“Segmental Anchoring” of F0 Under Changes in Speech Rate: Evidence from Russian

Yosuke Igarashi

The Graduate School of Area and Culture Studies, Tokyo University of Foreign studies
Japan Society for Promotion of Science
Yos_iga@ba3.so-net.ne.jp

Abstract

This paper reports the results of an experiment, which investigates what properties of Russian rising pitch accents are constant under changes in segmental duration brought about by modifications of speech rate. To be more specific, the experiment is aimed at examining whether the “segmental anchoring” of F0 movements is observed in this language. By segmental anchoring, I mean the phenomenon firstly found in Modern Greek that both beginning and the end of a rising pitch accent are anchored to specific points in the segmental string, which is regarded as a support for the framework that analyzes intonational contour as consisting of primitive level tones. The results revealed that 1) the duration of F0 rise was not constant but it increases as rate slows, 2) both the beginning and the end of the rise are anchored with specific points in the segmental string, regardless of the changes in speech rate, 3) for some speakers, rate had those effects on F0 excursion which suggest that the slope is constant. The results, partially replicating the findings in a similar experiment for English, confirm the existence of segmental anchoring in Russian on the one hand, and support a view that a given type of pitch accent has a constant slope, on the other.

1. Introduction

The alignment of fundamental frequency (F0) contours has been the concern of recent intonational research. Studies of several European languages have observed that the F0 valley at the beginning of a rising pitch accent consistently aligns with the onset of the accented syllable [1, 4, 12]. The location of the F0 peak at the end of the rise was reported to be more variable, being affected by a number of factors, such as speech rate, intrinsic segmental duration, upcoming word or phrase boundaries, and following stressed syllable [4, 12, 13].

An important finding along this line of research is a phenomenon found by Arvaniti et al. [1], which they call “segmental anchoring”. They found that in Modern Greek, when the location of nearby word boundaries and other accents is carefully controlled, *both* the F0 valley and the peak of the rising pitch accents are consistently anchored to specific points in the segmental string. Specifically, the valley is anchored at the onset of the accented syllable and the peak at the onset of the following unstressed vowel.

Segmental anchoring is of theoretical importance. First, it casts light on the issues concerning the invariant features of a given type of pitch accent. The fact that the alignment of both the F0 valley and peak of rising accents is invariant means that the rise duration is not invariant but it is determined by the segmental duration. Moreover, they found little effect of the rise duration on the F0 excursion. The constant alignment and excursion and the variable rise duration mean that the slope is not invariant either, but it is rather influenced by the segmental duration of the accented word. For example, Arvaniti et al. showed that in a word with shorter composition like [rodítiko], the rise duration is shorter and the slope is steeper than in a word with longer composition like [parémvasi].

Second, segmental anchoring has relevance to “the levels versus configurations debate” [7], a longstanding discussion that opposes those that analyze intonation contour as consisting of primitive level tones (“level view”) to those that see it as consisting of primitive movements or configurations (“configuration view”). Specifically, these two views make different predictions with regard to the slope and duration of F0 movements. In the configuration view, as Arvaniti et al. claim, it is reasonable to expect that a given F0 movement has relatively constant slope and/or duration. In the level view, by contrast, F0 movements are defined in terms of the alignment and F0 level of their beginning and ending points. Arvaniti et al. argue that segmental anchoring provides evidence in favor of the level view, which is widely assumed in current intonational work [7, 12]. In this framework (generally called “intonational phonology”) intonation contour is regarded as a sequence of phonological level tones such as highs and lows, occurring at specific points in the segmental string. For example, a rising F0 movement is, in this framework, taken as merely a transition from its beginning point (F0 valley) or “low tone”, to its ending point (F0 peak) or “high tone”.

In addition, the finding is interesting from a cross-linguistic point of view. The existence of segmental anchoring in other languages has been confirmed in various studies [2, 8, 9], suggesting that segmental anchoring is a universal phenomenon. For example, Ladd et al. [8] showed that the alignment of the valley and peak of an English rising pitch accent is unaffected by changes in segmental duration brought about by modifications of speech rate. The duration and slope of the accents become shorter and steeper as rate increases.

The aim of the present study is to contribute to this line of research. Specifically, an experiment was performed to examine what properties of Russian rising pitch accents (rise duration, the alignment of the F0 valley and peak, and slope) were constant and to examine whether the phenomenon of segmental anchoring can be observed in this language.

2. Experiment

2.1. Methods

This experiment[1] is meant to measure the rise duration, the alignment of the F0 valley and peak and the slope of rising pitch accents in prepared sentences read aloud at three

1 Preliminary results are given in my previous work [6].

different speech rates. This method was used by Ladd et al. in their study concerning English prenuclear rising accents [8]. I expect the present experiment to replicate the results of the above mentioned study: namely, regardless of changes in speech rate both F0 valley and peak should be anchored at specific points in the segmental string, and these points should be closer together as speech rate increases and the rise should therefore be shorter and steeper.

To test the segmental anchoring hypothesis (SAH), I took as null hypotheses "the constant duration hypothesis (CDH)", in which the rise duration of the accents is regarded as the constant property, and "the constant slope hypothesis (CSH)", in which the slope is seen as constant. I examined only non-sentence-final or "prenuclear" rising pitch accents, namely, the F0 rise occurring at lexically stressed syllable of the non-final content words, which appear in a neutral reading of a short declarative sentence [2]. The reason why I chose prenuclear pitch accents as subjects of investigation is that, as noted in Introduction, previous works have shown that the location of the F0 peak is affected by the following phrase boundary [1, 4, 13, 14].

2.1.1. *Materials, speakers and recording*

Twenty sentences were designed. A typical sentence is *Románova guljála v górode* (Romanova was walking in the city), in which the rising pitch accents were expected to occur on the first and the second words (see Figure 1). 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 one or two unstressed syllables, and the test word always had a lexical stress on the antepenultimate syllable. This criterion is adopted in order to avoid potential effects of the proximity of the word boundary and of the following accent on the alignment of the F0 peak. The consonants flanking the stressed vowel of the accented syllable are nasals or trills, because they would minimize segmentally-induced perturbation in the vicinity of the accented syllable.

The materials were read by eight native speakers of Russian, three females and three males. In what follows, the speakers are identified as FA, FL, FT MK, FM, MM, MN and MS, where F or M stands for female or male respectively, and the second letter is an initial. FA and MK are nineteen years old, and the other speakers were in their twenties. At the time of the recording they all were studying at institutes or universities in Tokyo. The results of MS had to be discarded, because he was too monotone to allow reliable marking of the F0 valley and peak of pitch accents. The speakers did not have any known speech or hearing problems and did not know the purpose of the experiment.

The recordings were made on Digital Audio Tape (DAT) in the recording studio of Tokyo University of Foreign Studies or in a quiet room at the speakers' 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; 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.

2.1.2. *Measurements*

All the measurements were performed manually in a simultaneous display of the waveform, wide-band spectrogram, and F0 track. The F0 valley was defined as the F0 minimum located in the vicinity of the onset of the accented syllable and was marked as "L". The F0 peak 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, and the onset of the following unstressed vowel. These points were marked as "C0", "V0", "C1" and "V1" respectively (see Figure 1).

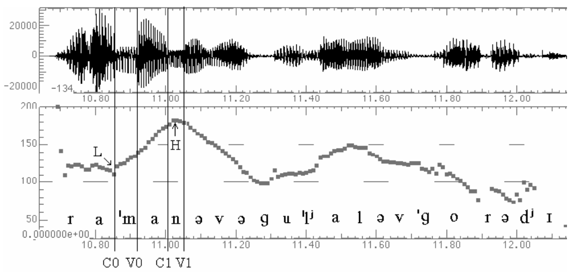

Figure 1*: Waveform and F0 track for one of the test sentences Románova guljála v górode (Romanova was walking in the city)*

2.2. Results and discussion

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

2.2.1. *Confirmation of speech rate manipulation*

First, I confirmed if the speakers had indeed produced significantly different speech rates for the three conditions. To examine this, I calculated the duration in ms of the accented syllable (Syllable duration) and conducted ANOVAs with Syllable duration as the independent variable. The means, *F*-ratios and *P*-values are presented in Table 1.

Table 1: *Syllable duration in ms (SE in parentheses).*

Speaker	Fast	Normal	Slow	$F(2, 59)$	P
FA	134.2(4)	146.5(4)	181.3(4)	54.156	<0.001*
FL	141.1(5)	166.1(5)	178.1(5)	36.392	<0.001*
FM	122.4(3)	151.1(4)	153.4(3)	52.825	<0.001*
FT	153.4(7)	191.4(5)	206.9(5)	31.689	<0.001*
MK	142.9 (4)	154.4(5)	184.1(5)	39.395	<0.001*
MM	137.3(3)	169.2(4)	197.9(7)	71.210	<0.001*
MN	114.3(3)	140.4(5)	166.0(8)	41.659	<0.001*

We can see that all the speakers produced longer syllable duration as rate slowed and that the effect of rate was significant for all speakers. This means that the speech rate manipulation was successful.

[2] This intonation pattern is called "sawteeth pattern" [11], "neutral intonation" [16], "intonation pattern of an isolated declarative sentence" [15] and "IK-1" [3]. In my own system, the pattern is transcribed as "L+H* L+H* H+L*" [5].

2.2.2. *Rise duration*

I investigated whether speech rate had an effect on the rise duration of the pitch accent. To examine this, I calculated the temporal distance in ms between L and H (Rise duration). First, I analyzed these data in ANOVAs with Rise duration as the independent variable. If the CDH is correct, there should be no significant effect. The SAH predicts that the rise duration should increase as rate decreases. The means, *F*-ratios and *P*-values are presented in Table 2.

Table 2: *Rise duration in ms.*

Speaker	Fast	Normal	Slow	*F*(2, 59)	*P*
FA	168.8(5)	201.3(4)	228.7(5)	40.895	<0.001*
FL	171.8(6)	191.8(8)	221.5(10)	11.667	0.001*
FM	168.2(7)	193.7(6)	200.2(6)	6.729	0.007*
FT	204.3(7)	229.6(8)	251.6(5)	15.636	<0.001*
MK	194.5(6)	214.1(5)	250.3(9)	11.988	<0.001*
MM	204.4(9)	233.3(12)	271.8(11)	15.525	<0.001*
MN	145.6(5)	184.8(7)	199.2(11)	14.953	<0.001*

It can be seen that for all the speakers, the rise duration is the shortest at fast rate, longer at normal rate and the longest at slow rate and that the effect of rate was significant for all the speakers.

Second, I correlated Rise duration with Syllable duration. The CDH predicts that there should be no significant correlation, whereas the SAH predicts a significant positive correlation. The results revealed a significant correlation for all the speakers [N=20; R=0.832 for FA, 0.677 for FL, 0.747 for FM, 0.812 for FT, 0.774 for MK, 0.688 for MM, 0.834 for MN; all significant at P<0.001].

Thus, the results presented clear evidence against the CDH: the rise duration is not constant, but it is highly dependent on the duration of the accented syllable, as predicted by the SAH.

2.2.3. *Alignment of the F0 valley and peak*

I then examined whether the speech rate had an effect on the alignment of the F0 valley and peak. I did these analyses separately for the valley and the peak.

To examine an effect of rate on the alignment of the F0 valley, I measured the temporal distance in ms between the onset of the accented syllable and the F0 valley (LtoC0), and ran ANOVAs with LtoC0 as a dependent variable. Means, *F*-ratios and *P*-values are indicated in Table 3.

Table 3: *LtoC0 in ms (negative values indicate that the valley precedes the onset of the accented syllable).*

Speaker	Fast	Normal	Slow	*F*(2, 59)	*P*
FA	-3.9(3)	-4.8(2)	-6.1(2)	0.122	0.886
FT	-10.9(4)	-9.0(4)	-5.3(3)	0.953	0.404
FL	-9.2(4)	-14.2(6)	-3.2(4)	1.099	0.355
FM	-0.2 (3)	4.7(2)	5.9(4)	1.030	0.377
MK	-17.6(5)	-13.9(5)	-5.9(5)	1.066	0.365
MM	-30.5(7)	-25.7(9)	-24.0(7)	0.368	0.698
MN	-10.1 (3)	- 6.7 (2)	1.4(3)	3.561	0.050*
Overall	-11.8(1)	-9.9(2)	-5.3(1)		

We can observe that for all the speakers but MN, there is no significant effect of rate. MN showed significantly earlier alignment of the valley at slow rate. Thus, it was shown that except for the case of MN's slow rate, the valley is constantly aligned relative to the onset of the accented syllable.

I used the same approach to investigate the alignment of the peak. I measured the temporal distance in ms between the offset of the accented syllable and the peak (HtoC1), and ran ANOVAs with HtoC1 as a dependent variable. The CDH predicts that, given the constant alignment of the valley, the peak should be aligned earlier as rate decreases. The SAH, on the other hand, predicts that there should be no significant effect. Means, *F*-ratios and *P*-values are shown in Table 4.

Table 4: *HtoC1 in ms.*

Speaker	Fast	Normal	Slow	*F*(2, 59)	*P*
FA	30.6(4)	50.0(3)	41.2(3)	8.860	0.002*
FL	21.4(6)	11.4(6)	40.1(6)	4.714	0.023*
FM	45.4(5)	47.2(4)	52.7(2)	1.531	0.243
FT	39.9(5)	29.2(4)	39.3(4)	1.340	0.287
MK	33.9(4)	45.8(3)	60.2(6)	8.119	0.003*
MM	36.5(6)	38.3(9)	49.9(7)	2.688	0.095
MN	21.1(2)	37.7(7)	34.7(5)	3.565	0.050*
Overall	32.7(2)	37.1(2)	45.4(2)		

We can notice that there is a significant effect of rate for four of the speakers, FA, FL, MK and MN. FL and MK showed the latest alignment at slow rate. For FA and MN, the alignment is the earliest at fast rate, later at slow rate and the latest at normal rate. While for these speakers the SAH seems to be not fully confirmed, we should notice that the effects are not what would be predicted by the CDH.

I then redid ANOVAs, this time with the temporal distance in ms between the onset of the following unstressed syllable and the peak (HtoV1) as the dependent variable. The rationale is based on the idea that the peak is not aligned relative to the offset of the accented syllable, but to the onset of the following unstressed syllable. The means, *F*-ratio and *P*-value are shown in Table 5.

Table 5: *HtoV1 in ms (negative values indicates that the peak precedes the onset of the following vowel).*

Speaker	Fast	Normal	Slow	*F*(2, 59)	*P*
FA	-4.4(3)	0.9(1)	-4.3(3)	1.317	0.292
FL	-28.9(7)	-40.1(8)	-20.4(6)	1.958	0.170
FM	-4.0(5)	-9.0(4)	-4.6(3)	0.426	0.660
FT	-6.6(5)	-24.0(5)	-16.6(5)	3.201	0.065
MK	-19.3(5)	-7.6(3)	-0.0(5)	3.330	0.059
MM	-4.1(5)	-9.3(8)	-1.2(8)	1.207	0.322
MN	-16.6(3)	-13.8(8)	-16.9(5)	0.048	0.953
Overall	-12.0(2)	-14.7(2)	-9.1(2)		

It can be seen that the effect for four speakers completely disappears: for all the speakers there was no significant effect of rate. Although the exact alignment point seems speaker specific, I could say that the location of the peak is anchored somewhere in the consonant of the following unstressed syllable.

In sum, there was no support for the CDH. Instead, the results revealed that the alignment of both the peak and valley are relatively constant, as predicted by the SAH.

2.2.4. *Slope*

Finally, I tested the CSH. If, for this hypothesis, the F0 changes at a constant rate, the F0 excursion should be greater as rate decreases, because there would be more time between segmentally anchored valley and peak. To investigate the effect of speech rate on the F0 excursion, I measured the F0 change in semitones between the valley and the peak

(Excursion). First, I conducted ANOVAs with Excursion as the dependent variable. If the CSH is correct, there should be a significant effect. The means are given in Table 6.

Table 6: *Excursions in semitones.*

Speaker	Fast	Normal	Slow	$F(2, 59)$	P
FA	4.8(0.2)	5.5(0.1)	6.0(0.1)	19.543	<0.001*
FL	5.0(0.2)	4.8(0.2)	5.0(0.2)	0.330	0.723
FM	5.2(0.2)	5.1(0.2)	5.2(0.2)	0.395	0.680
FT	4.0(0.2)	5.9(0.2)	5.6(0.2)	49.788	<0.001*
MK	4.1(0.2)	4.3(0.1)	4.8(0.1)	9.466	0.002*
MM	5.1(0.2)	5.7(0.2)	7.9(0.3)	48.172	<0.001*
MN	4.5(0.2)	6.9(0.2)	6.2(0.3)	19.617	<0.001*

There was no significant effect of rate for FL and FM, who showed relatively constant excursions across each speech rate. On the other hand, there was a significant effect of rate for other speakers. For FT and MN the 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 CSH. As for FA, MK and MM, they showed the effects that are predicted by the CSH: their excursions become greater as rate decreases.

Second, I correlated the excursion and the rise duration. The CSH predicts a significant positive correlation. The results are shown in Table 7.

Table 7*: Correlation between Rise duration and Excursion (N=20).*

Speaker	R	P
FA	0.496	<0.001*
FL	-0.055	0.679
FM	0.091	0.489
FT	0.277	0.032*
MM	0.541	<0.001*
MN	0.417	0.001*
MK	0.270	0.037*

Except FL and FM, speakers all show a significant correlation, and for FA, MM and MN correlation is relatively high. These results seem to support the CSH.

In sum, there were effects of the sort predicted by the CSH. The same kinds of effects were found in similar studies by Ladd et al. [8] concerning English pitch accents. They concluded that the effects were too small and too restricted to certain speakers to support the CSH. In Russian, however, the effects seems to be not as small as in English: five of seven speakers showed larger excursions as rate slows. It might be, therefore, plausible to assume that in Russian a given type of pitch accents has a constant slope.

3. Conclusion

In the present study, an experiment was conducted to examine what properties of Russian rising pitch accents were constant under changes in the segmental duration brought about by modifications of the speech rate. The results of the experiment partially replicated the findings of a similar study concerning English prenuclear accents; namely the rise duration of accents is not constant but it increases as rate slows, and both the beginning and the end of the F0 rise are anchored with specific points in the segmental string, regardless of the changes in speech rate. These results confirmed the existence in Russian of the phenomenon of segmental anchoring, which is regarded as a support for the view that treats level tones as intonational primitives. On the other hand, however, the results also presented evidence in favor of the view that a given type of pitch accents has a constant slope.

4. Acknowledgment

Parts of this work were supported by Grant-in-Aid for JSPS fellows.

5. References

[1] 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.

[2] Atterer, Micheala and Ladd, Robert D. forthcoming On the phonetics and phonology of “segmental anchoring” of F0: evidence from German. To appear in *Journal of Phonetics*.

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

[4] 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.

[5] Igarashi, Yosuke 2002 Tak nazyvaemaja "nejtralizacija intonacii" -Fonologičeskoe opisanie russkoj intonacii-. *Bulletin of the Japanese Association of Russian Scholars* 34, 15-21.

[6] Igarashi, Yosuke 2003 Variability and stability of F0 movements in Russian under changes in speech rate. *Proceedings of the 126th conference of the Linguistic Society of Japan*: 274-279

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

[8] Ladd, Robert D., 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.

[9] Ladd, Robert D., 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.

[10] Ladd, Robert D., and Schepman, Astrid 2003 “Sagging transitions” between high pitch accents in English: Experimental evidence. *Journal of Phonetics* 31. 81-112.

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

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

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

[14] 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.

[15] 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.

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