Proceedings of the 126th conference of the Linguistic Society of Japan: 274-279.

Variability and stability of F0 movements in Russian under changes in speech rate

Yosuke Igarashi
The Graduate School of Area and Culture Studies, Tokyo University of Foreign Studies
Japan Society for the Promotion of Science

1. **Introduction**

An important assumption that underlies recent intonational research (e.g. Pierrehumbert 1980, Ladd 1996) is that an intonation contour can be analyzed as a sequence of phonological level tones such as highs and lows, occurring at specific places in the utterance, and that pitch movements or "configurations" such as rises and falls *per se* are not primitives of the linguistic analysis. For example, a rising F0 movement is, in this view, taken as merely a transition from its beginning point (F0 minimum) to its ending point (F0 maximum). These points are accepted to call "tonal targets", which can be regarded as phonetic realizations of the phonological tones.

The levels versus configurations debate is a longstanding discussion (cf., Ladd 1996: 59-73). The relevant respect for the present study is that these two views entail different predictions as to variability and stability of F0 movements under different phonetic/ phonological conditions. In a view that treats configurations as primitives, duration and/or slope of the F0 movements should be stable. In a view that primitives of intonation contour are level tones, on the other hand, stable properties should be the temporal alignment and the F0 level of the tonal targets. While the debate is still in controversy, a number of recent works have presented evidence in favor of the level view.

Prieto et al. (1995) studied rising accents in Mexican Spanish and found that the low tonal target at the beginning of the rise stably aligns just before the onset consonant of the accented syllable (This finding is consistent with the finding of Casper and van Heuven 1993 for Dutch). The location of the high tonal target at the end of the rise is more variable, being affected by a number of factors such as the proximity of the following word or prosodic boundary, and the number of unaccented syllables intervening the test syllable and the following accent (this variability is reported in Casper and van Heuven 1993 for Dutch and Silverman and Pierrehumbert 1990 for English). They also found that the rise time and the slope are not constant properties of the accents: when the high target aligns earlier, the rise becomes shorter and steeper.

Arvaniti et al. (1998) showed that in Modern Greek, when the location of nearby word boundaries and other accents is duly controlled, *both* low and high targets of rising accents are anchored with specific points in the segmental string. Specifically, the low target aligns just before the onset of the accented syllable and the high target aligns at the onset of the following unstressed vowel. This clearly means that the rise time is highly depended on the *segmental* composition of the accented word.

Moreover, Arvaniti et al. found little effect of duration on F0 change between both targets (excursion, henceforth). The combination of variable temporal duration and constant F0 excursion of the rise indicates that the slope is also highly dependent on the segmental composition of the accented word. For example, in a word with shorter composition like [ro ditiko], the rise time is shorter and the slope is steeper than in a word with longer composition like [pa remvasi] (see figure 1).

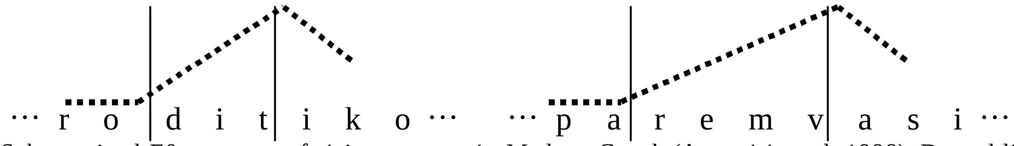

Figure 1 Schematized F0 contours of rising accents in Modern Greek (Arvaniti et al. 1998). Dotted lines indicate F0 contour and solid lines indicates the onset of the stressed syllable and the onset of following unstressed syllable of the words [ro ditiko] and [pa remvasi]. Note that the rise is shorter and steeper in the former word than in the latter.

The phenomenon that both low and high tonal targets of rising accents are anchored with specific places in the segmental string found by Arvaniti et al. is called “segmental anchoring”, and its existence in other languages was confirmed by Ladd et al. (1999) and Ladd et al. (2003) for English, Ladd et al. (2000) for Dutch. Arvaniti et al. argue that the finding of segmental anchoring provides evidence for a view that treats level tones as intonational primitives.

The aim of the present study is to contribute to this line of research. Specifically, an experiment was performed to examine whether the phenomenon of segmental anchoring can be observed in Russian, and to provide results from this language in favor of the level view.

2. Experiment

In this experiment, I examine whether the phenomenon of segmental anchoring can be observed in Russian rising accents. The independent variables used in the experiment are changes in segmental duration brought about by modifications of speech rate[1]. If there is the phenomenon of segmental anchoring, and if the F0 level of tonal targets of rising accents is unaffected by the experimental manipulations, the segmental points which the both targets align with should be closer together as speech rate increases, and the rise should therefore be shorter and steeper.

2.2. Method

The approach in this experiment is to measure rise time, alignment of low and high tonal targets and slope of rising accents in prepared sentences read aloud at three different speech rates. To test the segmental anchoring hypothesis (SAH), I took as null hypotheses “the constant duration hypothesis (CDH)” in which rise time of the accents is regarded as the stable property, and “the constant slope hypothesis (CSH)” in which slope is seen as constant. I examined only non-sentence-final or “prenuclear” rising accents, namely, F0 rise occurring at lexically stressed syllable of the non-final content words in a neutral reading of a short declarative sentence[2]. The reason why I choose prenuclear accents as subjects of investigation is that, as noted in Introduction, previous works have shown that the location of the high tonal target is affected by the following phrase boundary.

2.2.1 Materials, speakers and recordings

Twenty sentences were designed. A typical sentence is *Románova guljála v górode* (Romanova was walking in the city), in which the rising accents were expected to occur on the first and the second words (see Figure 2). The accent I measured is the one on the first word. Each sentence consists of three content words, which are incorporated into the Subject- Verb- Object or Adverbial syntactic structure. Each test word was followed by two or three unstressed syllables, and the test word always had a lexical stress on the antepenultimate syllable. This criterion is adopted because of the evidence that the location of tonal targets is affected by the proximity of word boundary and of following accent (see Introduction). The consonants flanking the stressed vowel of the accented syllable are nasals or trills in order to minimize microprosodic effects in the vicinity of the accented syllable.

The materials were read by four native speakers of Russian, two female and two male. All of them were in their twenties, and at the time of recording had been studied at universities in Tokyo. In what follows, the speakers are identified as FL, FT, MM and MN, where F or M stands for female or male respectively, and the second letter is an initial. The speakers had any known speech or hearing problems and were naïve as to the purpose of the experiment.

---

[1] This method is the one adopted in Ladd et al. (1999) for the study of English prenuclear rising accents.

[2] This intonation pattern is called “sawteeth pattern”, “neutral intonation”, “intonation pattern of an isolated declarative sentence” and “IK-1” in the systems of Odé (1989), Yokoyama (2001), Svetozarova (1998) and Bryzgunova (1980), respectively. While Bryzgunova regards the “prenuclear” rising accents as something incidental, the other authors see them as indispensable constituents of the pattern. In my own system, the pattern is transcribed as “L+H* L+H* H+L*” (Igarashi 2002).

Proceedings of the 126th conference of the Linguistic Society of Japan: 274-279.

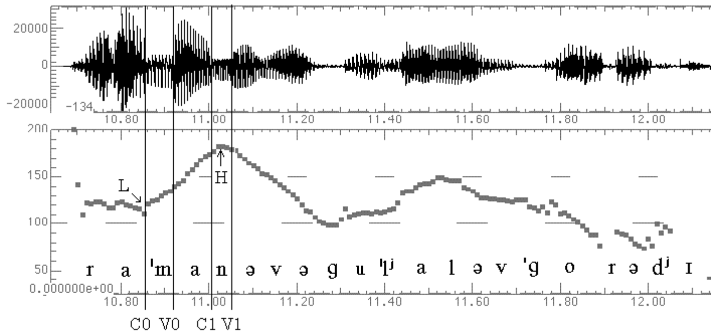

Figure 2. Waveform and F0 track for one of the test sentences *Románova guljála v górode* (Romanova was walking in the city), showing the segmental points obtained and the point at which low and high targets were measured.

The recordings were made on Digital Audio Tape (DAT) in the recording studio at Tokyo University of Foreign Studies or in a quiet room at the speaker's home. Speakers read the entire list of sentences once in each of three different speech rates. On the first reading, speakers were asked to read the list at normal speech rate. After the first reading, half of the speakers were asked to read fast on the second reading and slowly on the third and the other half were asked to read slowly on the second reading and fast on the third. Speakers were asked to repeat any misread sentences.

The recorded materials were digitized at a sampling rate of 16 kHz. The sentences were analyzed using ESPS Waves+ software.

2.2.2. Measurements

All measurements were performed manually in a simultaneous display of the waveform, wide-band spectrogram, and F0 track. Two tonal targets were measured as follows. The low tonal target was defined as the F0 minimum located in the vicinity of the onset of the accented syllable and was marked as "L". The high tonal target was measured at the highest F0 point around the offset of the accented syllable. This point was marked as "H". The four segmental points measured were the onset of the initial consonant, the onset of the stressed vowel, the offset of the stressed vowel (=the onset of the following syllable), and onset of the following unstressed vowel. These points were marked as "C0", "V0", "C1" and "V1" respectively (see Figure 2). On the basis of these segmental points, the following dependent variables were derived: 1) Syllable duration (C1 minus C0 in milliseconds); 2) Rise time (H minus L in milliseconds); 3) Alignment of L to C0 (L minus C0 in milliseconds); 4) Alignment of H to C1 (H minus C1 in milliseconds); 5) Alignment of H to V1 (H minus V1 in milliseconds); 6) Excursion (H minus L in semitones[3]); 7) Slope value (Excursion divided by Rise time).

2.3. Results and discussion

All data were analyzed by means of one-way repeated measures analyses of variance (RM-ANOVAs), with items as random factor, rate (slow, normal, fast) as repeated measures fixed factor, each speaker separately.

2.3.1 Confirmation of speech rate manipulation

First, I confirmed whether the speakers had indeed produced significantly different speech rates for the three conditions. Figure 3 shows the means of the duration of the accented syllable. As can be seen, speakers all showed longer syllable durations as rate slows. RM-ANOVAs with Syllable duration as a dependent variable showed a significant effect of rate for all speakers ($P<0.001$). This means that the speech rate manipulation was successful.

---

[3] Semitone is an appropriate scale for expressing the perceptual equivalence of F0 excursions in different voice range (see e.g., 't Hart et al. 1990).

2.3.3. Rise time

I investigated whether speech rate had an effect on rise time of the accent. If the CDH is correct, there should be no significant effect. The SAH predicts that rise time should increase as rate decreases. Figure 4 shows the means of rise time. It can bee seen that for each speaker, rise time is shortest at fast rate, longer at normal rate and the longest at slow rate. The RM-ANOVAs with Rise time as independent variables showed a significant effect of rate for all speakers (P<0.001). The effect is in the direction predicted by SAH.

In addition, I correlated Rise time with Syllable duration. The CDH predicts that there should be no significant correlation, whereas the SAH predicts a significant correlation. The result revealed that for all speakers, there is a significant positive correlation (for FL R=0.575, P<0.001; for FT R=0.693, P<0.001; for MM R=0.619, P<0.001; for MN R=0.770, P<0.001), showing that rise time increases as rate decreases. This means that the CDH is inadequate.

It was shown that rise time is not constant property of rising accents, and it is highly dependent on the duration of the accented syllable, which is predicted by the SAH.

2.3.4. Alignment of L and H tonal targets

I then examined whether speech rate had an effect on the alignment of tonal targets. I did these analyses separately for L and H targets.

To examine the effect of rate on the alignment of L, I ran RM-ANOVAs with Alignment of L to C0 as the dependent variable. The means, F-ratio and p-value are shown in Table 1. It can be seen that for three speakers, there is no significant effect of rate. For speaker MN showed significantly earlier alignment of L at slow rate.

In sum, it was shown that except in the case of speaker MN's slow rate, L targets is rather constantly aligned relative to the onset of the accented syllable.

I used the same approach to investigate the alignment of H. I ran RM-ANOVAs with Alignment of H to C1 as the dependent variable. The CDH predicts that, given the stable alignment of L, H should be aligned earlier as rate decreases. The SAH, on the other hand, predicts that there should be no significant effect. The means, F-ratio and p-value are shown in Table 2. It can be seen that, for all speakers, H aligns after the offset of the accented syllable, and that there is no significant effect of rate for three of the speakers. FL showed latest alignment at slow rate. While for this speaker the SAH is not fully confirmed, the effect is opposite direction to what would be predicted by the CDH.

I then, redid ANOVAs, this time with Alignment of H to C1 as the dependent variable. The means, F-ratio and p-value are shown in Table 3. It can be seen that the effect for speaker FL completely disappears: for all speakers there was no significant effect of rate. Although the exact alignment point seems speaker specific, the location of the H target is anchored somewhere in the consonant of the following unstressed syllable.

Thus, there is no support for the CDH. Instead, the results revealed that the location of H is relatively fixed, and support for the SAH.

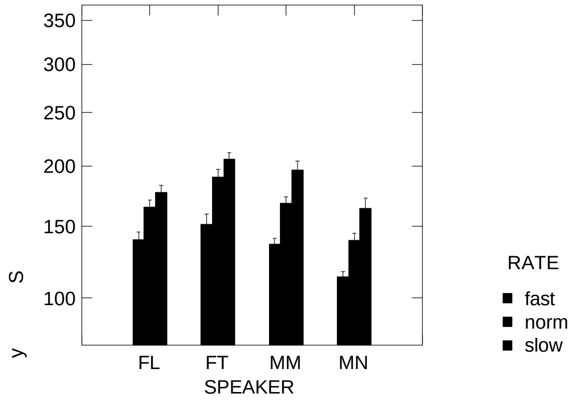

Figure 3. Means for syllable duration (milliseconds) with standard error bars.

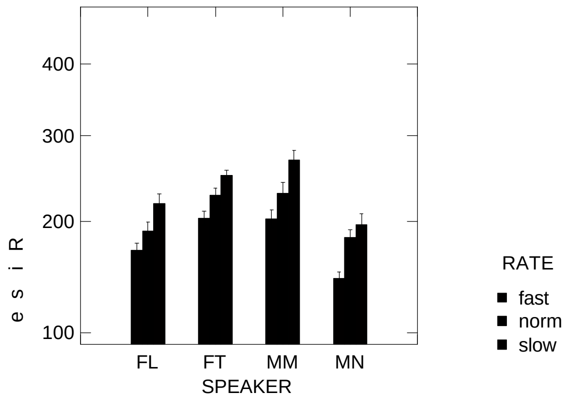

Figure 4. Means for Rise time (milliseconds) with standard error bars.

Proceedings of the 126th conference of the Linguistic Society of Japan: 274-279.

Speaker	FAST	NORM	SLOW	F(2,59)	P
FL	-9.2 (4.6)	-14.2 (6.5)	-3.2 (4.0)	1.067	0.3543
FT	-10.9 (4.0)	-9.0 (4.7)	-5.3 (3.1)	0.474	0.6261
MM	-30.5 (7.6)	-25.7 (9.1)	-24.0 (7.4)	0.221	0.8027
MN	-10.1 (3.4)	-6.7 (2.9)	1.4 (3.6)	3.755	0.0325*
Overall	-15.4 (2.8)	-14.4 (3.2)	-8.0 (2.6)		

Table 1 Means for Alignment of L to C0 (milliseconds) with standard errors in parentheses, for each of three speech rates. Negative values indicate that L precedes C0.

Speaker	FAST	NORM	SLOW	F(2,59)	P
FL	21.4 (6.5)	11.4 (6.8)	40.1 (6.7)	5.155	0.0105*
FT	39.9 (5.2)	29.2 (4.6)	39.3 (4.6)	1.329	0.2769
MM	36.5 (6.0)	38.3 (9.3)	49.9 (7.9)	1.713	0.1939
MN	21.1 (2.9)	37.7 (7.0)	34.7 (5.1)	2.776	0.0749
Overall	29.7 (2.8)	29.1 (3.7)	41.0 (3.1)		

Table 2 Means for Alignment of H to C1 (milliseconds) with standard errors in parentheses, for each of three speech rates.

Speaker	FAST	NORM	SLOW	F(2,59)	P
FL	-28.9 (7.1)	-40.1 (8.1)	-20.4 (6.7)	2.060	0.1415
FT	-6.6 (5.5)	-24.0 (5.4)	-16.6 (5.3)	2.598	0.0876
MM	-4.1 (5.0)	-9.3 (8.9)	-1.2 (8.5)	0.458	0.6359
MN	-16.6 (3.3)	-13.8 (8.0)	-16.9 (5.7)	0.077	0.9265
Overall	-14.1 (2.8)	-21.8 (4.0)	-13.8 (3.4)		

Table 3 Means for Alignment of H to V1 (milliseconds) with standard errors in parentheses, for each of three speech rates. Negative values indicate that H precedes V1.

2.3.5. Slope

Finally, I test the SAH against the CSH, i.e., the view that treats slope as constant property of accents. If, for CSH, F0 changes at a constant rate, F0 excursion should be greater as rate decreases, because there would be more time between segmentally anchored tonal targets.

To investigate the effect of speech rate on F0 excursion, I conducted RM-ANOVAs with Excursion as the dependent variable. The means are given in Figure 4. It can be seen that there was not a systematic tendency in the effects on excursion. There was no significant effect of rate for FL, who showed relatively constant excursions across each speech rate. On the other hand, there was a significant effect of rate for other speakers (P<0.001). For MM excursion becomes greater as rate decreases, which is predicted by the constant slope hypothesis. For FT and MN excursion is the smallest at fast rate, larger at slow rate and the largest at normal rate. These effects are not exactly in the direction that would be predicted by the constant slope hypothesis. Thus, except in the case of speaker MM, the CSH is not supported.

Figure 5 shows the means for Slope value. As can be seen, Slope values do not remain constant for all speakers. RM-ANOVA with Slope value as the dependent variable showed a significant effect of rate for all speakers (For FT P<0.001; for FL P<0.01; for MM and MN P<0.05). This result suggests that the CSH is inadequate.

In sum, it was shown that slope is not constant property of accents, and thus the CSH is not supported. On the other hand, the results presented is still compatible with the SAH. Under this hypothesis, non systematic tendency in the effect on excursion and slope can be attributed to speaker specific effects of the experimental manipulation of speech rate on F0 level of the segmentally anchored tonal targets: slope itself is derivable from the alignment and F0 level of the tonal targets.

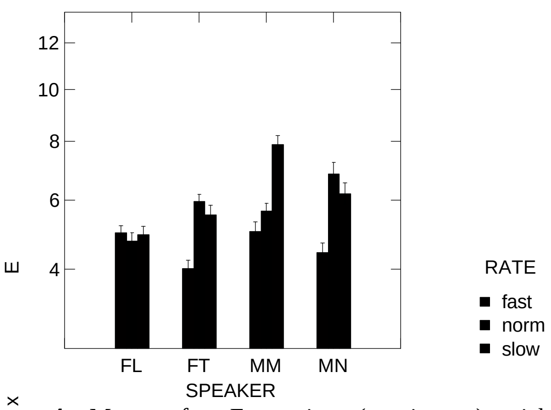

Figure 4. Means for Excursion (semitones) with standard error bars.

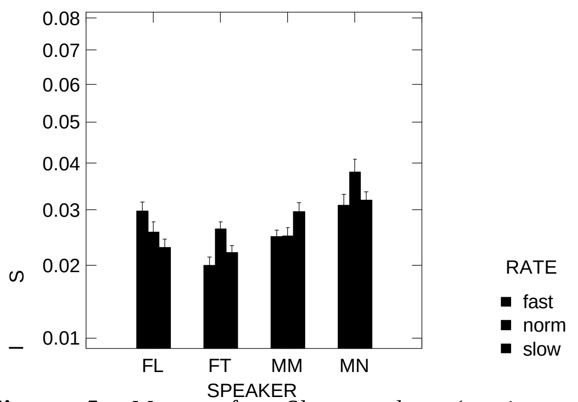

Figure 5. Means for Slope value (semitones/ milliseconds) with standard error bars. Higher value indicates steeper slope

3. Conclusion

The levels versus configurations debate is a longstanding discussion that opposes those that analyze intonational contour as consisting of primitive level tones to those that see it as consisting of primitive pitch movements or configurations. The present study treated this issue, and an experiment was conducted to investigate variability and stability of Russian prenuclear rising accents under changes in segmental duration brought about by modifications of speech rate. The results of the experiment showed that 1) duration of F0 rise was not fixed and was dependent on the duration of the accented syllable, 2) both beginning and end of the F0 rise (low and high tonal targets) are rather stably anchored with specific points in the segmental string regardless changes in rate, 3) rate had inconsistent and speaker specific effects on F0 excursion and slope of the F0 rise was not constant.

Showing that the phenomenon of segmental anchoring (Arvaniti et al. 1998) is observed in Russian and that duration and slope of Russian rising accents are not their constant properties but what can be derivable from the alignment and F0 level of the tonal targets, the results of the present study supports the view that treats level tones as intonational primitives.

Acknowledgement

This work was partially supported by Grant-in-Aid for JSPS fellows.

References

Arvaniti, Amalia, Robert D. Ladd and Menen, Ineke 1998 Stability of tonal alignment: the case of Greek prenuclear accents, *Journal of phonetics* 26, 3-25.

Bryzgunova, Elena A. 1980 Intonacija. In: Šedova, N. et al. (eds.) *Russkaja grammatika*, Tom I, 96-122. Moscow: Akademija Nauk SSSR.

Casper, J. and van Heuven, V.J. 1993 Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica* 50: 161-171.

Igarashi, Yosuke 2002 Tak nazyvaemaja "nejtralizacija intonacii" -Fonologiè e skoe opisanie russkoj intonacii-. *Bulletin of the Japanese Association of Russian Scholars* 34: 15-21.

Ladd, Robert D. 1996 *Intonational phonology*. Cambridge: Cambridge University Press.

————, Dan Faulkner, Hanneke Faulkner, and Schepman, Astrid 1999 Constant segmental anchoring of F0 movements under changes in speech rate. *Journal of Acoustic Society of America* 106, 1543-1554.

————, Ineke Mennen and Schepman, Astrid 2000 Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of Acoustic Society of America* 107. 2685-2696.

————, and Schepman, Astrid 2003 "Sagging transitions" between high pitch accents in English: Experimental evidence. *Journal of Phonetics* 31. 81-112.

Odé, Cecilia 1989 *Russian Intonation: A perceptual description*. Amsterdam: Rodopi.

Pierrehumbert, Janet. 1980 *The phonology and phonetics of English intonation*. Doctoral dissertation, MIT.

Prieto, Pilar, Jan van Santen and Hirschberg, Julia. 1995 Tonal alignment patterns in Spanish. *Journal of Phonetics* 23. 429-451.

Silverman, Kim. and Pierrehumbert, Janet 1990 The timing of prenuclear high accents in English. In J. Kingston and M. Beckman (eds.), *Papers in laboratory phonology* I. 72-106. Cambridge: Cambridge University Press.

Svetozarova, Natalija 1998 Intonation in Russian. In Hirst, D. & A. Di Cristo (eds.) *Intonation Systems. A Survey of Twenty Languages*: 261-274: Cambridge University Press.

't Hart, Johan., René Collier and Cohen, A. 1990 *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press.

Yokoyama, Olga T. 2001 Neutral and non-neutral intonation in Russian: A reinterpretation of the IK system. *Die Welt der Slaven* XLVI: 1-26.