

**This is an electronic reprint of the original article.  
This reprint *may differ* from the original in pagination and typographic detail.**

**Author(s):** Ullakonoja, Riikka

**Title:** How do native speakers of Russian evaluate yes/no questions produced by Finnish L2 learners?

**Year:** 2010

**Version:**

**Please cite the original version:**

Ullakonoja, Riikka. 2010. How do native speakers of Russian evaluate yes/no questions produced by Finnish L2 learners? *Rice Working Papers in Linguistics*, Vol. 2 : s. 92-105

All material supplied via JYX is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

# How do native speakers of Russian evaluate yes/no questions produced by Finnish L2 learners?\*

Riikka Ullakonoja  
University of Jyväskylä

## Abstract

This study analyzes native Russian speakers' evaluation of seven Russian yes/no-questions each produced by Finnish speakers in two sets of recordings (during a stay in Russia and after it). The Finnish speakers were six female university students of Russian. This research question is interesting because the two typologically unrelated languages differ in the prosody of yes/no-questions. In Russian a yes/no-question is created from a lexically and syntactically corresponding statement by means of intonation, whereas in Finnish the cue for questioning is an interrogative particle – ko/-kö instead of prosody. Hence, native Finnish speakers are likely to have difficulties in pronouncing Russian yes/no-questions. The aim was to find out how native Russian speakers recognise the intended questions produced by Finnish learners. First, the recognition rate of the different yes/no-questions was studied, and then the acceptability rating of questions was computed. The results show that in general students did not perform very well in producing a yes/no-question, but there was great variation depending on the question and learner. According to the successfulness of production two groups of utterances were established: successful and non-successful ones. The statistically significant difference between the two was explained by their syntactical and lexical content. The conclusions made are supportive of earlier findings, where Russian question intonation has been found difficult for Finns to learn.

**Keywords:** phonetics, speech perception, prosody, intonation, second language

## 1 Introduction

Perception of intonation in different languages has been the subject of extensive research and has been found to be a complex issue (Vaissière 2005). Yet a comprehensive theory of how intonation is perceived does not hitherto exist. The present study focuses on the perception of yes/no questions in Russian. The interest lies in the non-native Russian speech of advanced learners, whose mother tongue (L1) is Finnish. In this experiment native speakers of Russian were asked to evaluate a total

---

\* I would like to thank Docent Nina B. Volskaya from St. Petersburg State University for the help with recruiting the subjects and conducting the listening experiment.

of 84 utterances meant as questions by Finns. The stimuli were extracted from Finnish students' read-aloud dialogues. The study is a part of a set of papers, where students' performance prior to, during and following their stay in Russia are compared. In this paper, the students were recorded during and following a stay in Russia to allow for comparisons between the two sets of recordings. The purpose of the study was to determine whether the students' pronunciation of yes/no questions remained consistent between the two recordings (during the stay in Russia and following it) coupled with native Russians' evaluations of their questions. In this paper, the recognition rate of the utterances intended as questions will be studied and comparisons between the two recordings will be made. Secondly, the ratings of the acceptability of the stimuli as questions will be analyzed and compared between the speakers and the two recordings.

The motivation for comparing the two recordings comes from the question of retention in learning to produce a question. An earlier study (Ullakonoja 2008) showed that most Finnish students displayed development in their fluency between the recordings done prior to the stay in Russia and following it. However, it also indicated that there was a slight decrease in fluency for some students between recordings done during the stay in Russia and after it. Furthermore, Ullakonoja (2009) showed that speech rate of some students decreased between the two recordings. Hence, it may be suggested that the decline in the skills of some students can be due to the fact that they "lose" some of the skills that they had learned in Russia after returning home. The present study aims to compare the recordings to be able to determine degree of retention in learning to produce acceptable intonation in Russian yes/no questions.

In Finnish yes/no questions, prosody does not play a distinctive role. Instead, yes/no questions are produced by morphological means (the interrogative particle *-ko/-kö* is attached to the sentence initial finite verb in neutral word order). Finnish interrogative intonation is typically characterized by a high-initial pitch and an intensity contour that roughly follows the pitch contour (Hirvonen 1970). An intonational grammar does not exist as such in Finnish, and a typical pitch pattern for Finnish yes/no questions has not been authoritatively defined. A recent empirical study can, however, shed some more light on this issue. Anttila (2008:64) claims that in Finnish, the question type affects mostly the pitch distribution, not the shape of the contour. In Anttila's (2008:76-77) study the most typical pitch contours for yes/no questions in Finnish read-aloud speech were a fall or a rise-fall. According to Hirvonen (1970) yes/no questions can be characterized by a relatively high pitch before the nucleus. When Anttila (2008:79,82-83) used Hirvonen's categories to regroup her data she found that the most frequent pitch contour for yes/no questions was indeed this longer high pitch followed by a fall for both men and women for both read-aloud and spontaneous speech.

In Russian, a yes/no question differs from a corresponding declarative only by prosody. In written texts and when using a formal register – less common in everyday speech – it is possible to distinguish yes-no/questions also grammatically with an interrogative particle *li*. The Russian intonation research rests on Bryzgunova's (1977) description of Russian intonational constructions (IKs). According to this theory, there are seven different IKs, each of which has distinctive functions and uses. Each IK also has a typical or range of typical intonation patterns. As Bryzgunova's theory was essentially meant for teaching purposes and as it was mostly based on the auditory observations of the author it has recently evoked some criticism for its lack of empirical data (see e.g. Yokoyama 2001). However, it is still the most widely used theory on Russian intonation as only a few empirical studies exist to date. According to Bryzgunova (1977) yes/no questions are usually pronounced with IK-3. This

intonation pattern is characterized by a sharp rise on the nuclear syllable or, if the nucleus is in the phrase-final position, a final rising pitch contour (see also Svetožarova 1998; Volskaya 2009). Also a delayed peak seems to be characteristic to Russian yes/no questions (Igarashi 2006; Meyer & Mleinek 2006). As for Finnish, the traditional view is that final rising contours do not occur in interrogatives (Hirvonen 1970). However, some recent findings (Iivonen 1998; 2001; Ogden & Routarinne 2005; Anttila 2008:70-76) suggest that final rising pitch contours can also exist, even though their role is not only to signal an interrogative. They can also be used to indicate address, emotionality or continuation (Iivonen 1979; 2001; Mixdorff et al. 2002). Despite this, as Mixdorff et al. (2002) observed, the final rise in questions was perceived as fairly unnatural by native Finnish listeners.

So far, large-scale contrastive studies on Russian and Finnish intonation have not been conducted and studies on non-native Russian intonation are also scarce. The main difference seems to be that a Finnish declarative rarely differs from a question only by its pitch contour (see e.g. Iivonen 1979), whereas in Russian intonation has a distinctive function (Bryzgunova 1977, Svetožarova 1982). Because both the acoustic features and linguistic functions of Russian intonation are different from Finnish (de Silva & Ullakonoja 2009), Finnish students face a challenge in learning Russian intonation (de Silva & Volskaya 2005). Kuosmanen & de Silva (2003; 2007) found that Russian yes/no questions are difficult for Finnish learners to pronounce due to the differing pitch contours. They found that only 63% of the eight interrogatives produced by ten Finnish students were recognized as questions by native speakers of Russian. In contrast, Toivanen's (2001) study of Finnish university students of English showed that they were able to produce English short questions (for example *Agree?*, *OK?*) rather well with a final rising pitch contour, which seems to show that the difficulty is not purely related to L1 influence but also to other factors. The present paper partly replicates the studies by Kuosmanen & de Silva (2003; 2007). Here, however, more judges were used and one of the goals was to compare whether the students productions during versus following stay are different.

Native speakers have been found to ignore some acoustic markers in speech perception because they rely also on lexical information when listening to the intonation of the stimuli (Lieberman 1965). However, in Russian, there are cases in which native listeners have to disambiguate sentence types on the basis of prosodic cues only. Interestingly, Finnish native listeners (not knowing Russian) were shown to perceive most Russian interrogatives as emotional speech whereas Russian native listeners (not knowing Finnish) perceived Finnish interrogatives as declaratives (Shserbakova 2001).

## 2 Methods

### 2.1 Speakers

In collecting the data for the listening experiment, six female Finnish (L1) undergraduate students (aged 19-25) were recorded. They were majoring in Russian and had studied Russian as their third or fourth foreign language (L2) for three years prior to attending university as well as one year at university. During their second year at university they spent one semester (3.5 months) in Russia. Half of the students resided with a Russian host family and the rest in the dormitories for foreign students. When asked, half of the students reported they had practiced pronunciation

independently. All students had some contact with and thus the possibility of communication with native speakers during their stay in Russia.

## 2.2 Speech data

Students were asked to read aloud two Russian dialogues which were originally short texts designed as telephone conversations (dialogues 46 and 100). The texts were taken from Russian as a foreign language teaching materials (Shilova & Usmanova 1990). Seven yes/no questions (see Table 1) from the dialogues were used as stimuli in the listening experiment. These particular questions were selected from the dialogues, because they can be understood either as declaratives or questions depending on the pitch contour used by the speaker. According to Bryzgunova's (1977) classification of Russian intonational constructions (IKs), these questions would normally be produced with the same intonational construction (IK-3) by native speakers. Because the students saw the texts in their written dialogue form, it was not possible for them to interpret the questions as declaratives.

	Russian	Russian (in Roman alphabet)	English translation
Q1	У тебя совесть есть?	U tebya sovest' yest' <i>prep you conscience have</i>	Have you no conscience?
Q2	Соня?	Sonya? <i>sonya (proper name)</i>	Is that Sonya?
Q3	Чайный или столовый?	Chainyy ili stolovyy?  <i>tea or dinner</i>	Tea or dinner service?
Q4	Ты рада за меня?	Ty rada za menya?  <i>you happy for me</i>	Are you happy for me?
Q5	Да?	Da? <i>yes</i>	Yes?
Q6	Ты заболела?	Ty zaboleta? <i>you be-sick</i>	Were you taken ill?
Q7	Судя по торжественному тону, ты хочешь сообщить мне нечто важное?	Sudya po torzhestvennomu tonu, <i>judging prep festive tone</i> ty khochesh' soobshchit' mne <i>you want inform me</i> nechto vazhnoye? <i>something important</i>	Judging by your festive tone of voice, you want to tell me something important?

Table 1: The utterances used in the experiment.

## 2.3 Recordings

The speech data were recorded at two recording sessions: 1) during the students' stay in Russia (referred to as T1) and 2) following their return to Finland (T2). The interval between the two recordings was approximately three months. The T1 recordings were conducted in the middle of the students' stay in Russia, whilst the T2 recordings were done at the beginning of the spring semester about a month after their return to Finland. Most of the students had not spoken Russian since returning home. Different recording equipment was used for practical reasons. In T1, the recordings were done

with a Sony TCD-D3 DAT-recorder or a Roland Edirol 24-bit Wave/MP3 R-09 digital recorder with a Sony ECM-959A microphone in a hotel room. In T2 the students were recorded using a computer equipped with the program Adobe Audition 1.0 and 2.0, and AKG GN30 microphones. The students were recorded in pairs because the texts used were dialogues. The dialogues were chosen in order to give a more realistic setting for the study.

## 2.4 Listening experiment

The listening experiment was conducted in two parts: T1 and T2 recordings separately. Hence there were two groups of listeners, one listening to the samples of the T1 recordings and the other listening to the T2 recordings. This was done to avoid too lengthy a task, as there were a total of 84 stimuli to be rated. The stimuli were presented in the same randomized order with approximately an 11-second interval between the stimuli. Prior to each stimulus listeners heard a sound indicating the start of the next stimulus as well as its number in Russian. The total duration of the task was approximately 11 minutes for each group. The listening experiment was prepared in the computer program Praat (Boersma & Weenik 2009) from the recorded sound files.

The judges were asked to define: (1) whether they perceived the stimulus a question or not (categorization, 1=non question, 2=question) and (2) whether the stimulus was successful as a question (acceptability rating, 1-5 scale: 1=not a question, 5=a question). The judges were told that the speakers intended all the utterances as questions. This was done to avoid forcing them to rate some stimuli as declaratives. If a more traditional setting for the listening task had been used (not informing the judges of the speaker's intent), the stimuli should also have included utterances intended as statements by the speakers. This however, would have resulted in a listening task perhaps too lengthy for the listener's concentration span.

## 2.5 Judges

Judges were either students or staff members at the philological faculty of St. Petersburg State University. There were a total of 40 listeners (19 in the first group, 21 in the second), all native speakers of Russian. Most of them were under 20 years old (80%), female (90%), students (92%). About half of them (48%) reported that they were used to hearing foreigners speak Russian. The amount of their exposure to non-native Russian varied from everyday (7.5%) to once a week (25%), once a month (15%) and rarer than once a month (52.5%). Only three judges had ever taught Russian to foreigners.

Not all listeners rated all the stimuli in all aspects: hence, there were some missing values that were excluded from the analysis. Cohen's Kappa for the categorization task was 0.563 in T1 and 0.588 in T2 and Cronbach's alpha for the acceptability ratings yielded 0.960 in T1 and 0.858 in T2. For the categorization task, following the principles of Landis & Koch (1977), it can be concluded that the values of Cohen's Kappa indicate moderate interjudge agreement. Cronbach's alpha, on the other hand, shows the interjudge consistency for the acceptability ratings. As for T1 and T2 in this study Cronbach's alpha yielded over 0.8, it can be considered good (see e.g. Bryman & Cramer 2001:62).

### 3 Results

#### 3.1 Question categorization

First, I will discuss the results from the listening experiment from the point of view of question categorization. For this, the recognition rate (i.e. the proportion of the positive “yes” ratings) was calculated. When computing the recognition rates, missing values were excluded from the analysis. When looking at all the questions produced by all the speakers without distinguishing between the two recordings, it was found that the overall recognition rate was 57%. In other words, just over a half of the utterances intended as questions by Finns were recognized as such by native speakers of Russian.

Figure 1 depicts the mean recognition rate of the individual questions. As mentioned above, every utterance was pronounced twice by each of the six students. The utterance that was recognized the best as a question was Q4 (*Ty rada za menya?*). The overall recognition rate was not very high: questions Q1, Q2 and Q7 have a recognition rate of below 50%. Hence, the utterances can be grouped into two categories according to their recognition rate: (1) those mostly understood as questions with a recognition rate between 56% and 99%, i.e. Q3, Q4, Q5 and Q6 (henceforth S=successful questions) and (2) those mostly recognized as non-questions with a recognition rate between 29% and 38%, i.e. Q1, Q2 and Q7 (henceforth NS=non-successful questions).

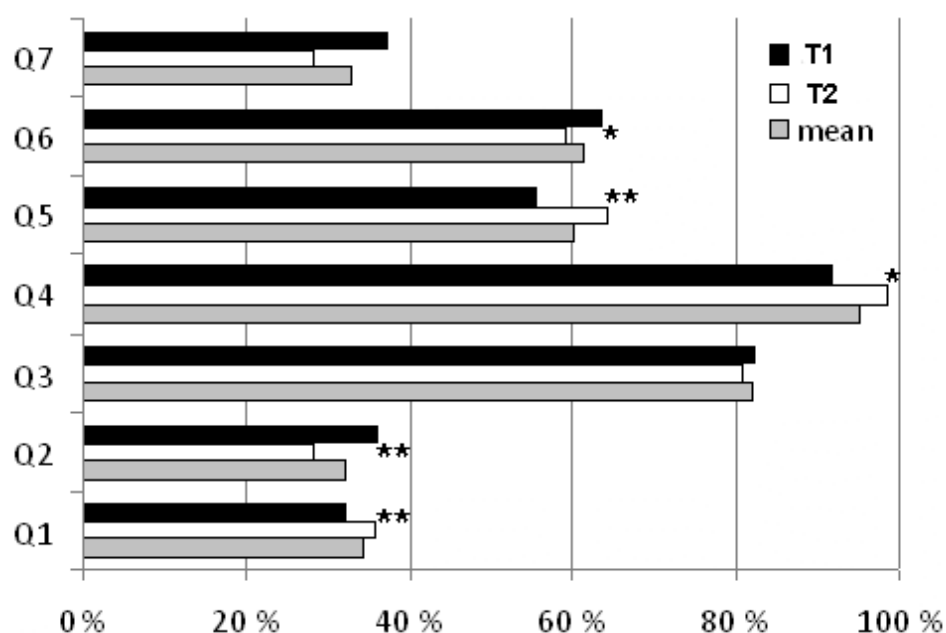


Figure 1: Native speakers' recognition rate of questions (n=84) produced by Finns during the stay (T1) and following it (T2) (\*  $p < 0.05$ , \*\*  $p > 0.0001$ ).

Moreover, Figure 1 compares the recognition rates in T1 and T2. The general recognition rate was slightly lower in T2 (56.4%) than in T1 (57.2%). The difference was statistically significant ( $\chi^2(1)=153.566$ ;  $p < 0.0001$ ). When looking at the individual questions, three out of seven questions were recognized better in T2 than T1. The difference was the greatest in the shortest question Q5 (*Da?*), which had 6% better recognition rate in T2 than T1. The statistical significance of the differences between

T1 and T2 was tested in the Chi-Square test, which indicated that the majority of the differences between T1 and T2 were significant. In three questions (Q1, Q4 and Q5) there was significant improvement, whereas for Q2 and Q6 there was decline. While most questions showed a statistically significant difference between students' production in T1 and T2 as perceived by native speakers, there is no general tendency whether the difference is positive or negative between T1 and T2.

In Table 2 the recognition rate of the questions is considered from the point of view of the individual learners. Great interspeaker variation was observed. It is useful to examine the data by separating questions that were mostly rated successful questions (S) and those that were mostly rated unsuccessful (NS). Table 2 indicates that despite the fact that NS questions were generally recognized rather weakly, there were individual students (like Fi1 and Fi3 in Q1 and Q7 and Fi4 in Q2) who received rather high recognition rates. Both Q3 (S) and Q4 (S) had a high recognition rate for all speakers in both T1 and T2. Greater variation can be seen in Q5 (S) and Q6 (S).

Speaker	Question and category						
	Q1	Q2	Q3	Q4	Q5	Q6	Q7
	(NS)	(NS)	(S)	(S)	(S)	(S)	(NS)
Fi1 (T1)	90%	21%	68%	<b>84%</b>	95%	100%	<b>68%</b>
Fi1 (T2)	50%	0%	68%	<b>100%</b>	95%	86%	<b>86%</b>
Fi2 (T1)	<b>16%</b>	11%	84%	100%	11%	<b>78%</b>	89%
Fi2 (T2)	<b>24%</b>	11%	55%	100%	0%	<b>95%</b>	0%
Fi3 (T1)	<b>74%</b>	<b>21%</b>	<b>90%</b>	<b>83%</b>	<b>33%</b>	95%	<b>53%</b>
Fi3 (T2)	<b>86%</b>	<b>43%</b>	<b>91%</b>	<b>90%</b>	<b>90%</b>	55%	<b>62%</b>
Fi4 (T1)	5%	<b>95%</b>	<b>90%</b>	<b>90%</b>	<b>95%</b>	<b>16%</b>	5%
Fi4 (T2)	5%	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>57%</b>	5%
Fi5 (T1)	5%	44%	<b>84%</b>	100%	95%	79%	<b>0%</b>
Fi5 (T2)	<b>10%</b>	14%	<b>90%</b>	100%	75%	57%	<b>10%</b>
Fi6 (T1)	5%	26%	<b>79%</b>	<b>95%</b>	5%	16%	11%
Fi6 (T2)	<b>43%</b>	5%	<b>81%</b>	<b>100%</b>	<b>25%</b>	5%	5%

Table 2. The recognition rate of questions of the individual speakers during the stay in Russia (T1) and following it (T2). The figures in bold indicate a higher recognition rate in T2 than T1. (NS=non successful questions, S=successful questions as perceived by native speakers).

When the development of individual speakers is compared, the student who scored a higher rate in most of the questions between T1 and T2 is Fi3. There is only one question in which she received a lower recognition rate in T2. There is also less variation in her utterances in T2 compared to T1, where the range was 43%-91%. Another successful speaker in T2 is Fi4, because four out of seven of her questions in T2 were recognized as an interrogative by all of the judges. There is, however, great variation in the speech of this subject: two of her intended questions were recognized as an interrogative by less than 6% of the judges. In fact, variation is typical for all speakers' productions: no single speaker achieved a high recognition rate in all questions, nor one did always have a low recognition rate. Hence, the results suggest that the target intonational construction IK-3 is difficult to learn to produce in all contexts. As mentioned above, according to the literature it is realized in at least two different kinds of pitch contours depending on the place of the nucleus. The students may struggle with the nucleus placement which would then lead to a contour choice, not perceived as interrogative by the listeners.



As Table 2 shows, only four out of 42 of the utterances intended as questions were *not* recognized as interrogatives by any of the judges, and 14 of the 42 utterances fall within the lowest 5% of recognition. It is also interesting to examine the utterances receiving 100% (or near) recognition rates. 11 out of 42 of the stimuli were interpreted as questions by all judges. Seven of these concern Q4 (*Ty rada za menya?*) which, as was shown in Figure 1, garnered the highest recognition rate for all speakers. It is also worthwhile to point out that over a half of the stimuli (27 out of 42) were recognized as questions by at least 90% of the listeners.

In Table 2 the comparison between the two recordings indicates that in about half (22 out of 42) of the questions the recognition rate is higher in T2 than T1, while in 16% (seven out of 42) it is the same and in 30% (13 out of 42) it is lower. The results would seem to suggest interspeaker differences. Some speakers have demonstrated improvement in their ability to pronounce questions during the last part of their stay in Russia, whereas for other speakers there was not such a significant change. As mentioned above speakers Fi3 and Fi4 seem to have benefited most from the study abroad: six out of seven stimuli of Fi3 have a better recognition rate in T2 than T1 and five out of seven stimuli of Fi4. The proportion of more successfully conveyed questions in T2 than T1 for Fi6 is about a half (four out of seven) but for Fi5 three out of seven and Fi1 and Fi2 only two out of seven. However, as in only 30% of the cases the recognition rate is lower in T2 than in T1, the students' performances show improvement and retention in learning. Hence, Table 2 clearly shows the complexity of the data, whereas Figure 1 simplifies it somewhat.

This raises the question as to why, then, are some questions (S) produced significantly more successfully than others (NS)? There are various factors that affect the success of the productions. For example, the possible explanations can be found in the syntax and the frequency of use of these constructions. Also fluency can offer an explanation; for instance, if a speaker struggles reading aloud a sentence (i.e. hesitates and pauses), it is difficult to produce acceptable sentence prosody.

The unsuccessfully produced questions in NS are in fact rather different from each other. For example, Q1 (*U tebya sovest' yest'*) is an idiomatic expression and resembles a rhetorical question. Another possible word order for a *yest'*-question would be *u tebya yest' sovest'*. This is more neutral and therefore might be used more often (Lobanova & Gorbachik 1976:6). It is possible students pronounced a pitch contour according to this word order which brought forward the nucleus. Perhaps this was then interpreted by the native speakers as a non-question, because the nucleus was on *sovest'* instead of *yest'*. To clarify, in order to be recognized as a question the nucleus should be on *yest'* despite the word order. The reasons for unsuccessful production of Q2 remain vague. The context *Allo! Sonya?* (Hello, is that Sonya?) clearly indicates an interrogative. However, as the line was in the very beginning of the second dialogue, the students were perhaps a bit out of touch or were merely using a L1 pitch contour. Using a L1 pitch contour in L2 could of course be an explanation for all the unsuccessful productions by the L2 speakers. Further analysis of the pitch contours will shed more light into this issue (Ullakonoja 2010). The unsuccessful production of Q7, on the other hand, can be also explained by its structure and lexicon: it is a very long question with some words that may be unfamiliar which may lead to the students focusing on the words rather than sentence prosody.

As mentioned above, the S questions were generally recognized well. Questions Q3 (*Chainyy ili stolovyy?*) and Q4 (*Ty rada za menya?*) are most likely to be interpreted as questions in any context perhaps due to their lexical content. Question Q5 (*Da?*) on the other hand is very frequent in everyday Russian, as a result students would have heard it regularly while in Russia. Utterance Q6 (*Ty zaboylela?*) is a short

question with a rather simple lexical content and syntax, which perhaps facilitated its production.

### 3 Results

#### 3.1 Acceptability ratings

Next, I will focus on the second goal of the listening experiment, i.e. determining how good the judges thought each stimulus was as a question. The acceptability rating of the question was investigated by examining the ratings of each question and each student at the two recordings on a scale of one to five (1=not a question, 5=a question). First, Figure 2 shows that in general the mean acceptability ratings were not very high. The overall mean of all questions was 2.95.

It is somewhat unexpected, when the results of Figure 2 are compared with Figure 1 that only question Q4 (*Ty rada za menya?*) reaches a mean acceptability of over four, while others are on average either between three and four or around two, which could be verbalized as “not very acceptable as a question”. From the results presented in Figure 1, one could anticipate that Q4, which had the highest recognition rate, would have received an acceptability rating of near five.

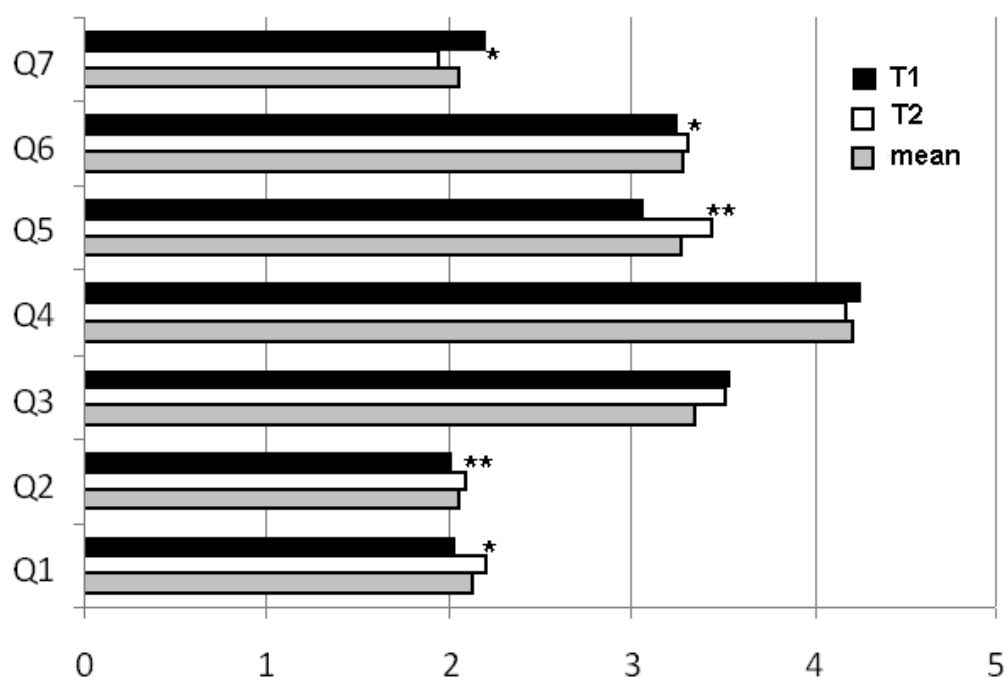


Figure 2: The mean acceptability rating of questions (n=84) during the stay in Russia (T1) and after it (T2) (0=not a question, 5=a question), \*)  $p < 0.05$ , \*\*)  $p > 0.0001$ .

The mean acceptability ratings (Figure 2) show a very similar pattern as the results from recognition rate measurement (Figure 1). The grouping of utterances into S and NS seems justifiable also in the acceptability ratings. In Figure 2 there are differences between T1 and T2 in acceptability ratings. The overall mean was slightly though reliably higher in T2 (2.99) than T1 (2.92),  $\chi^2(16)=204.970$ ;  $p < 0.0001$ . When the comparison of the acceptability ratings given by the judges for each question in the two recordings was further analyzed in Pearson’s Chi Square test, it was found that for Q1, Q2, Q5 and Q6 the acceptability rating was reliably better in T2 than in T1.

Conversely, for Q7 it was significantly lower. Hence, the acceptability ratings show that the students were evaluated in general as producing more acceptable yes/no questions following their stay in Russia.

Table 3 presents the acceptability ratings for each question by individual speakers. Comparison of the two recordings yielded an unexpected result. The mean acceptability rating was lower for about half (23 out of 42) of the utterances in T2 than T1. However, for slightly less than a half (19 out of 42) it was higher. The comparison between the ratings of different questions shows, not surprisingly, that there is a tendency for some questions e.g. Q4 (*Ty rada za menya?*) to be rated as a fairly successful (3.6-4.8) question for all speakers. In other questions, e.g. Q2 (*Sonya?*) (1.1-4.3) and Q6 (*Ty zabolela?*) (1.3-4.7), there is more interspeaker variation, or variation between the two recordings. Thus, it cannot be concluded whether utterances other than Q4 would always be rated consistently.

Speaker	Question and category						
	Q1	Q2	Q3	Q4	Q5	Q6	Q7
	(NS)	(NS)	(S)	(S)	(S)	(S)	(NS)
Fi1 (T1)	3.8	1.7	3.2	<b>3.7</b>	4.6	4.5	3.2
Fi1 (T2)	2.5	1.1	2.7	<b>4.1</b>	4.5	4.3	3.4
Fi2 (T1)	<b>1.5</b>	<b>1.1</b>	3.8	4.5	1.1	<b>3.6</b>	3.6
Fi2 (T2)	<b>1.6</b>	<b>1.4</b>	2.4	4.1	1.1	<b>3.9</b>	1.1
Fi3 (T1)	<b>2.9</b>	<b>1.8</b>	<b>3.8</b>	4.3	<b>2.1</b>	4.7	2.5
Fi3 (T2)	<b>3.7</b>	<b>2.5</b>	<b>4.1</b>	4.0	<b>4.7</b>	3.7	2.4
Fi4 (T1)	1.6	<b>3.8</b>	<b>3.6</b>	<b>4.4</b>	4.7	<b>1.4</b>	<b>1.4</b>
Fi4 (T2)	1.5	<b>4.3</b>	<b>4.4</b>	<b>4.7</b>	4.5	<b>3.1</b>	<b>1.5</b>
Fi5 (T1)	<b>1.1</b>	2.6	<b>3.8</b>	<b>4.2</b>	4.3	3.8	<b>1.2</b>
Fi5 (T2)	<b>1.2</b>	1.8	<b>4.3</b>	<b>4.8</b>	3.6	3.2	<b>1.4</b>
Fi6 (T1)	<b>1.3</b>	1.1	3.2	4.5	<b>1.4</b>	1.3	<b>1.2</b>
Fi6 (T2)	<b>2.4</b>	1.1	3.1	3.6	<b>1.7</b>	1.3	<b>1.5</b>

Table 3. The mean acceptability rating of questions of the individual speakers during the stay in Russia (T1) and following it (T2). The figures in bold indicate a higher recognition rate in T2 than T1. (NS=non successful questions, S=successful questions as perceived by native speakers.)

As Table 3 shows, utterances that were, in general, recognized poorly as questions (NS) were still produced successfully by some speakers. For example, Q1 (*U tebya sovest' yest'*) is rated fairly good as a question for Fi1 (T1) and Fi3 (T2) and Q2 (*Sonya?*) for Fi4 (T1 and T2). When comparing different speakers, there is one student who received overall good ratings (Fi3, T2) and one who received consistently poor ratings (Fi6). As all learners succeeded in producing some of the questions well, this seems to show that they in principle knew how to produce the pitch contour of a yes/no question in Russian, but did not always succeed in doing so.

In Table 3 the improvement between T1 and T2 is indicated in bold. Improvement is observed in about half of the (21 out of 42) cases. In three out of 42 cases there is no difference between T1 and T2. The comparison of Table 2 and Table 3 indicates that the acceptability ratings sometimes show a different result than recognition rate. If examining the improvement, there are nine out of 42 cases where there is a difference in acceptability ratings compared to the recognition rate.

### 3.2 Reliability of the ratings

The results seem to suggest that there is variation in the ratings between the judges. In fact, the judges were rarely completely unanimous in their judgments. Despite this, as the values of Cohen's Kappa and Cronbach's alpha showed, the interjudge consistency was rather good. To further verify the ratings made by the judges, the categorization of the question and its acceptability rating was subjected to comparison. Table 4 presents the results obtained from the comparison of question categorization and acceptability rating of the interrogative without taking into the account the time of recordings (T1, T2). From this table it becomes apparent that the relationship between the two ratings is statistically significant. This indicates that the judges gave similar acceptability ratings to the questions whilst giving different ratings to the non-questions. Hence, it can be concluded that the ratings by the judges are reliable.

Speaker	Pearson's Chi Square correlation	df	p
Fi1	93.820	4	<0.0001
Fi2	87.136	4	<0.0001
Fi3	106.192	4	<0.0001
Fi4	108.742	4	<0.0001
Fi5	99.215	4	<0.0001
Fi6	70.813	4	<0.0001

Table 4: Pearson's Chi Square correlation between the question categorization and acceptability rating of the interrogative.

## 4 Discussion and conclusions

It has been shown that native speakers can read aloud the same written text with different pitch contours (Brazil 1984). It would therefore be incorrect to presume that even native speakers of Russian would always pronounce the sentences in exactly the same manner. They are however all prone to using an interrogative pitch contour, whereas Finnish learners have been shown to struggle with this, possibly because of the fact that their L1 lacks such contours.

The results of this study show that a great number of utterances intended as questions by Finns were not perceived as such by native speakers. This leads to a number of conclusions. As Hirvonen (1967:42) suggests, one explanation for this might be in the different approaches the individual judges took in accepting intonation that differs from native production. Some judges are perhaps more ready to accept non-native production whereas others are not. It should be remembered that in this study the majority of the judges were not accustomed to hearing foreign-accented Russian, which may have resulted in strict ratings. Furthermore, as the students were also speaking slower than native speakers (see Ullakonoja 2009) the judges could have been disturbed by the inappropriate temporal structure of the pitch contour. As Russian yes/no questions are spoken faster than declaratives (Svetozarova 1982:111-112), the judges could have favored the "non-question" rating in some cases where the students were speaking slowly. The variation between T1 and T2 could also reflect the difference in strictness of the two groups of judges, not only differences in the learners' productions.

Kuosmanen & de Silva (2003; 2007) and de Silva & Volskaya (2005) have shown, that Russian interrogatives are difficult for Finnish speakers. These findings are

further substantiated by this study: only half of the utterances intended as questions were recognized by 90% of the judges, and the acceptability ratings remained at an average level. The overall recognition rate of all recordings was 57%, slightly lower than in Kuosmanen & de Silva (2003; 2007). Hence, it can be concluded that even reasonably proficient Finns often fail to produce acceptable pitch contours in questions in Russian.

The results from the overall recognition rate (Figure 1) of individual questions, however, are not consistent with the earlier studies by Kuosmanen & de Silva (2003; 2007), who found that longer Russian questions produced by Finns were harder to recognize by native speakers than shorter questions. In this study, the long question (Q7, *Sudya po torzhestvennomu tonu, ty khochesh' soobshchit' mne nehto vazhnoye?*) was not the hardest to recognize, nor was the shortest question (Q5, *Da?*) the easiest. The contradictory results can partly be explained by the fact that in the present study the panel of judges consisted of a greater number of participants.

The general recognition rate of the questions was only slightly lower for items recorded in T2 than T1. The difference was small, but statistically significant. Therefore, it seems that in general the learners are not as good at producing yes/no/questions following their stay in Russia compared to during it. However, even after a month in Finland (T2), with hardly using or hearing any Russian, they are capable of reading the yes/no questions in the text not as well as but almost as successfully as during their stay in Russia (T1). Some explanations for the unsuccessful productions were offered on the basis of syntax and lexicon; however, a further acoustical analysis of the pitch contours is conducted for more evidence (see Ullakonoja 2010). The acceptability ratings yielded similar results as those obtained from the recognition rate analysis. There was great interspeaker and intraspeaker variation. The contradictory result is that for the acceptability ratings, a statistically significant improvement was observed from T1 to T2. This could suggest that there is in fact no “loss” but retention in learning.

To conclude, the findings of this study show that only one of the utterances (Q4, *Ty rada za menya?*) was consistently judged as a question. The other items displayed no general tendency either in the categorization task or the acceptability ratings. Great interspeaker differences were also found. The almost unanimous judgments of Q4 as a question can partly be explained by its lexical content as the utterance would likely be used more often as a question than a declarative in daily conversation.

To conclude, I would like to mention an interesting finding by Kuosmanen & de Silva (2003). They found that in the Russian yes/no question *Mozhno?*, Finnish students who used an incorrect final-rise instead of a correct rising-falling contour in the nuclear syllable were more likely to be recognized as pronouncing a question than those who used the correct contour. Thus, in the future, it is also important to study the pitch contours of the students' questions experimentally in order to determine what changes in F0 (and where) function as important cues to perceiving a question. It also remains to be explored how the differences between the realization of pitch contours in yes/no questions in Finnish and Russian affect learners' production, i.e. are the learners, for example, relying on L1 (or other L2s) when learning Russian L2 prosody.

## References

- Anttila, H. 2008. The effect of interrogative function on intonation in spontaneous and read Finnish. Master's thesis. Department of speech sciences, University of Helsinki. Available at: <http://urn.fi/URN:NBN:fi-fe200807301740>
- Boersma, P. & Weenik D. 2009. *Praat: doing phonetics by computer*. (Versions 5.0-5.1) [computer program] Available at: [www.praat.org](http://www.praat.org)
- Brazil, D. 1984. *The intonation of sentences read aloud*. In: D. Gibbon & H. Richter (eds.). *Intonation, accent and rhythm*. Berlin: Walter de Gruyter. 46-66.
- Bryman, A. & Cramer, D. 2001. *Quantitative Data Analysis with SPSS Release 10 for Windows: A Guide for Social Scientists*. Great Britain: Routledge.
- Bryzgunova, E. A. 1977. *Zvuki i intonatsiya russkoy rechi*. Moscow: Russkiy yazyk.
- de Silva, V. & Ullakonoja, R. 2009. *Introduction: Russian and Finnish in contact*. In: V. de Silva & R. Ullakonoja (eds.). *Phonetics of Russian and Finnish. General description of phonetic systems. Experimental studies on spontaneous and read-aloud speech*. Frankfurt am Main: Peter Lang. 15-20.
- de Silva, V. & Volskaya, N. B. 2005. *Mesto fonetiki v standartakh RKI i v praktike vladeniya yazykom*. In: N. P. Isayev, A. Mustajoki & E. Protasova (eds.). *Russkiy yazyk kak inostranny: Problemy izucheniya, prepodavaniya, otsenki v svete obshcheyevropeyskikh kompetentsiy vladeniya inostrannym yazykom*. Moscow: MAKSS Press. 61-73.
- Hirvonen, P. 1967. *On the problems met by Finnish students in learning the rising interrogative intonation of English*. Publications of the Phonetics department of the University of Turku 2.
- Hirvonen, P. 1970. *Finnish and English communicative intonation*. Publications of the Phonetics department of the University of Turku 8.
- Igarashi, Y. 2006. *Intonational patterns in Russian interrogatives - phonetic analyses and phonological interpretations*. In: Y. Kawaguchi, I. Fónagy & T. Moriguchi (eds.). *Prosody and syntax: Cross-linguistic perspectives*. 175-196. Amsterdam: John Benjamins.
- Iivonen, A. 1979. *Is there interrogative intonation in Finnish?* In: E. Gårding, G. Bruce & R. Bannert (eds.). *Nordic prosody, papers from a symposium*. Lund: Department of Linguistics, Lund University. 43-53.
- Iivonen, A. 1998. *Intonation in Finnish*. In: D. Hirst & A. Di Cristo (eds.). *Intonation systems – a survey of twenty languages*. Cambridge: Cambridge University Press. 311-327.
- Iivonen, A. 2001. *Intonation of Finnish questions*. W. A. van Dommelen & T. Fretheim (eds.) *Nordic prosody: Proceedings of the VIIth conference Trondheim 2000*. Frankfurt am Main: Peter Lang. 137-151.
- Kuosmanen, A. & de Silva, V. 2003. *Why don't Russians answer my questions? Finnish students' problems in producing Russian interrogative intonation*. In: M. J. Solé, D. Recasens & J. Romero (eds.). *Proceedings of the 15th International Congress of Phonetic Sciences: Barcelona 3-9 August 2003*. 523-526.
- Kuosmanen, A. & de Silva, V. 2007. *Problemy finskih studentov pri ovladenii russkoy voprositel'noy intonatsiyey*. In: R. F. Kasatkina (ed.). *Problemy fonetiki V*. Moscow: Nauka. 297-307.
- Landis, J. R. & Koch, G. G. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33(1). 159-174.
- Lieberman, P. 1965. On the acoustic basis of the perception of intonation by linguists. *Word* 21. 40-54.
- Lobanova, N. A. & Gorbachik A. L. 1976. *Poryadok slov v russkom yazyke*. Moscow: Izdatel'stvo moskovskogo universiteta.
- Meyer, R. & Mleinek, I. 2006. How prosody signals force and focus - A study of pitch accents in Russian yes-no questions. *Journal of Pragmatics* 38(10). 1615-1635.
- Mixdorff, H., Vainio, M., Werner, S. & Järvikivi, J. 2002. *The manifestation of linguistic information in prosodic features of Finnish*. In: B. Bel & I. Marlien (eds.). *Proceedings of the Speech Prosody 2002, Aix-en-Provence 11-13 April 2002*. 515-518.
- Ogden, R. & Routarinne, S. 2005. The communicative functions of final rises in Finnish intonation. *Phonetica* 62. 160-175.
- Shilova, K. A. & Usmanova, E. E. 1990. *100 dialogov po telefonu*. Moscow: Russkiy yazyk.
- Shcherbakova, L. P. 2001. *Vospriyatie intonatsionnykh tipov nerodnogo yazyka*. In: L. V. Bondarko (ed.). *100 let eksperimental'noy fonetike v Rossii. Materialy mezhdunarodnoy konferentsii 1-4 fevralya 2001 g*. St. Petersburg: Filologicheskii fakul'tet SPbGU. 173-175.
- Svetozarova, N. D. 1982. *Intonatsionnaya sistema russkogo yazyka*. Leningrad: Izdatel'stvo Leningradskogo universiteta.

- Svetozarova, N. D. 1998. *Intonation in Russian*. In: D. Hirst & A. Di Cristo (eds.). *Intonation systems – a survey of twenty languages*. Cambridge: Cambridge University Press. 261-274.
- Toivanen, J. H. 2001. *Perspectives on intonation: English, Finnish and English spoken by Finns*. Frankfurt am Main: Peter Lang.
- Ullakonoja, R. 2008. Pausing as an indicator of fluency in the Russian of Finnish learners. In: P. A. Barbosa, S. Madureira & C. Reis (eds.). *Proceedings of the Speech Prosody 2008 conference*. Campinas, Brazil: Editora RG/CNPq. 339-342.
- Ullakonoja, R. 2009. Speech rate as an indicator of fluency in the Russian of Finnish learners. In: M. O'Dell & T. Nieminen (eds.). *Fonetiikan päivät 2008 – the Phonetics symposium 2008*. Tampere studies in language, translation and culture, series B. 97-109. Available at: <http://tampub.uta.fi/tup/978-951-44-7580-1.pdf>.
- Ullakonoja, R. 2010. Pitch contours in Russian yes/no questions by Finns. In: M. Hasegawa-Johnson, A. Bradlow, J. Cole, K. Livescu, J. Pierrehumbert & C. Shih (eds.) *Proceedings of the Speech Prosody 2010 conference*. Available at: <http://speechprosody2010.illinois.edu/papers/100072.pdf>
- Vaissière, J. 2005. *Perception of intonation*. In: D. B. Pisoni & R. E. Remez (eds.). *The handbook of speech perception*. Malden, MA; Oxford, Carlton: Blackwell publishing. 236-262.
- Volskaya, N. B. 2009. *Aspects of Russian intonation*. In: V. de Silva & R. Ullakonoja (eds.). *Phonetics of Russian and Finnish. General introduction. Spontaneous and read-aloud speech*. Peter Lang. 133-143.
- Yokoyama, O. 2001. A reinterpretation of the IK system in Russian. *Die Welt Der Slaven* XLVI. 1-26.