Psychology of Language and Communication 2005, Vol. 9, No. 2

ELENA E. LYAKSO, ALEXANDRA D. GROMOVA
Saint-Petersburg State University

THE ACOUSTIC CHARACTERISTICS OF RUSSIAN VOWELS IN CHILDREN OF 4 AND 5 YEARS OF AGE

The purpose of this investigation is to examine the process of first and second formants of the vowels from child speech approaching corresponding values in normal Russian adult speech. The vowels in the words of 5 Russian children of 4 and 5 years of age were analyzed. The vowel formant structure was examined. Word stress and palatal context influence on the formant structure of the vowels were taken into account. It was shown that word stress is formed by 4 years of age. The formant characteristics do not correspond to those in adult speech yet, but may be used to distinguish between the vowels.

Introduction

There are six vowel phonemes in the Russian language. Their main characteristics are as follows (Ganiev et al., 1990): /a/ is a back vowel, the most open one in Russian. It is a little moved to the center row as compared to back [ɑ] in other languages, but does not coincide with front [a]. /o/ occurs in stressed position only, otherwise is substituted by /a/. This is a labial middle vowel. In fact, it is a diphthong: it starts as an [u]-like vowel and in the end passes to a delabialized, almost [ɔ]-like sound. This feature is especially typical for women speech and is phonologically irrelevant. /u/ is a labialized, back, closed vowel. /i/ is a front closed vowel. It is pronounced after palatalized consonants only. /e/ is actually realized in two allophones that could be called basic ones. One of them is [ɛ], which occurs at the beginning of the word and rarely after some consonants with-

Elena Lyakso, Saint-Petersburg State University, Uchtomsky Institute of Physiology, Child Speech Research Group, 199034 University Emb. 7/9, S-Petersburg, Russia. E-mail: lyakso@mail.ru

out palatalized pairs (such as /ɔ/). Another one, [e], occurs only after palatalized consonants. It is more closed and moved to the front. This vowel occurs in stressed position only, just like /o/. Vowel /ɨ/ is the most closed vowel. It begins with a "non-front" articulation and during the pronunciation changes into a front sound. So this is a diphthong, but that is also phonemically irrelevant, as with /o/.This vowel is in fact sometimes considered as an allophone of /i/ pronounced after non-palatalized consonants. It is the hardest sound for learners of Russian to pronounce, and perhaps for children also. It occurs as an emotional reaction sound, but is absent in a linguistic sense.

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 the IPA trapeze, but neither do they reach the locations of more open vowels, such as [ɐ]. The small number of vowel phonemes in Russian make possible wide deviations of the allophones from the basic ones.

The first two formants (F1, F2) are used to describe vowels in adult speech, but it is known that these characteristics are inadequate when applied to child speech vowels (Galunov & Lyakso, 2001). Because of the F0 high values the first two formant absolute values and their location on the two-formant plane are not sufficient to describe the vowel-like sounds in children in the first year of life. The F1 and F2 values of infants' vowel-like sounds in the first month of life are shown to occupy the two-formant space areas that do not correspond to those occupied by the corresponding vowels in adults, lying in an area of significantly higher frequencies. The areas occupied by the sound pairs /o/, /u/ and /e/, /ɨ/ are almost joined together; those occupied by the vowel-like /i/, /a/ overlap in F1 values and partly overlap in F2 values. The difference between the formant values and their magnitudes may be used instead (Galunov & Lyakso, 2001).

As compared to the end of the second year of life, by three years of age the number of well pronounced words increases where pitch and formant values are close to those in adult speech. Still, F0 values are high (332±73 Hz) and make it impossible to describe the vowels /i/, /u/, /ɨ/ in terms of the first two formants' absolute values, but these vowels are identified clearly in the words (independently of their stressed or unstressed position) (Lyakso et al., 2003). Although the acoustic features of adult speech are not totally formed, it is shown that by the end of the third year of life words pronounced by children become comprehensible for adults in absence of a situational context.

As we showed before (Lyakso et al., 2003), during the third year of life the vowel stress is formed. In all the vowels the stressed vowel's duration tends to be longer than that of the unstressed one, but this difference is significant only in some of the vowels. The development of opposition palatalized versus non-palatalized consonants begins, as found in the characteristics of following vowels.

The purpose of the current investigation is to examine the further process of development of fundamental frequency (F0), first and second formants, duration

and intensity of the vowels in child speech approaching the corresponding values in normal Russian adult speech.

Methods

For this purpose a longitudinal audio recording of sound signals of five normally developing Russian children (two boys F and E and three girls D, S and Z) during the fourth and fifth years was performed. The children were recorded while interacting freely with their mothers and the investigator. The recordings were made by the "Marantz PMD222" recorder with a "SENNHEIZER e835S" external microphone. The sounds were instrumentally analysed in the Cool Edit 2.1 (Syntrillium Soft. Corp. USA) sound editor. 100 words were selected from each child's speech for analysis. The fundamental frequency and the first two formants were measured in the words, as well as the intensity of these harmonics, and the vowel and its stationary part duration.

To consider word stress development the vowel duration and its stationary part duration were compared in the stressed versus the unstressed vowels, as well as the pitch and formant values in the stationary parts. The same parameters were compared using the Mann-Whitney criterion in /a/, /i/ and /u/ after the following consonants: /k/ and /d/ for /a/, /b/ and /g/ for /u/ and /t'/ for /i/. These consonants cause minimal articulatory and hence acoustic influence on the corresponding vowels in Russian.

Phrasal stress development was considered. The stressed vowels were chosen from the words and the fundamental frequency values of the words remote from the phrasal stress were divided by that of the word which would be stressed in adult speech. Using the Wilcoxon criterion the median of this ratio was compared to 1.0. If the value thus obtained corresponded to 1.0, it was concluded that all the words in the phrase are emphasized to the same degree.

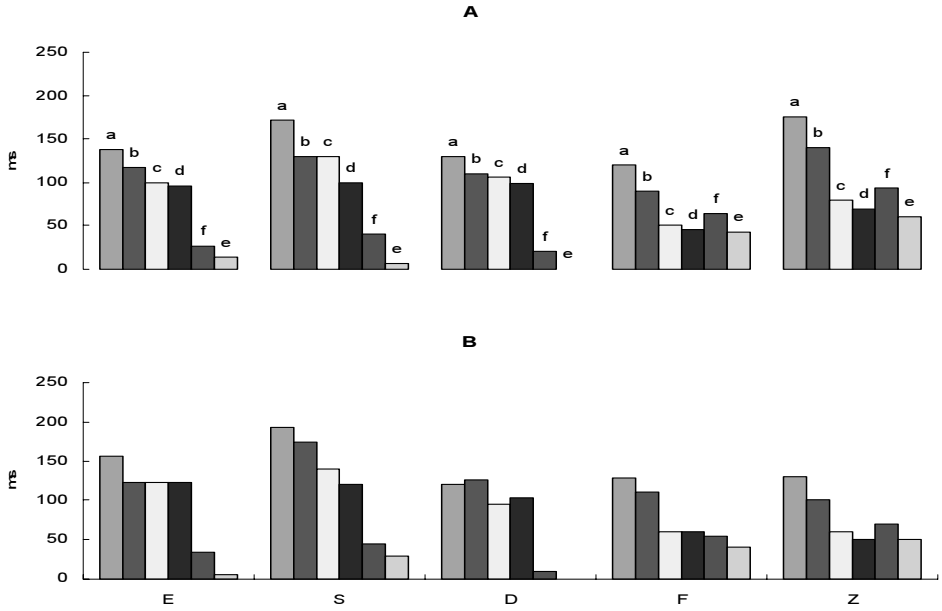
Vowel triangles were constructed for stressed vowels from the speech of each child at 4 and 5 years of age. The vowels were selected with the F0 up to 350 Hz. For /a/, /o/ and /o/ the vowels only in the non-palatalized context were included in the graphs, as palatalization causes these vowels to differ a lot from the base allophone. We did not consider the vowel /ɨ/ because of its complex acoustical nature (it is not quite comparable to other vowels). We considered [e] as the basic allophone of the phoneme /e/, as it is the more common one. Thus we analyzed the rest of the phonemes, namely /i/ and /e/, in palatalized context only.

Results

The median values of vowels F0 are higher at 4 years of age (296-431 Hz) as compared to those at 5 years of age (290-348). The F0 in stressed versus unstressed vowel comparison, not taking into account the vowel context and

Figure 1. The pitch median values in the vowels of 4 (A) and 5 (B) year old children. Grey column – F0 values of all the vowels; white column – F0 values in stressed vowels; shaded column – that of the unstressed vowels.

Figure 2. Vowel duration at 4 (A) and 5 (B) years of age (median values). a – stressed vowel; b – unstressed vowel; c – stressed vowel stationary part; d – that of the unstressed vowel; e – vowel and its stationary part duration difference in the stressed vowels; f – that in the unstressed vowels.



quality, showed that their difference is expressed more at four years of age than at five in the children E, S and Z. In the child D a reverse tendency was revealed (Fig. 1).

At 4 years of age in all the children the stressed vowel and its stationary part duration, as well as their difference, is higher in the stressed vowels than in unstressed ones. At 5 years of age this is true in all the children except for child D, however the difference of the stressed vowel duration and its stationary part duration is still higher than that in the unstressed vowels for all the children (Fig. 2).

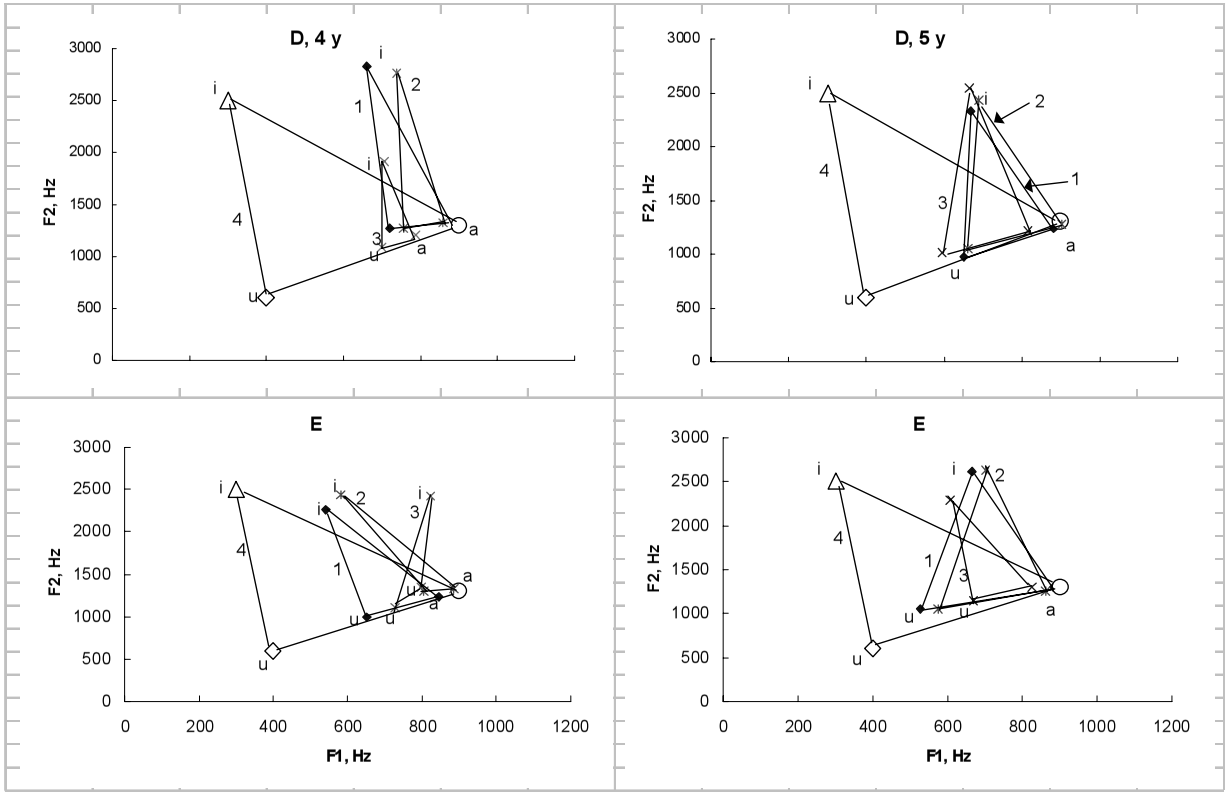
As to phrasal stress, only in one child – E – the median of unstressed word pitch to that of the stressed word ratio differed from 1.0 significantly ($p<0,05$). At 5 years of age this difference was not revealed. In two children (D and S) a tendency to this difference with age was revealed. In the child F a reverse tendency was shown, and in the child Z there were no changes with age. The median of the same ratio in an adult woman's speech (F's mother) addressed to an adult interlocutor differed very significantly from 1.0 ($p<0.0001$). The ratio median equaled 0.75.

The arrangement of the formant values on a two-formant plot is different for every child at every age. On the figure (Fig. 3) an example is given of formant triangles of vowels /a/, /u/ and /i/ in two children, compared to those in adult speech. The stressed vowel /a/ in these two children coincides with the basic allophone from adult speech. By the fifth year of life /u/ tends toward the normal values of adult speech by a decrease in the second formant. At 5 years of age the F2 of /i/ corresponds to its value in adult speech. The shape and orientation of the stressed vowel triangle where the vowels are chosen without regard to context and pitch (triangle 2) correspond to the stressed vowel triangle where the vowels were chosen with regard to these factors (triangle 1), but the former is moved up and right, perhaps as a result of the higher pitch values. The unstressed vowel triangle's shape and orientation usually differs from the stressed vowel triangles (at 5 years in D, and at 4 and 5 years in E). This might be evidence of a strong context influence on the characteristics of the unstressed vowels which changes them as compared to those of the basic allophones. The same happens in Russian adult speech (Bondarko, 1998).

In the child F at 4 years of age the F2 of all the vowels is higher than in adult speech. The shape of the formant triangle (triangle 1) based on the chosen criteria does not correspond to that in adult speech. By 5 years the F2 decreases in /i/, and all the other values remain higher than in the basic allophones. The stressed and unstressed vowel triangles built without taking into account the criteria differ greatly from the triangle 1 by shape and orientation. This is due to high pitch influence on the characteristics which is also seen in 3 year olds, and, perhaps, a non-final development of articulation skills.

In the child S at 4 years of age all the triangles differ by shape and orientation from the adult speech triangle. At 5 years of age the triangle of the stressed vow-

Figure 3. Vowel formant triangles with apexes /a/, /u/, /i/ in E and D and adult speech. 1 – stressed vowels chosen taking into consideration the context and pitch (see methods); 2 – the triangle of stressed vowels not taking into consideration the above factors; 3 – the triangle of unstressed vowels not taking into consideration context and pitch; 4 – the triangle of basic allophones in adult speech (Derkach et al., 1983). Horizontal axis values are F1, Hz, vertical axis values are F2, Hz



els chosen by the criteria corresponds to the adult speech vowel triangle by shape, but is oriented somewhat differently. The F1 values are higher than in adult speech, the F2 of /a/ and /i/ correspond to that of their basic allophones. The shapes and orientation of stressed and unstressed vowel triangles built without regard to pitch and context correspond to each other, but F1 is higher in stressed vowels than in unstressed ones. In this child a positive dynamics is seen by 5 years of age. The higher F1 of stressed vowels as compared to the unstressed ones is perhaps due to the higher pitch of the former.

In the child Z, all the triangles differ both in shape and orientation from the adult speech triangle. The stressed vowel triangle built without regard to pitch and context lies in an area that does not intersect with the adult vowel triangle. In the unstressed vowel triangle only /u/ lies in the area of the latter.

In general, at 4 years of age the second formant of vowel /i/ corresponds to that in adult speech (except for the child S). The vowel /a/ coincides completely with that in adult speech in two children only (D, E). In the same two children the lines /i/-/u/ are about parallel to those for adults, and in the child E even the /a/-/u/ line lies on that of the adult triangle. The /e/ and /o/ vowels lie on their lines in this child only.

At five years of age the second formant of /i/ corresponds to that in adult speech in all the children, and /a/ coincidence with adult /a/ in the child S occurs. As to the lines, /a/-/u/ is now about parallel to that in adults in the child D, and /i/-/u/ in the children F and S. In the child E only /o/ is on its proper line now.

So we can see that the line arrangement changes from 4 to 5 years in all the children, and not always in the direction of adult speech. In the child E a worsening is seen between these ages.

The /u/ formants never coincide with those in the adult speech. This may be explained by the low values of the latter and the high fundamental frequency in child speech which does not allow these low values to realize. Still, it lies in the right direction from /a/, so it tends to its proper values.

Conclusion and Discussion

Despite the still high fundamental frequency values, in our children they differ from stressed to unstressed vowels at 4 years of age. This is unusual for the Russian language where the stress is expressed by the duration of vowels. The higher pitch of the stressed vowel might be due to the fact that each word is designed both by word and phrasal stress, as a separate phrase, as it is pitch which expresses phrasal stress in Russian (Bondarko, 1998). This assumption is evidenced both by our data on phrasal stress and on formant structure of stressed versus unstressed vowels in the same context, since the unstressed vowel formants are closer to normal adult speech values, perhaps due to the lower fundamental frequency. This seems to be some deterioration as compared to the second year of

life, when children generally are known to pronounce whole "phrases" which are unintelligible but designed with intonation. Intonation is considered to discriminate between communicative and non-communicative pre-word babbling (Konopczynski, 1998). We suppose, however, that the later stage of development considered in our study is devoted to word formation and thus each word is designed as a whole phrase with its separate intonation, the instrument of which – the fundamental frequency changes – is formed earlier. The data on children acquiring intonation are evidence that prosodic development is subordinated to cognitive development of the categories the intonation expresses and thus is gradual and non-linear (Martel, 2002). Moreover, this development is different in different languages: at 4 years of age English-speaking children show an absence of adult-like phrase rhythmic structure, whereas French children have already acquired this trait (Watson, 1998). Thus it is quite possible that Russian children normally design their phrases with pitch at a later stage. On the other hand, it would be interesting to collect more data on intonation achievement with a special focus on language discrimination. Still, Jannedy shows (Jannedy, 1997) that English-speaking children start using adult-like phrasal stress prosody between 3 and 5 years, that is, at the age of our subjects, although this study has a weakness in using a somewhat subjective method of assessment through a psycholinguistic experiment, ignoring instrumental analysis.

The stressed vowel duration is higher than that of unstressed vowels at 4 and 5 years of age in all of our children, and the difference between the stationary part of the vowels and their whole duration is bigger in the stressed vowels than in the unstressed ones. As for the stationary part duration of stressed versus unstressed vowels, it is higher in the stressed vowels in all the children at all ages except for D at 5 years of age. This shows that word stress is already formed in these children. In fact, it has been already shown that it had developed by three years of age (Lyakso et al., 2003). We can also conclude that the unstressed vowel characteristics are more influenced by context than those of the stressed vowels, which correlates with the unstressed vowel duration reduction in Russian (Bondarko, 1998).

We chose the vowels that must be close to their basic allophones in the Russian language, namely, those not influenced by very high fundamental frequency, palatalization and stress. The formant triangle for these vowels at 4 and 5 years of age does not yet correspond to the basic allophone triangle in adult speech. It is surprising that in some children there is a deterioration seen from 4th to 5th year. Still, the vowels differ from each other by their characteristics and usually they form a triangle, although its orientation is different from that in adult speech. Attempts at automatic recognition of child speech described in the literature show that vowel acoustic characteristics become adult-like and are possible to recognize using adult speech recognition methods as late as at 12-13 years of age (Guiliani & Gerosa, 2003). The recognition is generally bad in children of 7-9 years of age,

although different children show different rates of success and, on the other hand, the improvement is evident from year to year (Elenius & Blomberg, 2004). Even if the recognizer is trained in child speech as opposed to adult speech, recognition at these ages is still unsatisfactory (Elenius & Blomberg, 2004). It might be that characteristics other than F1 and F2 values, not considered in the present study, are still used to discriminate between the vowels in children of our subjects' and older ages, just as at 3 years of age (Lyakso et al., 2003). The study of these alternative characteristics might be also useful in terms of automatic speech recognition, as it would show more ways in which humans perceive the acoustic characteristics of speech, and knowledge of natural perception mechanisms is necessary in this field (Hermansky, 1997).

Acknowledgements

The study is being performed with the financial support from the Russian Fund for Humanities (projects Nr 03-06-12024v) and the Russian Fund for Basic Investigations (project Nr 04-06-80334a).

References

Bondarko, L.V. (1998). *Fonetika sovremennogo russkogo yazyka* (Modern Russian phonetics). St-Petersburg: Izdatelstvo S.-Peterburgskogo universiteta.

Derkach M.F., Gumetskii R.Ya., Guba B.M., Chaban & M.E. (1983). *Dinamicheskie spektry rechevykh signalov* (The dynamic spectrograms of speech signals). Lviv.

Elenius D. & Blomberg M. (2004). Comparing speech recognition for adults and children. In P. Branderud & H. Traunmüller (Eds.), *Proceedings FONETIK 2004*. Stockholm: Stockholm University.

Galunov, V.I. & Lyakso, E.E. Formation the acoustic image of signals in early ontogenesis. In *Proceedings of the XI Session of the Russian Acoustical Society. Moscow, November 19-23, 2001.* Vol. 3 (p. 399-402). Moscow: RAO.

Ganiev, Zh.V. (1990). *Russkii yazyk. Fonetika i orfoepia* (Russian language: Phonetics and Orthoepy). Moscow.

Guiliani, D. & Gerosa, M. (2003). Investigating recognition of children's speech. In *Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing.* Vol. 2 (p. 137-140). IEEE Publications.

Hermansky, H. (1997). Auditory modeling in automatic recognition of speech. In A. Procházka, J. Uhlír, & P. Sovka (Eds.), *Proceedings of rhe First European Conference on Signal Analysis and Prediction* (p. 17-21). Prague: ICT Press.

Jannedy, S. (1997). Acquisition of narrow focus prosody. In A. Sorace, C. Heycock & R. Shillcock (Eds.), *Proceedings of the GALA '97 Conference on Language Acquisition*. Edinburgh: University of Edinburgh.

Konopczynski, G. (1998). Language development: typical and atypical. In S. Santi, I. Guaitella, C. Cavé, & G. Konopczynski (Eds), *Oralité et Gestualité: communication multimodale, interaction* (p. 684-689). Paris: L'Harmattan.

Lyakso, E.E., Gromova, A.D., & Frolova, O.V. (2003). An acoustic study of word production of Russian children during the third year of life. In R. Potapova (Ed.), *Specom 2003 – International workshop "Speech and Computer" Proceedings* (p.319-326). Moscow: MSLU.

Martel, K. (2002). Mutual knowledge and prosody in young children. In B. Bel & I. Marlien (Eds.), *Proceedings Speech Prosody 2002 conference* (p. 479-482). Pertuis: APPL.

Watson, I., Grabe, R., & Post, B. (1998). *The Acquisition of prosody in speech production: English and French.* Project Report, University of Cambridge.