Temporal "Declination"
in Different Types of IPs in Russian:
Preliminary Results

Tatiana Kachkovskaia[(✉)]

Saint Petersburg State University, Saint Petersburg, Russia
kachkovskaia@phonetics.pu.ru

Abstract. The paper explores temporal changes within an intonational phrase in Russian. The main question we aim to answer is whether we can speak about temporal "declination" in a similar way we speak about melodic declination. In order to answer this question, we analysed stressed vowel duration in intonational phrases (IPs) of different types using a speech corpus. We have found that (1) most intonational phrases in Russian do not have temporal "declination" or "inclination" in the pre-nuclear part: the tempo is relatively stable until the nucleus, where a noticeable lengthening is observed; (2) the rarely occurring temporal "declination" or "inclination" in certain types of IPs can be considered a specific speaker's trait; (3) the amount of lengthening on the last stressed vowel within the IP may play a role in distinguishing final and non-final IPs, rising vs. falling nuclei, but this is also speaker-specific.

Keywords: Prosody · Speech tempo · Segmental duration · Intonational phrase · Russian phonetics

1 Introduction

One of the most well-known prosodic universals is melodic declination, which is defined by Vaissiere in the following way: "ln relatively long stretches of continuous speech, there is a global tendency for the F0 curve[1] to decline with time, despite successive local rises and falls" [7][p. 75]. One of the consequences of this is the reset of the baseline, which occurs at the boundary between one stretch of speech and the following one. As a result, reseting the baseline functions as a boundary marker, and declination itself—as an organizing trend which joins a number or words into a longer speech unit.

In a similar way we could attempt to speak about "temporal" declination. What is known so far is that the last word within a large prosodic unit is longer due to the phenomenon called pre-boundary (final) lengthening, which is a universal itself. Less is known about the other words within the prosodic unit:

---

[1] fundamental frequency, or melodic, curve.

is duration evenly distributed among the other words, or we can speak about some "declination" (or "inclination") of tempo as well? The prosodic features—melody, duration, intensity—often interact with each other, either working in parallel (melody + intensity) or compensating each other (e.g., in whispered speech, where no clear melody is observed, especially for tone languages). Thus, given melodic declination, we might expect some accompanying temporal declination or inclination as well.

For Russian the melodic declination was described in detail in [3]. One of the key results of the paper is that melodic declination differs across types of utterance—almost zero in general questions (level tone), and quite significant in declaratives. Similar results were described for Dutch by van Heuven and Haan [1][p. 125]. In particular, this means that declination helps distinguish some types of utterances even before the utterance is finished—and perceptually this was proved for Russian in [6][p. 116–120], where speakers could successfully disambiguate general questions from non-final parts of declaratives before they heard the last word of the prosodic unit, probably relying on the melodic declination pattern. For speech tempo no research of this kind has been done yet.

In this paper we compared 4 types of intonational phrases (IPs) differing in the melodic movement within the nucleus and finality of the IP within the utterance: (1) low-falling, utterance-final (declaratives); (2) rising-falling[2], utterance final (general questions); (3) falling, utterance-medial (as, e.g., in cases of punctuation marks such as colon or semicolon); (4) rising-falling, utterance-medial (occurring usually with a comma or dash, or no punctuation mark at all to divide long stretches of speech). In the future more types of IPs may be analysed, but so far we have taken the ones that occur more frequently in Russian speech.

2   Materials and Methods

Temporal "declination", as well as melodic, can be calculated in several ways. When choosing a method for this paper, we aimed at obtaining results comparable with those found earlier for melodic declination in Russian as presented in [3]. In that study melodic declination was calculates by peak values for stressed syllables (upper declination). In a similar way, here we calculated temporal declination based on stressed syllables.

In Russian stressed syllables are longer than the unstressed. Within the stressed syllable, vowel duration is a more reliable measurement as consonants in the stressed syllables are not always longer than in the unstressed (see [2]). This is why we calculated temporal declination based on stressed vowel duration.

---

[2] Rise-fall—one of the most frequent types of nuclei in Russian speech—is used in general questions and non-final IPs; the rise is realized on the stressed syllable, and the melodic peak is close to its right boundary (or sometimes even later).

As vowel phonemes differ in their inherent duration, we calculated duration in z-scores using the formula suggested in [9]:

$$\tilde{d}(i) = \frac{d(i) - \mu_p}{\sigma_p}$$

where $\tilde{d}(i)$ is the normalized duration of segment i, $d(i)$ is its absolute duration, and μ_p and σ_p are the mean and standard deviation of the duration of the corresponding phone p. The mean and standard deviation values were calculated over the whole corpus for each speaker separately.

We also limited our data by only those vowels that occurred in CV syllables, in order to eliminate influence from syllable length.

The material is a subcorpus of CORPRES [5]: 20 h of segmented speech containing fictional texts recorded from 4 speakers, all native Russians with standard pronunciation. The recordings come with manual segmentation into sounds and prosodic annotation in terms of [8]. The basic large prosodic unit in the annotation is the IP defined as (1) the largest phonological chunk into which utterances are divided, (2) containing a single most prominent word (nucleus), (3) serving to join the words tightly connected with each other in terms of semantic or syntactic stricture (the definition consistent with the one given in [4][p. 311]).

Using a Python script, we retrieved all the IPs with a given type of nucleus and a given length in clitic groups (CGs). In this paper we only observe IPs with final position of the nucleus (which is not always the case in Russian, but still the vast majority). A clitic group is defined as one stressed lexical word plus (possibly) one or more adjacent unstressed clitics. Thus, the number of clitic groups within an IP equals the number of stressed syllables.

The data were analysed separately for each speaker in case individual strategies are found. Then, for each type of nucleus and each IP length average values were obtained for CG 1, CG 2, CG 3 etc. This was summarized in graphs and tables. Then each pair of adjacent words were compared using the two-tailed Student's t-test for independent datasets with unequal variances (e.g., to find out whether CG 1 is longer than CG 2 in 4-word declaratives, we compared the means for the respective stressed vowels' normalized durations using a t-test).

3   Results and Discussion

Table 1 summarizes the results of the analysis for four speakers. The table contains data for the four types of IPs, and for each type—for IPs made up of 3, 4 and 5 clitic groups (CGs). The first clitic group obtains the number 1. Thus, e.g., in an average 3-word IP of type 1 (utterance-final IP with a falling nucleus) for speaker C the first stressed vowel has the normalized duration of −0.33, the second −0.37, and the third 0.79.

The value of 1 corresponds to one standard deviation for vowel duration for the particular vowel phoneme and for the particular speaker. Standard deviation values in our data fall within the range of 23 to 35 ms. Thus, in our example the

Table 1. Mean normalized vowel duration in stressed CV syllables in IPs of different length (3 to 5 clitic groups) for each clitic group (CG) within the phrase. IP type 1 stands for utterance-final IPs with low-falling nucleus; type 2—utterance final, rising-falling nucleus; type 3—utterance-medial, falling nucleus; type 4—utterance-medial, rising-falling nucleus. Speakers C and K are females, speakers A and M are males. Asterisk (*) marks the value which is significantly different from the left neighbour; brackets around the asterisk mean that the p-value is above 0.05, but very close to this value (up to 0.065). Bold font marks the IP types where the value for the 1-st word differs significantly from the value for the penultimate word. Brackets mark a case of a small dataset. Empty cells mean a lack of a reliable dataset.

CG number		1	2	3	4	5	1	2	3	4	5
IP type	IP len	Speaker C					Speaker K				
1	3	−0.33	−0.37	0.79*			−0.42	−0.51	0.80*		
	4	−0.47	−0.30	−0.35	0.92*		−0.59	−0.68	−0.55	0.73*	
	5	−0.46	−0.40	−0.41	−0.46	0.86*	−0.68	−0.48	−0.59	−0.46	0.73*
2	3	−0.51	−0.64	0.43*			−0.69	−0.63	0.29*		
	4	**−0.25**	**−0.63(*)**	**−0.80**	**0.10***		−0.53	−0.71	−0.58	0.39*	
	5	**(−0.18)**	**(−0.39)**	**(−0.59)**	**(−0.79)**	**(0.42)***					
3	3	−0.46	−0.49	0.82*			−0.61	−0.57	0.96*		
	4	−0.54	−0.50	−0.42	0.92*		−0.69	−0.59	−0.65	0.91*	
	5	−0.48	−0.55	−0.36	−0.34	0.99*	−0.69	−0.52	−0.43	−0.35	1.18*
4	3	−0.53	−0.53	0.12*			−0.61	−0.50*	0.36*		
	4	−0.47	−0.44	−0.49	0.14*		**−0.72**	**−0.60(*)**	**−0.54**	**0.46***	
	5	−0.60	−0.53	−0.38	−0.51	0.24*	−0.73	−0.69	−0.55	−0.51	0.34*
		Speaker A					Speaker M				
1	3	−0.26	−0.33	0.30*			−0.45	−0.39	0.18*		
	4	**−0.46**	**−0.29***	**−0.23**	**0.37***		−0.31	−0.37	−0.33	0.38*	
	5	−0.67	−0.28*	−0.40	−0.56	0.24*	**−0.71**	**−0.49**	**−0.19(*)**	**−0.40**	**0.20***
2	3	−0.50	−0.54	−0.01*			−0.36	−0.60*	1.04*		
	4	−0.47	−0.53	−0.42	−0.06*		−0.44	−0.49	−0.43	0.68*	
	5										
3	3	−0.42	−0.47	0.79*			−0.30	−0.34	0.62*		
	4	−0.57	−0.47	−0.44	0.49*		−0.42	−0.28	−0.30	0.67*	
	5	−0.62	−0.69	−0.45	−0.49	0.95*	**−0.47**	**−0.41**	**−0.31**	**−0.24**	**0.52***
4	3	−0.52	−0.50	0.30*			−0.49	−0.42	0.60*		
	4	−0.54	−0.44	−0.45	0.23*		**−0.53**	**−0.38***	**−0.34**	**0.66***	
	5	−0.61	−0.54	−0.37	−0.46	0.35*	−0.48	−0.49	−0.34	−0.36	0.54*

first and the second words have almost equal stressed vowel duration, while on the last word (which is the nucleus) we observe a noticeable change in duration—1.16 standard deviations, i.e. more than 26 ms.

The results of statistical analysis are also shown in Table 1. An asterisk (*) marks the values that differ significantly from the left neighbour (p-value < 0.05). Those cases where the p-value was very close to 0.05 (up to 0.065) are marked by an asterisk in brackets.

In our data every last clitic group within the IP has significantly greater value; the p-values as for these cases were all below 0.002. This proves that the

nucleus in phrase-final words is much longer than other stressed syllables, but this result is not new, at least for Russian (e.g., see [2]).
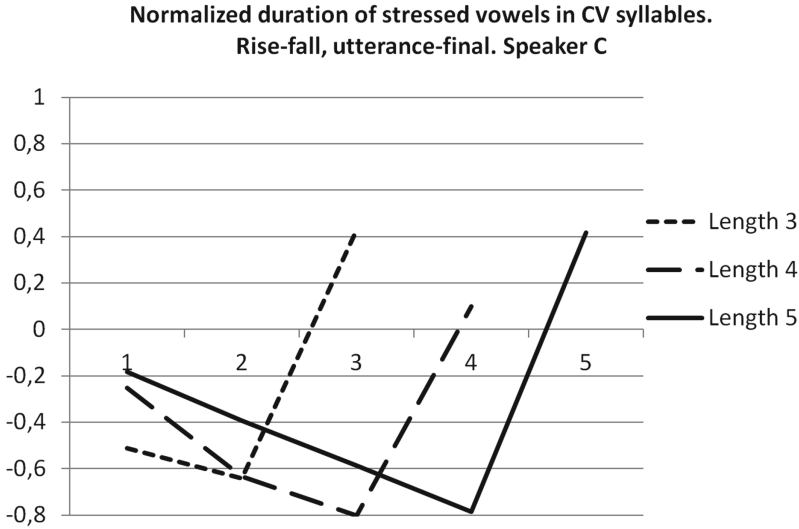

Fig. 1. Temporal pattern of a general question for speaker C showing temporal "declination". The graph presents data for IPs of different length (3–5 clitic groups); the x axis shows the index number of the clitic group within the IP.

In some cases IP-*medial* clitic groups also reveal a significant change in duration (see, e.g., 4-word IPs of type 1 for speaker A); the p-values are usually higher than for IP-final CGs (0.01 to 0.05). In most cases it is a decrease or an increase in duration on the second clitic group. In terms of IP type and length this is unsystematic.

We also analysed statistically the difference between the first CG and the penultimate CG. If the difference is significant, we may conclude that we observe temporal declination or inclination (not level tempo) in the pre-nuclear part of the IP. In Table 1 such cases are marked by bold font. We can see that there are only a few clear cases of declination, and they are very speaker-specific. Speaker C marks general questions (IP type 2) with temporal declination (see Fig. 1), while the other 3 speakers do not. Speakers K and M tend to mark non-final IPs with rising-falling nucleus with temporal inclination (see Fig. 2); speaker M also shows some temporal inclination in IPs with a falling nucleus; speaker A has a slight temporal declination in declaratives.

In the large majority of cases *no* temporal declination is observed (see, e.g., Fig. 3). This means that the temporal pattern of a typical Russian IP can be described in the following way: relatively stable tempo from the first CG to the penultimate CG, and a noticeable lengthening on the last CG.
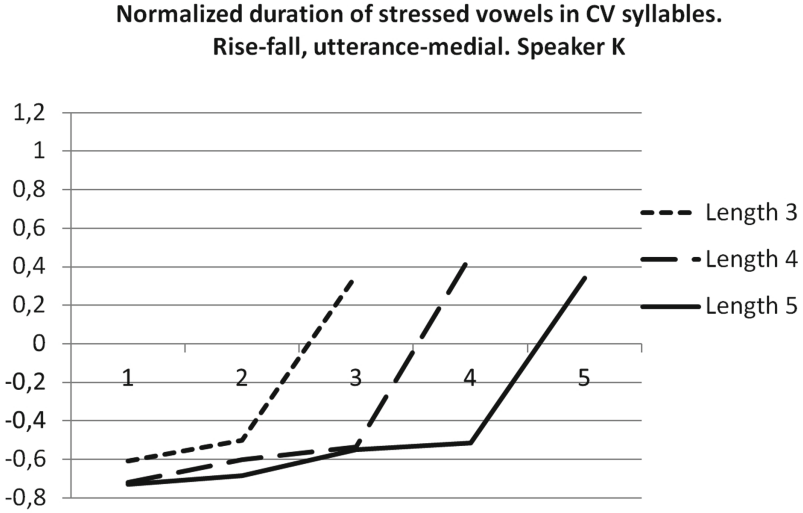
**Normalized duration of stressed vowels in CV syllables.
Rise-fall, utterance-medial. Speaker K**

Fig. 2. Temporal pattern of a non-utterance-final IP with a rising-falling nucleus for speaker K showing a slight temporal "inclination". The graph presents data for IPs of different length (3–5 clitic groups); the x axis shows the index number of the clitic group within the IP.
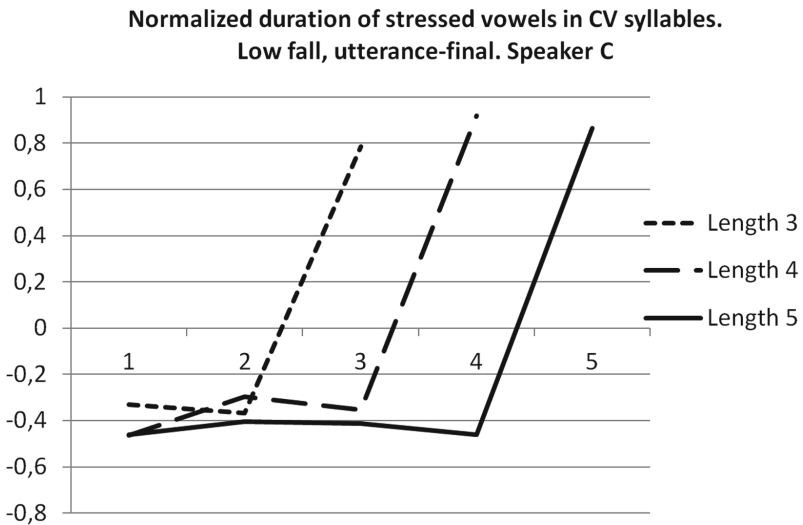
**Normalized duration of stressed vowels in CV syllables.
Low fall, utterance-final. Speaker C**

Fig. 3. Temporal pattern of a declarative for speaker C showing no temporal "declination" or "inclination". The graph presents data for IPs of different length (3–5 clitic groups); the x axis shows the index number of the clitic group within the IP.

However, the amount of lengthening on the last CG differs between IP types and between speakers. A number of individual strategies could be formulated.

1. Many speakers distinguish between IPs with falling and with rising-falling nuclei (speakers A, C, K)—falling nuclei have longer vowels.
2. Some speakers use duration to signal non-finality (speaker A)—vowels in nuclei of non-final IPs (types 3 and 4) are longer.
3. Some speakers tend to lengthen the nuclear vowel in general questions (speaker M).

The speaker-specific nature of the temporal pattern of the IP is in accordance with the findings presented in [3] for melodic declination. Still, the main result or this preliminary study is that temporal "declination" is absent in most types of IPs for most speakers. It might suggest that speech tempo is something that needs to be kept stable while other parameters are changing. But this required further proof. The next step of this ongoing research might be an analysis of those IPs which *do* have temporal "delination" or "inclination" using a series of perception experiments.

4   Conclusions

An analysis of 20 h of speech recorded from 4 speakers of Standard Russian enabled to obtain the following preliminary results.

1. Most intonational phrases in Russian do not have temporal "declination" or "inclination" in the pre-nuclear part: the tempo is relatively stable until the nucleus, where a noticeable lengthening is observed.
2. Temporal "declination" or "inclination" in certain types of IPs can be considered a specific speaker's trait.
3. The amount of lengthening on the last stressed vowel within the IP may play a role in distinguishing final and non-final IPs, rising vs. falling nuclei; this is also speaker-specific.

References

1. van Heuven, V., Haan, J.: Phonetic correlates of statement versus question intonation in dutch. In: Botinis, A. (ed.) Intonation. Text, Speech and Language Technology, vol. 15, pp. 119–144. Springer, Dordrecht (2000). https://doi.org/10.1007/978-94-011-4317-2_6
2. Kachkovskaia, T., Nurislamova, M.: Word-initial consonant lengthening in stressed and unstressed syllables in Russian. In: Karpov, A., Jokisch, O., Potapova, R. (eds.) SPECOM 2018. LNCS (LNAI), vol. 11096, pp. 264–273. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99579-3_28
3. Kocharov, D., Volskaya, N., Skrelin, P.: F0 declination in Russian revisited. In: Proceedings of 18th International Congress of Phonetic Sciences (2015)
4. Ladd, D.R.: Intonational phrasing: the case for recursive prosodic structure. Phonol. Yearb. **3**, 311–340 (1986)

5. Skrelin, P., Volskaya, N., Kocharov, D., Evgrafova, K., Glotova, O., Evdokimova, V.: CORPRES. In: Sojka, P., Horák, A., Kopeček, I., Pala, K. (eds.) TSD 2010. LNCS (LNAI), vol. 6231, pp. 392–399. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15760-8_50
6. Svetozarova, N.: The Intonation System of Russian [Intonacionnaya Sistema Russkogo Yazyka]. Leningrad University (1982)
7. Vaissiere, J.: Language-independent prosodic features. In: Cutler, A., Ladd, D.R. (eds.) Prosody: Models and Measurements, Springer Series in Language and Communication, vol. 14, pp. 53–66. Springer, Heidelberg (1983)
8. Volskaya, N., Skrelin, P.: Prosodic model for Russian. In: Proceedings of Nordic Prosody X, pp. 249–260. Frankfurt am Main: Peter Lang (2009)
9. Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M., Price, P.: Segmental durations in the vicinity of prosodic phrase boundaries. J. Acoust. Soc. Am. **91**, 1707–1717 (1992)