



# Russian Vowels System Acoustic Features Development in Ontogenesis

*Elena E. Lyakso, Olga V. Frolova*

The Child Speech Research Group  
 Saint-Petersburg State University, Saint-Petersburg, Russia  
 lyakso@mail.ru

## Abstract

These data form a part of a longitudinal study of vowel system formation for Russian language. The goal of this investigation is to examine the process of vowels spectral characteristics dynamics from child speech to the corresponding values in normal Russian adult speech. The vowels from vocalizations and words of 5 Russian children from 3 to 60 months of age were analyzed. The spectral features of the vowels recognized by native speakers as corresponding Russian phonemes were taken into account. It was shown that the formants characteristics do not correspond to those in adults' speech yet, but they differ from vowel to vowel at 3-5 years of age. Many sounds at each age are impossible to describe by means of classical acoustic keys, but native speakers attributed these sounds to certain categories with high probability. To describe these sounds we use an additional system of distinctive features.

**Index Terms:** child speech, formant structure, child speech perception

## 1. Introduction

There are six vowel phonemes in Russian language (according to Saint-Petersburg phonetics school) [1]. Russian articulation is somewhat flabby, so the "closed" vowels are actually less closed than in some other languages. They all are somewhat moved to the center of IPA trapeze, but neither do they reach the locations of more open vowels, such as [i]. The small amount of vowel phonemes in Russian causes a possibility of big deviations of the allophones from the basic ones. The formants triangle for children's vowel-like sounds and vowels does not correspond to the basic allophones triangle in adult speech up to 5 years of age. Still the vowels differ by their characteristics from each other and usually they form a triangle, although its orientation for some children differs from that in adult speech [2]. The classical acoustic keys for adult vowels are the two first formants values. [3]. It was determined for different languages that the two first formants values could not be used for children's vowel-like sounds description [4]. It is caused by size and form difference of a child's vocal-tract as opposed to that of an adult. According to classical theories of speech production [3], the formants values are caused by the length of vocal-tract and its resonance characteristics. As we supposed before [5], a distinctive features system development in the ontogenesis may differ in time and is organized hierarchically in accordance with the evolving genetic program of development, which, in turn, is socially modified by the environmental factors. The goal of this investigation was to reveal the dynamics of vowels acoustic features attributed by native speakers to the phoneme categories of

corresponding vowels, with children's age. We had taken into account both the two first formants frequencies values and other characteristics that allowed describing the dynamic spectrum of vowels more completely. We supposed that the quantity of additional characteristics required for the description of children's sounds would decrease with children's age, and the amount of sounds described with classical acoustic keys would increase.

## 2. Methods

For this purpose a longitudinal recording of sound signals of five normally developing Russian children from 3 to 60 months of life has been performed. The recordings have been made with using a "Marantz PMD222" recorder with an external microphone "SENNHEIZER e835S". For the analysis we have not taken isolated vowels, but indicative vocalizations and words (from the point of view of child speech complication). In every child 10 vocalizations were chosen at 3 and 12 months (all the vocalizations contained several vowel-like sounds), 10 words or speech-like sequences of syllables at 24 months, 10 phrases with 1-3 words at 36 months of life, and 100 words at 48 and 60 months of life were taken to form test sequences for every child. The test sequences were presented to Russian-speaking adults (auditors). The amount of native speakers was: 3 mos, 12 mos – 133 auditors, 24 mos – 103 auditors, 36 mos – 137 auditors, 48-60 mos – 165 auditors. Each word was presented three times in the tests, the pause between presentations being 5 s, while the pause between diverse words was 15 s. The signals sequence was presented to the auditors to listen to using a computer or a tape recorder. If more than 75% auditors recognized a vowel, these vowels have been analyzed in the "Cool Edit 2.1" sound editor (Syntrillium Soft.Corp.USA). To build the formant triangle we measured the pitch (F0), the formant frequencies (F1, F2, F3) and their amplitude (E1, E2, E3) in each signal. Firstly we measured the amount of vowels with "classical" formant structure for [a], [i], [u]. Then we analysed the spectrums of the vowels that had no formant structure corresponding to that of the adult speech vowels. We chose the two maximums (from formants and pitch maximums) that were expressed more clearly. Then we estimated the amount of vowels whose maximums (M1, M2) values were in the frequency areas corresponding to every vowel, based on the data on Russian vowels perceptive phoneme borders [6]. Secondary we have taken the vowels that were impossible to describe on the base of the above characteristics. To describe these vowels we used several features: the difference between maximums, including pitch maximum, spectral maximums amplitude ratios. Mann – Whitney, Kruskal-Wallis test were used.

### 3. Results

Native speakers started recognizing words at the beginning of the second year of children's life. 60-100% of children's words were recognized by native speakers by 48-60 months. All the stressed vowels in 48-60 months olds words were recognized by native speakers. Adults detected vowels in 3 months olds vocalizations. The vowels of 36 months olds were recognized with high probability (Table 1).

Table 1. The amount of children vowels recognized by native speakers as corresponding Russian phonemes (probability 0,75), %.

Age, mos	a	u	i
3	53,5	54	0
12	71	0	0
24	51	22,6	20
36	87	100	67

The amount of vowels recognized by native speakers: 3 mos – 59 [a], 3 [u]; 12 mos – 33 [a]; 24 mos – 37 [a], 4 [i], 6 [u]; 36 mos – 20 [a], 6[u], 6 [i]; 48 mos – 189 [a], 30[u], 76 [i]; 60 mos – 234 [a], 50 [u], 105 [i].

Formant values medians of the vowel [a] were in the same area of the two-formant plot as the formant values

of adult's [a] at all the ages described. First formant median values of [u] were in a higher frequency area then those of an adult [u]; the first and second formant medians of [u] were in a higher frequency area at the age of 24 months. The first formant medians of [i] were in a higher frequency area then those of an adult [i] (fig. 1). P values for significant differences of [a], [i], [u] formants in the 3-60 months olds vocalizations are in the table 2.

Table 2. P values for significant differences of [a], [i], [u] formants in vowels recognized by native speakers.

Age, mos	a\i	a\i	i\i
3	-	-	-
12	-	-	-
24	-	-	-
36	F1: P<0.05; F2: p<0.01	F2: p<0.05	F2: p<0.001
48	F2: p<0.01	F1: p<0.05; F2: p<0.001	F2: p<0.001
60	F1: p<0.01; F2: p<0.01	F1: p<0.001; F2: p<0.001	F2: p<0.001

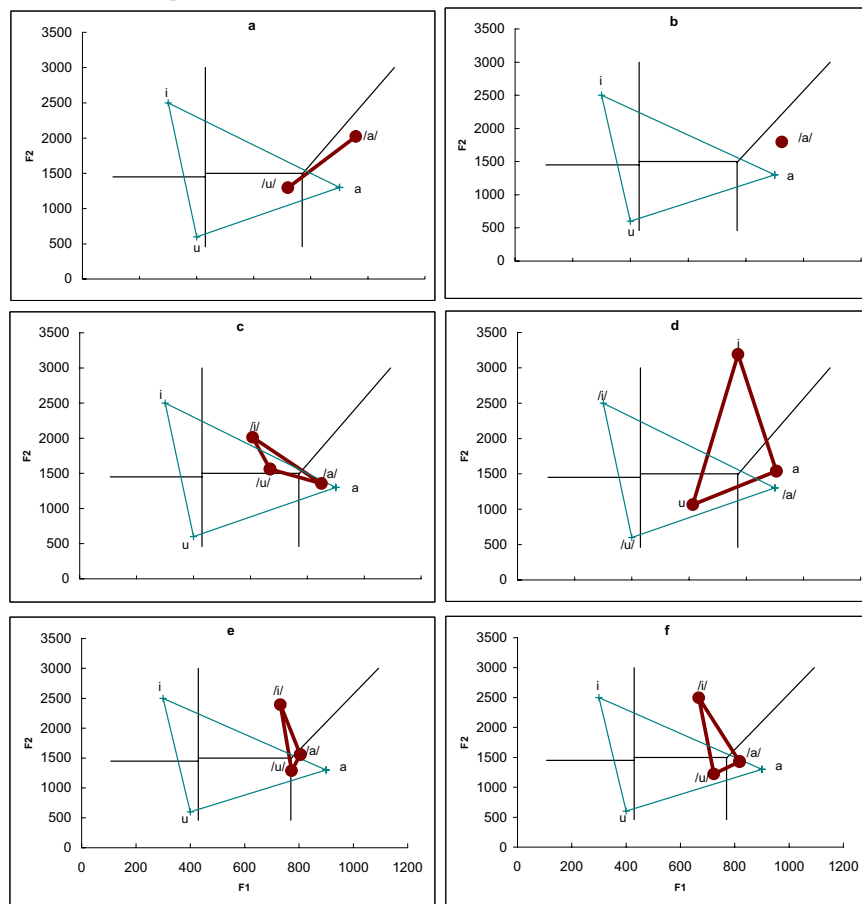


Fig.1. Formants vowel triangles (medians) in children from 3 to 60 mos of age (firm lines) versus those in adult speech. Perceptive bounds for adult Russian vowels [6]. a - data for 3 mos; b - data for 12 mos; c - 24 mos; d - 36 mos; e - 48 mos; f – 60 mos.

The differences between vowels [a], [i], [u] formant values increase with children's age and become the same as those in an adult by 60 months.

The formant triangle orientation approaches to that in adults by the age of 60 months, but the first formant values are shifted in a higher frequency area. It could be caused by the high pitch values (table 3).

Table 3. Pitch values (medians, minimum and maximum) for vowels recognized by native speakers (Hz).

3 mos	12 mos	24 mos	36 mos	48 mos	60 mos
389 (210-1459)	432 (279-771)	307 (216-604)	318 (252-551)	345 (170-658)	340 (172-638)

The pitch values decrease significantly from the first year of children's life by 60 months. Significant differences were revealed between ages 3 mos – 24 mos and 3 mos - 48 mos ( $p < 0.05$ ); between 3 mos - 60 mos and 12 mos – 24 mos, 36 mos, 48 mos, 60 mos ( $p < 0.001$ ).

Thereby the sounds recognized by native speakers were impossible to describe based on the two first formants (table 4). To describe these vowels we used the spectral maximums – M1, M2 (see methods).

Table 4. The amount of vowels (% of all the vowels recognized by auditors) with formants values within the perceptive phoneme borders of Russian vowels (column 1), and other vowels, without a corresponding formant structure (column 2).

Age, mos	a		u		i	
	1	2	1	2	1	2
3	59	41	0	100	0	100
12	73	27	0	100	0	100
24	32	68	0	100	0	100
36	80	20	0	100	0	100
48	48	52	0	100	0	100
60	56	44	2	98	4	96

Vowels [u] and [i] are described mainly on the base of spectral maximums. For the vowel [a] the amount of vowels describable in terms of the two first formants is 56% or less by 48-60 months. It could be caused by the effect of

coarticulation and particularities of perception of vowels as parts of words recognized by native speakers. The frequency values medians of M1 and M2 in all the vowels lay in the corresponding areas of the two-formant plot (fig 2).

There were some sounds recognized by native speakers with maximums and formants outside the perceptive phoneme borders of Russian vowels (Table 5).

Table 5. The amount of vowels (% of all the vowels recognized by auditors), with maximums and formants outside the perceptive phoneme borders of Russian vowels.

Age, mos	a	u	i
3	29	33	-
12	6	-	-
24	46	50	25
36	20	33	17
48	38	23	47
60	37	14	10

For [a] it were mainly the [a]-like vowels with formants and maximums in the [e] area on the two-formant plot, for [u] the maximums were in the [o] area and for [i] – in the [e] area. These sounds were described based on the system of extra characteristics [5]. We used the differences between the spectral maxima most clearly expressed. These were different for every pair of vowels: F2-F1 for [a] and [e]; F2-F0 - for [i], [e]; F1-F0 for the vowels [o] and [u]. This feature made it possible to reveal significant distinctions within the pairs [a], [e] and [I], [e]. However it is not informative for the vowels [o] and [u]; to differentiate between [o] and [u] the following characteristic could be used: the difference between the spectral maxima most clearly expressed (F1-F0) and the ratios of the amplitudes of these spectral maxima (E1/E0). The [a] and [e] vowels can be described both based on spectral values and on amplitude ratios. It is significant to note that the spectral maxima amplitudes ratio (relative to the pitch amplitude) is a usable feature for the [a] and [e] vowels (fig 4).

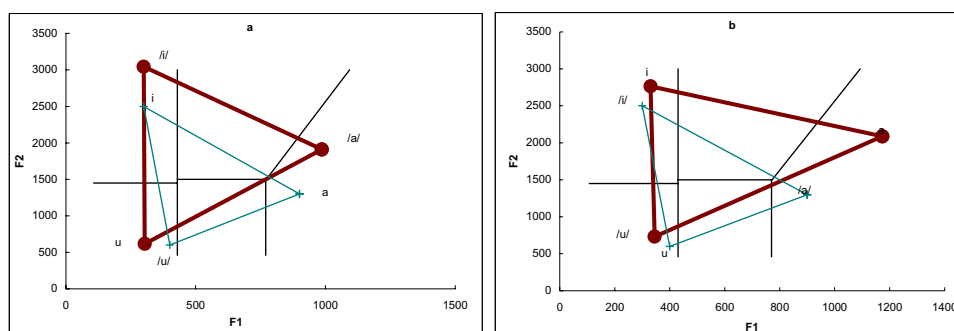


Fig.2. Maximum vowel triangles (medians) in children from 3 mos to 60 mos of age (firm lines) versus those in adult speech. Perceptive bounds for adult Russian vowels [6]. a - data for 24 mos; b – 60 mos.

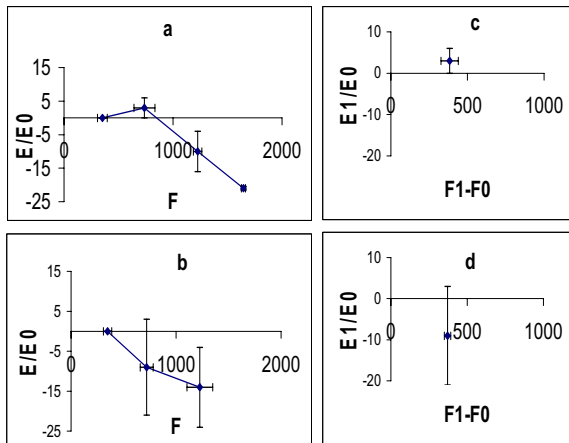


Fig.3. Amplitudes ratios for the spectral maxima (a, b) and the combined feature (c, d) representing the difference between the spectral maxima most clearly expressed (F1-F0) (the horizontal axis) and the amplitudes ratio of these spectral maxima (E1/E0). All vowels data are for 36 months olds [7]. a, c – vowel [o]; b, d – vowel [u].

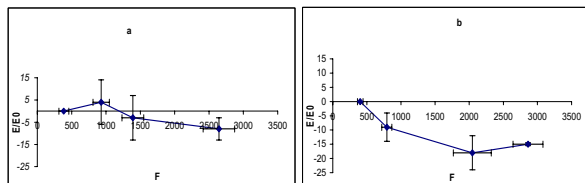


Fig.4. Amplitudes ratio of the spectral maxima in all the stressed vowels [a] [e] in the words of three-years-olds [7]. Horizontal axis – formant frequencies values (Hz) located consecutively on the horizontal axis: F0, F1, F2, F3. Vertical axis – amplitude values for corresponding formants relative to the fundamental frequency amplitude. a – [a] vowel, b - [e] vowel.

The data presented in the figures 3, 4 imply that the amplitudes ratio of the second spectral maximum is a distinctive feature. Vowels [a], [i], [u] from words and vocalizations of 3-60 months olds recognized by native speakers with no maximums and formants within the perceptive phoneme borders of Russian vowels could be distinguished based on this characteristic. Significant differences between [a] and [i] based on  $E1/E0$  ( $p < 0.001$ ),  $E2/E0$  ( $p < 0.001$ ) were revealed; these between [u] and [a] were based on  $E1/E0$  ( $p < 0.01$ ).

#### 4. Discussion

Usually the speech production theory [3] is used to describe vowels based on the two first low-frequency formants. Our data show that in children a vowel formant structure does not correspond to the formant structure of the same adult vowel even when the pitch becomes lower. However native speakers recognized vowels of 3 months olds. Our investigation results do not confirm the hypothesis. In children under 5 years of age there was no significant increase of vowels possessing a formant structure characteristic of the corresponding adult speech vowels. Our data could be interpreted from the motor theory point of view. First, a motor sound articulation program is developed

in a child, its realization determined by the articulation patterns maturity. Auditors recognize a child's sounds if the "acoustic image" of the latter corresponds to the auditors' sound standard. If the child articulates clearly, the auditory system extracts the formants from frequency bands (frequency-time encoding principle). Otherwise it is probable that features are extracted successively allowing to pick a vowel category for the indistinct sound. If this is the case, the features described above are significant, namely the spectral maxima difference and their intensity.

#### 5. Conclusion

The recognition of child speech by adult native speakers increases with children's age. Formant structure of Russian vowels is not formed completely by the age of 5 years. The vowel dynamic spectrum description based on the additional characteristics is a more informative one for the majority of sounds in the age period analyzed.

#### 5. Acknowledgements

The study is being performed with the financial support from the Russian Fund for Humanities (projects № 06-05-12623v and № 06-06-00519a).

#### 6. References

- [1] Ganiev, Zh. V. *Russkii yazyk: Fonetika i orfoepiya* (Russian language: Phonetics and Orthoepy), Moscow, 1990.
- [2] Lyakso E., Gromova A. The acoustic characteristics of Russian vowels in children of 4 and 5 years of age, *Psychology of Language and Communication*, 2005, Vol. 9, № 2, p. 5 - 14.
- [3] Fant G. *Akusticheskaya teoriya recheobrazovaniya*. (The acoustic theory of speech production), M, 1964.
- [4] Kent R.D. The psychophysiology of speech development: co emergence of language and movement system, *American Journal of Physiology*, 1984, Vol. 246, p. 888-894.
- [5] Lyakso E, Chelibanova O., Gromova A., Novikova I., Galunov V. Noise immunity development in the early ontogenesis, *Proceeding X IV session of the Russian acoustic Society*, Moscow, Russia, 2003, Vol. 3, p. 101-106.
- [6] Slepokurova N.A. *Protsedura raspoznavaniya stacionarnykh glasnnykh*. (The procedure of recognition of stationary vowels), *Journ. of Sensory System*, 1972, Vol. 3, p. 156-168.
- [7] Lyakso E., Gromova A., Frolova O., Romanova O. Acoustic aspect of the formation of speech in children in the third year of life, *Neuroscience and Behavioral Physiology*, 2005, Vol. 35, N 6, p. 573-583.