

# Relevance of F0 peak shape and alignment for the perception of a functional contrast in Russian.

Tamara Rathcke

Institute of Phonetics and digital Speech Processing  
Christian-Albrechts-University of Kiel, Germany

tkh@ipds.uni-kiel.de

## Abstract

This paper reports a perception experiment carried out to investigate the perceptually relevant properties of *yes/no*-questions and contrastive emphasis in modern Russian spoken by young people in Kaliningrad. Only melodic cues were involved in the test stimuli such as alignment and shape of F0 peaks as well as presence of a peak plateau. A semantic congruity test was performed to investigate these form-function relations. Results indicate that peak alignment is the strongest cue for the perceptual distinction of the investigated categories. Contour shape (including plateau property) serves as a secondary cue, whereas the effect of a plateau seems to be very small. Results are discussed in terms of phonological modeling of Russian intonation based on an experimental approach including the investigation of intonational forms in relation to linguistic functions.

## 1. Introduction

Modern Russian belongs to the group of languages in which *yes/no* questions usually have no lexical or grammatical markers. In most cases of everyday speech, an utterance can be realized and interpreted as a question only by using an intonational code. The problem of how the functional contrast between statements and questions with identical lexical and syntactic structures is borne out by intonation is treated by a number of authors [3, 8, 12, 13]. In general, the basic difference can be described as that of relative F0 peak alignment with questions being aligned later and statements being aligned early. Furthermore, Rathcke [12] indicates that the shape of the F0 peak has a strong effect on perception of the category distinction: Whereas the steep slope of the rise is a major perceptual cue for questions, the gradual rise is more characteristic of statements.

Additionally, the above-mentioned study [12] shows that the perceptual cues of *yes/no* questions and statements with emphasis of contrast are closely related, so that a single-accent utterance with a medial pitch peak alignment on the penultima can be perceived both as a question and as a statement with contrastive emphasis. According to Odé [11], only one phonological rising pitch accent (classified as H\*L) fulfills the communicative function of both questions and statements with a contrastive or emphatic prominence. But from a linguistic point of view it can be assumed that the signalling of a question as opposed to the signalling of a statement with contrastive emphasis is communicatively important and should be warranted by the system of forms of a given language. Since the formal difference between both functions is not indicated by grammatical means in modern Russian, the intonational cues have to be present.

In the model of Russian intonation proposed by Svetozarova [13], the only formal difference between statements with so-called logical accent (which can be

interpreted as contrastive or emphatic prominence) and questions seems to be the presence of a high F0 peak plateau in questions. Furthermore, for other languages it has been shown that both gradients of the peak shape (i.e. differently rising *and* falling slopes) can influence the perception of linguistic categories covered by intonation [4, 9]. In the previous study of Russian questions and statements done by Rathcke [12], the role of falling slope in F0 peak shapes was not investigated. Therefore, the list of possible intonational features relevant to the perception of functional contrast should be expanded by analysing the effect of a falling slope.

Another way of expressing a linguistic distinction by an intonational contrast is by means of the pitch peak alignment [cf. 6, 8]. So far, only two phonological alignment features have been differentiated in Russian (early vs. late) [10]. Against the background of [6, 8], it can be assumed that the functional distinction between statements with or without contrastive emphasis and questions may be a matter of different peak alignments, whereby a more fine phonological classification of the alignment feature (viz. early vs. medial vs. late) will be needed.

In the study presented here, the problem of whether there are any acoustic cues that differentiate between question and contrastive emphasis is approached by the following hypotheses: (1) the functional difference is assisted by an intonational distinction, (2) the formal difference lies in the shape of pitch peak contour, whereby different slopes of rises and falls and/or presence of a peak plateau may have a strong effect on the perception of the functional contrast, and (3) pitch peak alignment is assumed to be a secondary cue for this perceptual distinction.

## 2. Method

### 2.1. Speech material and stimulus generation

The Russian sentence “*Ona ranjshe ne jela malinu*” (She didn’t eat raspberries earlier) was chosen as speech material. The sentence was realized by a native speaker of Russian (the author) as a statement with an accent on the verbal phrase “*ne jela*” (she didn’t eat).

The original F0-contour of the utterance was systematically replaced by other contours, using *praat* [1] as the experimental tool. The differences in F0-values were converted into semitones (st) by the formula:

$$f_{st} = 12 \log_2(F0_1/F0_2) \quad (1)$$

with F0<sub>1</sub> the peak height maximum and F0<sub>2</sub> the respective endpoint of the gradient. Four acoustic parameters were manipulated for the perception experiment (see Table 1 for the outline). The height of the peak was not manipulated: peak and preceding F0-contour differed by 6 st. The values of slopes as well as peak height were chosen against the background of the

previous study [12]. All combinations of parameters given in Table 1 were tested (i.e. 2 risings x 2 fallings x 2 plateaus x 2 alignments). Thus, the procedure resulted in 16 stimuli.

Table 1: Outline of the parameters manipulated for the perception experiment.

parameter	specification	
rising slope	slow (10 st/s)	fast (50 st/s)
falling slope	slow (30 st/s)	fast (70 st/s)
peak plateau	-p (0 ms)	+p (40 ms)
peak alignment	medial (centre of the accented vowel)	late (centre of the following lateral)

Figure 1 shows the completed manipulations: The combinations of two rising and two falling slopes result in four peak shapes (*fast/fast*; *fast/slow*; *slow/fast*; *slow/slow*), which can be modified by the presence of a peak plateau. Both alignments are marked in relation to the segmental string.

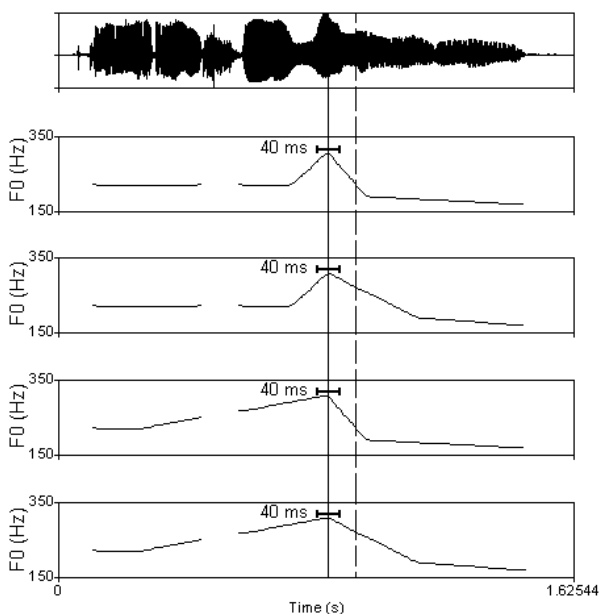


Figure 1: Waveform (top), F0-contour shapes *fast/fast*, *fast/slow*, *slow/fast*, *slow/slow* from top to bottom, respectively. The contour modification by plateau is shown schematically. The vertical lines indicate the medial (solid line) and the late (dashed line) peak alignments in relation to the speech signal.

## 2.2. Procedure and participants

Two different dialog contexts were designed: (1) a context for question and (2) a context for statement with contrastive emphasis. The dialog structure contained following three parts:

- a statement about a person X,
- the test item: a continuative question about X for (1) or contrastive revision of the given statement for (2),
- a reaction to the question or to the contrastive statement.

The dialog contexts were presented on paper sheets, the wording of the auditory presented test utterances were printed in grey.

A CD comprising the auditory stimuli was prepared. Each stimulus was preceded by a short beep and a 4 s pause for reading the context sentences. The stimuli were followed by a 3.5 s pause for decisions. All stimuli and context sentences were presented in randomized order. Each stimulus occurred 4 times in each context. A session was preceded by twelve dummy dialogs (six for each context) to familiarize the subjects with contexts and procedure. So, the total number of experiment stimuli amounted to 140 items (16 stimuli x 2 contexts x 4 occurrences + 12 dummies).

Each subject was tested separately. An experimental session took about 20 minutes. Stimuli were presented from a portable CD player via headphones. The subjects were asked to imagine the dialog situation and to judge on a five point scale whether stimuli and contexts matched semantically. The following scores for judgments were used: 1 = *definitively non matching*; 2 = *doesn't quite match*; 3 = *not sure*; 4 = *matches quite well*; 5 = *a definite match*.

Fifteen subjects (six male and nine female) with no known speech or hearing disorders, aged between 14 and 15 years participated in the experiment. Their and their families' native language was Russian. All participants were students of the Kaliningrad gymnasium № 49.

## 3. Results

The data from two subjects (one male and one female) were excluded from the following analysis because they lacked any judgment profiles. That is, their response variance for each stimulus in each context lay considerably above the variance found for the entire sample. The following results are therefore based on the data from thirteen subjects.

For each subject, personal score values for the four presentations of the 16 stimuli and the two contexts were calculated. Mean values of these data are given graphically: Figures 2 and 3 show the 'matching' responses for each context as a function of peak shape for medial and late alignments separately. The higher the scale value, the better the stimulus fits the given context.

On the basis of the personal score values of 'matching' judgments from all thirteen subjects, a multifactorial analysis of variance (ANOVA) with repeated measures (within-subject design [2]) was performed. The factors were: (1) peak shapes (*ff*; *fs*; *sf*; *ff*); (2) plateau (-p; +p); (3) peak alignment (*m*; *l*); (4) context (*q*; *c*). Since the variances were not homogeneous, the conservative Greenhouse-Geisser-test was used. Table 1 shows the significant results of the statistical analyses. Other findings were not significant.

Table 2: Results of multifactorial analysis of variance with repeated measures (within-subject effects, n=13)

	F	df <sub>1</sub> /df <sub>2</sub>	p
alignment*context	34.570	1/ 12	< 0.001
shape*context	18.773	1.7/ 19.9	< 0.001
plateau*context	19.908	1/ 12	< 0.001

The statistical results show that peak alignment has the strongest effect on the perception of intonational contrast related to linguistic categories: Medial alignment is shown to be associated with the expression of contrastive emphasis, whereas late alignment is more linked with the expression of

questions. For different peak shapes and plateau types, the results seem to be more complex: According to the statistical analysis, both parameters have a similar effect on perception, but the response profiles given in Figures 2 and 3 suggest that the influence of slopes resulting in different peak shapes have a greater effect than a plateau. Whereas the change of peak slopes may produce a strong change of perceived category in medial as well as in late peak position, the addition of a plateau can only make the perception of a question more likely (and impair the perception of contrastive emphasis) for all peak shapes.

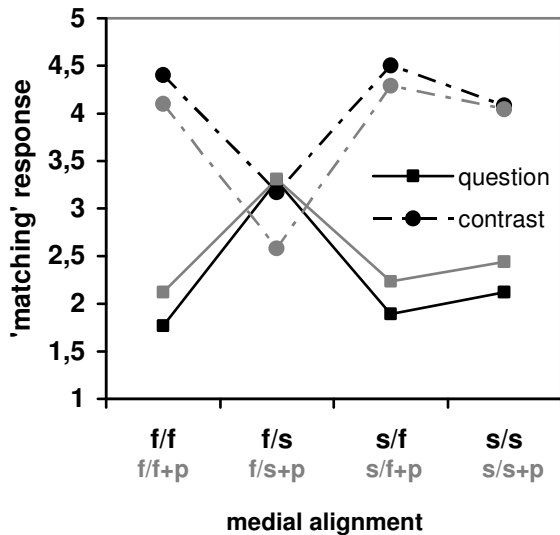


Figure 2: Response profiles for stimuli with (grey) and without (black) a plateau at the medial alignment. Solid lines show the results for question context, dashed lines show those for contrastive context.

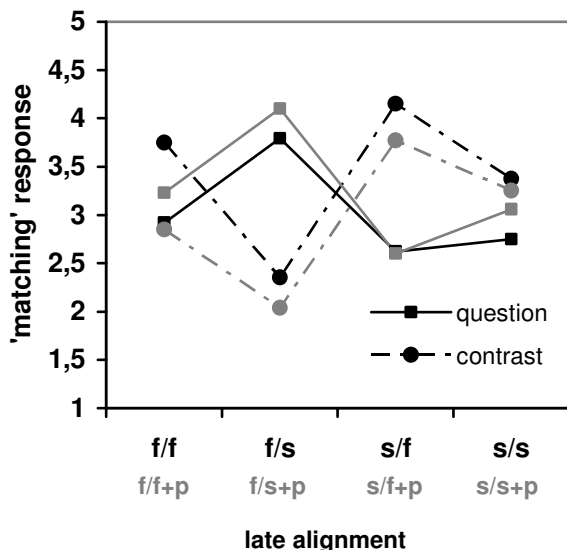


Figure 3: Response profiles for stimuli with (grey) and without (black) a plateau at the late alignment. Solid lines show the results for question context, dashed lines show those for contrastive context.

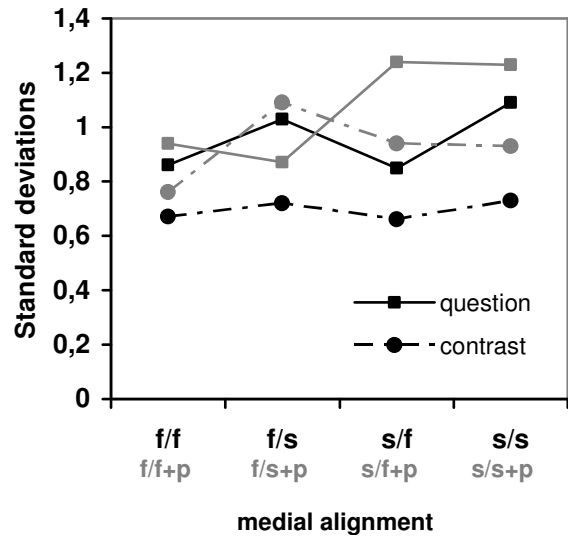


Figure 4: Standard deviations of 'matching' judgments for stimuli with (grey) and without (black) a plateau at the medial alignment. Solid lines show the results for question context, dashed lines show those for contrastive context.

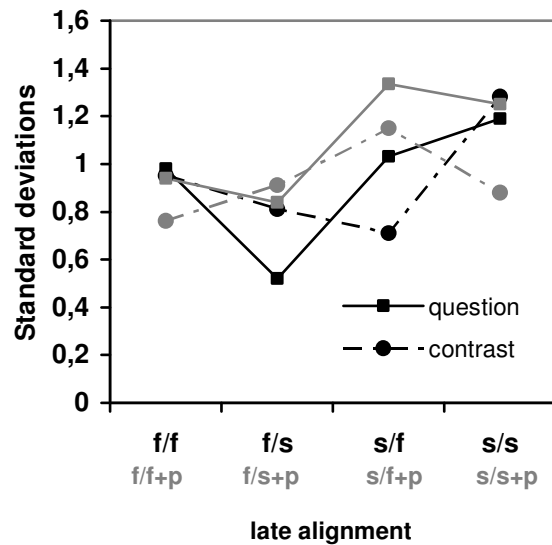


Figure 5: Standard deviations of 'matching' judgments for stimuli with (grey) and without (black) a plateau at the late alignment. Solid lines show the results for question context, dashed lines show those for contrastive context.

The results for peak shape show that combinations of different falling and rising slopes are relevant for a categorical distinction rather than the falling slope alone. So, a combination of a fast rise and slow fall causes the perception of question, whereby a slow rise followed by a fast fall is a strong indication of contrastive emphasis. At this point, the variance of responses can give some more information about degree of certainty and agreement between subjects. As shown in Figures 4 and 5, the shapes *fast/fast* and *slow/fast* with medial alignment as well as *slow/fast* with late alignment

cause the lowest level of variance for the emphasis category. Similarly, the shape *fast/slow* with a plateau in medial position and the shape *fast/slow* (without a plateau) in late position have the lowest level of variance. The peak shape with slow rises and falls in both alignment positions is neither a clear case of question nor of emphasis.

#### 4. Discussion and Conclusions

As the results of the experiment show, F0 provides cues for the discrimination between question and contrastive emphasis in Russian. The perceptual difference depends predominantly on the alignment property of the F0 peak. The shape of the peak contour has a secondary relevance for the perception of the investigated distinction, whereby the slopes of peak gradients have a stronger effect on the perception of the functional contrast than a high peak plateau. So, a plateau property cannot be regarded as phonologically relevant, but can instead be assumed to influence the formation of perceptual prototypes (cf. [12]) or to play a role in cases of untypical alignment.

The present findings for peak alignment and shape lead to a reconsideration of the hierarchy of phonological features for Russian intonation phonology that was proposed by Odé [10]. The formal difference between questions on the one hand and statements with and without contrastive emphasis on the other can be described in terms of peak alignment (cf. also [12]) with the consequence that alignment (or ‘timing’ in terms of Odé [10]) is not just binary. The methodical approach of Odé [10, 11] is characterized by a sharp division between investigation of intonational forms and that of their functions. Therefore, the method of her intonational research is divided in two separate steps: (1) the investigation of the form and (2) the filling of the postulated forms with linguistic functions. As the present study indicates, a form-oriented approach can lead to incorrect conclusions about the distinctive features in intonational forms. To conclude, semantically motivated tests are indispensable in the investigation of intonational forms.

Recent investigations of Russian intonation [5, 11, 14] are mostly inspired by the autosegmental-metrical model [7] and the possibility of reducing the intonational phonology to a small number of contrasts (cf. [11]). In this framework, phonological categories of pitch peaks in Russian may be expressed in terms of the following three tone accents: H(+L)\*, H\*(+L) and L\*(+H), which are primarily related to the basic linguistic functions of a simply statement, a statement with contrastive emphasis and a question, respectively. However, in this type of notation the characterization of an F0 peak contour is effectively defined by only one property, viz. a steeper slope – so that it does not really contain assertions about the relevance of the form for perception (and probably production). The present study suggests that a contour with an F0 peak should be treated as a *gestalt* rather than a succession of two tone targets, since both rising and falling gradients of a peak appear to be salient for the perception of linguistic functions related to intonational forms.

Furthermore, the intended functional differences are to be realized on variable segment structures, resulting in variable acoustic F0-forms. So the contour shape properties as well as alignment should be classified as a feature bundle belonging to an accent type. Depending on the character of segment structures, each of those properties may get a stronger weight for the perception (and therefore production).

#### Acknowledgments

I would like to thank the headmistress Tatyana Mishurovskaya of Kaliningrad gymnasium № 49 for her help in the organization of experiment and the students for their participation. Further, I thank Ernst Dombrowski for his help with statistical analysis as well as Jonathan Harrington and Michel Scheffers for their help in the preparation of this paper.

#### 5. References

- [1] Boersma, P.; Weenink, D. Praat: doing phonetics by computer. <http://www.fon.hum.uva.nl/praat/>
- [2] Brosius, F., 2002. SPSS 11. Bonn: MIT-Verlag.
- [3] Bryzgunova, E. A., 1980. Intonatsija. In *Russkaja grammatika, I*, Shviedova, N. J. (ed.). Moskva: Nauka, 96-122.
- [4] Gósy, M.; Terken, J., 1994. Question marking in Hungarian: timing and height of pitch peaks. *Journal of Phonetics* 22, 269-281.
- [5] Igarashi, Y., 2004. “Segmental anchoring” of F0 under changes in speech rate: Evidence from Russian. *Proceedings of Speech Prosody*, Nara, 25-28.
- [6] Kohler, K. J., 1991. Terminal Intonation Patterns in Single-Accent Utterances of German: Phonetics, Phonology and Semantics. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK 25)*, 115-187.
- [7] Ladd, D. R., 2001. Intonational phonology. Cambridge: University Press.
- [8] Makarova, V., 1999. Pitch Peak Alignment in Russian Declaratives, Interrogatives and Exclamations. *Proceedings of the XIV International Congress of Phonetic Sciences, Vol. 2*, San Francisco, 1173-1176.
- [9] Niebuhr, O., 2003. Perceptual study of timing variables in F0 peaks. *Proceedings of the XVth International Congress of Phonetic Sciences*, Barcelona, 1225-1228.
- [10] Odé, C., 1989. Russian intonation: a perceptual description. Amsterdam: Rodopi.
- [11] Odé, C., 2005. Towards a description of communicative functions and prosodic labeling of Russian rising pitch accents. *Between Stress and Tone. Abstracts of IAS Conference*, Leiden, 80-82.
- [12] Rathcke, T., forthcoming. A perceptual study on Russian Questions and Statements. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK 37)*.
- [13] Svetozarova, N. D., 1998. Intonation in Russian. In *Intonation Systems. A Survey of Twenty Languages*, D. Hirst, A. Di Cristo (eds.). Cambridge: University Press, 261-275.
- [14] Yokoyama, O. T., 2001. Neutral and non-neutral intonation in Russian: a reinterpretation of the IK-system. In *Die Welt der Slaven XLVI*, 1-26.