On some acoustic features of spontaneous speech and reading in Russian (quantitative and qualitative comparison methods)

Olga Bolotova
Department of Phonetics, St. Petersburg State University, Russia
E-mail: obolotova@gg7414.spb.edu

ABSTRACT

This paper deals with comparative characteristics of spontaneous speech and reading in Russian. The material used was obtained in the course of fulfilling the INTAS-915 research project. High quality sound recordings of spontaneous speech and reading of five standard speakers of Russian were made (the procedure is described in detail in Bondarko et al. in this volume). Identical fragments of both types of speech were chosen for comparison. To obtain relevant results, large amounts of data were analyzed with the help of computer software developed for that purpose (Praat scripts). The methods to reveal the phonetic differences between the two types of speech are described. The characteristics under analysis are segment durations, stress dependent vowel duration and different types of transcription.

1. INTRODUCTION

The material was segmented into intonation units, accent groups, open syllables and phones. At the segmental level three types of segmental transcription were performed, acoustic, perceptual and "ideal" phonemic. The acoustic transcription was performed on the basis of acoustic characteristics of the sound, the perceptual transcription is based on the expert evaluation of word-like length units. The "ideal" transcription is a phonemic representation of theoretically expected sound structure. The syllables were marked as unstressed or stressed (being in word stress position). The stressed syllable of the intonation center was marked as principal pitch prominent during segmentation into intonation units.

With the help of Praat scripts the following characteristics for each sound were obtained: three types of transcription of each sound segment, the syllable containing it, the stress mark of the syllable, the length of the sound and the corresponding syllable, the syllable position within the intonation unit, the number of syllables in the intonation unit.

One script was used to compare the transcription of the same audio segments on different transcription levels (ideal phoneme, perceptual and acoustic transcription) and to compare features of different speaking styles. Another script was used to measure vowel characteristics, and to compare vowel durations in different positions with regard to the stress for the two types of speaking. The data were further processed with Statistica software package.

2. COMPARISON OF THE THREE TYPES OF TRANSCRIPTION

The frequency of occurrence of different allophones in the three types of transcription were compared. The data presented in this paper show the results obtained for one female speaker. The number of occurrences of vowel allophones was calculated. Then the vowel allophones were divided into seven main groups, each group approximately representing one phoneme (/i/, /e/, /ɨ/, /ə/, /o/, /u/ and /a/). For the perceptual or the "ideal" transcription analyses, combinations other than vowel and /j/, another vowel or sonorant were not included in the statistics. The statistics for the acoustic transcription included all vowel sounds.

a-group	æ, ɑ, ʌ, ɤ, a, ɐ, combinations /a/ + j, /a/ + V, /a/ + sonorant
i-group	i, ɪ, j, j + V, combinations /i/ + j, /i/ + V, /i/ + sonorant
o-group	ɔ, o, œ, ɶ, ø, ɵ, combinations /o/ + j, /o/ + V, /o/ + sonorant
u-group	ʊ, u, ɯ̞, ʏ, y, combinations /u/ + V
ɨ-group	ɨ, ʉ, combinations /ɨ/ + V, /ɨ/ + j
e-group	e, ɛ, combinations /e/ + V
ə-group	ɜ, ə, 9

Table 1: The list of quasi-allophones of the seven main vowel groups.

For instance, Table 2 shows the frequency of vowel /a/ quasi-allophones in three types of transcription in spontaneous speech and reading.

ISBN 1-876346-48-5 © 2003 UAB

a-group	Spontaneous			Reading		
	Acou	Perc	Ideal	Acou	Perc	Ideal
æ	39	5		35	6	
ɑ	21	27	346	39	60	375
ʌ	16	58		21	67	
ɤ	1	2		1	15	
a	170	198	5	200	101	1
ɐ	14	2		4		
ɑj		1	3			4
ɑV		4	7	4	2	12
ɑ-son		1	1			
Total	261	298	362	304	251	392

Table 2: Number of /a/ quasi-allophones occurrences in acoustic, perceptual and "ideal" phonemic transcriptions in spontaneous speech and reading.

Figure 1 represents the data on vowel groups frequency of occurrence according to the acoustic transcription as compared to the expected realisation based on "ideal" phonemic transcription data. The number of realized quasi-allophones of /i/ and /a/ is considerably smaller in both types of speech (45% of the expected /i/ realisations in spontaneous speech and 59 % in reading, and 72 % and 78 % of the expected /a/ realisations respectively). It is worth mentioning that the number of /o/ and /e/ quasi-allophones is unexpectedly great. Theoretically, these vowels occur only in stressed positions. The data show that in both types of speech, especially spontaneous, they can also occur in unstressed positions. Unstressed /e/ mostly replace expected /i/ allophones. The significant number of /o/ and also /u/ vowel groups is due to the "labial harmony" typical for connected speech, when instead of unstressed /a/ before unstressed labial /u/ a labialized vowel is realised ([2], [4]). The o-like vowels in this position were centralised (θ).

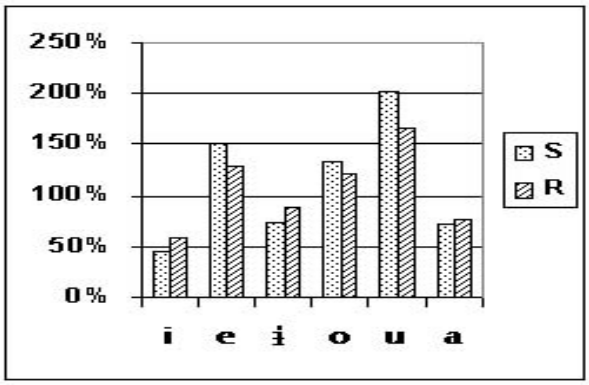

Figure 1: Ratios of the actual vowel group frequencies to the expected ones in spontaneous speech and reading.

Table 3 summarises the data on the distribution of vowel sounds among the seven vowel groups in 3 types of transcription, the total number of the actual vowels in each group being 100%.

	Acoustic		Percept.		Ideal	
	S	R	S	R	S	R
i	**11%**	13%	**21%**	19%	**25%**	23%
e	**11%**	9%	**10%**	8%	**8%**	7%
ɨ	**4%**	4%	**5%**	4%	**5%**	5%
ə	**14%**	10%	**7%**	14%		
o	**14%**	14%	**11%**	12%	**11%**	11%
u	**18%**	17%	**13%**	14%	**10%**	10%
a	**29%**	33%	**34%**	28%	**42%**	43%
Total	**909**	921	**870**	900	**864**	906

Table 3: The percentage of the quasi-phonemes in three types of transcription.

The table shows that the number of vowels of each type is not the same for different types of transcription [1]. The considerable part of realisations is composed of schwa-like vowels, especially in spontaneous speech [3].

Perceptual transcription tends to preserve the "ideal" ratios more than acoustic as the perceptual labelling was performed on the bases of word-like units.

3. VOWEL AND CONSONANT CHARACTERISTICS

The sound duration data according to the type of speech were obtained for five speakers. To compare the duration characteristics of reading and spontaneous speech, the allophones of the cardinal vowels /a/, /i/, /u/ were selected. The duration values were measured within each vowel group. The absolute values below 0,030 s and over 0,250 s were not included in the following analysis. For every speaker the absolute duration values were divided by the average sound duration in order to receive relative duration data comparable for all speakers. The relative duration values for each speaker were grouped together.

The statistical data obtained show that the mean values of the vowel duration are greater in reading (see Table5, Bondarko et al. in this volume).

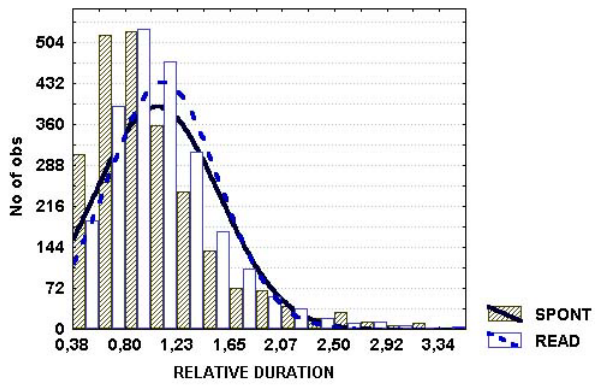

Figure 2: The distribution of /a/ relative duration values in spontaneous speech and reading.

The greater values of variance and standard deviation in spontaneous speech show greater variation of vowel duration (Figure 3).

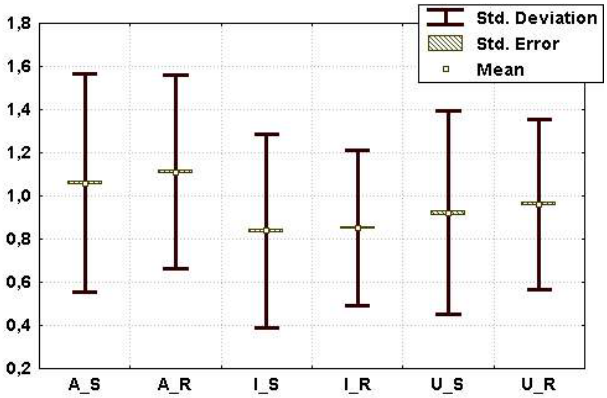

Figure 3: Statistic characteristics of vowel /a/, /i/, /u/ durations in spontaneous speech (S) and reading (R).

To compare the means of cardinal vowel duration values in spontaneous speech and reading the t-tests for independent samples were computed. The resultant t-value shows that there is no statistically significant difference between the mean values in spontaneous speech and reading for /i/ and /u/. However, the means for vowel /a/ in spontaneous speech and reading are different. This is caused by the fact that in Russian the duration of /a/ is characterized by greater variability in different positions with regard to the stress, this tendency being especially evident in spontaneous speech. The closed vowels /i/ and /u/ tend to be less variable.

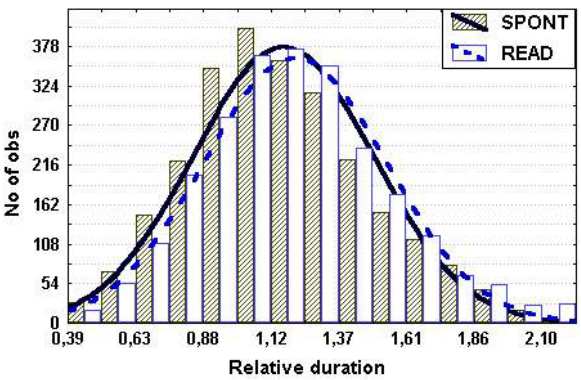

Figure 4: The distribution of voiceless plosives relative duration values in spontaneous speech and reading.

The statistical data were also obtained for the main five consonant groups: voiceless and voiced obstruents and sonorants. The mean duration values for obstruents and sonorants are greater in spontaneous speech as compared to reading (see Table 6 Bondarko et al. in this volume). The variance and standard deviation values for sonorants are greater in spontaneous speech, but those for obstruents show no considerable difference in spontaneous speech and reading (Fig. 5). Thus the sonorants duration characteristics resemble those of the vowels.

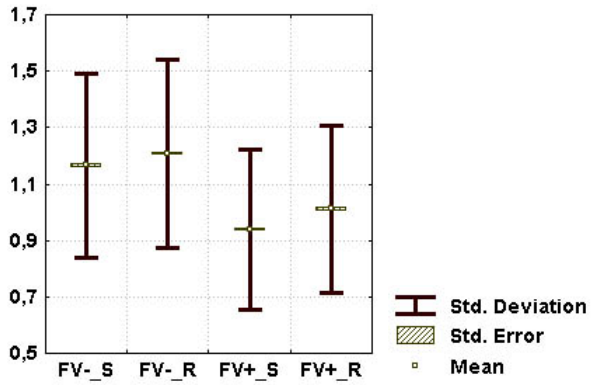

Figure 5: Mean, standard deviation and standard error values for voiceless and voiced fricatives in spontaneous speech and reading.

The computed t-values for independent samples show statistically significant difference in consonant duration means between spontaneous speech and reading for the sonorants and all obstruent types except for the voiced fricatives.

4. THE VOWEL DURATION VARIATION DEPENDING ON STRESS

Table 4 shows the statistical characteristics obtained for the duration values for stressed and unstressed vowels. The STR column contains the data on all vowels in stressed syllables regardless of the position of the syllable in the intonation unit. WSTR column contains the data on duration values of the stressed vowels, except for the stressed vowels in the intonation centre. SYNTSTR column includes the data on the duration values of the pitch prominent vowels (in the intonation center syllable). The average duration values in reading are greater than in spontaneous speech. The fact that the standard deviation value for the unstressed vowels as well as for the non-prominent stressed vowels is greater in spontaneous speech also shows its greater variability. The standard deviation values of all stressed vowels and those of prominent vowels is greater in reading.

The number of WSTR vowels in reading and spontaneous speech appeared to differ significantly. Quite unexpectedly, this number is greater for spontaneous speech, where auxiliary words and pronouns were in stress position more often than in reading.

The mean values for vowels of each type (UNSTR, STR, WSTR, SYNTSTR) in spontaneous speech and reading were compared. The t-test for independent samples shows statistically significant differences between the means of vowel duration values in spontaneous speech as compared to reading at the significance level of 95 % for all types of vowels except for the WSTR type (see Fig. 6). The probability that stressed vowel means are equal in both types of speech is p = 0,177584.

ISBN 1-876346-48-5 © 2003 UAB

		UNSTR	STR	WSTR	SYNTSTR
Mean	**S**	**0,059**	**0,084**	**0,075**	**0,099**
	R	0,064	0,096	0,079	0,118
Conf. Interval	**S**	**0,056-0,062**	**0,080-0,088**	**0,070-0,079**	**0,093-0,106**
	R	0,062-0,067	0,091-0,100	0,075-0,083	0,111-0,125
Min	**S**	**0,011**	**0,020**	**0,020**	**0,028**
	R	0,011	0,032	0,032	0,038
Max	**S**	**0,258**	**0,285**	**0,285**	**0,240**
	R	0,194	0,338	0,201	0,338
Variance	**S**	**0,001**	**0,001**	**0,001**	**0,001**
	R	0,001	0,002	0,001	0,002
Std. Dev.	**S**	**0,035**	**0,038**	**0,035**	**0,038**
	R	0,030	0,040	0,027	0,044
Std. Error	**S**	**0,002**	**0,002**	**0,002**	**0,003**
	R	0,001	0,002	0,002	0,004
Number	**S**	**550**	**359**	**225**	**134**
	S	580	331	189	142

Table 4: Stressed (word stress and pitch prominent) and unstressed vowels characteristics.

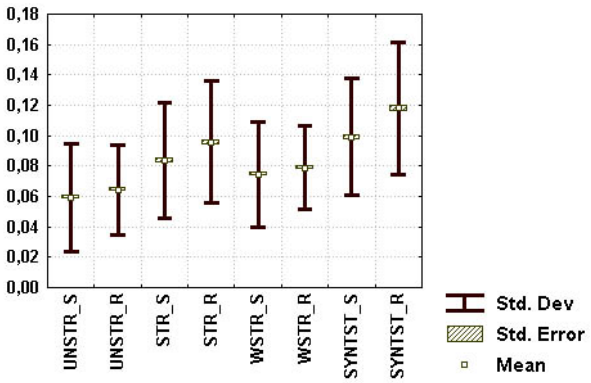

Figure 6: Mean, standard deviation and standard error values for /a/, /i/, /u/ in different positions with regard to the stress.

The p-level values were calculated for stressed vs. unstressed vowel duration value samples within each type of speaking. The data show that the stressed vowel and prominent vowel duration values are different, the latter being greater. This feature is typical both for spontaneous speech and reading.

4. CONCLUSIONS

1. Both spontaneous speech and reading are characterized by some degree of allophonic variation. Some of the allophones appear in the contexts where they are theoretically impossible (o- and e- like vowels in unstressed positions). This may be caused either by the provoking phonetic context, or by the weakening of articulation.

2. The vowel and consonant duration mean values are greater in reading as compared to spontaneous speech. There is a statistically significant difference in means for the vowel /a/, as well as for sonorants, voiceless and voiced plosives, and voiceless fricatives for spontaneous speech and reading samples. Spontaneous speech is characterized by greater variability of vowels and sonorants duration.

3. There are significant differences in vowel duration values depending on the position of the sound with regard to the stress. The duration means of vowels in the word stress position differ from those of the vowels in the pitch prominent syllables. The data show differences between the unstressed vowel duration values in the two types of speech, as well as between the durations of the vowels in the pitch prominent position. However, there is no significant difference between the duration values of the stressed vowels outside the intonation centre in the two types of speech.

ACKNOWLEDGMENTS

This study was supported by INTAS (project number 00-915).

REFERENCES

[1] L. Bondarko. Spontaneous speech and language system organization. *Russian Speech Phonetic Fund Bulletin*. *Phonetic feutures of Russian spontaneous speech,* pp. 17-23. St. Petersburg-Bochum, 2001. *(in Russian)*

[2] V. Kuznetsov. Connected speech vowel system. St. Petersburg University Press, 1997. *(in Russian)*

[3] Spontaneous speech phonetics. St. Petersburg, 1988. *(in Russian)*

[4] N. Svetozarova. On some features of Russian spontaneous speech phonetics. *Russian Speech Phonetic Fund Bulletin*, **8**. *Phonetic features of Russian spontaneous speech,* pp. 7-15. St. Petersburg-Bochum, 2001. *(in Russian)*

.