Typical and rare post-nuclear melodic movements in Russian

Tatiana Kachkovskaia, Anna Mamushina, Alyona Portnova

Saint Petersburg State University, Saint Petersburg, Russia

kachkovskaia@phonetics.pu.ru, mamushina.anna@mail.ru, alice.kate2310@gmail.com

Abstract

Russian intonation is traditionally described in terms of nuclei, pre-nuclei and post-nuclei. Nowadays accurate data is available on the melodic movement within the nucleus and pre-nucleus. The post-nuclei (tails) lack detailed descriptions, probably because (a) long post-nuclei are quite rare and (b) post-nuclei are often treated as automatic and unable to affect the meaning of the phrase or add extra connotations. In this paper we describe the variability of post-nuclear melodic movements for the most frequent types of nuclei. The material was a large labelled Russian speech corpus (CORPRES). We analyzed IPs with long post-nuclei in terms of direction of melodic movement within the post-nucleus and intervals. This enabled us to find typical and rare tail movements. Then, we performed a perception experiment to determine how native speakers perceive phrases with non-typical movements within the post-nucleus. For this, typical realizations were modified into falling to low, level high and rising to very high. The experiment showed that in most cases the modified signal differed from the original. Modifications into rising movements often contained additional connotations, mostly "non-finality". Modifications into falling movements were rarely described as having additional connotations.

Index Terms: intonation, Russian, post-nucleus, melody, perception of intonation

1. Introduction

The traditional system used to describe Russian sentence prosody was originally suggested by E. A. Bryzgunova [1] and is similar to what is called the British School (as, e.g., in books by O'Connor and Arnold [2]). The basic segmental unit is an intonational phrase (IP), such that: each IP usually contains one main word (the nucleus); other words are joined by a single declination (or inclination) trend; certain prosodic phenomena occur at IP boundaries. The traditional E. A. Bryzgunova description is made up of 7 basic types called "intonation constructions" (ICs); there also exists an extension suggested later by N. Volskaya [3] [4] containing 13 basic "melodic types" with further subtypes.

Since the development of the Bryzgunova system, some details on their realizations and variability have appeared in literature. Accurate data on the intervals and timing within the nucleus were obtained by C. Odé [5]. Melodic declination in Russian is described in [6]; pre-nuclear accents bearing additional perceptual prominence are described in [7].

These descriptions are concerned with the melody of pre-nucleus and nucleus. But when it comes to post-nucleus, no accurate data is available so far. This might be due to the following reasons:

- Long post-nuclei are quite rare (only around 8 % of IPs in read speech contain at least one prosodic word in the post-nucleus [8]).

- The post-nucleus is often treated as "automatic" and not able to change the meaning of the phrase or add extra connotation.

At the same time, a description of post-nuclear melody is necessary for a number of reasons. Apart from the theoretical value, these new data could be used to improve the naturalness of speech synthesis as well as in teaching Russian as a foreign language.

Thus, the principal aim of the present study is to describe the variability of post-nuclear melodic movement for the most frequent types of nuclei for read speech. This is done based on a large speech corpus (CORPRES [9]), where we can guarantee a sufficient number of phrases with long post-nucleus.

The secondary aim is to estimate the role of post-nuclear melodic movement for the listener. This is performed via a perceptual experiment with artificially modified F0 in the post-nucleus.

For the convenience of the readers that are not familiar with the Bryzgunova system, in this paper we will use the ToBI-like notation suggested by C. Odé [5] (ToRI). The two notations for the types of nucleus within the focus of this paper are matched in Table 1.

2. Method

2.1. Experiment 1: Variability of post-nuclear melody

As the material we used the Corpus of Russian Professionally Read Speech (CORPRES [9]). In total, we analyzed 20 hours of read speech recorded from 4 speakers aged 30–50.

CORPRES contains manual prosodic annotation: boundaries of intonational phrases (IPs), position of the nucleus within the IP, melodic type on the nucleus. The melodic type was assigned according to the system suggested by N. Volskaya [3], which is an extension of the Bryzgunova system and therefore can be easily reduced to the latter, if needed.

Based on the prosodic annotation provided with the corpus, we automatically found all the IPs with long post-nucleus— i.e., IPs where the post-nucleus contained at least one lexically stressed word. E.g., the IP "...and she swallowed this unfairness"[1] with the nucleus on the word "swallowed" and lexical stress on two following words, counted as an IP with long post nucleus.

In the present paper we analyzed four frequent types of nuclei. They are briefly listed in Table 1.

1. L*: used in utterance-final IPs of declarative sentences, in utterance-medial IPs with a weak syntactic link with the next IP (such as with a semicolon), before lists (such as with a colon).

2. H*L: used in yes/no questions and non-utterance-final IPs with a rather strong syntactic link with the next IP.

---

[1] translated form Russian, keeping the original word order

Table 1: *The four types of nuclei analyzed in this paper in different notation: "intonation constructions" by E. A. Bryzgunova and the ToBI-like notation by Cecilia Odé; column 3 presents post-nuclear melodic movement according to traditional descriptions.*

Nucleus, Bryzgunova	Nucleus, Odé	Post-nucleus
IC-1	L*	low
IC-3	H*L	falling
IC-4	L*H	high
IC-6	H*H	high

3. L*H: used in contrastive questions (such as "And you?") and non-utterance-final IPs.

4. H*H: used in exclamations (often expressing admiration or dreaminess) and non-utterance-final IPs.

During the preliminary analysis we found that in COR-PRES there were not enough examples of IPs with long post-nucleus for yes/no questions (H*L), contrastive questions (L*H) and exclamations (H*H). Therefore, for these three types of nuclei we were so far limited by non-utterance-final IPs only.

Then for each type of nucleus and each of the four speakers, 30 to 50 examples of IPs with long post-nucleus were selected and analysed manually to determine:

1. the direction of melodic movement after the nuclear syllable: falling, rising or level (changes below 3 semitones were counted as level);

2. the interval, in semitones;

3. any specific phenomena, such as an additional small rise-fall in the post-nucleus.

2.2. Experiment 2: Perceptual significance of post-nuclear melody

For the perceptual experiment 4 recordings from the same speaker (female) were selected—one for each of the types of nuclei. Each recording contained standard post-nuclear melody (as found within Experiment 1).

Then the post-nuclear melody was modified in Praat [10] (resynthesized using the command "ToManipulation..."). Low post-nuclei (as in L*) were modified into rising; level high post-nuclei (as in L*H and H*H)—into falling and rising; falling post-nuclei (as in H*L)—into level high and rising.

In F0 manipulation we intended to achieve a strong difference that could easily be noticed by the listener. This is why in modifications we used extreme F0 values: for modifications into falling movement—the F0 minimum observed within the given IP; for modifications into rising movement—the F0 maximum observed within up to 3 IPs to the left from the given IP.

In total, we obtained 7 modified recordings. The listeners were presented 14 recordings (along with the similar non-modified ones) in random order. Listeners were asked to pay special attention to intonation.

After specifying their age and gender, the listeners were asked the following questions:

• Are the two recordings identical?

If not, for each recording:

    – Does it sound natural?
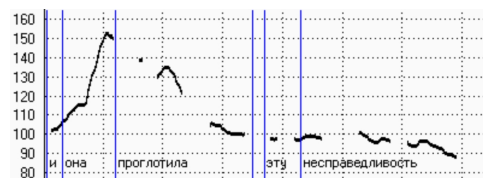
    – Does it sound unnatural ("not Russian")?

Figure 1: *Level melodic movement after L* nucleus (typical). The text reads "...and she swallowed this unfairness"; the nucleus is "swallowed".*
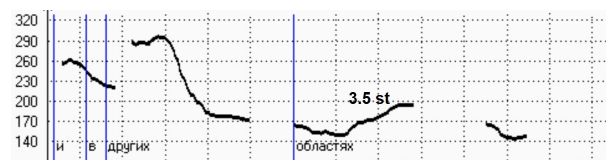

Figure 2: *Extra rise-fall after L* nucleus (rare). The text reads "...and in other regions"; the nucleus is "other".*

    – Does it have additional connotations? (If yes—please specify.)

21 native Russian speaker (8 males, 13 females) aged 20-60 participated in the experiment.

3. Results and discussion

3.1. Experiment 1: Variability of post-nuclear melody

3.1.1. Post-nucleus in L*

The frequencies of melodic movements after L* are given in Table 2. The most typical post-nuclear melodic movement for L* is level. An example is given in Figure 1. Another quite frequent case is falling post-nucleus. On average, the falls are around 3–4 semitones; in rare cases—up to 8 semitones.

Rising post-nuclei are never observed for L*. A very rare, and speaker-specific, case is an extra rise-fall in the post-nucleus. An example is given in Figure 2. Compared with the nucleus itself, such rises are rather small: 2–4 semitones; on the other hand, such intervals are big enough to be noticeable to the listener. A possible interpretation is that in these cases a post-nuclear word carries additional prosodic prominence. Thus, e.g., the phrase presented in Figure 2 ends a paragraph, and the additional prominence could serve here to signal finality. This might also be the case when a speaker needs to fall very low, but the inability to go beyond his/her individual F0 range makes him/her make an additional fall on the last word.

3.1.2. Post-nucleus in H*L

The frequencies of melodic movements after H*L are given in Table 3. The typical post-nuclear melodic movement for H*L is

Table 2: *The frequency of post-nuclear melodic movements after L* nuclei, for four speakers: M (male), C (female), K (female), A (male).*

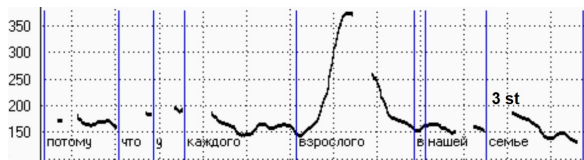|	M	C	K	A
level (low)	83%	74%	65%	31%
falling	17%	26%	33%	52%
extra rise-fall	0%	0%	2%	17%

Figure 3: *Extra rise-fall after H*L nucleus (rare). The text reads "...because every adult in our family"; the nucleus is "adult".*

Table 3: *The frequency of post-nuclear melodic movements after H*L nuclei, for four speakers: M (male), C (female), K (female), A (male).*

	M	C	K	A
falling	94%	100%	95%	91%
extra rise-fall	6%	0%	5%	9%

falling (followed by level low), and these cases are the absolute majority. The interval may be as high as 18 semitones.

In a similar way as for the L* nuclei (see above), after H*L there might be an additional small rise-fall—as in Figure 3. This example may have an alternative interpretation in terms of IP boundary placement: one may argue that the post-nucleus forms an IP on its own, and the small rise-fall is a nucleus itself. The counter-argument for this is that such "extra rise-fall" is too small compared with the 12-semitone rise-fall on the nucleus.

3.1.3. Post-nucleus in L*H

The frequencies of melodic movements after H*L are given in Table 4. The typical post-nuclear melodic movement for L*H is level high (i.e. staying on the same level where the nuclear movement has ended). Another quite frequent possibility is a rise (as in Figure 4), although it is speaker-specific, as speaker C rarely uses it.

A much more rare post-nuclear movement is falling, and it is speaker-specific. The falls do not exceed 5 semitones, which is rather small compared with the interval for the nuclear rise. Such post-nuclear falls can be explained by melodic declination—which, as we see here, may go on after the nucleus.

3.1.4. Post-nucleus in H*H

The frequencies of melodic movements after H*H are given in Table 5. There are two typical post-nuclear melodic movements for H*H: level high (i.e. staying on the same level where the nuclear movement has ended, or falling slightly—within the interval not exceeding 3 semitones) and falling. The interval for the fall may be up to 11 st, but still the fall never comes close to the lowest F0 for the given.

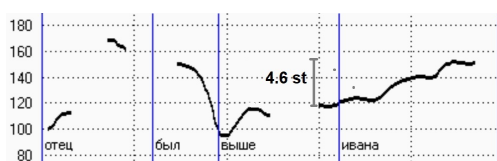An example of falling post-nucleus after H*H is given in

Figure 4: *Extra rise-fall after L*H nucleus (frequent). The text reads "...father was taller [than] Ivan"; the nucleus is "taller".*

Table 4: *The frequency of post-nuclear melodic movements after L*H nuclei, for four speakers: M (male), C (female), K (female), A (male).*

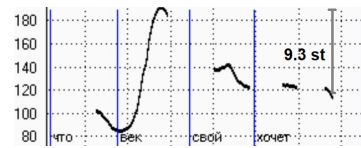|	M	C	K	A
level (high)	55%	62%	69%	50%
falling	-	31%	-	8%
rising	45%	8%	31%	42%

Figure 5: *Extra rise-fall after H*H nucleus (typical). The text reads "...that days of him wants... [long to live]"; the nucleus is "days".*

Figure 5. In a similar way to the case of L*H (see above), these cases can be explained by melodic declination which can go on even after the nucleus.

3.2. Experiment 2: Perceptual significance of post-nuclear melody

Table 6 provides the list of recordings and describes the F0 modifications within the post-nucleus. The listeners' responses are presented in Table 7. The phrases themselves are as follows (the nucleus is underlined).

  1 мы не стали бы спрашивать (we wouldn't ask)

2–3 а я как правило (and me as a rule)

4–5 прямота многое искупает (frankness a lot [can] compensate)

6–7 чем старей человек (the older the man)

The modification of post-nucleus for L* (level low into rising) is perceived very well, but sounds unnatural only for half of the listeners; if unnatural, it reveals additional connotations typical for phrase-final rises (non-final, question-like) and counted as universal [11].

Both modifications for H*L are also perceived well and sound unnatural in half of the cases, but fail to contain connotations. It should be borne in mind that when the post-nucleus for H*L is modified into level high, what we get is practically an H*H—another type of nucleus that is quite frequent in Russian speech. This could explain the absence of connotations, but not the unnaturalness; the latter might be because the artificially level high melody lacks any declination.

Modifications for L*H are also perceived well, but they only sound unnatural when the post-nucleus becomes rising; the connotations in this case are again typical for rising phrase ends. A falling post-nucleus does not cause a feeling of unnaturalness

Table 5: *The frequency of post-nuclear melodic movements after H*H nuclei, for four speakers: M (male), C (female), K (female), A (male).*

	M	C	K	A
level (high)	61%	43%	28%	35%
falling	39%	57%	72%	65%

Table 6: *The modifications of the original recordings for the perception experiment.*

	Nucleus	Melody in the original rec.	Modification of the post-nuc.
1	L*	level low	rising
2	H*L	falling	level high
3	H*L	falling	rising
4	L*H	level high	rising
5	L*H	level high	falling
6	H*H	level high	rising
7	H*H	level high	falling

Table 7: *Listeners' responses for the 7 modified stimuli.*

	Heard the difference	Sounds unnatural	Has connotations	Connotations
1	21/21	10/21	9/21	non-final, question emotional
2	18/21	11/18	2/18	emotional; sounds as an excuse
3	18/21	11/18	1/18	pushy
4	21/21	19/21	6/21	non-final, question, emotional; expecting "but"
5	20/21	3/20	2/20	with extra prominence; stating a fact
6	15/21	11/15	-	-
7	14/21	9/14	1/14	emotional (sad)

or additional connotations. At first glance, this asymmetry does not match very well with the data presented in Table 4, as both post-nuclear movements do appear in real speech. However, if we look carefully at the example in Figure 4, we will find that the post-tonic rise there is rather small—only 3 semitones (according to [5], the average rise within L*H is 9 semitones), but the melody continues to rise after the nucleus and reaches the intended interval only by the end of the IP, as though the H pitch target is delayed (this realization might even be described as L*M H%, a sequence that is not allowed in the ToRI model [5]). We do not know so far how frequent this is, and it requires further investigation.

Both modifications for H*H are perceived by around 70 % of the listeners, sound unnatural in around half of the cases, but do not contain additional connotations. The reason why listeners are not that sensitive to post-nuclear movement after H*H but more sensitive to post-nuclear movement after L*H is not that obvious so far. However, the H pitch target in H*H is usually higher than in L*H (see [5]), and listeners can be more sensitive to F0 changes within the lower level.

4. Conclusions

To sum up, the following post-nuclear melodic movements are observed in Russian.

- For L* (IC-1) the post-nucleus is level low or falling. A rare and speaker-specific case is an additional small rise-fall within the post-nucleus.

- For H*L (IC-3) the post-nulceus is typically falling. A rare case is an additional small rise-fall within the post-nucleus.

- For L*H (IC-4) the post-nucleus is level high or rising, for some speakers, falling may also be quite frequent.

- For H*H (IC-6) the post-nucleus is level high or falling.

The *additional small rise-fall* within the post-nucleus (for nuclei ending in a low pitch target, i.e. L* and H*L), has not been discussed in literature before. Given its low frequency of occurrence, it might be caused by some contextual or speaker-specific factors. The factors themselves are yet to be discovered. One of the possibilities is that a fall followed by another fall causes a sensation of a deeper fall (which we might need, e.g., to signal end of a paragraph [3]).

Our data also show that *melodic declination* can go on *after the nucleus*. This becomes clear when we analyze nuclei ending in a high pitch target (i.e. L*H and H*H) where there is enough space to go down.

An unexpected case, with regard to traditional descriptions, is *rising post-nucleus after L*H*, which is, moreover, very frequent (see Table 4). In terms of boundary tones, this would be marked as H%—which, according to C. Odé [5], does not exist in Russian.

In fact, the evidence for greater variability of boundary tones in Russian was discussed in [12] where different boundary tones were found to correlate with turn allocation in spontaneous dialogue speech. Our results, however (1) were obtained on read speech, and thus turn allocation could not explain this variability, (2) are based on a much larger dataset for L*H and H*H (as opposed to [12] where the material contained only a few examples), (3) include inter-speaker variability.

The perception experiment presented in this paper has crucial implications for speech synthesis technologies: one should be careful with the post-nucleus, as non-typical melodic movements will probably cause a feeling of unnaturalness and possibly—extra connotations.

5. Acknowledgements

6. References

[1] E. A. Bryzgunova, "Intonation," in *Russian Grammar*, N. Shvedova, Ed. Moscow: Nauka, 1980, vol. 2, pp. 96–122.

[2] J. D. O'Connor and G. F. Arnold, *Intonation of colloquial English*. Bristol, U.K.: Longman Group Ltd., 1973.

[3] N. B. Volskaya and P. A. Skrelin, "Prosodic model for Russian," in *Proceedings of Nordic Prosody X*. Helsinki: Frankfurt am Main: Peter Lang, 2009, pp. 249–260.

[4] N. Volskaya and T. Kachkovskaia, "Prosodic annotation in the new corpus of Russian spontaneous speech CoRuSS," in *Proc. of Speech Prosody 8*, 2016, pp. 917–921.

[5] C. Odé, "Transcription of Russian intonation, ToRI, an interactive research tool and learning module on the internet," in *Dutch Contributions to the Fourteenth International Congress of Slavists*, ser. Linguistics, H. P. Houtzagers, J. Kalsbeek, and J. Schaeken, Eds. Netherlands: Rodopi, 2008, vol. 1, pp. 431–450.

[6] D. Kocharov, N. Volskaya, and P. Skrelin, "F0 declination in Russian revisited," in *Proceedings of the 18th International Congress of Phonetic Sciences*, T. S. C. for ICPhS 2015, Ed. Glasgow, UK: the University of Glasgow, 2015, pp. 293.1–5.

[7] N. Volskaya and T. Kachkovskaia, "Tonal specification of perceptually prominent non-nuclear pitch accents in Russian," in *Speech and Computer*, A. Ronzhin, R. Potapova, and G. Németh, Eds. Cham: Springer International Publishing, 2016, pp. 699–705.

[8] Authors, "Prosodic phrasing in Russian spontaneous and read speech: evidence from large speech corpora [submitted]," in *Conference Proceedings*, 2020.

[9] P. Skrelin, N. Volskaya, D. Kocharov, K. Evgrafova, O. Glotova, and V. Evdokimova, "CORPRES — Corpus of Russian Professionally Read Speech," in *Text, Speech and Dialogue*, ser. Lecture Notes in Computer Science, P. Sojka, A. Horák, I. Kopeček, and K. Pala, Eds.   Springer Berlin Heidelberg, 2010, no. 6231, pp. 392–399.

[10] P. Boersma and D. Weenink, "Praat: doing phonetics by computer (version 5.1.13)," 2009. [Online]. Available: http://www.praat.org

[11] D. Hirst and A. DiCristo, *Intonation Systems: A Survey of Twenty Languages*.   Cambridge University Press, 1998.

[12] L. Paschen, "Boundary tones indicate turn allocation in Russian," in *Proceedings of ConSOLE XXIII*, 2015, pp. 493–508.