Social variability of peak alignment in Russian rise-fall tunes

Tatiana Kachkovskaia, Svetlana Zimina, Alena Portnova, Daniil Kocharov

Department of Phonetics, Saint Petersburg State University, Russia

Abstract

In Russian, rise-fall tunes (H*L) are very typical in yes-no questions and non-utterance-final clauses. In standard descriptions of Russian intonation, the melodic maximum in this tune is located late in the stressed vowel. However, studies of modern Russian intonation, especially within the younger age group, report on cases of “displaced” melodic peaks—shifted significantly to the right, so that the F0 maximum occurs on the post-stressed syllable. In this paper we analyse the frequency of such misplaced peaks in Russian dialogue speech, with respect to the factors of gender, age and social distance between the interlocutors. The research is based on the SibLing speech corpus: 90 dialogues with varying relationship between the interlocutors.

Index Terms: Russian intonation, melodic peak alignment, social variability

1. Introduction

The Rise-fall tune is one of the most frequent melodic patterns in Russian speech [1]. It is used mostly in non-final intonational phrases and yes-no questions. According to traditional descriptions, neutral variant of the Rise-fall is produced with the melodic peak late in the stressed vowel. Shifting the peak much to the right or somewhat to the left is possible, but it is typical for emotional speech—e.g., very late peak may signal high degree of surprise [2].

Recent studies carried out by Nina Volskaya showed that melodic peak shifted to the right—that is, to the post-tonic syllable—are rather frequent in speech of the young generation. But more importantly, such intonation patterns are perceived by older generation as impolite and disrespectful, thus causing the “conflict of generations” [3].

At the same time, our data based on 15 dialogues from the spontaneous speech corpus CoRuSS [4] have shown that such “misplaced” peaks are observed in the speech of the older generation as well, just more rarely than in the young generation’s speech. Thus, on average, speakers aged 16–30 shifted the peak in 38% of rise-falls, 31–45 in 24%, 46–77 in 17%. (However, the percentage of shifts also depends on gender: on average, it is 1.5 times higher within the female group.) [5] Based on these results, we could try to re-interpret the reasons for the “conflict of generations” mentioned above: it may be caused by not only the peak locations, but also some other prosodic features (e.g., voice quality).

Another research [6] threw more light on the sources of this variability. An analysis of speech recorded in various regions of Russia (not only in Moscow and Saint Petersburg, as before) showed that such peak shifts are typical for many regional dialects. Thus, we may assume that the younger generation of the capitals acquires these patterns because of close contact with representatives of regional varieties—in colleges and universities.

Still, it is not quite clear why female speakers tend to shift the peaks more often. In an attempt to answer this question, it seems reasonable to compare the nature of “misplaced” melodic peaks with the phenomenon of the uptalk (final tone rising at the end of declarative utterances) since both phonetic phenomena are relatively recent and are found more often, though not exclusively, in the speech of the young. Up until recently, the uptalk was strongly stigmatized and believed to be a trait of the young females’ speech. However, recent studies [7] [8] [9] refute this view as, according to experimental data, females use uptalk more often than males only in conveying certain pragmatic meanings, namely intention to hold the floor [8] and to minimize positions of power [10]. At the same time, the uptalk is used more or less equally often by men and women to express some other meanings (e.g., uncertainty). Thus, it appears reasonable to address the uptalk not as the gender-specific but rather as a situational phenomenon—much as the shifted melodic peaks in Russian, which, similar to the uptalk, may be caused, e.g., by the intention to keep leadership in a collaborative task.

Given the high variability of peak alignment in Russian rise-falls, we assumed that a particular speaker may alter his/her speech based on situational factors. If the “shifted to the right” variant is indeed considered impolite or at least somewhat colloquial, we might suppose that one may use it more rarely in dialogues with strangers, especially when there is an age gap. We tested this hypothesis using the SibLing speech corpus, where various types of relationship between the interlocutors are represented. We also took into account the factors of gender and age.

2. Method

2.1. Material

This research is based on the recently created corpus of Russian dialogue speech SibLing [11]. The corpus was designed for research on speech entrainment, with special attention to variability caused by social distance between the interlocutors. For higher reliability of the results, the corpus design was as follows. There were two groups of participants: the basic set of speakers—10 pairs of same-gender siblings aged 23–40, and their interlocutors, aged 20–70. Each of the speakers from the basic set was recorded in five dialogue settings: (1) with a sibling, (2) with a close friend of the same gender and similar age, (3) with a stranger of the same gender and similar age, (4) with a stranger of the opposite gender and similar age, and (5) with a stranger of the same gender, greater age and higher job position.

The conversations were organized in the form of collaborative tasks: first, a card-matching game, and then a classical map task. For this research we only used the card-matching game, as only this part of corpus was annotated prosodically.

The card matching game was based on searching for sim-
Figure 1: Manual segmentation of the word /'robat/ (robot) in Wave Assistant: pitch period boundaries (green), word boundaries (yellow), and stressed vowel boundaries (blue). Broad context (translated from Russian, word-by-word): “like [is] sitting [a] robot /PAUSE/ and [the] camera [is] like behind it”.

Figure 2: Manual segmentation of the word /'parj inj / (guy) in Wave Assistant: pitch period boundaries (green), word boundaries (yellow), and stressed vowel boundaries (blue). Broad context (translated from Russian, word-by-word): “/PAUSE/ there’s [a] guy which /PAUSE/ [is] somewhat bearded”.

similarities in two decks of ten Dixit-like cards. The interlocutors were instructed to switch turns after finding a suitable pair of cards, but in most cases the conversation was not much role-dependent and more like a free conversation. The game was time limited and lasted for around 12 minutes.

The dialogues were recorded in a soundproof studio. Interlocutors wore individual headsets with microphones and were separated by a non-transparent screen which prevented them from seeing each other.

The corpus contains orthographic, phonetic and, partly, prosodic annotation, and is segmented automatically into inter-pausal units and turns. Within the orthographic tier the annotators registered lexical stress and all non-speech events (hesitations, false-starts, word abruptions) and marked lexical stress. Prosodic annotation was performed manually using the annotation scheme suggested by Nina Volskaya [1]. Prosodic annotation included segmentation into intonational phrases, determining the nucleus and its melodic type. Within the annotation scheme, there are 13 melodic types plus up to four subtypes in each type; the melodic types were defined in both form and function.

2.2. Method

Based on prosodic annotation, we looked through all IPs with rise-falls, but limited ourselves to only those used in non-final clauses. (That is, we did not analyse rise-falls in yes/no questions, due to rather low frequency of this type of utterance in spontaneous speech.)

For the selected target words, boundaries of pitch periods were placed automatically in Wave Assistant software, with subsequent manual correction. For precise detection of melodic peaks, we selected only those words where (1) there was at least one post-tonic syllable (providing space for the peak to shift), (2) after deleting the pitch period boundaries on fragments with significant F0 fluctuations (including microprosodic events, such as those observed in the Russian trilling [r]), the peak was still visible on the intonogram. We also labelled boundaries of the target words and their stressed vowels.

Based on this annotation, peak locations were measured automatically with a Python script. Timing was calculated in ms relative to the right boundary of the stressed vowel.

Figures 1 and 2 illustrate manual segmentation of the target words produced by the same male speaker S11 (24 y.o.). In Figure 1, the word /'robat/ (robot) is produced with no melodic peak shift: −16 ms relative to the right boundary of the stressed vowel. In Figure 2, the word /'parj inj / (guy) is produced with the peak shifted to the post-stressed syllable: 30 ms relative to
the right boundary of the stressed vowel.

Of all the 90 dialogues and 20 siblings (basic set of speakers), not all the dialogues provided enough reliable pitch curves for rise-falls. In some cases, speakers rarely used this particular pattern (replacing it with a synonymous high rise, or H'H H'). In other cases, there were no post-stressed syllables (in other words, a truncated rise-fall, where only the rise was produced). Often, the post-tonic part of the target word contained a voiceless consonant where the peak, presumably, occurred—but the precise location of the peak could not be calculated. In the end, we obtained reliable data for 14 speakers (9 female, 5 male), each communicating with five (or, in rare cases, 3 or 4) interlocutors of various social distance.

Detecting F0 based on boundaries of pitch periods results in detection errors. The error depends on the period values themselves—that is, about 10 ms in male speech and about 5 ms in female speech.

Statistical analysis was performed in R using the \texttt{anova} function (package \texttt{rstatix}). The speakers were divided into three age groups: 23–26, 28–31, 36–38. In our mixed anova design, age and gender were between-subjects variables, and social distance (i.e., the interlocutor)—a within-subjects variable.

3. Results and discussion

In total, we were able to retrieve 937 words with rise-fall tunes appropriate for further analysis. In terms of gender, the groups were not equal in size (due to frequent creaky voice occurring in male speech): 635 tokens within the female group and 302 tokens within the male group. For the youngest age group (23–26) we retrieved 173 tokens from female speakers and 97 tokens from male speakers; for the middle age group (28–31), 187 and 104, respectively; for the eldest age group (36–38), 275 and 101, respectively. After further grouping of the data with respect to the type of interlocutor, the average cell size of the resulting data frame was 14.4 tokens.

3.1. Frequency of misplaced peaks

Figure 3 shows the percentage of rise-falls with peaks shifted to the right. Here a peak was considered shifted if it was located anywhere to the right from the end of the stressed vowel. Statistical analysis has shown that there is significant influence of age (p=0.003) and gender (p=0.001), but no influence of social distance (p=0.714). Interactions did not reveal statistical significance. Average peak shifts frequency for females is 73 %, for males—26 %; within the three age groups, from younger to elder—84 %, 58 %, 53 %.

We also tried to analyse peak shifts relative to another reference point: not the right boundary of the stressed vowel, but 15 ms to the right from it. The motivation for this was to take into account the listener’s sensibility to durational changes. For these data, statistical analysis also showed that there is significant influence of age (p=0.008) and gender (p<0.001), but no influence of social distance (p=0.442) or interactions.

Figure 3 illustrates the findings supported by statistical analysis: (1) in general, female speakers shift the peaks more often than male; (2) younger speakers shift the peaks more often. Additionally, we may notice that inter-speaker variability in the frequency of misplaced peaks seems to “run in the family”: see, e.g., the pair of male twins aged 24, or the pair of sisters aged 23 and 25.

Concerning social distance, we may suppose that there are individual strategies in how speakers use shifted peaks in dialogues with familiar people and strangers. Both siblings from the pair of male twins aged 24, e.g., use more shifted peaks with strangers, especially with those who are elder (termed “bosses”). A male speaker aged 36, on the other hand, uses fewer peak shifts with strangers. The pair of sisters aged 23 and 25 use a lot of shifted peaks with all types of interlocutors.

3.2. Peak shift time

Statistical analysis of peak locations (in ms) among the cases of shifted peaks has shown that there is significant influence of gender only (p=0.008), but no influence of age (p=0.308) or social distance (p=0.725). Interactions did not reveal statistical significance. Average shift for females is 42 ms, for males—26 ms.

Figure 4 shows median peak shift to the right among the cases of shifted peaks. The graphs confirm the results obtained by statistical analysis. In contrast with peak shift frequency, here we would not infer that siblings may show similar behaviour in terms of social distance variability.
However, our data failed to prove this dependency in a system-traditionally described as non-neutral in terms of connotations. We tors: namely, “social distance” between the interlocutors. We tion, and also in some regions of Russia apart from Moscow and known to be observed more frequently in the younger genera-
2. Conclusions
It has been discussed in literature that peak alignment in speech may vary. In Russian rise-falls, peak shifts to the right were known to be observed more frequently in the younger genera-
3.3. Entrainment in peak shifts
One may suppose that a speaker would make more peak shifts when his/her interlocutor uses more shifts, and vice versa. This could be one of the manifestations of speech entrainment, or accommodation [12]. Using this speech corpus, speech entrainment may be estimated in the following way. Since the same speaker took part in several dialogues, we may hypothesize that the peak shift frequency in his/her speech varies due to the peak shift frequency in the interlocutor’s speech. To test this hypo-
thesis, we may try using correlation analysis. Figure 5 shows the frequency of melodic peaks shifted to the right for one of the speakers, a female aged 38 (speaker S10, blue line), and her interlocutors (5 different speakers, red line). It is worth noting that all the values along the blue line belong to the same speaker, while the values along the red line belong to different speakers. There is significant Pearson’s correlation between these two sets of values ($r=0.88$, $p=0.048$), which implies that the variation observed in S10’s speech might be due to variation in her interlocutors’ speech. However, this approach would not enable us to find out whether speaker S10 entrained to her interlocutors or vice versa.

Of course, statistically significant correlations are not al-
ways the case. E.g., the S10’s twin sister, S09, does not show similar tendencies. Unfortunately, in our material there were not many examples that would provide the full picture for both interlocutors, as mentioned in the Method section—because some speakers rarely used rise-falls that provided precise data on peak locations. Still, we may assume that at least in some conversations interlocutors do entrain in this feature.

5. Acknowledgements
This work is supported by the Russian Science Foundation (grant 19-78-10046).

6. References
ferece on Language Resources and Evaluation (LREC’16). Port-
[5] S. V. Zimina, “Main tendencies in the dynamics of the melodic curve for the intonation construction 3 [Osnovnyje tendenci di-
"itskaya, “Rhythmical structure of protological word as a function of phrase tonal accent’s timing in regional varieties of Russian,” in Topicus in theoretical and applied phonetics [Aktualnije voprosy teoreticheskoj i prikladnoj fonetiki] (in Russian). Buki-
lish,” in Proc. of ICPWS, 2015.
[8] A. Ritchart and A. Arvaniti, “The form and use of uptalk in south-