Russian Infants and Children’s Sounds and Speech Corpuses for Language Acquisition Studies

Elena E. Lyakso, Olga V. Frolova, Anna V. Kurazhova, Julia S. Gaikova

The Child Speech Research Group
Saint-Petersburg State University, Saint-Petersburg, Russia

lyakso@gmail.com

Abstract
«INFRANTRU» and «CHILDTRU» are the first Russian child speech database. The corpus «INFRANTRU» contains longitudinal vocalizations and speech records (n=2967) of 99 children from 3 mos to 36 mos. by long utterance sequences and separate utterances in different psychoemotional state of the child. The database “CHILDTRU” contains the records (n=28079, 13956Mb) of 150 children’s speech at the age from 4 to 7 years. Speech material are presented by the following situations: spontaneous speech, answers to questions, reading, poetry or retelling a tale, count and alphabet, play. Speech files format is Windows PCM, 22050 Hz, 16 bit.

Index Terms: child speech database, spectrographic analysis, phonetic transcriptions, vocalizations, child speech

1. Introduction
Nowadays speech databases are important integral part of speech investigations. Their structure and technical characteristics depend on the investigation tasks. Database size, speech material: reading speech – pre-defined words-commands, sentences for phonemic representatives; induced, spontaneous speech; communication channel – stationary phone communication, mobile communication, broadband channel etc – are defined by the particular aim of use [1]. The modern speech databases are created mainly for the purposes of automatic adults speech recognition (for instance, SpeechDat-I, SpeechDat-II, SpeechDat-E, SpeechDat-Car, Speecon databases) and speaker verification [2]. Databases format defines by existing generally accepted technologies of speech recognition (Hidden Markov Models). The great amount of different speakers (several thousand), the presence of all phonemes (monophones) of the language, valid presence of dyphones and triphones are necessary. Such characteristics as utterance prosodic organization, dialogues structure are not considered, as a rule.

As far as it is known child speech databases are very few. This is due to a low level of child speech signal acoustics knowledge and, consequently, problems of existing technologies appliance to child speech recognition. Consequently there are some problems when using the existing technologies for the recognition of child speech. So generally accepted approaches to format of early child (from the first year of life) vocalizations databases and to the materials contained in these databases are absent.

For example, school children reading speech corpuses are known for English, Italian, Dutch languages [3, 4, 5]. These corpuses are used for reading mastering estimation and correction. Corpuses of induced, spontaneous and emotional child speech are used for speech disorders diagnostics and correction and for child-computer dialogues. For example such corpuses are known for Hungarian, Swedish, English, German pre-school and school children [6, 7, 8, 9, 10]. Vocalizations and speech corpus of children during the first four years of their life in different interaction situation in English-speaking families are used for studies of input’s meaning in speech development [11]. Well-known child speech corpus «CHILDES» [12] is used for linguistic investigations. Databases containing sounds produced by Russian children of the first year of life and spontaneous speech utterances of children on the first seven years of life do not exist.

The goals of our work are the collection of vocal and speech material for databases and analysis of these data for different studies of speech ontogenesis.

2. Method

2.1. Data collection
The corpus «INFRANTRU» is the first database containing vocalizations of children of the first three years of life in Russian language. The database was developed since 1999 to 2005 years.

The creation of the database was based on the following: a sufficient amount of informants (children, mothers, families) and their speech signals; the specificity of vocalizations in early childhood (in particular the absence of speech-like signals in children of the first year of life); the conditions of psychoemotional state of the child.

Material: The database contains 2967 recordings (70 records hour) of 99 children of the first three years of life, beginning from the third month of life, 83 children were normally developing, 16 children with neurological disorders according medical reports. For 76 children corpus contain the longitudinal vocalization’s data from 3 mos to 36 mos. The recordings were made every three months.

The database includes audio recordings of children born and living in Saint-Petersburg (the parents of the children are born in St-Petersburg of have been living there for more than ten years).

The child’s vocalizations and speech in following situations are represented: spontaneous utterances of children, mother-child vocal-speech interaction in the situations: «face to face», play, reading, spontaneous interaction. Each recording fragment corresponding to one situation at the given age is up to 5 min long. For some of the children of the first year of life separate utterances are represented pronounced in calm, comfortable and emotional condition. Additional sound signals registration of 88 orphans from 1,5 mos to 3 years olds was executed. They were 54 children of the first year of life, 7

1878

Copyright © 2010 ISCA

1878
children of the second year of life, 27 children of the third year of life. Sound signals were registered during children interaction with researchers and during playing with toys. It was impossible to collect sounds and speech data for orphans in different situations as we had made for children from families. Sound and speech data for orphans were presented separately.

The corpus «CHILDRU» is the first database containing speech material for 4 -7 years old Russian children. The child speech database was developed since 2003 to 2007. The database contains 28079 recordings (20 records hour) of 150 children: 142 children growing in family and 8 orphans - in Child Home. Database contains speech material of those children, whose sound and speech material was in database «INFANTRU», and other “new” children that were recorded from 4 years of age.

Records conditions: Place of record were home, kindergarten and Child Home. The following situations are represented: spontaneous speech (child speaks about subjects, chosen by himself without adult’s help), answers to questions, reading, poetry or retelling a tale, count and alphabet, play (interaction in play situation), interaction with parents and another adults, interaction between children. All situations were recording for 5 min. Every record is accompanied by detailed protocol and parallel videotaping.

The recordings were made by the “Marantz PMD222” recorder with a “SENNHEIZER e835S” external microphone and digital recorder “Marantz PMD660". The sounds were instrumentally analyzed, selected and described in the Cool Edit Pro (Syntrillium Soft. Corp. USA) sound editor and the Praat v. 4.3.

2.2. Corpus structure

Two corpuses have a similar structure. “INFANTRU” and “CHILDRU”

<table>
<thead>
<tr>
<th>FILE</th>
<th>CONTENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>VDB.EXE</td>
<td>Database program shell</td>
</tr>
<tr>
<td>BORLNDM.DLL</td>
<td>Backup file</td>
</tr>
<tr>
<td>CC3260MT.DLL</td>
<td>Backup file</td>
</tr>
<tr>
<td>RTL60.BPL</td>
<td>Backup file</td>
</tr>
<tr>
<td>VCL60.BPL</td>
<td>Backup file</td>
</tr>
<tr>
<td>VCLX60.BPL</td>
<td>Backup file</td>
</tr>
</tbody>
</table>

The VDB.EXE program was created for effective work with database. “INFANTRU”: The program allows to choose and to play speech recordings using the following characteristics: child number, child gender, the child’s number in the family, child’s illnesses, mother’s age at child birth, recording situation, emotional state, child’s age, family type (complete / incomplete), the person bringing up the child. It is a shell for database search and allows to choose and to play speech recordings according to the features above.

Different situations amount for every children age in “INFANTRU” is represented in table 1. From data presented in table 1 we can conclude that interaction in mother-child dyads in situation face-to-face decreases with children age. Interaction in mother-child dyads in situation “reading book” increases with children age. Possibility of recording sounds and speech in situations “child alone” (O) decreases with children age.

Table 1. Different situations amount for every children’s age are represented in “INFANTRU” (percentage of common amount of situations for every age).

<table>
<thead>
<tr>
<th>Age, mos</th>
<th>Or</th>
<th>F</th>
<th>P</th>
<th>O</th>
<th>B</th>
<th>S</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>148</td>
<td>23</td>
<td>22</td>
<td>2</td>
<td>1</td>
<td>20</td>
<td>32</td>
</tr>
<tr>
<td>6</td>
<td>147</td>
<td>14</td>
<td>18</td>
<td>5</td>
<td>2</td>
<td>29</td>
<td>32</td>
</tr>
<tr>
<td>9</td>
<td>146</td>
<td>9</td>
<td>19</td>
<td>4</td>
<td>5</td>
<td>30</td>
<td>33</td>
</tr>
<tr>
<td>12</td>
<td>141</td>
<td>8</td>
<td>25</td>
<td>2</td>
<td>11</td>
<td>26</td>
<td>28</td>
</tr>
<tr>
<td>15</td>
<td>124</td>
<td>3</td>
<td>25</td>
<td>1</td>
<td>15</td>
<td>36</td>
<td>20</td>
</tr>
<tr>
<td>18</td>
<td>128</td>
<td>0</td>
<td>29</td>
<td>2</td>
<td>14</td>
<td>33</td>
<td>22</td>
</tr>
<tr>
<td>21</td>
<td>117</td>
<td>1</td>
<td>26</td>
<td>0</td>
<td>17</td>
<td>41</td>
<td>15</td>
</tr>
<tr>
<td>24</td>
<td>117</td>
<td>0</td>
<td>24</td>
<td>1</td>
<td>22</td>
<td>39</td>
<td>14</td>
</tr>
<tr>
<td>27</td>
<td>115</td>
<td>0</td>
<td>36</td>
<td>0</td>
<td>18</td>
<td>33</td>
<td>13</td>
</tr>
<tr>
<td>30</td>
<td>116</td>
<td>0</td>
<td>31</td>
<td>1</td>
<td>21</td>
<td>34</td>
<td>13</td>
</tr>
<tr>
<td>33</td>
<td>108</td>
<td>0</td>
<td>43</td>
<td>0</td>
<td>21</td>
<td>29</td>
<td>7</td>
</tr>
<tr>
<td>36</td>
<td>110</td>
<td>0</td>
<td>43</td>
<td>0</td>
<td>16</td>
<td>33</td>
<td>8</td>
</tr>
</tbody>
</table>

Or – original records, F – face to face interaction, P - playing with a toy, O - the child is alone, B - book reading, S - spontaneous interaction, N - emotional state – neutral.

“CHILDRU”: This program allows to choose speech material in accordance with following parameters: original record; phrases; dialogue; questions; articulation mistakes:
mistakes in word in phoneme «r»; other mistakes in word –
different variants of gap, change, transposition of phonemes
and syllables; grammatical mistakes; separately pronounced
sounds; syllables; reading of words and phrases; words with
different syllables amount – one, two, three, four, five or more
syllables. Choice of speech material could be made as for all
children, as for every child. It is a shell for database
“CHILDRU” search and allows to choose and to play speech
recordings according to the features above (table 2).

Table 2. Different situations amount for every children’s age
are represented in “CHILDRU” (percentage of common
amount of situations for every age).

<table>
<thead>
<tr>
<th>Age</th>
<th>Codes of situations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Or</td>
<td>S</td>
</tr>
<tr>
<td>4y</td>
<td>236 4 9 34 33 7 13</td>
</tr>
<tr>
<td>4y6m</td>
<td>97 9 35 14 13 20</td>
</tr>
<tr>
<td>5y</td>
<td>350 3 11 42 11 12 21</td>
</tr>
<tr>
<td>5y6m</td>
<td>150 9 13 33 11 13 21</td>
</tr>
<tr>
<td>6y</td>
<td>128 6 9 38 20 19 8</td>
</tr>
<tr>
<td>6y6m</td>
<td>93 5 11 28 14 14 28</td>
</tr>
<tr>
<td>7 y</td>
<td>56 11 12 23 11 12 31</td>
</tr>
</tbody>
</table>

Or – original records, amount of original records, n; S -
Spontaneous speech, R – read, A - Answers to questions, Play,
C - Count and alphabet, T - Poetry and tale

3. Result

We use the data from “INFANTRU” and “CHILDRU”
speech corpuses for studies of Russian children speech
ontogenesis. Some data that were selected from databases are
presented below.

“INFANTRU”: Example: The description and comparing
of acoustic characteristics of different types of vocalizations
(cry, cooing, babbling and first words) were made for
normally developed children, preterm newborns,
neurologically disordered infants and children from Child
Homes. It was shown that vowel-like sounds prevail in norm
(normally development children) and risk group (children with
neurological disorders) children’s sounds repertory during the
first year of life. Different vocalization categories (vowels,
consonants and syllables) of 12 mos children with different
neurological state are in figure 1. Amount of syllables in norm
group exceeded amount of syllables in risk group. First words
in norm group children appeared at the age of 12 mos, in risk
group children the first words were absent up to 24 mos.
These data indicated lower speech development level in risk
group children vs. norm group children.

“CHILDRU”: 4-7 years olds active lexicon, their
articulation mistakes in words, vowels acoustic characteristics
in different contexts were estimated. Some examples of these
data are presented below.

Example 1: Two syllables words are prevail in child
lexicon (fig. 2). Amount of words with 3 or more syllables
increased with child age. Since 5y 6m to 7y of child age the
amount of words with different syllables have been permanent
in children lexicon.

Example 2: The analysis of articulation mistakes in speech
in different situations shown that the prevalent amount of
children had mistakes when answering for adult’s questions
then in spontaneous speech. We could suggest that adult
stimulated the child to use in communication process the
greater variety of more complex articulation words.

![Figure 1: The ratio of vowels, consonants and syllables in the 12 mos olds sounds repertory](image1)

![Figure 2: Words with different amount of syllables in 4-7 years old children lexicon](image2)
The classification of mistakes connected with consonants substitutions are based on the place of formation: fricative, affricate, lateral sonant, implosive, other (unclear types of consonant mistakes). In 4 years old speech more often frequency of fricative substitution: /sh/ for /s/ (loshadki – lasadki); /j/ for /s/ (knijka – kniska); /j/ for /z/ (lojitsja-lozitsja); in 5 years old child’s speech – substitution /sh/ for /s/ (fig. 3) were revealed. Reduction of the variety of mistakes before age of 6y6m possess more difficult in words pronouncing with consonant /sh/ articulation (fig. 3).

Example 3: Acoustic characteristics of 6 and 7-year-old Russian children vowels [13].

F1, F2 median values of vowels /a/ were chosen with regard to factors of context and pitch are situated in the area of two-formant plot typical for F1, F2 values of adult’s /a/ (fig. 4), but they are not coincide with the basic allophone from adult speech. F1 median values of /i/ and /a/ vowels are higher than in adult speech. The unstressed vowels triangle’s shape and orientation differs from the stressed vowels triangle. The shape and orientation of the stressed vowels triangle where the vowels are chosen without regard to context and pitch is not correspond completely to the stressed vowels triangle where the vowels were chosen with regard to these factors mainly in 6 years old children. 6-years-old children’s stressed vowels triangle where the vowels are chosen with regard to context and pitch occupies larger area on the two-formant plot then the same triangle built for vowels of 7-year-old children (fig. 4). So these examples reflect main direction of our researches in Russian language ontogenesis.

5. Conclusion

From our point of view, children speech material from the databases could be the basis for scientific projects about Russian language mastering by Russian children. These corporuses could be used in developing of child speech automated recognition systems.

6. Acknowledgements

The study is being performed with the financial support from the Russian Fund for Humanities (projects №06-06-12623b) and Russian Fund for Basic Investigations (project № 09-06-00338a).

7. References


