INTONATION CONTOUR PROCESSING
BY JAPANESE AND RUSSIAN LISTENERS

June-ko Matsui
Veronica Makarova

Meikai University (Akemi 8, Urayasu, Chiba, Japan, 279)

ABSTRACT

This study deals with the production and perception of short Japanese and Russian declaratives, interrogatives and exclamations and its application to speech recognition systems. Parameters such as frequency and duration differ for Japanese and Russian perception of one-word sentences with similar segments and rise-fall or fall contours. Russian listeners were able to correctly identify all three sentence types in their native language, and distinguished between sentence types in Japanese, while Japanese subjects only differentiated declaratives and non-declaratives.
Significant parameters for each set of listeners are as shown below:

	Japanese declartv	Japanese interrgt	Japanese exclamtn	Russian declartv	Russian interrgt	Russian exclamtn
Rise duration	○			○		
Fall duration					○	○
Total duration		○		○	○	
Initial pitch height					○	
Average Pitch	○		○	○		○
Standard deviation Pitch	○		○	○		○
Maximum Pitch	○	○	○	○	○	
Pitch Range	○	○	○	○		○
Total number of parameters	5	3	4	6	4	3

Table 1

The table seems to indicate that duration is more important for the Russian listener than the Japanese, while the Japanese appear to rely more heavily on pitch.
The results of the experiment can further be applied to speech recognition systems with automatic interpretation of sentence type based on pitch, duration, velocity of the pitch rise/fall, and other relevant parameters.
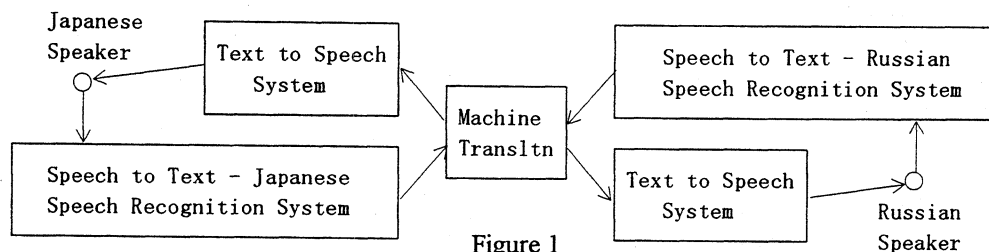

Figure 1

1. INTRODUCTION

The present study focuses on Japanese and Russian 2 syllable fall contours and 3-syllable rise-fall contours. The chosen patterns are fairly common in many languages and are considered archetypal (Vassiere, 4). Both Russian and Japanese have rise-fall and fall contours for declaratives and exclamations (Miura, Hara, 1; Ohde, 2; Svetozarova, 3). However, while interrogatives with the rise-fall contour are typical for Russian, they are more marked in Japanese, and frequently function as echoes.

2. MATERIALS AND METHODS

2.1   Informants and Stimuli

A 32-year old male standard Japanese speaker and a 26-year-old male standard (St. Petersburg) Russian speaker recorded a list of 24 two-syllable and 13 three-syllable words, each prononunced as an exclamation, interrogative and a declarative. The Japanese and Russian sets of data were matched for vowel and consonant type. Japanese words were 'high, low' for the 2-syllable and 'low, high, low' for the 3-syllable structures. Russian words had a stressed first syllable for 2-syllable words and a stressed second syllable for the 3-syllable words.

2.2 The equipment

The one-word phrases were processed on a Visipitch 6095/6097 and SONA programme (37 words X 3 sentence types X 2 informants = 222 words) to obtain pitch, duration, intensity and statistical, which were then input on an Excel chart for sorting and analysis.

2.3 The stimuli and subjects for the listening
   experiment

The original recording was edited into 2 random-order lists. The Japanese list had 32 two-syllable and 26 three-syllable phrases, and the Russian, 35 and 22, respectively. Each stimulus was recorded three times with intervals of about 2 seconds. The lists were presented to 20-21 Russian subjects, and 63-64 Japanese subjects, male and female, aged 17 to 23, who were requested to a make a forced choice between a declarative, interrogative, or exclamation.

3. RESULTS AND DISCUSSION

3.1 The listening experiment
Russian and Japanese subjects,   employ different listening strategies when listening to their native language intonation contours. Russians have a higher percentage of correct recognition responses than the Japanese. Possible contributing factors are: the subjects' linguistic experience and expectations based on the frequency of occurrence for each sentence type, and the function of grammatically unmarked utterances in each language.
Russian and Japanese subjects also manifest different perception patterns when listening to the intonation contours of an unknown foreign language. While the Japanese subjects tend to misidentify both Russian exclamations and interrogatives as declaratives, the Russian subjects are apparently able

to hear the differences between Japanese interrogatives and exclamations, but they 'label' them incorrectly, systematically perceiving Japanese interrogatives as exclamations (48.5%), and exclamations as interrogatives (62.8%).

Type	Percvd	Japanese Listeners						Russian Listeners					
		Japanese Speaker			Russian Speaker			Russian Speaker			Japanese Speaker		
		2-syl.	3-syl.	total	2-syl.	3-syl.	total	2-syl.	3-syl.	total	2-syl.	3-syl.	total
!	!	45.2	33.2	f 40.1	36.1	61.3	f 45	87.1	91.4	g 88.7	24.6	e 16.9	21.3
!	?	29.4	43.6	35.4	14.3	15.3	14.7	0.4	2.1	1	58.3	68.8	62.8
!	.	25.4	23.3	24.5	49.6	23.4	40.3	12.5	6.4	10.3	17.1	14.3	15.9
?	!	40.6	54.7	46.6	34.5	14.4	26.9	17.3	0	10.6	50.4	46	48.5
?	?	36.1	30.9	c 33.9	30.2	28.2	c 29.5	66.4	93.6	c 76.9	c 22.6	d 36	28.3
?	.	23.3	14.5	19.5	35.3	57.8	43.6	16.4	6.4	12.5	27	18	23.1
.	!	3.5	3.2	3.3	11.1	8.3	10	4.2	b 11.3	7	11.3	11.9	b 11.6
.	?	1.4	1.8	1.6	12.3	10.7	11.6	0	1.3	0.5	4.8	6	5.4
.	.	95.1	95	a 95.1	76.7	81	a 78.4	95.8	87.5	a 92.5	83.9	82.1	a 83

Table 2

Percentage of subjects who perceived each sentence type as an exclamation (!), interrogative(?) or declarative (.). Russian listeners are able to differentiate the three sentence types, while Japanese only perceive the difference between declarative/non-declarative.

The percentage of correct recognition is highest for **declaratives** (Table 1 a). This can be attributed to gradual rise/fall and low peaks in both languages.

Correct recognition is lowest for **interrogatives** (Table 1 c) (with the exception of 3-syllable Japanese words perceived by Russian listeners - Table 1 d - where the percentage of correct recognition for exclamations is lower - Table 1 e).

Correct recognition for **exclamations** is lower than declaratives but better than interrogatives for the Japanese subjects (both languages) (Table 1-f) and for Russians listening to Russian (Table 1-g).

3.2 The instrumental analysis

Some intonation parameters for which there are notable differences between the Japanese and Russian listeners are: rise/fall/total duration, initial pitch height, average/standard/maximum pitch, and pitch range. Other possible factors are velocity of the pitch rise, the total relative/absolute duration of the utterance, pitch peak timing, length of the last syllable, devoicing, early endings, extended endings and similarity in Russian and Japanese overall contour shape. Japanese interrogatives are extended at the end of the utterance possibly because the typical Japanese question rises in the end. This rise is most probably replaced by duration.

Rise/fall/total duration appears to be more important for the Russian listener than the Japanese, while the Japanese seem to rely more heavily on pitch.

Data obtained from the experiment can be incorporated into a speech recognition system, enabling automatic discrimination of sentence types. Statistical probability data for different parameters such as pitch and timing will serve as the basis for evaluation.

SUMMARY

The use of phonetic parameters such as pitch, relative/absolute duration, velocity of rise/fall (angle of slope), shape of contour, and devoicing were somewhat different for interrogatives and exclamations, but similar for declaratives, for the Russian and Japanese informant. Interaction of the parameters and listener expectations lead to correct/incorrect perception of the three sentence types. In some cases, numerous parameters were substantially similar for both languages, resulting in high positive transfers for that sentence type. Results indicate that a combination of native language parameters was employed by the subjects to interpret the foreign language.

For instance, the Russian and Japanese informants use similar phonetic parameters to realize sentence types such as declaratives (gradual rise/fall, low peak). Perception tests indicate that this common parameter usage leads to a high rate of accurate recognition for that particular sentence type across languages. On the other hand, infrequently used sentence types such as the Japanese rise-fall-rise interrogative, were not accurately perceived by the Japanese listeners.

Therefore, a combination of phonetic parameters, listener expectations, and native language listening strategies seem to account for the majority of the positive/negative transfers and native language perception observed in this experiment.

The results of the experiment can be applied to a speech recognition system, where interacting phonetic parameters serve to identify sentence types based on statistical data and probability. Such systems are important because intonational as well as segmental information is vital to correct semantic interpretation.

REFERENCES

1. Miura, I., Hara, N. Production and Perception of Rhetorical Questions in Osaka Japanese, *Journal of Phonetics*, Vol. 23: 291-303, 1995.
2. Ohde, C. *Russian Intonation: A Perceptual Description,* Amsterdam; Atlanta, GA, 1989.
3. Svetozarova, N.D. *Intonatsionnaya sistema russkogojazika*, LGU, Leningrad, 1982.
4. Vassiere, J. "Natural explanations for prosodic cross-language similarities", *ICPhS, Vol. 4, 1995, 654-657.*