A perceptual study on Russian questions and statements

Tamara Rathcke
Institute of Phonetics and Digital Speech Processing, Christian-Albrechts University Kiel, Germany

This paper presents the results of perception experiments designed to investigate the contribution of f0-peak rise, height and alignment to signaling interrogative vs. declarative sentence mode in Russian. The results of the study show that the major perceptual cues for this category distinction are the f0-peak alignment and the slope of the rise. According to the results, the primary perceptual cues for questions are a steep rise and a late peak alignment at the offset of the accented vowel, whereas the more gradual rise and early f0-peak alignment at the onset of the accented vowel are strong cues for a declarative mode. The height of the f0-peak has no influence on the category distinction. The results are discussed in terms of the phonological modeling of Russian intonation as well as in terms of the frequency code for universal meanings in intonation.

1. Introduction

1.1. Theoretical backgrounds

It has been reported for a great number of languages that intonation in the so-called yes/no questions has a final rise in contrast to statements which have a final fall (cf. Hermann 1942; Bolinger 1978). This finding was often generalized as a language universal (cf. the strong universalist hypothesis, Ladd 1981).

However, this universal has some counter-examples like Hungarian (Gósy and Terken 1994), Neapolitan Italian (D’Imperio and House 1997) or Australian English (Fletcher and Harrington 2001). In these languages the intonational contrast between a question and a statement does not depend on the final contour, but a specific f0-pattern of the nuclear syllable. In Australian English final rises are used in both statements and questions, but the starting point of the rise is different in the accented syllable: It is low in statements and high in questions. In Neapolitan Italian as well as in Hungarian, grammatically and lexically identical utterances can be realised as statements vs. questions by changing the alignment of a final falling f0-peak from early to medial or late.

Moreover, Gósy and Terken (1994) as well as D’Imperio and House (1997) demonstrate that there are more tonal cues which can interact in the perception of these pragmatic categories. Shifting the f0-peaks from early to medial or late alignment results in a categorical change from the perception of statements to that of questions. In addition, the peak shape or height can have a strong effect on the categorisation. In this way, an increase of peak height, steeper slopes in a rise and more gradual slopes in a fall are both likely to elicit the perception of questions in Hungarian and thus can be interpreted as secondary tonal cues for the pragmatical distinction. Similarly, in Neapolitan Italian a blunt as opposed to a pointed peak is a perceptual cue for statements.
Modern Russian belongs to the group of languages in which the distinction between statements and yes/no questions (in the following: 'questions') is not marked by lexical or grammatical means (cf. Kirschbaum 2001: 309f.). The intonational difference between these two pragmatic categories is not related to the typical rising pitch movement at the end of the utterance. Earlier acoustical analyses have shown that there are a number of intonational differences between statements and questions in Russian (Wenk 1975; Nikolajeva 1977; Bryzgunova 1980; Makarova 1999; Svetozarova 1982, 1998). These differences can be summarized as follows: (1) Utterances containing a single accented syllable in the nuclear alignment have a falling pitch in statements and a rising pitch in questions. (2) If the nuclear accent is not realized on the last syllable of the utterance, questions have a rising-falling pitch. Makarova (1999) describes these differences as a property of pitch peak alignment which is early in statements and late in questions. (3) Questions have a greater excursion size and higher peaks than statements (Bryzgunova 1980). (4) In the model of Russian intonation given by Svetozarova (1982, 1998) the only difference between statements with so called logical accent and questions is described in terms of presence of a high pitch peak plateau in questions.

The only well-known perceptual study on Russian intonation was done by Odé (1989). In the framework of the IPO approach (‘t Hart et al. 1990), she tested the perceptual relevance of rising and falling pitch movements in Russian and analyzed their acoustical properties. The perceptually most relevant features were found to be excursion, timing and slope of prominence lending pitch movements. No semantic interpretations of the postulated accent types were given – however, against the background of the studies cited above it can be expected that the intonational form of questions is characterised by a rising accent with great excursion, late timing and low posttonic part, i.e. the accent type Rl- in terms of Odé (1989). This accent type contains an obligatory peak plateau of 60 ms.

The investigation by Odé (1989) extended the number of distinctive features which are perceptually relevant for the intonational phonology of Russian, but the communicative functions of the proposed accent types were not discussed. One of the gaps in Odé’s (1989) study identified by Kodzasov (1992) was that there was no analysis of a rising-falling accent type without a plateau which is characteristic for some question utterances.

The studies on Russian intonation cited above give some information about the potential features for the contrast between questions and statement. The empirical question of the present study is: Which acoustic cues are established in Russian to discriminate between question and statement utterances? Perception studies on other languages (Gósy and Terken 1994, D’Imperio and House 1997) show that the pragmatic language categories like statements and questions can be formalised by a hierarchical network of tonal cues. There are no comparable investigations on Russian.

Four hypotheses are to be examined:

(1) Following the above discussion, is the f0-peak alignment critical for the perceptual distinction between statements and questions?

(2) Are questions distinguished from statements by having higher f0-peaks?

(3) To what extent does the f0-slope play a role in distinguishing questions from statements? More specifically, is a steep late peak more likely to give rise to the perception of a question than a high late peak with a gradual rise?
Are the f0-peak parameters alignment, height and rising slope sufficient for the perceptual distinction between statements and questions or are there more perceptual cues at work?

Since the slope of fall has been showed to be highly variable and dependent on the post-nuclear segmental composition (cf. Wenk 1975: 188f.), it will not be analyzed in the present study.

2. Method

2.1. Material and stimulus generation

The sentence 'Jejo zovut Jeljena' (/jijo zavut jiljena/) was produced by a native speaker of Russian (the author) as a statement ‘Her name is Helena.’ vs. as a question ‘Is her name Helena?’ and used as test material. The sentence accent was realized on the penultimate syllable /jelj/. The naturally produced contours were stylized in Praat using the PSOLA algorithm (Moulines and Carpentier 1990) with 5 relevant points in terms of the IPO method (’t Hart et al., 1990). The f0-peak was defined by three points at the onset, turning-point and offset of the peak contour. The stylized contours of the naturally produced sentences are shown in Figures 1 and 2.

The following two f0-parameters for peak shape were manipulated for the experiment: (1) the height (high vs. low) and (2) the slope of the rise (steep vs. gradual). Table I gives an overview of the four contour shapes carried out for the experiment. Notice that slope and peak height were varied independently. Additionally, the alignment of the f0-peak was progressively changed from early to late with 11 equidistant steps of 30 ms for each contour shape. The resynthesized contours are illustrated in Figure 3, the systematically manipulations of peak alignment are marked by arrows.

The stylized contour of the naturally produced statement served as a basis for all manipulations of the resynthesis described above. The stylization of the question utterance was used only as a control stimulus.

Figure 1: Synchronized time-waveform and f0-contour of the naturally produced statement time-aligned with the stylized f0. The five points that were fixed for stylization are marked. The dashed lines give the duration of the accented vowel (/e/).
Figure 2: Stylized contour from the naturally produced question (see Figure 1 for further details).

Table I: Scematic overview of the four contour shapes used in the experiment.

<table>
<thead>
<tr>
<th>stimulus series</th>
<th>peak height</th>
<th>slope of rise</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. low</td>
<td>(320 Hz like in statement)</td>
<td>gradual (133 Hz/s like in statement)</td>
</tr>
<tr>
<td>2. high</td>
<td>(360 Hz like in question)</td>
<td>gradual (133 Hz/s like in statement)</td>
</tr>
<tr>
<td>3. low</td>
<td>(320 Hz like in statement)</td>
<td>steep (885 Hz/s like in question)</td>
</tr>
<tr>
<td>4. high</td>
<td>(360 Hz like in question)</td>
<td>steep (885 Hz/s like in question)</td>
</tr>
</tbody>
</table>

Figure 3: Schematic description of four experiment contours. The arrow marks the temporal interval over which f0-peak was shifted.
2.2. Procedures and subjects

The experiments were designed within the paradigm of categorical speech perception (for an overview see Repp 1984). The following three tests were performed:

(1) a quick serial discrimination test: the entire set of unrandomized stimuli as well as control stimuli (both stylizations, Figures 1 and 2) were presented; no categories for decision were given; the listeners were asked to rate the perceived changes of the sentence melody and to paraphrase their meanings.

(2) an identification test: the stimuli were played in randomized order with a pause of 3.5 seconds after each presentation; listeners responded to the melody using a five-point scale extending from ‘a clear statement’ to ‘a clear question’.

(3) a formal randomized AX discrimination test: the stimuli pairs for this test differed in two steps of the peak alignment from left to right and also included (identical) control pairs; the listeners were asked to rate the melody of the stimuli pairs as 'same' or as 'different'; the randomized pairs were presented with a pause of 4 seconds for decisions. No time-order effects were tested.

The stimuli were repeated ten times in the identification test and five times in the discrimination test. The number of repetitions for the discrimination tests was reduced in order to limit the total experiment time to one hour. Four different stimulus series (corresponding to peak alignment continua of four different peak shapes, s. Table I) were tested separately.

A total of 38 native speakers of Russian (31 female and 7 male) with no known speech and hearing disorders participated on the experiment. The subjects were aged between 20 and 56 and came from various cities of Russian Federation. At the time of experiment they worked or studied in Kiel. All of them claimed to be native speakers of Standard Russian (rus. literaturnoje proiznoshenije, cf. Ozhegov and Shvedova, 1993: 336). The subjects were divided in four groups for each of the stimulus series. The tests took place in the sound treated room of the Institute of Phonetics (IPdS) Kiel.

3. Results

The data from only 30 subjects (24 female and 6 male) were analyzed. The data of eight subjects were excluded from analysis for two reasons: (1) One subject did not perceive any change in the melody of the utterance during the quick serial discrimination test. (2) In the quick serial discrimination test, 7 subjects did not interpret the perceived changes in the speech melody as a question but as an emphasis of contrast, even in the case of the stylized contour of the natural produced question utterance. The latter finding will be discussed in Section 4.

3.1. Identification

Listener responses were separated into two categories (question vs. statement). Only the responses 'more like a question then a statement' and 'a clear question' (scale values 4 and 5) were added to the category 'question', because this should exclude ambiguous stimuli. The other response categories ('a clear statement', 'more like a statement then a question' and 'not sure' corresponding to scale values 1, 2 and 3) were categorized as 'statement'. The results are
presented in Figure 2. Identification functions show percentage of 'question'-responses depending on the f0-peak alignment of the four tested contours. The f0-peak alignment is given in relation to the onset of the accented vowel (0 ms). Ratings of the control stimuli are given on the left and on the right edges of the Figure 4 for all subject groups separately.

The identification functions demonstrate that the shift from early to late peak alignment increases the frequency of 'question' judgements. The greatest increase in 'question' responses took place between the middle (+60 ms) and final third (+90 ms) of the accented vowel. However, the degree of the change in judgements varies. Results for peak shapes low/steep and high/steep are markedly different from that for the peak shape low/gradual or high/gradual. The late peak with the shape high/steep shows the highest rate of 'question' judgements (more than 90%, Figure 2). Additionally, early alignment of the peak shapes low/steep and high/steep also evokes an increase of 'question' judgements. The stylized natural productions of statement and question utterances which served as control stimuli are nearly always identified as statement and question respectively.

The data were tested statistically. A repeated measures ANOVA was run in SPSS (Brosius 2002) with independent variables corresponding to the group factors were slope of the rise (gradual vs. steep) and height of the peak (low vs. high); the repeated measure was peak alignment with 6 levels (viz. alignments of 0; +30; +60; +90; +120; +160 ms related to the accented vowel). The number of factor levels was reduced for the statistical analysis because the present study was primarily concerned with the perceptual changes associated with the accented vowel.

Tables II and III present the results of the statistical analysis. The probability of alpha error was set to p = 0.01%. The factor slope shows a significant between-subjects effect on the tested dependent variable ('question' judgements, p<0.001, Table II). For the factor of peak alignment there are two significant findings: the change of peak alignment from earlier to medial (+30 vs. +60) and from medial to later one (+60 vs. +90) causes a significant increase of 'question' judgements with the later alignment producing a stronger effect (F=56.169 vs. F=8.852, Table III).

<table>
<thead>
<tr>
<th>group factors</th>
<th>contour shapes</th>
<th>df₁ / df₂</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>slope</td>
<td>(l/g, h/g) vs. (l/s, l/s)</td>
<td>1/26</td>
<td>15.975</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>height</td>
<td>(h/g, h/s) vs. (l/g, l/s)</td>
<td>1/26</td>
<td>3.361</td>
<td>n. s.</td>
</tr>
<tr>
<td>slope and height</td>
<td>l/g vs. h/g vs. l/s vs. h/s</td>
<td>1/26</td>
<td>.000</td>
<td>n. s.</td>
</tr>
</tbody>
</table>

Table II: Results of one-way multifactorial analysis of variance for identification tests (between subjects)

<table>
<thead>
<tr>
<th>repeating factor</th>
<th>df₁ / df₂</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 vs. +30</td>
<td>1/26</td>
<td>.851</td>
<td>n. s.</td>
</tr>
<tr>
<td>+30 vs. +60</td>
<td>1/26</td>
<td>8.852</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>+60 vs. +90</td>
<td>1/26</td>
<td>56.169</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>+90 vs. +120</td>
<td>1/26</td>
<td>.897</td>
<td>n. s.</td>
</tr>
</tbody>
</table>

Table III: Single comparisons of the repeating factor for identification tests (repeated contrasts)
Figure 2. Identification functions of the four tested contours depending on a change of f0-peak alignment, in comparison with the judgements of stylized original productions of statement vs. question (different judgements of the latter concern with different groups of subjects). The number of subjects per stimulus series is included in the legend.

3.2. Discrimination

Figure 3 displays percentage of 'different' judgements for the non-identical stimuli pairs of each peak shape as a function of peak alignment. The x-axis presents pairs of stimuli with different peak alignments in relation to the onset of accented vowel (0 ms). The pairs have a time interval of 60 ms between their peak alignments.

The main result of the discrimination tests is that for all of the tested peak shapes, the stimuli pairs in the middle of the continuum can be better discriminated in comparison to the stimuli from both edges of the continuum. The identical control stimuli are predominantly judged as the same (discrimination rate under 30%, without figure).

<table>
<thead>
<tr>
<th>group factors</th>
<th>contour shapes</th>
<th>df1 / df2</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>slope</td>
<td>(l/g, h/g) vs. (l/s, l/s)</td>
<td>1/26</td>
<td>4.708</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>height</td>
<td>(h/g, h/s) vs. (l/g, l/s)</td>
<td>1/26</td>
<td>1.884</td>
<td>n. s.</td>
</tr>
<tr>
<td>slope and height</td>
<td>l/g vs. h/g vs. l/s vs. h/s</td>
<td>1/26</td>
<td>.057</td>
<td>n. s.</td>
</tr>
</tbody>
</table>
TABLE V: Single comparisons of the repeating factor for discrimination tests (repeated contrasts)

<table>
<thead>
<tr>
<th>repeating factor</th>
<th>df1/ df2</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>-60/ 0 vs. -30/ +30</td>
<td>1/26</td>
<td>.110</td>
<td>n. s.</td>
</tr>
<tr>
<td>-30/ +30 vs. 0/ +60</td>
<td>1/26</td>
<td>2.888</td>
<td>n. s.</td>
</tr>
<tr>
<td>0/ +60 vs. +30/ +90</td>
<td>1/26</td>
<td>1.082</td>
<td>n. s.</td>
</tr>
<tr>
<td>+30/ +90 vs. +60/ +120</td>
<td>1/26</td>
<td>7.620</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>+60/ +120 vs. +90/ +150</td>
<td>1/26</td>
<td>10.810</td>
<td>&lt; 0.01</td>
</tr>
</tbody>
</table>

Figure 3. Discrimination functions of the four tested contour shapes for non-identical stimuli pairs. The number of subjects per stimulus series is included in the legend.

The tested peak shapes affected the judgements in three different ways: (1) Discrimination maxima for the shapes high/steep and low/steep are located between an early and a medial peak alignment, whereas the shape low/gradual shows a discrimination peak between a medial and a late peak alignment. The contour high/gradual has two discrimination maxima without being very distinct (under 50%, Figure 3). (2) Best discrimination is given by contour shapes high/steep (88%, Figure 3) and low/steep (ca. 78%, Figure 3). The discrimination maximum of the contour shape low/gradual exceeds the 50%-boundary. (3) In addition, the stimuli series of shapes low/gradual, high/gradual, low/steep (but not high/steep) indicate a secondary discrimination peak at the edge of the vowel onset.

These data were tested by means of a one-way ANOVA with independent variables slope of the rise and height of the peak and with pairing of peak alignments as a repeated measure (Tables IV and V). At the alpha level p < 0.01, there were no significant findings for any of the group factors, but there was a significant trend for the factor slope (p < 0.05, Table IV).
The factor of peak alignment revealed two significant findings: the changes of peak alignments from early to medial alignment (+30/ +90 vs. +60/ +120 ) and from medial to late alignment (+60/ +120 vs. +90/ +150) caused a significant decrease of 'different' judgements, whereas the finding for pairs with later peak alignment was stronger (F=10.810 vs. F=7.620, Table V).

3.3. Informal results

The results of the quick serial discrimination test, which was always run first in the experiment sessions, are inconclusive. There are some suggestive interpretations of the subjects to be noted.

The contours with early peak alignments were predominantly interpreted as statements or assertions, the contours with late alignment as questions or uncertainty, doubt of the speaker. Besides the pragmatic difference between statement and question, one additional category was found for the peak shapes high/gradual, low/steep and high/steep with early or medial alignments, namely the category of contrast or emphatic accent. For some of the subjects, the contour with the shape high/steep and medial peak alignment sounded like a question, while others judged the same stimuli as an emphasis of contrast.

4. Discussion and Conclusions

4.1. Summary

A categorical perception change in the traditional sense (cf. Repp 1984) was found for the contour shapes low/steep and high/steep. The series with shape high/gradual did not show any distinct discrimination peak and the series with shape low/gradual did not show a clear discrimination maximum nor a strong identification jump. However, all tested shapes showed a strong tendency of change from the perception of question to that of statement accompanied by the peak shift from medial to late alignment within the accented vowel. This tendency was borne out by the statistical results.Compatibly with these findings in identification, the statistical results of discrimination tests showed that the pairs of stimuli aligned after the middle of the accented vowel evoked a significant drop of 'different' judgments. This may lead to the conclusion that there is a categorical boundary at the middle of accented vowels. Thus, the first hypothesis of the study (Section 1.3.) was supported: The late peak alignment (later then the last third of accented vowel) is a primary strong cue for perceiving a question in Modern Russian.

At the late peak alignment, the presence of each additional parameter of contour shape produced stronger 'question' identification. The results of discrimination showed a similar tendency: Except for the shape high/gradual, the discrimination ability of the subjects improves the more parameters of the naturally produced question contour are integrated in the tested contour shapes. Consequently, the parameter slope of the rise has to be interpreted as the stronger cue for question than the parameter peak height: the effect of slope was significant for identification and showed a significant tendency for discrimination. These results were consistent with the hypothesis 3 whereas hypothesis 2 was not supported.

Furthermore, the results lend support to the idea that there is an interaction between the f0-peak alignment and the contour shape: Peaks with a gradual slope of rise should be aligned later than peaks with steep slopes to create a perception of question. Compatibly, a perception study by Niebuhr (2003) has shown for German that a more gradual rise can shift the
perceptual boundary between early and medial peaks towards the syllable onset. A similar interaction was observed in the perception of questions and statements in Hungarian (cf. Gósy und Terken 1994): At the medial peak alignment, a slow rise can weaken the perception of question, in contrast to a slow fall which facilitates it. Similar results from different languages indicate a perceptual sensibility to tonal changes in relation to the accented vowel and the tone bearing segments of accented syllables respectively. Thus, the alignment as well as the shape of rising-falling pitch peaks can be broadly used for establishing intonational categories in languages.

The current linguistic research indicates that language categories are organized prototypically. A prototype is defined as a central or most representative member of a category (cf. Saeed, 1997: 37). In the experiments presented here, a prototype effect can be observed for late peaks with the shape high/steep as well as for the naturally produced stylized contour of question. This result has to be interpreted in terms of hypothesis 4 which claims that in utterances with a single focus on the final word, the intonation difference between a statement and a question is carried by the three tested parameters (height, slope of the rise and alignment of the f0-peak). In contrast to this, a prototypical intonation contour of a statement has an early alignment before the vowel onset and a gradual slope of the rise. Contours with steep rises are untypically and evoke far less clear 'statement' judgments.

In the tested continua, an additional discrimination maximum for contour shapes high/gradual and low/steep indicates the presence of another category. Such interpretation is corroborated by the results of informal tests: The intuitive reactions of subjects show that high peaks as well as peaks with a steep slope of the rise can be interpreted as an expression of contrast or emphasis. Also in many other languages, the emphasis can be related to the height and extension of pitch peaks (cf. Gussenhoven (2002) for an overview). More empirical investigations are needed in order to shed light on this problem.

The results of the experiments suggest that the perceptually based intonational phonology proposed by Odé (1989) should be expanded. This phonological system lacks the accent type found to be prototypical for a question (Rl+ in terms of Odé). Further, in the intonational phonology of Russian, slope of the rise in rising-falling pitch accents is assumed to be distinctive. This assumption argues against the concept of Odé (1989) that the slope of rising pitch accents is not an independent phonological feature, but is dominated by excursion and timing. However, the results of the present study show that the pitch excursion size (which incorporates peak height in terms of Odé) is inferior. Indeed, the parameters of alignment and slope of rise can interact, but they seem to be distinctive for a differentiation between statement, question and emphasis of contrast. This assumption is currently being investigated (Rathcke, 2006).

4.2. Universal hypothesis from a perceptual point of view

The results of perceptual studies from various languages summarized above suggest that very subtle changes of intonation contours in accented syllables are perceptually discriminable and may be linked to language concepts (like question and statement) by creating prototypical contours and specific perceptive cues (cf. also Kohler 1991: 156). Thus, in addition to a non-early alignment, a steep rise (like in Russian) or a high peak height (like in Hungarian) can function as dominant cues for the perception of a question. The form-function relations are on the one hand conventional and language-specific but at the same time some of these have been shown to be almost universal (e.g., Gussenhoven 2004).

The results of the experiments can be interpreted to be in accord with the universal hypothesis in terms of the frequency code proposed by Ohala (1984). Against the background
of the frequency code the linguistic universal hypothesis should be formulated differently as done by Ladd (1981). It should be understood as a strong tendency of using high or rising tonal patterns for questions and low or falling tonal patterns for statements, namely without a concrete reference point in the utterance, because it would offend a conventional principle of language (cf. de Saussure 1931). From a perceptual point of view, the hypothesis involves a contrast between a relative tonal height vs. tonal depth in relation to perceptually relevant landmarks such as an accented syllable or the end of a sentence. Actually, for languages with variable forms of questions and statements there is some evidence that rising-falling contours with high and/or late peaks tend to be perceived as ‘inquiring’ (cf. e.g. Gussenhoven and Chen 2000; Dombrowski 2003).

4.3. Conclusions

The results of the present investigation are consistent with findings in other languages which show that pragmatic categories like question and statement can be considered as a hierarchical network of tonal cues with language-specific compounds of cues being prototypical. Further, the results are consistent with the universal tendency of form-function relations proposed by the frequency code. Finally, the results underline the importance of including a semantic and/or pragmatic analysis in modeling a language’s intonational phonology.

I thank Ernst Dombrowski and Michel Scheffers for their valuable comments on the method of the study and Klaus J. Kohler for fruitful discussions.

References


