we must not forget protruding lips. In practice this means that, together with a decrease in cross-sectional area perturbation (shallow articulation), there is a smaller range of VT length (less lips protruding). Finally, a simulation of different larynx height index (LHI) (Goldstein, 1980) is done. The process of converting VT with different LHI is illustrated in Figure 1. The first diagram shows the initial (unchanged) VT shape during the pronunciation of vowels /a/ (Fant, 1970). On the second diagram, the lengths of all cylindrical segments in the range from 2 cm to 10 cm are increased 0.7 times, and the length of remaining cylindrical segment is increased 1.4 times. The lengths of the first four cylindrical segments (range from 0 cm to 2 cm) are unchanged (Vojnović, 2013a).

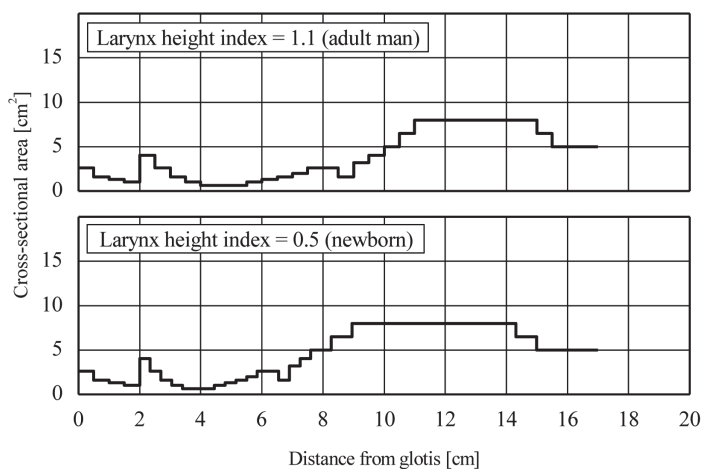


Figure 1 – Dependence of the VT cross-section area in function against distance from glottis for two larynx height index values (pronunciation of Russian vowel /a/)

The second formants for the middle and back vowels (/a/, /o/ and /u/) have the biggest changes in the simulation of different LHI values. The changes of these formant frequencies are about 10%. Accordingly, this parameter has significant influence on vowel formant frequencies and it should be incorporated in the process of generating VT acoustical model of a child.

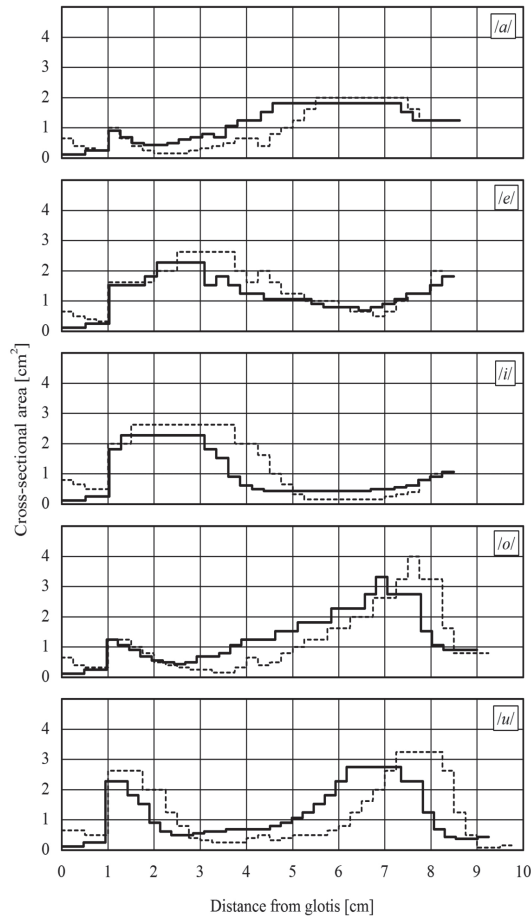


Figure 2 – Estimated VT shapes of one-year-old child in the case of vowels pronunciation (thick line) and the scaled VT shapes of adult male (thin dashed line)

All analyzed parameters shows significant influence on formant frequencies and they must be considered in the estimation of VT geometrical shape in a one year old child. 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 less than some referent (mean) value and vice versa. Mean cross-sectional area in an adult male is 5 cm^2 and 1.25 cm^2 ($5/4 = 1.25$) in a one year old child. 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 children, a correction of the VT length was made in the sense of simulating less ability of protruding lips in children. The average length of an adult male VT is about 17.5 cm. According to the criteria adopted in this paper, the average length of a one year old child VT is 8.75 cm ($17.5/2=8.75$). Limited lips protruding, in some way, involve equalizing the VT length in the case of vowel pronunciation. We use the following principles of equalization of VT length: If the length of 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 its length is increased.

Figure 2 shows estimated VT shapes of a one year old child (solid line), for five vowels pronunciation. VT shapes for an adult male are presented 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 compare VT shapes easily. As it can be seen, there are differences caused by different LHI and differences in articulation (less range of changes the cross-sectional area and the total length of VT in the case of a one year old child).

The shape of VT for a newborn (Goldstein, 1980) is slightly different from the results presented in Figure 2. In (Goldstein, 1980) only three VT configurations are presented: for vowels /i/, /a/ and /u/. Figure 3 shows VT configuration for these vowels together with VT configurations estimated in Figure 2.

The biggest differences between the estimated VT shapes are the values of cross-sectional area. They are considerably higher in estimated configurations of (Goldstein, 1980). There are significant differences in the VT length, for example vowel /i/. In principle, in (Goldstein, 1980) VT is shorter because it is a newborn VT model. There is a big difference in cross-sectional area of oral cavity, even three times greater for vowel /a/. Vowel /u/ also shows differences in cross-sectional area of oral cavity. Cross-sectional area of the pharynx cavity is different, according to (Goldstein, 1980) it is two times greater, for vowel /i/. It is clear that VT configurations from Figure 2 are hypothetical and can be changed. They are introduced so that MVS can be simulated. Based on the similarity of simulated and

real MVS, validity of hypothetical configurations should be estimated. If these two areas of formant frequencies are overlapping, that is enough for accepting the proposed configurations. On the other hand, it is possible to correct them so that these two MVS can be complied.

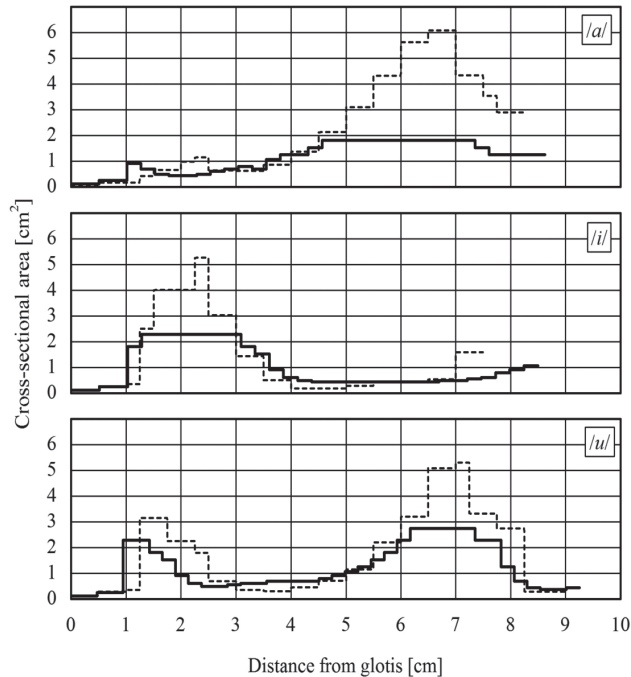


Figure 3 – Estimated VT shapes in a one year old child (solid line) compared with data from the results (Goldstain, 1980.) (thin dashed line)

Maximal vowel space

MVS represents the frequency area of the first three vowel formant frequency. It should distinguish two main types of MVS estimation:

- Estimation based on the analysis of real speech (pronunciation of vowels) and
- Estimation based on simulation of vowel pronunciation.

In order to define the borders of vowel formant frequencies in real speech, the first step is to do MVS estimation based on simulation of vowel

pronunciation. This requires knowing the shape of the VT and its acoustical model. It is the proposed process for the VT in a one year old child, and Figure 2 shows VT configurations for pronunciation of five vowels.

The problem of MVS simulation will be considered on a case of an adult male. There are an enormous number of configurations, so the strategy for selection must be defined. VT configurations for an adult male during the pronunciations of Russian vowels are well known (Fant, 1970). It is easier to verify simulation procedure for this particular case. When this procedure is verified for an adult male, it is applied to the case of a one year old child, thus the case configurations from the Figure 2.

If we assume that the average length of an adult male VT is 17.5 cm, this means that his VT can be modelled with 35 cylindrical tubes with cross-sectional area in the range from 0.16 to 16 cm². According to (Fant, 2004), it should take the discrete logarithm distribution of cross-sectional area, so that this range has 16 different values of the cross-sectional areas. This means that we should analyze about $16^{35} \approx 1.4 \cdot 10^{42}$ different acoustic configurations, which is, of course, impossible. From this huge number of possible VT configurations we should choose a small part, while at the same time it remains representative of the entire set. This is the main problem in estimating the MVS.

MVS estimation of an adult male

Before we get to the problems of MVS estimating for accurate VT models, a four-tube VT model (Fant, 1970; Flanagan, 1972) will be considered. This is an acoustic VT model presented on the left side of Figure 4. The acoustic model is completely determined by length (l) and cross-sectional area (A) of every cylindrical segment particular. Each of these cylindrical segments can be transformed into an electrical model by symmetrical T-network. T-network impedances (Z_a and Z_b) are defined by VT physical dimensions: cylindrical segments lengths and cross-sectional areas. In this model (Figure 4), the first tube simulates mouth opening, the second tube mouth cavity, the third tube represents narrowing the „tongue-palate“ and the fourth tube simulates pharyngeal cavity.

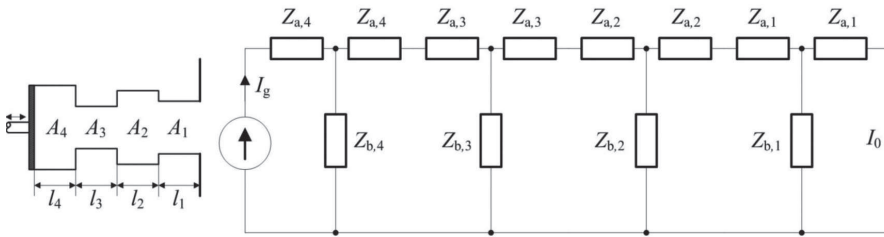


Figure 4 – Acoustic and electric four-tube VT model

If we assume that, the total length of VT is 17.5 cm and the lengths of four cylindrical segments are equal, then each tube shall be the length of 4.375 cm. For 16 different values of the cross-sectional areas, the total number of possible combinations is $16^4 = 65536$. This is not a huge number of configurations; therefore all of them can be used in the MVS simulation process. A more detailed analysis can be found in Vojnović (2013b).

On the example of a four-tube VT model, it is shown that at MVS simulation:

- VT model should incorporate mouth opening radiation impedance,
- Simulation should be done for three different VT lengths minimum, which correspond to the vowels: /u/, /a/ and /i/.
- It should use more precise VT modeling (with more cylindrical segments, 5 mm length) and
- Cross-sectional area values should be chosen by logarithmic distribution.

A more accurate VT model implies shorter cylindrical segments, the length not exceeding 5 mm. The number of configurations for this VT model is huge so the some kind of random sampling method has to be used (Monte Carlo method).

The results of the first examples of MVS estimation for a more accurate VT model are shown in Figure 5. VT shapes were chosen randomly and there are 3'50,000 in total. The three groups of estimated resonant frequencies are presented that correspond to three VT lengths: 16.5, 19.5 and 17 cm. These lengths correspond to the pronunciation of Russian vowels: /i/, /a/ and /u/, respectively. Three spaces of resonance frequencies are marked with different shades of grey:

- Light grey – VT length of 16.5 cm (vowel /i/),

- Dark grey – VT length of 17 cm (vowel /a/) and
- Black – VT length of 19.5 cm (vowel /u/).

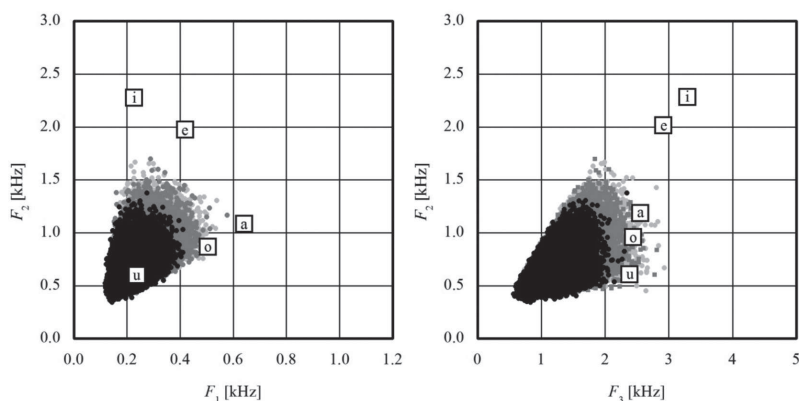


Figure 5 – VT resonant frequencies modelled with four cylindrical tube 5 mm length and 16 different values of the cross-sectional area. Three VT lengths: 16.5 cm (vowel /e/), 17 cm (vowel /a/) and 19.5 cm (vowel /u/)

There is a clear distinction between the spaces of the resonant frequencies for VT length of 19.5 cm (marked with black dots) and the length of 17 cm (dark grey marked points). The space of resonant frequencies for 16.5 cm VT length (light grey marked points) and 17 cm is very similar so it can be seen as a clear boundary between them. It is more important to say that vowels /e/ and /i/ are not covered with estimated spaces. Vowels /a/ and /o/ are on the very border. Figure 5 shows the reason for doing MVS estimation for an adult first. It can be seen that choosing VT configuration roughly is not adequate. Real formant frequencies (white squares) are not included in MVS, and they should have been. The conclusion is that VT configurations that were picked are not representative.

Due to the huge number of possible VT configurations and their random sampling, the analysis does not include the real (or close to them) VT configurations that correspond to the pronunciation of the back and middle vowels. It just shows that estimation should be adapted to the real situation, i.e. allow sampling of only those configurations that are possible in real speech.

After series of analyses (Vojnović, 2013b) all of the five vowels are in the simulated space. Only the third formants of the vowels /u/ and /i/ are

not „deep“ in the simulated space. The other formant frequencies of the Russian vowels are „deep“ in the simulated spaces. Figure 6 shows the results that coincide with (Carré, 2009) although the resonant frequencies area was analyzed and simulated with a different method.

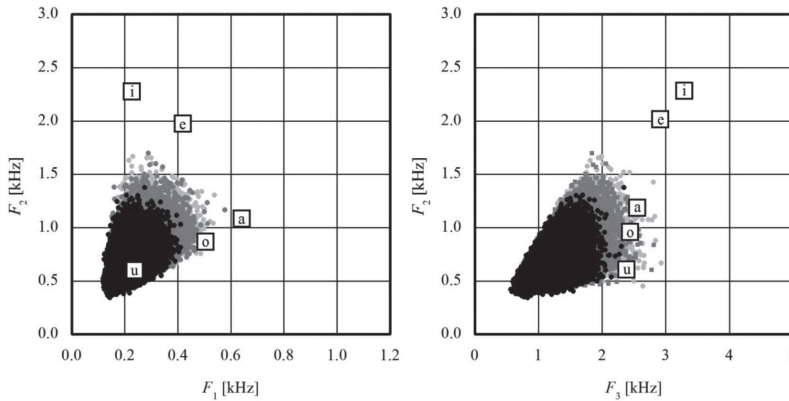


Figure 6 – Resonant frequencies modelled with four cylindrical tubes 5 mm length with disabled rapid change in the values of the cross-sectional areas

MVS estimation of a one year old child

Previously shown MVS stimulation procedures were applied to the case of a one year old child whose VT configurations are shown in Figure 2. The results of this simulation are shown in Figure 7. As it can be seen, formant frequencies are within the simulated space. Therefore, there is good compatibility between suggested VT shape of a one year old child (Figure 2) and simulated MVS (Figure 7). This compatibility implies that this sampling strategy of VT configurations is good. That still does not confirm that the proposed VT shapes (Figure 2) are good and match the real situation. The research presented in (Goldstein, 1980) gives a somewhat different VT configuration during pronunciation of three vowels: /u/, /a/ and /i/. The VT configurations in Figure 3 are related to the infant VT. The biggest difference between VT configurations in Figure 2 and Figure 3 refers to cross-sectional area. In some regions of VT, the cross-sectional area is about two to three times higher in (Goldstein, 1980), although they relate to an infant. Larger changes in a newborn VT cross-sectional area caused a significantly higher MVS (Figure 8).

First frequency formant changing area is greater in (Goldstein, 1980) for 450 Hz. The second and third formant have greater changing, about 950 Hz – 1100 Hz, respectively. In percentage, that is 37% for the first formant, 28% for the second and 21% for the third formant. Roughly speaking, MVS area of the first and second formant is greater for 75% in (Goldstein, 1980). For the second and third formant MVS area is 50% greater. These numbers confirms significant differences in the MVS estimation when VT configurations are used from Figure 3.

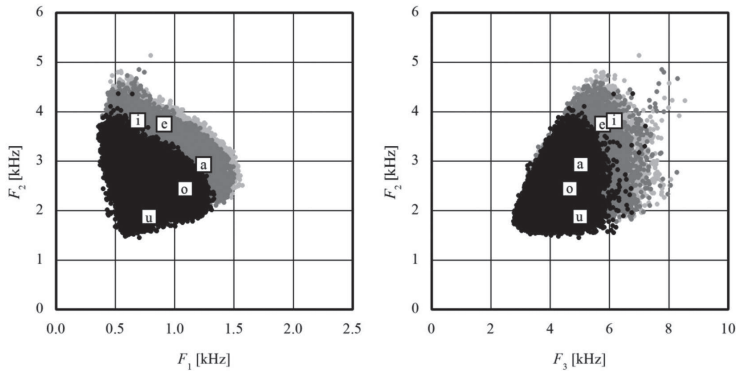


Figure 7 – Estimated MVS of a one-year-old child

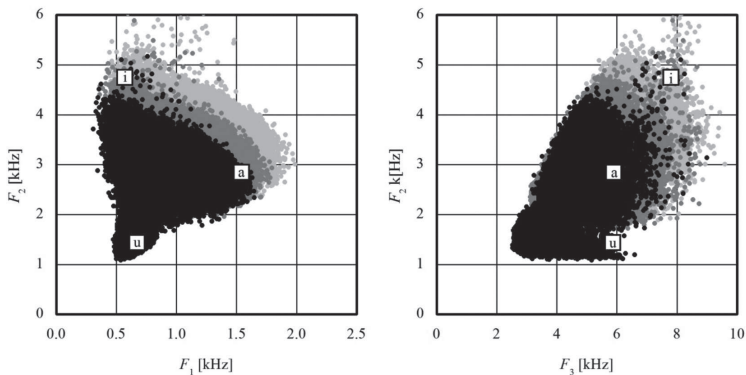


Figure 8 – Estimated MVS for a newborn according to the data of the VT shape (Figure 3) (Goldstein, 1980)

Real confirmation of the validity of hypothetical VT models can be verified by analyzing the voice of a one-year-old child and the MVS estimation based on these data.

RESULTS AND DISCUSSION

Simulated MVS from Figure 7 is strongly linked with VT configurations in Figure 2. These results can be used for program estimation of formant frequencies. If the vowels are pronounced by a one year old child than F1 is in the range from 400 to 1600 Hz, F2 is in the range from 1350 to 4000 Hz and F3 is in the range from 2500 to 8000 Hz. Mean values of the first, second and third formant in estimated MVS are: 1000, 3175 and 5250 Hz respectively. For the VT length 8.25 cm and sound speed 35300 cm/s, the first four quarter wave resonant frequencies are: 1070, 3210, 5350 and 7490 Hz, what is relatively close to the estimated MVS. If it is necessary to analyze the first three formant frequencies of vowels that a one year old child pronounced, then the upper limit frequency needs to be set up on 6500 Hz.

Figure 9 shows the results of formant frequencies estimation for real speech of a one year old child. Every single chart represents theoretical and estimated (real) MVS of the child between two and twelve months of age. Gray dots represent space of simulated resonant frequencies, and black dots represent estimated resonant frequencies of the child's speech recordings.

It should be noted that all charts are very similar, except the chart for the seventh month. This chart has small number of dots (recordings are short), which can be the reason for the differences. Chart similarity in Figure 9 shows that articulation capability of a child is very big and that it does not change much during the first year of a child's life. Matching is good for the first formant, slightly lower for the second formant and the weakest for third formant. Theoretical and experimentally obtained MVS are well matched in general. This fact indicates that estimated VT shapes (Figure 2) are very similar to real VT shapes of a one year old child.

In some charts (fourth and eighth month) slightly wider first formant frequency space can be seen, comparing with simulated (theoretical) space. The reason for this is recording of the child with higher fundamental frequency (F_0): squeal, scream, and cry. It is known that in these cases reliability of formant frequency estimation decreases (van der Stelt et al., 2005). Partial discrepancy of the second and third formant frequency space should be analyzed in more detail. That analysis should define is the real MVS more narrow than simulated MVS, or the child at that age hardly or rarely pronounce front vowels: /e/ and /i/.

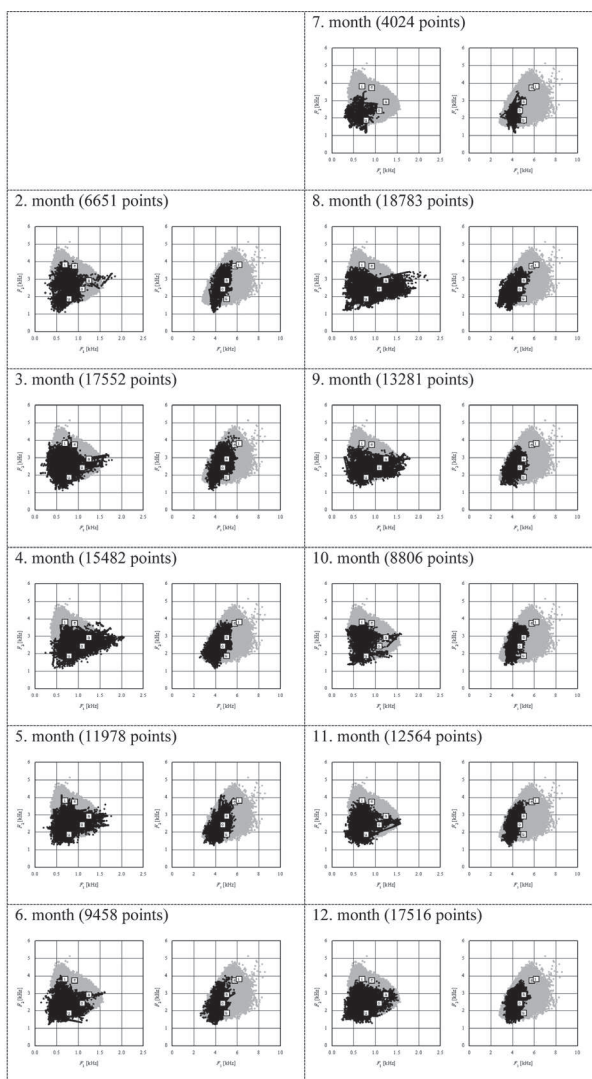


Figure 9 – Simulated MVS (gray dots) and estimated MVS (black dots) for a child from two to twelve months of age

Formant frequencies of vowels have a tendency to decrease in time. As a child gets older, the VT length gets longer and consequently formant frequencies decrease.

CONCLUSION

The obtained results show that MVS can be used in monitoring children's speech development. MVS depends on VT shape. On the other hand, the shape of VT depends of articulatory organs movement. The idea of proposed monitoring system is: if the theoretical MVS (gray dots on Figure 9) is similar to the MVS obtained from children's real speech (black dots on Figure 9) we can conclude that there is no typicality in the movement of children's articulatory organs. With a more detailed study of relations between MVS and children speech disorders, the diagnostic method can be developed. This diagnostic method has specific significance because it can be used from a child's earliest age. It should be noted that the method itself is appropriate for that age because a child pronouncing individual phonemes rather than responding to speech stimuli is enough.

This is a pilot study, which means that the results shown needs to be confirmed on a larger sample. The extended research should be based on establishing the criteria for discrimination of typical and atypical formant frequencies. Only when MVS is linked with different atypical phenomenon in a child's development, will it be possible to see the deeper meaning of the proposed method.

REFERENCES

1. Anderson, N. (1978). On the calculation of filter coefficients for maximum entropy spectral analysis. In D. G. Childers, *Modern Spectrum Analysis*, (pp. 252-255). Wiley, IEEE press.
2. Badin, P., & Fant, G. (1984). *Notes on vocal tract computation*. STL-QPSR 2-3/1984, 53-108.
3. Boë, L. J., Perrier, P., & Guérin, B. (1989). *Maximal vowel space*. Eurospeech 89, 281-284.
4. Boersma, P., & Weenink, D. (1992-2005). PRAAT: *A system for doing phonetics by computer*. Retrived from <http://www.praat.org/>.
5. Carré, R. (2009). Dynamic properties o an acoustic tube: Prediction of vowel systems. *Speech Communication*, 51, 26-41.

6. Dobrijević, Lj., Vujović, M., & Sovilj, M. (2013). Intellectual ability and emotional maturity in children from high risk pregnancies. *Abstract Book: 15th International ESCAP Congress* (pp. 253). Dublin.
7. Fant, G. (1970). *Acoustic theory of speech production*. The Hague: Mouton & Co. N. V.
8. Fant, G. (2004). *Speech acoustics and phonetics*. Dordrecht: Kluwer Academic Publishers.
9. Flanagan, J. (1972). *Speech analysis, synthesis and perception*. New York: Springer-Verlag.
10. Goldstein, U. G. (1980). *An articulatory model for the vocal tract of the growing children*. (PhD dissertation, MIT, Cambridge, Massachusetts). Retrieved from <http://hdl.handle.net/1721.1/22386>
11. Kostić, Đ., & Vladislavljević, S. (1995). *Govor i jezik deteta u razvoju*. Beograd: Zavod za udžbenike i nastavna sredstva.
12. Ménard, L., & Boë, L. J. (2001). Perceptual categorization of maximal vowel space from birth to adulthood. *European Conference on Speech Communication and Technology (Eurospeech)*, (pp. 167-170). Aalborg, Denmark.
13. Ménard, L., Schwartz, J. L., & Boë, L. J. (2007). Articulatory-acoustic relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model. *Journal of Phonetics* 35, 1-19.
14. Schwartz, J. L., Abry, C., & Boë, L. J. (2005). Asymmetries in vowel perception, in the context of the Dispersion-Focalisation Theory. *Speech Communication*, 45, 425-434.
15. Soquet, A., Lecuit, V., Metens, T. & Demolin, D. (2002). Mid-sagittal cut to area function transformations: Direct measurements of mid-sagittal distance and area with MRI. *Speech Communication*, 36, 169-180.
16. Sovilj, M. (2002). *Dečji govor – kvantitativni pratioci govora*. Beograd: Zadužbina Andrejević.
17. Story, B. H., Titze, I. R. & Hoffman, E. A. (1996). Vocal tract area functions from magnetic resonance imaging. *Journal of the Acoustical Society of America*, 100, 537-554.

18. van der Stelt, J. M., Zajdó, K., & Wemp T. G. (2005). Exploring the acoustic vowel space in two-year-old children: Results for Dutch and Hungarian. *Speech Communication*, 47, 143-159.
19. Vojnović, M., & Mijić, M. (2005). An improved model for the acoustic radiation impedance of the mouth based on an equivalent electrical network. *Applied Acoustics*, 66, 481-499.
20. Vojnović, M. (2008). *Uticaj maske na akustičke i artikulacione karakteristike govora* (Unpublished PhD dissertation), Beograd: Elektrotehnički fakultet, Univerzitet u Beogradu.
21. Vojnović, M. (2013a). Vocal tract shape estimation for children ages to one year. M. Sovilj, & M. Subotić (Eds.) *Proceedings Speecha and Language 2013 – 4th International Conference on Fundamental and Applied Aspects on Speech and Language* (pp. 55-64). Belgrade: LAAC & IEPSP.
22. Vojnović, M. (2013b). Maximal vowel spaces. M. Sovilj, & M. Subotić (Eds.) *Proceedings Speecha and Language 2013 – 4th International Conference on Fundamental and Applied Aspects on Speech and Language* (pp. 137-147). Belgrade: LAAC & IEPSP.

METOD MAKSIMALNOG OPSEGA VOKALA U ANALIZI VOKALA TOKOM PRELINGVALNE FAZE

Milan Vojnović*, Ivana Bogavac**, Ljiljana Jeličić Dobrijević**
*Centar za unapređenje životnih aktivnosti, Beograd**
Institut za eksperimentalnu fonetiku i patologiju govora, Beograd
*Centar za unapređenje životnih aktivnosti, Beograd***

Sažetak

Glavni problemi pri analizi vokala koji se javljaju u prelingvalnoj fazi su centralizacija izgovora i nepoznate dimenzije vokalnog trakta. Većina istraživača u ovoj oblasti bazira analizu na maksimalnom opsegu vokala (MOV) jer je njihova diskriminacija otežana kod dece na najranijem uzrastu. Analiza MOV uključuje estimaciju fizičkih dimenzija vokalnog trakta (VT). Cilj istraživanja je da se utvrde i definišu promene pri izgovoru vokala tokom prelingvalne faze. Snimanje i analiza su urađeni tokom prve godine života (od drugog do dvanaestog meseca). Snimanje je obavljeno u 42 sesije, u proseku 4 sesije u mesecu. Zvučni segmenti koji su ličili na izgovor vokala su izolovani i korišćeni pri estimaciji u PRAAT softveru. Za estimaciju formantnih frekvencija korišćen je Burgov metod. Rezultati ukazuju da se MOV analiza može koristiti na najranijem uzrastu jer nije neophodno da dete odgovara na dati stimulus. Ovaj rezultat treba potvrditi na većem uzorku, pri čemu bi proširena analiza definisala i kriterijume za diskriminaciju tipičnih i atipičnih formantnih frekvencija.

Ključne reči: prelingvalna faza, vokalni trakt, maksimalna oblast vokala, formantne frekvencije vokala

Primljeno: 5.5.2014.

Prihvaćeno: 16.9.2014.