Developing a Russian vowel space in infancy: The first two years

van der Stelt, J.M.; Lyakso, E.E.; Pols, L.C.W.; Wempe, A.G.

Published in:
Proceedings of the 11th International Conference "Speech and Computer" SPECOM 2006

Citation for published version (APA):
Developing a Russian vowel space in infancy: the first two years

Jeannette M. van der Stelt\textsuperscript{1}, Elena Lyakso\textsuperscript{2}, Louis C.W. Pols\textsuperscript{1} & Ton G. Wenpe\textsuperscript{1}

\textsuperscript{1} Institute of Phonetic Sciences/ACLC
Universiteit van Amsterdam, Amsterdam, The Netherlands
\texttt{J.M.van.der.stelt@uva.nl}

\textsuperscript{2} Uchтомsky Institute of Physiology
St. Petersburg State University, St. Petersburg, Russia
\texttt{lyakso@mail.ru}

Abstract

In the first year of life infants do not yet produce vowels in a full sense of the word. Yet, for the tuned-in listener the infant’s sound can elicit a vowel-like perception that, with age, becomes more adult-like. Mastering the vowels from the ambient language also means dealing with developmental changes in the speech mechanism and in perception. An enormous variability in vocalic productions is found.

The Principal Component Analysis method on band filter data employed in this study focuses on the spectral envelope, and thus accounts for frequency as well as intensity information, much like the information used by the ear. Results for Russian-learning infants indicate that between 6 months and two years of age the infants explore the vowel space in quite different manners. Data on 5 Dutch boys, acquiring 12 vowels, are compared to the Russian results, to interpret the influence of the number of vowels per language.

1. Introduction

The development processes in infant’s vowel productions are not yet thoroughly explored. Kuhl and Meltzoff (1996) suspect that imitation leads the infant from its universal starting point to the language-specific vowel productions. As early as 19 weeks old, infants can perceive adult corner vowels to belong to their mother tongue and they can link patterns for vowel productions of 10-month-old infants that starting point to the language-specific vowel productions. As early as 19 weeks old, infants can perceive adult corner vowels to belong to their mother tongue and they can link patterns for vowel productions of 10-month-old infants that can explore the vowel space in quite different manners. Data on 5 Dutch boys, acquiring 12 vowels, are compared to the Russian results, to interpret the influence of the number of vowels per language.

1.1. Automated analysis

The first step is the selection of infant utterances, defined as a sound production during the expiratory part of the respiration cycle. In infants the recordings are made mostly in naturalistic situations. In selecting the utterances, it is important to avoid overlapping mother-infant sound parts. Then a PRAAT script [8] is used that reads the utterances from the file. Each utterance is divided in ten parts of equal duration, thus accounting for articulatory drifting. In each part the vowel-like nucleus will be analyzed automatically.

Selection criteria for the measure points in the ten parts of one utterance are set for \( F_0 \) to be below 425 Hz and for intensity to be between 0.5 and 10 dB below the maximum to avoid clipping and to ensure measuring the vowel-like parts of the utterance. The \( F_0 \) criterion determined the filter bandwidth (1.1 x 425 Hz), a good compromise between frequency resolution and occurrence of \( F_0 \) “ripple”. Spectral information is present in one single \( F_0 \) period. One single period is chosen from 5 contiguous pitch frames around the measurement point. This period is recycled up to 50 ms. The thus constructed sound is multiplied by a Kaiser window, and pre-emphasis and level normalization (0.3 Pa RMS) is applied. A stepped Gaussian broad bandpass filter analysis, interpolating between spectral “samples”, results in linear intensity values per filter bin of 175 Hz that are used for further analysis. So, a “whole spectrum approach” (0 – 7 kHz) of the 40-filters output is chosen.

1.2. Data reduction

Data reduction of all 40-filters outputs is achieved via Principal Component Analysis (PCA). Each row in the band filter matrix is seen as a point in a 40-dimensional space. The data are reduced to the first two principal components that in this study explained about 65% of the variance. Previous research suggests that the first two principal components are related to the \( F_1 \) and \( F_2 \) values of vowels [4]. Comparison of different utterance data sets is reproducible since it is fully automated.

2. The target vowel system

In Russian, there are six vowel phonemes, although some linguists consider them to be only five. This is mostly a matter of definitions.

\(/a/\) – a front vowel in children, back in adults, most open one.

\(/o/\) – a labial middle vowel, seen only in stressed position, it otherwise is substituted by \(/a/\). It is in fact a diphthong that
starts as an /u/-like vowel that passes to an unlabialised, /ə/-like sound. This feature is especially typical for women speech and phonologically irrelevant. /a/ – labialised, back, closed vowel. /\i/ – front, closed vowel. It is pronounced after palatalized consonants only. /e/ – there are two basic allophones. One is /\i/, occurring in the beginning of a word and rare after some consonants (such as /c/). The other one, /e/, occurs after palatalized consonants only. It is more closed and moved to the front. This vowel occurs in stressed position only. /y/, or /ɨ/ in SAMPA – the most closed vowel, begins with a “non-front” articulation and during the pronunciation changes into a front sound. So, this is also a diphthong, which is also phonologically irrelevant. In the 5-vowel system, this vowel is seen as an allophone of /i/ pronounced after non-palatalized consonants.

Figure 1. A Russian adult vowel system [9] with 5 vowels is represented on the left. To the right a 6-vowel system is given [10].

3.1. Vowels in child speech

In child speech /a/ and /e/ are the first vowels to occur. Then come /\i/, /o/ and /u/, and /y/, the last one, is the hardest sound to pronounce for learners of Russian, and perhaps for children too.

Phoneticians can describe about 30% of infant speech sounds both phonetically and phonemically by the end of the first year of life [11].

3.2. Russian pronunciation

Russian articulation is somewhat sloppy, so the “closed” vowels are actually produced less closed than in some other languages. They all move somewhat to the center of IPA trapeze, but none of them reach the locations of more open vowels, such as /a/. The most striking difference occurs in the /a/ vowel. It results in quality change: the articulation becomes even less definite than in usual sloppy Russian articulation. The stationary part of the vowel may vanish, leaving only the transition from the previous consonant immediately followed by a transition to the next consonant. The /y/ sound is also very difficult to recognize as was found in speech perception experiments [12].

The duration is irrelevant in Russian, but /a/ and /\i/ are usually longer that the other vowels.

The stress is considered to be temporal: stressed vowels differ from unstressed ones only in duration.

4. Data collection

From a database of 6 infants, 4 subjects are chosen in this study, since for these 4 infants recordings were available over all months involved. They are healthy, Russian learning boys that were recorded as soon as possible after birth with a frequency of every three months in their naturalistic home situations during ‘face-to-face’ and ‘play’ (months 3 and 6) and during ‘face-to-face’ and ‘reading’ situations (9, 12, 18, and 24 months). No duration limit is set for the recordings. The recordings for boys 3 to 5 were made with a Marantz PMD222 recorder with an external microphone Sennheiser E835S. For boy 2, a Teac W-800R recorder with a Lomo-82A.02 microphone is used. The Cool Edit 2.1 sound editor digitized the recordings at a sample rate of 44100Hz, mono, and a resolution of 16 bits. The sound file of each recording of boys 2 to 5 is saved in a wav-format.

4.1. Selection of 50 utterances per infant and per month

The second author and her assistants selected per infant and per recording 50 utterances, using a transcription system described by [13]. This transcription system focuses on speech apparatus movements. The duration of an utterance depends on the expiration cycle of the respiration. Phonation usually is present in infant vocalizations, and can be continuous or interrupted. Further, the infant may make no articulatory movements at all or one or more per utterance. In this manner the linguistically unlabeled utterances are divided in syllables. The criteria set in the automated acoustic analysis are chosen to explicitly analyze the syllable nucleus.

5. Results

The automated acoustic analysis by means of the PRAAT script ideally can result in 10 parts x 50 utterances = 500 measurement points per infant per recording. The script indicates for what reason an utterance part is not analyzed: intensity and/or pitch. Further, when the signal volume is too low (below 0.025 Pa), an utterance part is not analyzed either. In table 1 the numbers of measurement points per infant and per recording are given.

Table 1. Number of measurement points (mps.) per infant in the 4 selected months (from 50 utterances per infant). The lowest number of mps. is found for boy 2 when 18 months old. The total number of 677 mps. is used for the construction of the vowel reference plane for two-year-old infants.

<table>
<thead>
<tr>
<th>Month</th>
<th>Infant</th>
<th>6</th>
<th>12</th>
<th>18</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>boy 2</td>
<td>91</td>
<td>156</td>
<td>72</td>
<td>137</td>
<td></td>
</tr>
<tr>
<td>boy 3</td>
<td>128</td>
<td>358</td>
<td>244</td>
<td>160</td>
<td></td>
</tr>
<tr>
<td>boy 4</td>
<td>116</td>
<td>256</td>
<td>262</td>
<td>213</td>
<td></td>
</tr>
<tr>
<td>boy 5</td>
<td>211</td>
<td>248</td>
<td>357</td>
<td>157</td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>546</td>
<td>1018</td>
<td>935</td>
<td>677</td>
<td></td>
</tr>
</tbody>
</table>

5.1. The construction of the reference plane

Assuming a gradual approach in vowel development, the utterances of the two-year-old infants are taken to be the closest to the adult vowel system. This reference plane is constructed by means of the 677 measurement points that resulted from analyzing the utterances of the 24-months olds.

The acoustic analysis of the spectral envelope in each measurement point results in a row with intensity measures, distributed over the 40 bins with a width of 175 Hz, thus covering a range from 0 – 7000 Hz. Formant-like maxima can be detected, but these vary greatly over the different utterances, and even the parts. In addition, we wanted to use the intensity information of the distribution as well as it plays an important role in auditory perception.
Table 2. Percentages of variance explained by each of the first 4 eigenvectors in the PCA for the utterances of 4 two-year-old Russian boys (667 mps).

<table>
<thead>
<tr>
<th>Eigenvector</th>
<th>38.8%</th>
<th>38.8%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenvector 2</td>
<td>26.5%</td>
<td>65.3%</td>
</tr>
<tr>
<td>Eigenvector 3</td>
<td>10.8%</td>
<td>76.1%</td>
</tr>
<tr>
<td>Eigenvector 4</td>
<td>0.1%</td>
<td>76.2%</td>
</tr>
</tbody>
</table>

5.3. Mean developmental patterns

The 4 infants’ vowel productions, balanced for all 4 months to randomly chosen 546 mps per month (the overall lowest number: month 6) are displayed in the figures 4-7, in which the ellipses indicate 1 standard deviation from the mean. Over the four age periods, the ellipses slightly move from one area to another, which is especially present when the infants are 12 months old. Then the ellipse has moved towards the /i-/a/ dimension, which is probably related to the mouth opening. When 2 years old, the ellipse is nicely in the middle of the reference plane, but only 546 out of 667 mps are displayed.

5.2. The eigenvectors and the explained variance

The eigenvectors display the contribution over the frequency bins. In figure 3, the first two eigenvectors for the reference plane are given, in table 2 the explained variance is presented per vector as well as cumulatively.
change is given in lines, drawn between the individual means. (see table 1). For the individual comparison we balanced the pc1-pc2 reference plane for the two-year-olds. The number of measurement points per infant per recording is not at all equal (see table 1). For the individual comparison we balanced the number of measurement points for all infants and all recordings to 72, randomly taken from the original matrix.

In these figures 8-11 we have used colors to distinguish between the four infants, which will not show in print. Boy 2 is represented in red, boy 3 in green, boy 4 in blue, and boy 5 in black. The rationale behind these figures is that the individual variance still shows, that overlap in month 12 is the smallest, and the biggest when the infants are 24 months old. In figure 12, the individual contribution in developmental change is given in lines, drawn between the individual means.

5.4. Individual contributions to the mean patterns

We reckoned that infants could still show certain preferences for vowel productions at the different ages, which may have to do with their ambient environment. In the figures 8 to 11 the individual contributions of the infants are displayed in the pc1-pc2 reference plane for the two-year-olds. The number of measurement points per infant per recording is not at all equal (see table 1). For the individual comparison we balanced the number of measurement points for all infants and all recordings to 72, randomly taken from the original matrix.

In these figures 8-11 we have used colors to distinguish between the four infants, which will not show in print. Boy 2 is represented in red, boy 3 in green, boy 4 in blue, and boy 5 in black. The rationale behind these figures is that the individual variance still shows, that overlap in month 12 is the smallest, and the biggest when the infants are 24 months old. In figure 12, the individual contribution in developmental change is given in lines, drawn between the individual means.
Individual contributions for 4 boys when 24 months old (ellipse is 1 s.d).

Figure 11. The 72 measurement points per infant when 24 months old in the two-year-olds reference plane (667 mps).

5.5. Individual developmental patterns

In the present section we will focus on the individual contributions of the 4 boys. In figure 12, the dotted inner box from figure 4 is enlarged and the overall mean changes for all 4 infants (must be read clockwise: 6-12, 12-18, 18-24) are given in dotted black lines. The mean developmental changes for the 4 infants are given: the lines are connecting the individual means per infant and per month. (In print the lines are in black or gray, but colors on line: blue for 6-12, green for 12-18, and red for 18-24 months).

For the lines in figure 12 we used the same random selection of 72 mps as was used for the figures 8-11. A random selection of mps from a matrix hardly changes the distribution in comparison to the overall distribution [7], but for a visual balance in the figures, equal numbers per infant are better. As a reminder for looking at figure 12: figures 2 and 13 show that the infants’ realizations of the /a/ vowel is in the (2;1) (pc1;pc2) area. Three out of the 4 infants start their vowel productions along the /i/-/u/ dimension, and they all end up in that same area when two years old.

5.6. Comparing infant and adult corner vowels.

Figure 13 shows the reference plan for two-year-olds with the areas for corner vowels of infants and of an adult. For the time being we have neglected the speaker normalization problem. The six ellipses represent 1 s.d. of the distributions of the /a/, /i/, and /u/ that were perceived by the second author, a native speaker of Russian, to be correctly pronounced by 24 month old infants, as well as by an adult speaker of Russian (female student). For the infants only 6 measurement points for /a/ resulted from the band filter analysis, 29 mps. for /i/, and 19 mps. for /u/. For the adult, 10 mps. per corner vowel are selected randomly from the matrix that contained 25 mps. for /a/, 30 for /i/, and 27 for /u/.

The corner vowel triangle for adult is smaller than for the infants, which is understandable since the smaller size of the infant’s oral cavity consequently produces higher frequencies. For the infants a corner vowel triangle is found as well, although the number of /a/ measurement points is rather low. This possibly is caused by the strict selection criteria: an utterance (a selected vowel) must have a duration of at least 60 ms. to be analyzable, and the vowel part from a whole utterance can be shorter than that duration.

The /i/ vowels of the infant occupy a larger area than in the adults, but the overlap is clearly present. It is tempting to interpret this phenomenon in the light of the palatalized vowel pronunciation in Russian: do infants hear a lot of palatalized vowels? The non-overlap of /a/ and /u/ possibly can be blamed partly on lack of speaker normalization.
6. Discussion and future work

If the general impression holds that Russian pronunciation is sloppy (reduced) then it is understandable that the infants alike do not vary much in their vowel space exploration. Certain sounds may pose articulatory difficulties as well, and then probably are acquired later than others. But this lack of variation may also come from the number of vowels that are distinctive in a language.

Comparing Russian (n=4) and Dutch infants (n=5) in their exploration of the vowel spaces of their ambient languages, we found similarities and differences. The dispersion, for example in 6 months old Dutch infants is large and remains quite large with a slight exception for month 12 [14]. Dutch has 12 vowels (and 3 of them are diphthongized) as well as three real diphthongs. For Russian infants the choices clearly are less large. A similarity between the two groups of infants is found in the production of the /i/ vowel, which is “correctly in place” like in this study when comparing to the respective adults. Probably, Dutch infants do hear a lot of palatalized vowel onsets as well, since the use of the Dutch diminutive (e.g. ’je’, ’tje’, ’kje’ and ’pje’) in Infant Directed Speech is frequent.

The relatively low number of measurement points derived from the 50 selected utterances per infant per recording is probably due to technical complications. The home recordings are made with a fixed microphone position: the older infant was allowed to move around, and mothers often were talking to other persons present. Even older siblings sometimes were in the room. This far-from-perfect sound quality limits the selection of analyzable utterances [e.g. 15].

An important step to be made is that we further develop insight in the relation between the PCA reference plane and the F1-F2 plane. Synthetic vowels will be needed in which the parameters, adjustable for infant articulation, have been well defined. We already have contacted the speech research group Institut de Communication Parlé in Grenoble, France [16], to start a common project for these research matters.

The developmental path for vowel acquisition in young infants is not yet a clear-cut story. Many aspects have not been considered at all, such as the number of vowels, the influence of imitation, the pathological processes in speech development. We mentioned above a comparable dataset for Dutch infants, which includes 5 normally hearing and 5 deaf infants that were followed to 24 months of age [6,7]. For Hungarian and Dutch, we compared two-year-old infants [17]. In the near future, we hope to present data for Dutch, Hungarian, and Russian children [18], and to cooperate with researchers involved in vowel development in still other languages.

7. Conclusions

The automated band filter analysis applied to unlabeled utterances holds promises, since the bias that a researcher may introduce in ‘formant picking’ is eliminated. The method however still needs refinement with regard to technical aspects such as critical signal-to-noise ratios as well as methodological aspects.

8. References

[10] www.phon.ucl.ac.uk/home/sampa/russian.html