

PAPER

Acoustic cues of surprise in Russian questions

Veronika Makarova

*Electrotechnical Laboratory,
1-1-4, Umezono, Tsukuba, 305-8568 Japan ;
Meikai University,
8, Meikai, Urayasu, 279-0014 Japan*

(Received 24 September 1999)

This study investigates the production of neutral (unemotional) and surprised (emotional) 'yes/no' questions in spoken Russian. The paper compares acoustic parameters across these two questions types in three-syllable one-word sentences with the terminal rise-fall tone. The study takes account of traditional intonation parameters as well as pitch event alignment data. The study demonstrates the existence of statistically significant differences between the following parameters of neutral and emotional questions: total word (=sentence) duration, pitch peak height, characteristics of the accented syllable (duration, pitch heights at the onset and offset of the stressed vowel, intervals of pitch movements). The study suggests the existence of language-specific acoustic cues in the expression of emotions. The study also shows that despite the existence of speaker-independent ways of expressing surprise, there is also a considerable degree of intra-speaker variability. The results have direct implications for speech processing and communication systems, natural language processing, speech synthesis and recognition.

Keywords: Speech production, Prosody, Intonation, NLP, Speech synthesis and recognition

PACS number: 43.70.+i

1. INTRODUCTION

1.1 The State of the Problem

The development of modern multilingual communication and machine translation systems requires adequate processing and decoding of all the semantic layers of a human message including the emotive meaning which is largely rendered by intonation. It is known that various emotional states are acoustically signalled by changes in the mean value, range, contour type of F_0 and voice quality as well as by changes in the mean value of intensity.^{1,2)}

One of the current goals of speech technologies is to 'model natural speech on a global level'.³⁾ It is now becoming generally understood that attitudes and emotions are such an intrinsic part of human speech that they have to be accounted for in speech synthesis, since without them the quality of synthe-

sized speech is perceived as poor.⁴⁾ Processing and modelling of attitudes and emotions has been therefore addressed in a number of speech synthesis and text-to-speech systems.^{3,5,6)}

The automatic recognition, synthesis and modelling of emotional states cannot be accomplished without a clear understanding of the acoustic correlates of emotions in real speech data. The current state of the description is insufficient since only a few emotional states have been addressed and the scope of languages included into investigation has been very limited. For example, there are studies of emotional intonation in English^{7,8)}; German^{9,10)}; Japanese,¹¹⁾ and a few other languages. Some fundamental theoretical issues also remain unclear. Emotive contrasts rendered by intonation are described either within the 'covariance view' according to which intonation parameters change along grad-

ual rather than categorical continuum, or within the 'configurational view' whereby different emotions are associated with distinct contour configurations or tones.¹²⁾

Both the range of languages and the amount of real speech data have to be increased for a better understanding of the functioning of emotive intonation across languages.

This paper addresses acoustic correlates of surprise in Russian 'yes/no' questions.

Very little information has been available so far on the acoustic clues of expressing emotions in Russian. The most famous classification of Russian intonation by Bryzgunova¹³⁾ does not consider emotional variants of F_0 contours. Emotional variants of contours are only discussed very briefly in some works related mostly to grammatical functions of Russian intonation.¹⁴⁾ A paper by Holden and Hogan¹⁵⁾ compares the perception of emotive contours by Russians and Americans, but does not address the actual acoustic parameters of Russian emotional contours versus neutral contours. To the best of my knowledge, no specific studies of surprised and neutral 'yes/no' questions in Russian have been performed. The current paper is aimed to contribute to the description of acoustic parameters of emotional variants of Russian intonation contours.

1.2 Approach

This study is based on the following theoretical assumption:

Emotions can be expressed by modifications both across and WITHIN intonation units (within the rise-fall accent in this study). It is possible to find acoustic correlates of the expression of emotions. Two kinds of intonation parameter variability related to emotions can be distinguished. The paralinguistic aspects of emotion expression (in particular, the voice quality) are not controlled by the speaker, they are physiologically determined and are presumably universal.^{16,17)} The linguistic aspects are controlled by the speakers, can be easily produced by untrained subjects in ordinary conversation and in laboratory conditions, and may be language-specific.

Although the actually measured parameter was F_0 (Hz), in accord with the listener's perception-oriented linguistic tradition which emphasizes auditory correlates of acoustic properties, in this paper I

shall use the term 'pitch' as a synonym to F_0 .

1.3 Rising-Falling Contour in Russian 'Yes/no' Questions

'Yes/no' questions in Russian bear no grammatical or lexical markers of sentence type, and are only distinguished from other sentence types by intonation. They have a terminal rise if the word bearing the phrasal accent is phrase-final, and the stress in it falls on the last syllable. If any unstressed post-accented syllables follow, the accented syllable has a rising pitch movement, whereas the post-accented syllables are pronounced with a fall. Whenever post-accented syllables are present, the pitch movement thus becomes a rise-fall.¹⁸⁾

2. MATERIALS AND METHODS

2.1 Material

One-word three-syllable sentences were selected as the material for current investigation. This limitation allowed me to concentrate on the subtle differences in the realization of nuclear accent between neutral and emotional contours by means of excluding all possible distinctions between the sentence types in the pre-nuclear part. A list of 22 Russian words was prepared by selecting the best matches satisfying the following criteria: a noun with the lexical stress on the penultimate syllable; balanced by the type of the stressed vowel (i, e, a, o), CVCVCV type and containing preferably sonorant consonants, to avoid interruptions in the F_0 contour. The number of words was determined by the above restrictions and the requirements of the statistical analysis. I shall further refer to the words' segments as c1, c2, c3 (consonants of the first, second and third syllables), and v1, v2, v3 (vowels of the first, second and third syllables). The list of words is given below in transliterated spelling followed by the phonemic transcription and English translation in italics; palatalized consonants are indicated by an apostrophe, e.g., /n'/. In all the words the second (penultimate) syllable bears the lexical stress.

List of one-word sentences employed in the experiment

1. Marina /mʌr'inʌ/, *female name*; 2. malina /mʌl'inʌ/, *raspberry*; 3. manera /mʌn'erʌ/, *manner*; 4. maraly /mʌrʌlʌ/, *moor*; 5. romany /rʌmʌnʌ/, *novels*; 6. galera /gʌl'erʌ/ 7. minory /m'inorʌ/, *minors*; 8. rulony /rulʌnʌ/, *rolls*; 9. volany /vʌlʌnʌ/, *volans*; 10. Galina /gʌl'inʌ/, *female name*; 11. binomy

/ b'inomʉ/, *binomials*; 12. limony / l'imonʉ/, *lemons*; 13. varany / vʌranʉ/, *giant lizards*; 14. lavina / lʌv'inʌ/, *avalanche*; 15. Elena / jel'enʌ/, *female name*; 16. Emelja / jemel'ʌ/, *male name*; 17. minëry / m'in'orʉ/, *minelayers*; 18. barany / baranʉ/, *ram*; 19. Dalila / dʌl'ilʌ/, *female name*; 20. nedelja / n'ed'el'ʌ/, *week*; 21. dolina / dʌl'inʌ/, *valley*; 22. banany / bʌnanʉ/, *bananas*.

2.2 Speakers and Method

Five female native Russian subjects aged between 32 and 43 participated in the experiment. The speakers were presented with the two lists consisting of the same one-word phrases, but with different punctuation ("?" for questions and "???" for surprised questions, in accordance with commonly used Russian punctuation). The speakers read the first list as a question, and then the second list as a surprised question. After practicing reading the lists, a recording was made, whereby the speakers read the lists once unless they made a mistake and a repetition was required.

The digital recording was made in the recording studio and analysed using ESPS Waves on a SUN workstation at the Department of Linguistics, the University of Edinburgh. The F_0 contours were obtained for all the data. The data were manually segmented. The following parameters were used for the analysis: segmental durations (ms), pitch height at the beginning and end of each segment (Hz), pitch height at the half of each vowel duration (Hz), pitch peak height (Hz), pitch peak alignment (ms from the stressed vowel offset), pitch minimum in the last syllable (Hz), alignment of pitch minimum (ms from the final vowel offset), pitch peak plateau duration (ms). Pitch peak alignment was measured in ms from the offset of the vowel in the penultimate stressed syllable as well as in relative units normalised regarding segmental durations. I define pitch peak plateau as a stretch of F_0 contour in proximity of the pitch peak with relatively stable pitch height, F_0 fluctuations not exceeding + or - 2 Hz in each successive pair of F_0 periods. It should be pointed out that such parameters as intensity and voice quality which are known to contribute to emotional distinctions^{2,19} have been deliberately left outside the present investigation since I wanted to fully concentrate on pitch contour characteristics.

The data were put on EXCEL 97 worksheet to obtain descriptive statistics. Parametric differences

between the contours were analyzed using a single-factor ANOVA (analysis of variance) whereby the null hypothesis was tested that the values for the mean of each intonation parameter (listed above) are significantly different across the two groups: neutral and emotional questions for every speaker separately. By-speaker ANOVA analysis is chosen since the limited prosodic material of the experiment (3 syllable sentences) allows no normalization of data, and consequently no direct multi-factor procedure could be justified.

3. RESULTS

Average parameters of the F_0 contours produced by the five speakers are represented in Fig. 1, panels a-e (In the graphs as well as in the text speakers are indicated by their first name initials: Speaker 1 — by V, Speaker 2 — by O, Speaker 3 — by L, Speaker 4 — by M, Speaker 5 — by S). The graphs demonstrate the existence of some distinctions between intonation parameters of neutral and emotional F_0 contours in the pronunciation of all the five speakers.

ANOVA comparison of intonation parameter means between neutral and emotional F_0 contours for each speaker is represented in Table 1.

3.1 Duration

Total sentence duration for all the five speakers is significantly longer in emotional than in neutral F_0 contours, as shown below in Table 2.

The increase in the total duration occurs mostly due to the extension of the accented vowel. The average accented vowel length in neutral and emotional questions is represented below in Table 3. In the graphs it can be seen as a considerable delay of the pitch peak (regarding the word onset) in the emotional questions by speakers V, O, L and M (Fig. 1, panels (a) through (d)). For speaker S the vowel lengthening in emotional questions is insignificant.

Other differences in duration are either insignificant or inconsistent across the speakers.

3.2 Pitch Height

The beginning of the F_0 contour (pitch height within the pre-accented syllable) has no significant differences across the two question types. However, the onset of the accented vowel is lower in emotional than in neutral questions (Ref. Table 4

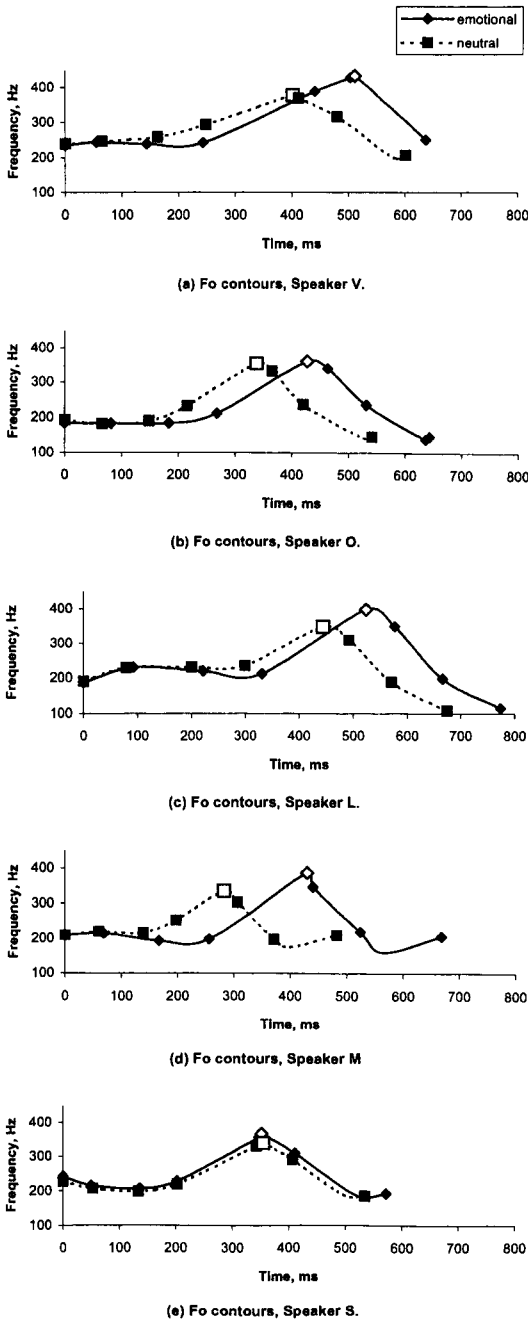


Fig. 1 Average F_0 and duration parameters of emotional and neutral interrogatives pronounced by the five speakers. (Point markers with black background indicate segmental borders, point markers with white background indicate pitch peak alignment.)

below), although the difference is insignificant for speaker S (Ref. Table 1).

Pitch peak and the offset of the accented vowel are higher in emotional than neutral contours pronounced by all the speakers (the difference is insignificant for speaker O). Pitch peak height distinctions are shown below in Table 5. No other speaker-consistent differences in pitch height were observed.

3.3 Pitch Intervals

In the pre-accented syllable pitch movement direction, interval and the difference between emotional and neutral contours vary across the speakers. Pitch rise starts in the second (accented) syllable and its interval is significantly greater in emotional than in neutral questions for all the speakers (Ref. Table 6). Pitch movement and its interval in the post-accented syllable differ across the speakers.

3.4 Alignment of Pitch Events

As can be seen from Fig. 1 (a-d), the rising movement in emotional questions is delayed as compared to the neutral questions: in the former the rise begins from the accented syllable onset, whereas in the latter — from the accented vowel onset. This delay manifests in the differences of the height of the accented vowel offset (reported above in Section 3.2). This difference along with the increased accented vowel duration (reported in Section 3.1) leads to the later absolute timing of the pitch peak in emotional questions regarding the word onset. However, no differences in the relative alignment of pitch peak regarding the accented vowel onset and offset have been observed: for all the speakers pitch peak in both interrogative types is located close to the accented vowel offset (with the exception of emotional interrogatives by V with a very prominent shift of the pitch peak to the right).

4. DISCUSSION

4.1 Duration Patterns

Duration in general is known to be affected by emotions.^{18,20} This experiment has shown that as compared with the neutral questions, emotional questions have greater sentence (word) duration, duration of the accented syllable and the duration of the accented vowel. The increase is particularly prominent in the accented vowel, and ranges from 41 to 5% of the vowel length depending on the

Table 1 Significance of parameter differences between emotional and neutral questions.

Parameters		SPEAKERS									
		V		O		L		M		S	
		<i>F</i> value	<i>E</i> vs <i>N</i>	<i>F</i> value	<i>E</i> vs <i>N</i>	<i>F</i> value	<i>E</i> vs <i>N</i>	<i>F</i> value	<i>E</i> vs <i>N</i>	<i>F</i> value	<i>E</i> vs <i>N</i>
DURA-TIONS	total	6.89	<i>E</i> > <i>N</i>	56.65	<i>E</i> > <i>N</i>	34.03	<i>E</i> > <i>N</i>	216.17	<i>E</i> > <i>N</i>	11.63	<i>E</i> > <i>N</i>
	c1	3.4		9.13	<i>E</i> > <i>N</i>	7.1	<i>E</i> > <i>N</i>	1.41		0.25	
	v1	0.01		15.75	<i>E</i> > <i>N</i>	2.57		8.76	<i>E</i> > <i>N</i>	1.3	
	c2	0.23		8.47	<i>E</i> > <i>N</i>	1.59		17.54	<i>E</i> > <i>N</i>	0.21	
	v2	22.14	<i>E</i> > <i>N</i>	51.44	<i>E</i> > <i>N</i>	70.02	<i>E</i> > <i>N</i>	94.33	<i>E</i> > <i>N</i>	0.31	
	c3	1.43		10.7	<i>E</i> > <i>N</i>	2.6		13.69	<i>E</i> > <i>N</i>	0	
	v3	1.75		2.75		0.53		19.3	<i>E</i> > <i>N</i>	26.14	<i>E</i> > <i>N</i>
	c1v1	1.75		19.5	<i>E</i> > <i>N</i>	14.69	<i>E</i> > <i>N</i>	8.32	<i>E</i> > <i>N</i>	0.24	
	c2v2	11.27	<i>E</i> > <i>N</i>	45.94	<i>E</i> > <i>N</i>	49.41	<i>E</i> > <i>N</i>	83.48	<i>E</i> > <i>N</i>	0.03	
	c3v3	0.03		0.31		2.86		48.57	<i>E</i> > <i>N</i>	18.87	<i>E</i> > <i>N</i>
<i>F</i> ₀ VAL-UES	c1f	1.96		11.9	<i>E</i> < <i>N</i>	0.393		0		4.68	<i>E</i> > <i>N</i>
	v1f	1.85		0.38		0.04		2.33		3.94	
	v1/2f	6.62	<i>E</i> < <i>N</i>	0.63		4.84	<i>E</i> < <i>N</i>	12.53	<i>E</i> < <i>N</i>	5.44	<i>E</i> > <i>N</i>
	c2f	44.1	<i>E</i> < <i>N</i>	3.5		12.12	<i>E</i> < <i>N</i>	28.99	<i>E</i> < <i>N</i>	5.59	<i>E</i> > <i>N</i>
	v2f	118.3	<i>E</i> < <i>N</i>	8.45	<i>E</i> < <i>N</i>	30.58	<i>E</i> < <i>N</i>	64.41	<i>E</i> < <i>N</i>	1.57	
	v2/2f	59.5	<i>E</i> < <i>N</i>	0.12		8.17	<i>E</i> > <i>N</i>	4.56	<i>E</i> < <i>N</i>	21.97	<i>E</i> > <i>N</i>
	h	187.57	<i>E</i> > <i>N</i>	3.6		113.26	<i>E</i> > <i>N</i>	15.34	<i>E</i> > <i>N</i>	9.96	<i>E</i> > <i>N</i>
	c3f	11.06	<i>E</i> > <i>N</i>	1.92		37	<i>E</i> > <i>N</i>	5.51	<i>E</i> > <i>N</i>	9.64	<i>E</i> > <i>N</i>
	v3f	263.98	<i>E</i> > <i>N</i>	0.03		1.74		2.73		8.22	<i>E</i> > <i>N</i>
	v3/2f	449.31	<i>E</i> > <i>N</i>	0.17		3.28		5.57	<i>E</i> < <i>N</i>	0	
	v3fe	40.9	<i>E</i> > <i>N</i>	0.29		1.66		0.3		1.96	
	v3min	70.68	<i>E</i> > <i>N</i>	0.36		1.42		4.9	<i>E</i> < <i>N</i>	2.16	
<i>F</i> ₀ INTER-VALS	c2-c1	4.33	<i>E</i> < <i>N</i>	0.81		1.75		11.35	<i>E</i> < <i>N</i>	0.77	
	c3-c2	44.75	<i>E</i> > <i>N</i>	5.57	<i>E</i> > <i>N</i>	41.77	<i>E</i> > <i>N</i>	10.9	<i>E</i> > <i>N</i>	6.6	<i>E</i> > <i>N</i>
	c3-min	17	<i>E</i> < <i>N</i>	1.63		13.4	<i>E</i> > <i>N</i>	13.39	<i>E</i> > <i>N</i>	7.69	<i>E</i> > <i>N</i>
	h-c1	117.81	<i>E</i> > <i>N</i>	1.55		64	<i>E</i> > <i>N</i>	11.65	<i>E</i> > <i>N</i>	7.07	<i>E</i> > <i>N</i>
	h-min	0.49		1.86		26.94	<i>E</i> > <i>N</i>	35.5	<i>E</i> > <i>N</i>	7.88	<i>E</i> > <i>N</i>
	risev2	92.5	<i>E</i> > <i>N</i>	17.57	<i>E</i> > <i>N</i>	148.49	<i>E</i> > <i>N</i>	30.38	<i>E</i> > <i>N</i>	5.9	<i>E</i> > <i>N</i>
	fall v2	10.89	<i>E</i> < <i>N</i>	0		2.2		0.5		2.06	
Align-ment	mintim, ms	19.08	<i>E</i> < <i>N</i>	1.37		0.107		5.43	<i>E</i> > <i>N</i>	2.89	
	plat, ms	0.25		0.125		0.37		2.79		0.01	
	htim, ms	65.58	<i>E</i> > <i>N</i>	4.21	<i>E</i> > <i>N</i>	0.79		4.46	<i>E</i> > <i>N</i>	1.91	
	htim, ru	159.7	<i>E</i> > <i>N</i>	0.01		2.5		13.23	<i>E</i> > <i>N</i>	0.33	

The difference is significant if the *F* value in the Table 1 exceeds *F* critical of 4.078 at *p* value set at less than 0.05. To make the interpretation of ANOVA results easier, '*E* vs *N*' columns in the Table 1 indicate by '>' and '<' sign if the given parameter is significantly larger or smaller in emotional (*E*) or neutral (*N*) contours. In case where no significant difference was found the cell is left blank.

The following abbreviations were used in the table: c1, v1, c2, v2, c3, v3 — durations in ms of the first consonant, first vowel, second consonant, etc; c1v1, c2v2, c3v3 — durations of the respective syllables; c1f, v1f... — pitch heights in Hz at the onset of the first, second, etc. consonant and vowel; v1/2f, v2/2f, v3/2f — pitch heights at half of the respective vowel length; h — height of the pitch peak; v3fe — pitch height of the last vowel offset; v3min — pitch height of the pitch height minimum; *F*₀ intervals denote the difference in Hz between the pitch heights of the respective points; Mintim, ms — alignment of the *F*₀ minimum regarding the last vowel offset; plat, ms — the duration of pitch plateau; htim — alignment of the pitch peak regarding the offset of the second (accented) vowel in ms; htim, ru — alignment of the pitch peak regarding the offset of the second (accented) vowel in relative duration units (alignment in ms over the vowel duration).

speaker. Stressed vowel lengthening is known to be one of the major means of making a word prominent, emphasizing its semantic value.¹⁸⁾ Even in a

tone language like Mandarin Chinese duration of a stressed vowel is salient for the perception of stress.²¹⁾ Accented/unaccented vowel duration

Table 2 Total sentence duration (ms) in neutral and emotional questions.

Question type	SPEAKER									
	S ₁ (V)		S ₂ (O)		S ₃ (L)		S ₄ (M)		S ₅ (S)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
neutral	602.5	26.7	541.4	46.9	674.2	35.6	482.6	23.8	524.2	37.8
emotional	631.9	43.8	643.2	42.7	762.0	61.0	668.3	52.9	582.2	36.2

Table 3 Accented vowel duration (ms) in neutral and emotional questions.

Question type	SPEAKER									
	S ₁ (V)		S ₂ (O)		S ₃ (L)		S ₄ (M)		S ₅ (S)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
neutral	163.6	23.8	150.1	20.5	193.3	18.9	109.1	15.5	220.8	11.4
emotional	198.1	23.6	194.9	20.9	245.9	22.5	184.8	32.2	230.3	29.6

Table 4 F_0 (Hz) at the accented vowel onset in neutral and emotional questions.

Question type	SPEAKER									
	S ₁ (V)		S ₂ (O)		S ₃ (L)		S ₄ (M)		S ₅ (S)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
neutral	293.7	14.1	233.5	23.9	235.8	12.6	252.9	22.9	229.3	29.6
emotional	242.5	16.2	212.1	24.8	212.2	15.5	198	21.9	220.8	11.4

Table 5 Pitch peak height (Hz) in neutral and emotional questions.

Question type	SPEAKER									
	S ₁ (V)		S ₂ (O)		S ₃ (L)		S ₄ (M)		S ₅ (S)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
neutral	378.1	8.0	355.0	11.6	349.0	15.1	335.0	31.2	339.9	15.6
emotional	434.2	16.9	362.1	13.0	398.1	15.4	387.4	53.1	366.5	53.6

Table 6 Pitch rise interval (Hz) in the accented syllable in neutral and emotional questions.

Question type	SPEAKER									
	S ₁ (V)		S ₂ (O)		S ₃ (L)		S ₄ (M)		S ₅ (S)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
neutral	81.9	2.6	121.4	4.2	113.2	3.5	86.2	9.4	111.7	3.6
emotional	147.3	6.1	149.9	5.2	185.9	4.7	169.5	11.4	134.1	8.4

ratios obtained in my experiment for both interrogative types are similar to the earlier reported ones, whereby unstressed vowels are about 1.5–2 times shorter than stressed.²²⁾

Total sentence duration has been also reported to increase in emotional questions (18), although the type of emotion was not specified in the source. This experiment confirmed that ‘yes/no’ questions

with the emotion of ‘surprise’ have longer duration than the neutral ones.

4.2 Pitch Patterns

There are some differences in the overall contour pattern between surprised and neutral interrogatives: emotional F_0 contours tend to have a slight falling movement before the beginning of the rise.

The rise in the emotional contours starts later — from the accented vowel onset, whereas in neutral F_0 contours it starts from the c2 onset. It has been earlier reported by Ode²³⁾ that 'yes/no' questions have a very steep rise early in the tonic syllable. In my experiment this is true of neutral, but not of surprised questions.

The rise in surprised questions is accomplished later than in the neutral ones but NOT by changing the alignment of pitch peak (which is understood as location of pitch peak in a particular part of a particular segment), but by increasing the accented vowel length. Alignment of pitch peak remains more or less constant for all the speakers. Only speaker V shows a greatly significant shift in the alignment of the pitch peak. It is possible to suggest that this change in alignment may be speaker-specific or could be caused by the employment of somewhat different attitude, *e.g.* 'astonished' rather than 'surprised'.

Pitch range has been reported to be important for expression of emotions.^{10,24)} In my experiment the overall interval of the rise as well as the interval of the rise within the accented vowel are more important than the range for the 'surprised' vs. 'neutral' interrogative distinction.

It has been suggested that pitch height may be important for emotional distinctions.²⁵⁾ Although on the whole it is true that higher pitch peak is associated with emotional involvement, it should also be pointed out that this does not hold for all the subjects in this experiment. Speaker O does not show any significant difference in pitch height between emotional and unemotional sentences, and for speaker S the difference in pitch peak height is significant, but not very strong (27Hz). Besides the pitch peak height, other parameters important for the neutral-surprised interrogative distinction include the height of the accented vowel onset and offset as well as duration and the interval of the rise in the accented vowel and syllable.

4.3 Alignment Patterns

The experimental results support the idea that alignment and association of pitch events are different phenomena: although pitch accent is associated with the stressed syllable it may fall in its proximity, not necessarily within the syllable.¹²⁾ This is particularly true of speaker V's surprised interrogatives where pitch peak is regularly located

in the post-accented, but not accented syllables. Pitch peak in interrogatives produced by other speakers also sometimes occurred within the third syllable. Although the pitch peak alignment distinction has been reported to be used in Russian to distinguish between declaratives, exclamations, interrogatives and enumerations,²⁶⁾ with the exception of speaker V, alignment differences are not employed for the distinction between neutral and surprised interrogatives.

Timing of the pitch minimum as well as the length of pitch plateau are not employed for these distinctions either.

4.4 Intersubject Variability

Many speech production studies found considerable intersubject variability in acoustic cues of emotions.^{9,27)} It is attributed to differences in the style of acting and the ability to simulate emotions in studies where professional actors were involved as subjects.²⁾ In the experiment reported in this paper ordinary subjects were employed, and the emotion of surprise was easy for simulation. Nevertheless, significant intersubject variability was observed. In the choice of parameters employed for the emotional/unemotional question distinctions, the subjects differed not only by parameter values, but by a set of parameters as well. For example, Speaker S employs 15, whereas Speaker V-22 of the 33 analyzed parameters. It is possible to suggest that the language system itself provides a considerable degree of individual freedom in ways of expressing emotional states.

5. CONCLUSION

I investigated the production of one-word emotional and neutral 'yes/no' questions with a rise-falling contour in Russian and found some statistically significant differences between the two groups of questions. The following parameters show statistically significant differences in values: total word/sentence duration, duration of the accented vowel and syllable, pitch height of the accented vowel onset, height of the pitch peak, pitch height of the accented vowel offset, interval of pitch movement in the accented syllable and vowel. Employment of other parameters is speaker-dependant. Two of the investigated parameters (the pitch height of the first vowel onset, and the duration of pitch plateau in ms) are not used by the speakers for the

neutral/surprised interrogative distinction.

The experiment described in this paper questions the validity of a well established fact about intonational expression of emotions: emotion equals increased pitch height and range. In Russian other factors (duration and interval of pitch movement in the accented vowel) also contribute to the differentiation of emotional and unemotional interrogatives. These results indicate that acoustic aspects of the expression of emotions can be language-specific. These findings can be employed for automatic recognition and synthesis of Russian 'yes/no' questions.

ACKNOWLEDGEMENTS

This paper became possible due to Speech Signal Processing project conducted at the Electrotechnical Laboratory. I would like to express my special gratitude to Dr. Kazuyo Tanaka for his advise and support.

REFERENCES

- 1) A. Kappas, U. Hess, and K. R. Scherer, "Voice and emotion," in *Fundamentals of Nonverbal Behaviour*, R. S. Feldman and B. Rimé, Eds. (Cambridge University Press, Cambridge, 1991), pp. 200-238.
- 2) A. Ni Chasaide and C. Gobl, "Voice source variation," in *The Handbook of Phonetic Sciences*, W. J. Hardcastle and J. Laver, Eds. (Blackwell, Oxford, 1997), pp. 427-461.
- 3) R. Carlson and B. Granstrom, "Speech synthesis," in *The Handbook of Phonetic Sciences*, W. J. Hardcastle and J. Laver, Eds. (Blackwell, Oxford, 1997), pp. 768-788.
- 4) I. R. Murray and J. L. Arnott, "Toward the simulation of emotion in synthetic speech: A review of literature on human vocal emotion," *J. Acoust. Soc. Am.* **93**, 1097-1108 (1993).
- 5) J. E. Cahn, "The generation of affect in synthesized speech," *J. Am. Voice I/O Soc.* **8**, 1-19 (1990).
- 6) I. R. Murray, J. L. Arnott, N. Alm, and A. F. Newell, "A communication system for the disabled with emotional synthetic speech produced by rule," *Proc. Eur. Conf. on Speech Communication and Technology*, Genova, Italy, 311-314 (1991).
- 7) M. Schubiger, *English Intonation: Its Form and Function* (Niemeyer Verlag, Tübingen, 1958).
- 8) L. Cosmides, "Invariances in the acoustic expression of emotion in speech," *J. Exp. Psychol.: Hum. Percept. Perform.* **9**, 864-881 (1983).
- 9) K. R. Scherer, D. R. Ladd, and K. Silverman, "Vocal cues to speaker affect: Testing two models," *J. Acoust. Soc. Am.* **76**, 1346-1356 (1984).
- 10) D. R. Ladd, K. Silverman, F. Tolkmitt, G. Bergmann, and K. R. Scherer, "Evidence for the independent function of intonation contour type, voice quality and F_0 range in signalling speaker affect," *J. Acoust. Soc. Am.* **78**, 435-444 (1985).
- 11) T. Adachi, "Sarcasm in Japanese," *Stud. Lang.: International Journal sponsored by the 'Foundations of Language'* **20**(1), 1-36 (1996).
- 12) D. R. Ladd, *Intonational Phonology* (Cambridge University Press, Cambridge, 1996).
- 13) E. A. Bryzgunova, *Zvuki I intonatsiya russkoy rechi*. (Russkij yazik, Moscow, 1977).
- 14) I. G. Torsueva, *Intonacija i Smysl Vyskazyvanija* (Nauka, Moscow, 1979).
- 15) K. T. Holden and J. T. Hogan, "The emotive impact of foreign intonation: An experiment in switching English and Russian intonation," *Lang. Speech* **36** (1), 67-88 (1993).
- 16) B. Rimé and D. Giovanni, "The physiological patterns of reported emotional states," in *Experiencing Emotion: A Cross-cultural Study*, K. R. Scherer, H. G. Wallbott, and A. B. Summerfield, Eds. (Cambridge University Press, Cambridge, 1986), pp. 84-97.
- 17) P. Ladefoged, "Linguistic phonetic descriptions," in *The Handbook of Phonetic Sciences*, W. J. Hardcastle, and J. Laver, Eds. (Blackwell, Oxford, 1997), pp. 589-618.
- 18) N. D. Svetozarova, *Intonatsionnaya Sistema Russkogo Yazyka* (Leningrad University, Leningrad, 1982).
- 19) P. Lieberman and S. B. Michaels, "Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech," *J. Acoust. Soc. Am.* **34**, 922-927 (1962).
- 20) R. Collier, "On the communicative function of prosody: some experiments," *IPO Annu. Prog. Rep.* **28**, 67-75 (1993).
- 21) X. S. Shen, "Relative duration as a perceptual cue to stress in Mandarin," *Lang. Speech* **36**(4), 415-433 (1993).
- 22) L. V. Zlatoustova, "Dlitel'nost' glasnykh I soglasnykh zvukov russkogo yazyka," *Uch. Zap. Kazan. Univ.* **114**(6), 95-123 (1954).
- 23) C. Ode, *Russian Intonation: A Perceptual Description* (Rodopi, Amsterdam, 1989).
- 24) J. Hirschberg and G. Ward, "The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English," *J. Phonet.* **20**, 241-251 (1992).
- 25) J. D. O'Connor and G. F. Arnold, *Intonation of Colloquial English*, 2nd ed. (Longman, London, 1973).
- 26) V. Makarova, "Rising-falling intonation contour across Russian sentence types," *Forthcoming in Rusistika* (2000).
- 27) C. E. Williams and K. N. Stevens, "Emotions and speech: some acoustical correlates," *J. Acoust. Soc. Am.* **84**, 1250-1261 (1972).