

PERCEPTION OF INTONATIONAL CONTRASTS BY JAPANESE AND RUSSIAN LISTENERS

Veronika Makarova
The University of Edinburgh, UK

ABSTRACT

This paper addresses the problem of universal and language-specific features in intonation perception. It reports the results of a sentence type (declarative, interrogative, exclamatory) identification task performed by Japanese and Russian subjects on re-synthesised one-word two-syllable stimuli with modified rising-falling contours. The effect of parameter manipulations on sentence type perception is investigated. An increase in the height or interval of the rise leads to a decrease in declarative judgement. Stimuli with a rise exceeding 5.2st are not perceived as declarative, which suggests the existence of a categorical boundary in perception. Raising the fall start pitch increases exclamatory and decreases declarative judgement. Raising the fall end pitch has the same effect on the perception by Japanese subjects, but does not influence the perception by Russian subjects. Some significant differences between the two subject groups are found in the perception of stimuli as interrogative and exclamatory, but not as declarative.

1. INTRODUCTION

Intonation frequently performs a sentence-disambiguation (grammatical) function in a number of languages, such as Greek, Hungarian, Italian, Spanish, Portuguese, Russian, and others [2, 3, 1, 6]. Specifically, intonation in these languages serves as the only marker of sentence type in “yes/no” questions. Intonation was shown to be effectively used by listeners as the only clue for sentence type disambiguation not only in speech production, but in perception as well [3, 4, 9]. An experiment by Repp and Lin [8] has shown some uniformity in the perception of tonal contrasts by speakers of different languages. An earlier experiment with real-speech stimuli perception [5] has demonstrated that Japanese and Russian subjects have both similarities and differences in sentence type identification relying on prosodic clues.

However, still very little is known about perceptual clues that help listeners to identify sentence types by intonation means, and in particular, which of these clues are language-specific, and which may be universal. To compare the effect of gradual changes in intonation parameters on listeners' perception of sentence type we designed an experiment with re-synthesised rise-fall contours.

2. MATERIALS AND METHODS

2.1 Manipulations.

One-word interrogative “tata” with a rise in the first and a fall in the second vowel was chosen as the basis for re-synthesis since it had a high percent of correct sentence type identification in our previous experiment [5, 6], and had good vowel quality allowing

easy manipulations. The original formant values were preserved, while pitch and duration parameters were modified. The choice of manipulations was determined by intonation parameters which had maximal and minimal salience for perception in our previous experiment with real-speech data [5]. Modifications were obtained on a Kay Elemetrics CSL 4300.

Six series of re-synthesised contours were produced with the following manipulations: 1) interval of the initial rise; 2) height of the initial rise with constant interval of 4.2 st; 3) height of the initial rise with constant interval of 0.5 st; 4) duration of the rise and the fall; 5) pitch height of the end of the fall and interval of the fall; 6) pitch height of the beginning of the fall and the interval of the fall. The durations of segments (with the exception of series 4 where durations were the subject for manipulation) constituted: /t/ - 97 ms, /a/ - 55 ms, /t/ - 145 ms, /a/ - 115 ms. The exact values of manipulated parameters are given in Table 1.

The six series constituted stimuli for a perceptual experiment in the course of which 40 Russian and 39 Japanese male and female subjects identified each stimulus as an exclamation, interrogative or declarative.

2.2. Analysis

The responses of subjects were tabulated using SPSS software. Percent of identification of each stimulus as a declarative, interrogative, or exclamation was calculated, and represented in Table 1. The chi-square test was used to determine the significance of the difference in stimuli identification between the two groups of listeners (at $p < 0.05$). Significant differences between the perception of each stimulus by Japanese and Russian subjects are represented in Table 1 by asterisks.

3. RESULTS

3.1. Series 1 (manipulated interval of the rise in the first syllable).

For both groups of listeners the percent of stimuli perception as declarative decreases with the increase of the interval of the rise. At stimulus 4 (7.2 st) we find a transition in the listeners' perception from declarative to non-declarative. The transition is associated with exclamatory judgement for the Japanese subjects, and with interrogative judgement for the Russian subjects. There is a significant difference in perception of stimuli 4 and 5 by Japanese and Russian subjects (chi-sq=26.99, AS=0.000; chi-sq=12.783, AS=0.002). Japanese subjects prefer “exclamatory”, and Russian subjects – “interrogative” labels for non-declarative responses to stimuli 7 and 8. The perception of all stimuli in this series as interrogative (chi-sq=7.535, AS=0.006) and exclamatory (chi-sq=16.216, AS=0.000) is also significantly different between the two groups of listeners. The perception of all stimuli in this series as belonging to one of the three sentence

types is significantly different between the two groups of listeners (chi-sq=25.868; AS=0.000).

3.2. Series 2 (manipulated height of the steep rise).

The lowest stimulus (1) is identified by Russian subjects as declarative, while Japanese subjects believe it can be either declarative or interrogative. There is a subsequent peak in declarative responses at stimulus 2 (156-199 Hz), followed by a small decrease in declarative identification at stimulus 3 (175-223 Hz), and an abrupt decrease at stimulus 4 (220-280 Hz). The latter is identified as exclamatory by Japanese subjects, and has random recognition by Russian subjects. The difference in perception of stimulus 4 by Japanese and Russian subjects is significant (chi-sq=24.624, AS=0.000). Subsequent “non-declarative” stimuli on the whole tend to be perceived as exclamatory by Japanese subjects, and as interrogative by Russian subjects.

The perception of all stimuli in this series as interrogative (chi-sq=6.551, AS=0.010) and exclamatory (chi-sq=9.966, AS=0.002) is significantly different between the two groups of listeners. The perception of all stimuli in this series as belonging to one of the three sentence types is also significantly different between the two groups of listeners (chi-sq=19.549; AS=0.000).

3.3. Series 3 (manipulated height of the gradual rise).

Perception strategies of both groups of subject are similar, with the exception of the significant difference in the first (lowest) stimulus labelling (chi-sq=12.65; AS=0.002). It is believed to be interrogative by the Japanese, and declarative by the Russian subjects. On the whole, similarly to other series, Japanese subjects produce more exclamatory than interrogative judgements, while the opposite is true of the Russian subjects. As compared to series 2, Series 3 shows more gradual and steady change in the perception of sentences as interrogative (for Russians) and exclamatory (for Japanese) with the increase of the height of the rise.

The perception of all stimuli in this series as exclamatory is significantly different between the two groups of listeners (chi-sq=8.363, AS=0.004). The perception of all stimuli in this series as belonging to one of the three sentence types is also significantly different between the two groups of listeners (chi-sq=16.008; AS=0.000).

3.4. Series 4 (manipulated vowel durations).

Three duration values were resynthesised for the first and the second vowels. The first vowel had a “short” duration of 55 ms, “medium” – of 90 ms, “long” – 135 ms; the second vowel had the duration of 60, 120 and 180 ms respectively.

On the whole, perception strategies of both groups of listeners are extremely similar. Durational changes cause changes in the perceived category. “Short+short” and “medium+short” stimuli were perceived as either an interrogative or an

exclamation. The “long+short” stimulus is perceived as interrogative. The “short+medium” stimulus is perceived as declarative. “Medium+medium” and “long+medium” stimuli are less likely to be perceived as declaratives and more likely to be perceived as interrogatives. Long rise is associated by both groups of listeners with interrogativity. No significant differences were found in the perception of this series by the two groups of listeners.

3.5. Series 5 (manipulated height of the fall end).

The only significant difference (chi-sq=11.761, AS=0.003) was found in the perception of stimulus 1 with the lowest end. While Japanese speakers positively identify it as a declarative, Russian speakers label it either a declarative, or an exclamation. While Russian subjects tend to identify all the stimuli in this series as declarative, irrespective of the varied parameter, Japanese subjects’ react to the raised pitch of the fall end by producing fewer declarative labels.

3.6. Series 6 (manipulated height of the beginning of the fall).

Raising the pitch of the beginning of the fall leads to an increase in stimuli identification as exclamations by both groups of listeners. No statistically significant differences are found between the responses of Japanese and Russian subjects.

4. DISCUSSION

Listeners’ perception of sentence type is affected by the manipulated intonation contour parameters. The increase in the height and interval of the rise leads to a decrease in declarative judgement, which agrees with earlier findings for other languages [3]. The results suggest the existence of a categorical threshold in listeners’ perception whereby stimuli with a pitch rise of over 5.2st do not elicit declarative judgement. Changes in the pitch characteristics of the rise on the nuclear syllable are more significant for the perception of Japanese and Russian subjects than changes in duration of the nuclear and post-nuclear syllable, or changes in the pitch characteristics of the post-nuclear syllable.

We found similar strategies in the perception of intonation contrasts by Japanese and Russian listeners. Little difference was also found earlier in tone perception by Chinese and English subjects [8]. Perception of stimuli as declarative, as well as the “positive identification” zones for declaratives show no significant difference between the subjects, which suggests that there exists a uniform mechanism of perception of declaratives across languages. There are, however, some significant differences in stimuli identification as interrogative and exclamatory, which suggests language-specific nature of intonation perception in these sentence types.

The cause of these differences can possibly be found in the functional load of rise-fall interrogatives in Japanese and Russian, and subjects’ linguistic expectations connected with it.

Table 1. Percent and significance of subjects' identification of each stimulus as belonging to one of the three sentence types

SERIES 1. Manipulated INTERVAL OF THE RISE									SERIES 4. Manipulated VOWEL DURATION										
N	varied param.		% of sentence type identification						JvsR sig	N	varied param.		% of sentence type identification						JvsR sig
	Hz	st	Russ. subjects			Japan. subjects					v1,ms + v2, ms	Hz	st	Russ. subjects			Japan. subjects		
			.	?	!	.	?	!				.	?	!	.	?	!		
1	196-210	1.2	70	15	15	67	10	23		1	55 + 60		40	16	40	51	18	31	
2	196-236	3.2	67	10	22	64	13	23		2	90 + 60		47	25	28	43	18	38	
3	196-265	5.2	55	30	15	49	26	26		3	135 + 60		32	42	25	28	46	25	
4	196-297	7.2	27	52	20	23	5	72	*	4	55 + 120		72	15	12	90	10	0	
5	196-336	9.3	25	32	42	15	18	67	*	5	90 + 120		60	25	15	77	23	0	
6	196-376	11.3	22	32	45	13	43	43		6	135 + 120		50	40	10	41	51	8	
7	196-424	13.3	17	47	35	18	31	51		7	55 + 180		35	20	45	43	36	20	
8	196-500	15.3	27	40	32	15	28	56		8	90 + 180		45	25	30	33	33	33	
										9	180 + 180		48	20	32	28	51	20	

SERIES 2. Manipulated HEIGHT OF THE 4.2st RISE									SERIES 5. Manipulated HEIGHT OF THE FALL END										
N	varied param.		% of sentence type identification						JvsR sig	N	varied param.		% of sentence type identification						JvsR sig
	Hz	st	Russ. subjects			Japan. subjects					Hz	st	Russ. subjects			Japan. subjects			
			.	?	!	.	?	!				.	?	!	.	?	!		
1	139-177	0	60	27	12	41	43	15	*	1	217-72	+12	40	22	38	77	13	10	*
2	156-199	+2	72	12	15	69	28	2		2	217-81	+10	57	20	23	82	13	5	
3	175-223	+4	62	17	20	59	20	20		3	217-92	+8	53	10	37	69	15	15	
4	220-280	+6	32	35	32	8	5	87	*	4	217-103	+6	45	27	27	57	15	28	
5	247-315	+8	17	37	45	8	18	74		5	217-114	+4	52	25	22	64	23	13	
6	277-354	+10	7	52	40	10	28	61		6	217-128	+2	47	25	28	25	33	15	
7	312-397	+12	2	25	72	2	8	90		7	217-144	0	52	20	28	28	15	56	
8	349-447	+14	5	70	25	8	43	49											

SERIES 3. Manipulated HEIGHT OF THE 0.5st RISE									SERIES 6. Manipulated HEIGHT OF THE FALL START										
N	varied param.		% of sentence type identification						JvsR sig	N	varied param.		% of sentence type identification						JvsR sig
	Hz	st	Russ. subjects			Japan. subjects					Hz	st	Russ. subjects			Japan. subjects			
			.	?	!	.	?	!				.	?	!	.	?	!		
1	153-158	0	72	25	2	36	64	0		1	177-127	0	57	22	20	28	36	36	
2	172-177	+2	72	20	8	67	33	0		2	199-127	+2	37	25	37	64	10	25	
3	193-199	+4	57	27	15	77	13	10		3	223-127	+4	30	40	30	23	20	56	
4	216-223	+6	65	15	20	56	20	23		4	250-128	+6	25	37	37	13	20	67	
5	243-250	+8	32	25	42	25	15	59		5	280-127	+8	7	35	57	8	20	72	
6	273-280	+10	22	33	45	5	23	72		6	315-128	+10	0	35	65	2	33	64	
7	306-315	+12	12	42	45	5	20	74											
8	343-354	+14	0	52	48	0	31	69											
9	385-397	+16	2	58	40	0	31	69											
10	433-447	+18	5	35	60	0	13	87											

In "J vs R sig" columns an asterisk indicates cases when the perception of stimulus was significantly different between Japanese and Russian listeners.

In Russian rise-fall is the most common way of intoning “yes/no” questions (with the exception of the accented words with the stress on the last syllable), and questions bear no other markers of sentence type [9]. Russian subjects have therefore high expectancy of rise-falling interrogatives, hence the preference of “interrogative” judgement by Russian subjects. In Japanese, on the other hand, rising interrogatives are more common, and the rising contour is mostly used in combination with the interrogative particle “ka” [7]. Rise-falls are used mostly as echoes, and they are therefore peripheral in the linguistic expectations of the subjects, who produce relatively few interrogative judgements while listening to isolated word-sentences with a rise-fall. We found also another example of the differences in perception determined possibly by the respective language systems. For Russians, a long rise in combination with a medium and short fall sounds interrogative, but not a long rise in combination with a long fall, whereas for Japanese subjects a long rise leads to the perception of a stimulus as an interrogative irrespective of the fall duration. This perception may be explained by the characteristics of interrogative contours in production: interrogatives in both languages are characterised by lengthening of the rise (as compared to declarative and exclamation) [6]. However, Japanese interrogatives with a rise-fall have a long fall [5], whereas Russian interrogatives have a relatively short fall, and the total duration for Russian interrogatives shorter than the total duration for either declaratives, or exclamations [9].

5. CONCLUSION

Japanese and Russian subjects show very similar perceptual strategies. There is no significant difference in the perception of stimuli as declaratives between the two groups of subjects. There are, however, some significant differences in the perception of stimuli as exclamations and interrogatives. Japanese subjects are more likely to react to the increase in height and interval of the rise and the increase of the interval of the fall (when the ending pitch is constant) by labelling contours as exclamatory, while Russian subjects use the “interrogative” label more frequently (Series 1-3, 6). The opposite is true for varied duration (Series 4).

An increase in the height or interval of the rise leads to a decrease in the perception of stimuli as declarative by both groups of listeners. Stimuli with the interval of the rise of over 5.2 st do not elicit declarative judgements, which may be interpreted as evidence of the categorical nature of declarative/non-declarative contrast perception.

Changes in duration alone are less salient than interval and height of the rise; nevertheless, duration differences do lead to changes in stimuli identification.

ACKNOWLEDGMENTS

I would like to thank Professor Natal'ya Svetozarova for discussions of the research results, Professor Hakutaro Jou for his support, and all lecturers of the Department of Phonetics, St. Petersburg University and Department of Foreign Languages, St. Petersburg Electro-Technical University who helped with recruiting listeners.

REFERENCES

[1] Arvaniti, A., Ladd, D.R., and Mennen, I. 1999. Quantitative F0 descriptions require reference to local targets: The intonation of Greek polar questions and emphatic statements. *Journal of the Acoustic Society of America*, forthcoming.

[2] Cruttenden, A. 1986. *Intonation*. Cambridge: Cambridge University Press.

[3] Gossy, M. and Terken, J. 1994. Question marking in Hungarian: timing and height of pitch peaks. *Journal of Phonetics*, 22, 269-281.

[4] Hadding-Koch, K. and Studdert-Kennedy, M. 1964. An experimental study of some intonation contours. *Phonetica*, 11, 175-85.

[5] Makarova, V. and Matsui, J. 1998. Sentence type perception strategies in a native language/foreign language. *Journal of the English Phonetics Society of Japan*, 2, 331-352.

[6] Makarova, V. and Matsui, J. 1996. Perception of potentially ambiguous intonation contours by speakers of Russian and Japanese. *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP 96)*, Philadelphia, 2-7 October 1996, 13-16.

[7] Miura, I. and Hara, N. 1995. Production and perception of rhetorical questions in Osaka Japanese. *Journal of Phonetics*, 23, 291-303.

[8] Repp, B.H. and Lin, H.B. 1990. Integration of segmental and tonal information in speech perception: a cross-linguistic study. *Journal of Phonetics*, 18, 481-495.

[9] Svetozarova, N. D. 1982. *Intonatsionnaya sistema russkogo yazyika*. Leningrad: Leningrad University.