

Phonetic Properties of Russian Spontaneous Speech

Liya V. Bondarko, Nina B. Volskaya, Svetlana O. Tananaiko and Ludmila A. Vasilieva

Department of Phonetics, Saint-Petersburg State University,

Universitetskaya emb., 11, Saint-Petersburg, Russia

E-mail: LVBon@LB1082.spb.edu

Abstract

This article deals with the results of the systematically performed comparative study of read and spontaneous Russian speech. The data described were obtained in course of fulfilling the RFBR project No 01-06-80188 "Phonetic Properties of Russian Spontaneous Speech". Recordings of Russian spontaneous speech were made from 5 Russian standard speakers (the speakers were chosen from different age groups, from 18 to 55 years old). Then the read versions of spontaneous dialogues based on a transcript of spontaneous speech were recorded from the same speakers. Lexically identical fragments of spontaneous speech and reading were chosen and compared. For each speaker the duration of the sound material was at least 10 minutes. A multilevel segmentation of the material into intonation units, accent groups, open syllables and phonemes was performed, followed by prosodic and segmental analysis. A hierarchically presented data lets study every unit or segment in the context of a given and higher level of presentation. Further statistical processing gave the data on the comparison of three types of transcription, on the sound duration and on formant characteristics of vowels in two types of speech.

1. Procedure

At the prosodic level intonation unit boundaries were determined. The most important criterion for the demarcation of intonation units (IU) is the occurrence of the nucleus. One of the most important clues used for the detection of IU boundaries is the assumption that it should contain only one principal pitch prominence (intonation center), though analysis showed that in spontaneous speech there are many IUs, having two pitch prominences. Number of intonation units, their length, number of pauses and their duration for each type of speech were determined. At the segmental level three types of segmental transcription were performed: "ideal" phonemic, perceptual (based on the experts' evaluation) and acoustic (with the acoustic data used). Their comparison provided data on vowel modifications. Vowel duration data were obtained. Segment boundaries were defined using the acoustic data of the oscillograms and spectrograms. Vowel transitions were included into the vowel segment boundaries. Identified vowel insertions (inside consonant groups), if not long and intense enough to be classified as independent sounds, were

considered as a part of the preceding consonant. In the acoustic transcription each sound was transcribed regardless of its phonetic context. The perceptual transcription was obtained through listening to segments no shorter than a phonetic word. Syllable boundaries were defined on the basis of the "open syllable" principle. Final consonants preceding the pause were considered as a separate syllable, with the exception for /j/ in adjective endings, which formed part of the same syllable as the preceding vowel. Segmental and prosodic analyses were performed using PRAAT and EDS programs.

2. Results of the Prosodic Analysis

Prosodic features of read and spontaneous speech in the experiment are characterized primarily by differences in prosodic phrasing. Table 1 shows that the number of IUs is regularly greater for spontaneously produced texts for all the 5 speakers (in the tables letters F, M indicate gender; figures stand for age category).

	Speaker	F<20	F>20	M>50	M>40	M<20	Means	SD
Num IU	Spontan	257	123	188	296	256	224	68.5
	Read	222	117	162	271	222	198,8	59.9

Table 1. Number of intonation units in spontaneous and read texts.

Intonation unit length. The value is 1,5 sec both for spontaneous and read speech (Table 2), if measurements include pauses. It should be noted though, that in calculations all types of pauses, occurring inside the IU and following it were considered. With pauses excluded from calculations, the data is 1,1 both for spontaneous and read texts (averages over all speakers). Averages over all speakers conceal the inter-speaker differences: for 3 of the

	Speaker	F<20	F>20	M>50	M>40	M<20	Means	SD
1	Spontan	1.7	1.5	1.5	1.7	1.3	1.5	0.17
	Read	1.5	1.6	1.8	1.4	1.2	1.5	0.22
2	Spontan	1.2	1.2	1.0	1.2	1.0	1.14	0.09
	Read	1.1	1.1	1.3	1.1	1.2	1.12	0.11

Table 2. Average IU length with and without pauses for spontaneous and read texts.

5 speakers the IU length with pauses in spontaneous speech is greater, than in the read text.

Phrasing and IU length are closely connected with **pausing strategy** and **pause duration**. Spontaneous speech is characterized by two types of pauses: syntactic and hesitation ones, the latter can be realized either like filled or unfilled ones. These pauses may appear in various combinations with each other, along with repetitions and false starts, within one unit or at unit boundaries, which makes the analysis of spontaneously produced text extremely difficult and error-prone. In our material pauses are shorter in spontaneous speech than in reading (Table 3). In the read text prosodic and syntactic boundaries coincide more often than not, which is quite natural, since speaker's pausing strategy is guided by the syntactic structure of the text.

Speaker	F<20	F>20	M>50	M>40	M<20	Mean	SD
Du-	445	312	664	543	517	496	129
rat.	543	507	568	509	443	514	47

Table 3. Average pause duration in spontaneous and read texts.

Prosodic boundaries can be marked by different prosodic parameters: tonal changes, pre-boundary lengthening, declination reset and pauses. **Zero pauses** and very short pauses with duration below 200 ms at the boundaries of syntactic units in fluent reading and speech are very common, but the number of IUs linked to each other without a pause (with a zero pause) is greater for spontaneous speech than for reading for all the 5 speakers in the experiment (see Table 4). Zero pauses are typical for prosodic boundaries that separate non-final phrases and subordinate clauses from the main clause.

Speaker	F<20	F>20	M>50	M>40	M<20	Mean	
IUs with							
0 pause	Spont	66	43	33	31	58	46.2
	Read	49	20	15	26	62	34.4
Pauses							
<200ms	Spont	14	19	15	11	11	14
	Read	19	26	22	17	22	21.2
Virtual							
pauses	Spont	80	62	48	42	69	60.2
	Read	68	46	37	43	84	55.6

Table 4. Percentage of units with "zero" and very short pauses for 5 speakers.

These pauses can be regarded as virtual from the point of view of the listener, since there is no discontinuity of the speech signal at the perceived boundary of the IUs. Their number in spontaneous speech is greater than in reading for all the speakers (Table 4).

3. Vowel and Consonant Characteristics

The hierarchical representation of the data enabled us to obtain the following characteristics for each sound: three types of transcription of each allophone, the syllable that contains it, its position relative to the stress, the length of the sound and the corresponding syllable, the syllable position inside the intonation unit, the number of syllables in that intonation unit, the vowel formant values in three points (the onset, the middle and the end of the sound).

The data were further processed with Statistica software package. The types of transcription for each sound were compared (the cases when one acoustic sound was perceived as several and visa versa were treated separately). The changes of sound duration relative to the type of speech were obtained. To compare the duration characteristics of reading and spontaneous speech only /a/, /i/, /u/ were selected, i.e. the cardinal vowels of the articulation triangle; besides, these vowels occur in standard Russian both in stressed and unstressed positions. For these vowels the relative duration values in reading and spontaneous speech were received. The absolute duration values below 0,030 sec and over 0,250 sec were not included into the analysis. For relative duration values the absolute duration values were divided by the average sound duration for every speaker. The data for 5 speakers were compiled. The statistical processing shows greater variation of vowel duration in spontaneous speech than in reading, because the variance and the standard deviation values for spontaneous speech are higher (see Table 5).

		<i>i</i>	<i>u</i>	<i>a</i>
Mean	Spont	0.838	0.921	1.060
	Read	0.852	0.962	1.110
Minimum	Spont	0.381	0.380	0.381
	Read	0.385	0.421	0.381
Maximum	Spont	3.682	3.428	3.557
	Read	3.713	2.512	3.830
Variance	Spont	0.202	0.223	0.255
	Read	0.131	0.156	0.201
Standard Deviation	Spont	0.449	0.472	0.505
	Read	0.364	0.395	0.449
Valid Number	Spont	1007	670	2345
	Read	1098	575	2309

Table 5 Statistical data on relative duration values of cardinal vowels in reading and spontaneous speech.

The statistical data for the main consonant groups (voiceless and voiced obstruents and sonants) were also obtained. The absolute duration values below 0,030 sec and over 0,150 sec for plosives, below 0,030 sec and over 0,200 sec for fricatives and below 0,030 sec and over 0,250 sec for sonants were not included into the analysis. This explains the fact that the number of the consonants in spontaneous speech and in reading is different. The study of consonants gave the following results (Table 6). Average

duration values for obstruents in reading are greater than in spontaneous speech, and the difference between the variance and the standard deviation values for the spontaneous speech and reading isn't significant. The distribution of sonant duration values resembles that of the vowels, the variance and the standard deviation values for the spontaneous speech being higher than for reading with the result that vowels and sonants can be included in the same group. For these sounds the relative duration values in spontaneous speech vary more than in reading.

		<i>PV-</i>	<i>PV+</i>	<i>FV-</i>	<i>FV+</i>	<i>Son</i>
Mean	S	1.165	0.939	1.193	0.891	0.833
	R	1.208	1.011	1.264	0.920	0.870
Minimum	S	0.385	0.383	0.386	0.388	0.381
	R	0.386	0.450	0.419	0.390	0.382
Maximum	S	2.227	2.201	2.625	2.432	3.269
	R	2.369	2.323	3.016	2.427	3.023
Variance	S	0.108	0.081	0.201	0.111	0.115
	R	0.111	0.087	0.220	0.094	0.092
Std Dev	S	0.328	0.285	0.448	0.334	0.340
	R	0.333	0.296	0.469	0.307	0.303
Valid Num	S	2528	971	1404	650	2933
	R	2461	870	1338	826	2914

Table 6 Statistical data on relative duration values of consonant types in reading (R) and spontaneous speech (S), *PV-* stands for plosive voiceless, *PV+* for plosive voiced, *FV-* for fricative voiceless, *FV+* for fricative voiced, *Son* for sonants.

4. Formant Values in Different Syllables

The formant characteristics of reading and spontaneous speech were calculated for vowels /a/, /i/, /u/. The analysis was restricted to the first two formants. The areas of formant values for all three vowels (Figures 1-3) were built on the coordinate plane, with values of the first and the second formants shown on axes X and Y. The peak coordinates for each square are the extreme values of each formant. The information was collected separately for pre-stressed, stressed and post-stressed vowels. On the same coordinate planes there are triangles, which peaks are mathematically expected formant values for the same vowels. The figures show that the deviation of formants in reading is less noticeable than in spontaneous speech. Vowel realization in spontaneous speech is more varying as the formant value area is broader. As the triangles for post-stressed vowels are almost identical, the quality reduction of the vowels in this position is permanently great regardless of the type of speech.

5. Conclusions

Results of the described experiment allow us to draw some general conclusions about differences in speakers' behavior

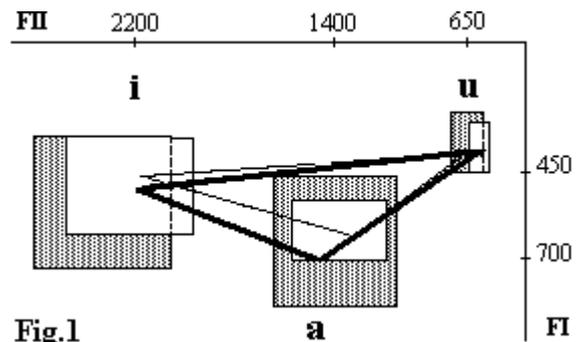


Fig.1

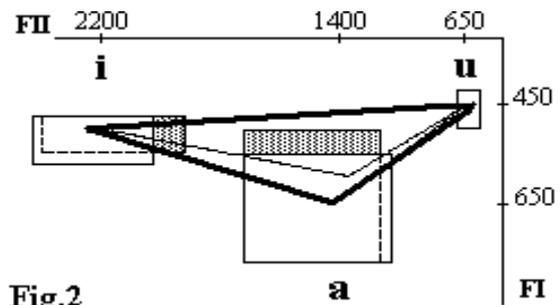


Fig.2

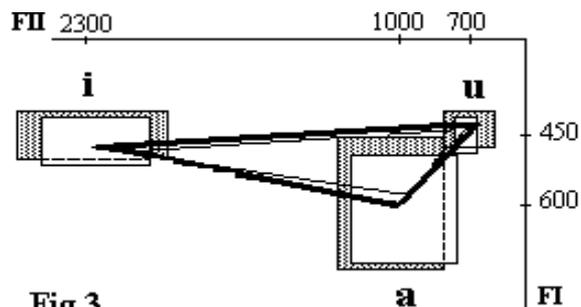


Fig.3

Figures 1-3. Formant value areas for vowels /i, a, u/ in the stressed (Fig.1), pre-stressed (Fig.2) and post-stressed (Fig.3) positions. The formant value areas for spontaneous speech are shaded, the formant value areas for reading shown above them, are white, the spontaneous speech triangle is **—**, the reading triangle is **---**.

in producing spontaneous text and in reading. Some parameters are more informative in this respect.

1. The problem of the **intonation unit length** has been discussed in prosodic literature for quite a long time, first of all in connection with the development of prosodic rules for speech synthesis. The data reported for spontaneous and read speech for different languages reveal certain similarities in the IU length. For English, German and French, for example, 1,5 sec seems to be the most frequent "interpause distance". Results of our experiment show that

Russian is not an exception. This looks like a universal, reflecting the temporal structure of the text and the speaker's strategy in generating it.

2. Average pause duration is greater in the read text, than in spontaneous speech, at the same time the number of IUs and the number of virtual pauses at the boundaries of the IUs is considerably greater in spontaneous speech than in reading.

The following conclusions were made **at the segment analysis level**.

1. The quantity difference recorded in various types of transcription is not significant, the acoustic transcription identifies about 95% of sounds registered in the perceptual transcription. The difference between the number of sounds recorded in the same type of transcription in reading and spontaneous speech is about 2-3%. Yet, there are considerable differences in the way the same segment units are realized in reading and spontaneous speech.

2. The study of the first and the second formant values has confirmed that spontaneous speech is characterized by greater variability of the acoustic properties of vowels. Besides, it supports the point of view that the Russian vowels fall into two groups according to their acoustic characteristics: the stressed and pre-stressed vowels constitute one group, the post-stressed vowels – the other. The differences between these two are greater than between the traditionally distinguished stressed and unstressed vowels in general.

3. The vowel transitions can be identified in all the vowels both in reading and in spontaneous speech, while the stationary parts are often absent from spontaneous speech as the vowels in this case are more reduced. In its turn, a greater degree of reduction is closely related to uneven tempo: spontaneous speech is rich both in hesitations, sound prolongation and increases of tempo as well.

4. Vowel insertions inside the consonant groups is obviously a universal characteristic of the modern standard pronunciation: they are common both in reading and spontaneous speech.

5. Another typical characteristics of the modern pronunciation standard is partial or total devoicing of voiced consonants in various positions where devoicing isn't allowed by standard assimilation rules, and also devoicing of the sonants surrounded by voiceless consonants.

6. Omission of unstressed, especially post-stressed vowels characterizes both types of Russian speech. It should be noted that the syllable structure of the word is preserved at the perception level when syllable-forming vowels of this word are completely dropped (as registered at the acoustic level). In this case the syllable may be formed by a sonant, a voiced consonant or a voiceless fricative. If the omission of

the syllable-forming vowel takes place after a voiced consonant before a voiceless plosive, the first consonant remains voiced and there is no assimilation necessary in Russian pronunciation standard (see Figure 4). In this case the syllable is still perceived as such, because a voiced consonant becomes a syllable-forming sound. This result makes essential further research into the tendencies of consonant assimilation in standard Russian.

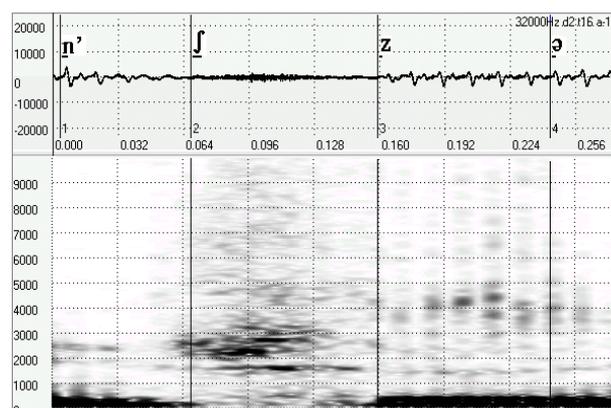


Figure 4. The omission of the post-stressed vowel in the word group /ran'ʃə zan'imaləs'/ from the spontaneous speech. The vowel /ə/, marked in bold type, is omitted, so the voiced /z/ follows the voiceless /ʃ/, and there is no assimilation. The syllable /ʃə/ is still perceived as such, the voiced /z/ being the syllable-forming sound.

Summing up, we should state that similar phonetic phenomena, observed in both types of speech, are caused by generating long speech fragments. They lead to considerable deviation of the real speech data from what is theoretically expected. The characteristics of spontaneous speech as compared to reading at this stage of research can be rather described as quantitative: the probability of a certain phenomenon in spontaneous speech is higher than in reading.

Acknowledgments

The authors thank the researchers of the Experimental Phonetics Laboratory (Saint-Petersburg State University) Prof. Pavel Skrelin, Prof. Natalia Svetozarona, Dr. Svetlana Stepanova and PhD students Ekatherina Volskaya and Olga Bolotova, who carried out the study along with the authors.

The study is supported by INTAS -- project No915 "Spontaneous Speech of Typologically Unrelated Languages (Russian, Finnish and Dutch)".