CROSS-LINGUISTIC ASPECTS OF INTONATION PERCEPTION

Veronika Makarova
Meikai university, Electrotechnical Laboratory, Japan

ABSTRACT

This paper compares the perception strategies employed by the speakers of British English, Japanese and Russian in sentence type processing relying on prosodic clues. The paper reports the results of an experimental phonetic study investigating the effect of the manipulations of pitch and duration in a phonetically rising-falling contour on sentence type perception by the three groups of subjects. The results of the experiment indicate the existence of similar perception patterns and of a common threshold of declarative judgement across the three groups of listeners.

1. INTRODUCTION

In many languages of the world a lot of linguistic information including the sentence or utterance type can be indicated by intonation means, which determines the necessity of accounting for prosodic information in automatic speech recognition systems [1]. The development of these systems can become more effective by relying on prosodic parameters relevant for the perception of sentence or utterance type by humans. The question of sentence type identification by human listeners has been a subject of extensive research, e.g., [2,3,4,5,]. However, earlier research has mostly dealt with the distinction associated with the falling/ rising pitch movement contrast, and usually involved one language or related languages, e.g., [6]. Yet, it is known that some prosodic universals can be found in the perception of prosody even across unrelated languages [7,8].

The novelty of the described experiment is, firstly, in addressing three sentence types (statements, yes/no questions, exclamations); secondly, in choosing the rise-fall contour as the basis for comparison, and, thirdly, in investigating the perception of the speakers of unrelated (Japanese vs. English and Russian) or very distantly related languages (English and Russian) with different prosodic systems and the functioning of the investigated intonation distinction.

2. MATERIALS AND METHODS

One two-syllable word ‘tata’ pronounced by a male speaker of Japanese was re-synthesized with manipulated parameters of the rise, fall and duration. The word was chosen because it had good vowel quality, a rise in the first vowel and a fall in the second vowel, and F0 values could therefore be easily manipulated.

The manipulations were produced by raising Fo values at the start and end of the rise and fall in 2-st steps with the subsequent linear smoothing of the Fo contour. Durations of the rise and fall were manipulated in 30ms steps whereby the Fo values of the beginning and end of the rise and fall were kept constant. The original formant values were preserved. Modifications were obtained on a Kay Elemetrics CSL 4300.

Re-synthesized stimuli with modified intonation parameters were grouped into 6 series for a perception experiment whereby each series contained one manipulated parameter as follows: 1) pitch height of the ending point of the rise (with a constant height of the rise start); 2) pitch height of the starting and ending points of the rise with the constant pitch interval of 4.2st (pitch interval of a real speech question); 3) pitch height of the starting and ending points of the rise with the constant pitch interval of 0.5st (pitch interval of a real speech declarative); 4) pitch height of the end of the fall (with constant height of the fall start); 5) pitch height of the beginning of the fall (with constant height of the fall end); 6) duration of the rise (short - 55ms, medium – 90ms, long – 135ms) and the fall (60, 120 and 180ms respectively). Parameters manipulated in each serious are represented in Figure 1.

Forty British, forty Japanese, and thirty eight Russian subjects, both male and female, took part in a perception experiment whereby they performed a forced-choice utterance (sentence) type identification task. The subjects were requested to listen to the six experimental series in succession, and identify each presented stimulus as a statement, a question or an exclamation. The subjects indicated their choice by circling a corresponding sign in the answer sheet.

Percent of identification of each stimulus as a statement, question or exclamation (also referred to as declarative, interrogative and exclamatory judgement) is shown in Figure 2. Chi-square test was used to determine the significance of the difference in stimuli identification between the three groups of listeners (at p<0.05).
3. RESULTS

Series 1. The results demonstrate the decrease in declarative, and an increase in interrogative and exclamatory judgements with the increased height of the ending point of the rise for all the three groups of subjects. A threshold of declarative judgement is observed at stimulus 3 (196 to 265 Hz). Subsequent stimuli tend to be perceived as exclamatory by British and Japanese subjects, whereas for Russian subjects they sound either interrogative or exclamatory. These differences between the subjects’ performance are statistically significant at stimuli 4 and 5.

Series 2 and 3 with manipulated height of the rise (with a constant interval of the rise) yield similar results despite the difference in the magnitude of the rise (a gradual rise of 0.5 st in Series 2 and a steeper rise of 4.2 st in Series 3). An increase in the pitch height of the rise leads to the decrease in declarative and an increase in non-declarative judgement by all the listeners. However, percent of exclamatory perception of the stimuli is higher for the Japanese and British than for Russian subjects.

The results of listeners’ perception in Series 4 demonstrate a small effect of the height of the fall end on listeners’ perception as compared to the parameters of the rise manipulated in series 1-3. Raised height of the fall end leads to a very gradual decrease in the British and Japanese listeners’ declarative judgement, and has no effect on the perception of Russian subjects (this difference is significant at Stimulus 1: 217 to 72 Hz).

Perception of stimuli in Series 5 shows that the increase in pitch height of the fall start decreases declarative and increases exclamatory judgements for all the three groups of listeners.

Results in Series 6 (manipulated duration) show that this factor has little effect on listeners’ perception of utterance (sentence) type. All the stimuli in this series tend to be perceived as statements.

4. CONCLUSION

The comparison of sentence type identification results across the three languages shows similar perception strategies with almost no statistically significant differences in the subjects’ performance. The thresholds of declarative judgement are the same for the speakers of British English, Japanese and Russian. Some language-specific features are found in the labeling of non-declarative stimuli (question or an exclamation). Russian subjects produce more interrogative judgements than the other two groups of listeners because rise-fall questions are more common in Russian.

Parameters of the rising part of the rise-fall contour (in particular, the pitch height of the end of the rise) are crucial for the identification of sentence type, whereas the parameters of the fall have less, and of duration -- little effect on the subjects’ perception.

The results of the experiment suggest that the speakers of unrelated or distantly related languages with different types of prosodic systems and different employment of intonation contrasts for utterance/sentence type distinctions can rely on the same clues while performing an utterance/sentence type disambiguation task. These results can be employed for synthesis of sentence types in the three languages.

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

Figure 1: Manipulated parameters in Series 1-6.
Figure 2: Listeners’ perception of the stimuli
Vertical axis indicate percent of stimuli identification as belonging to one of the three sentence types; horizontal axis indicate numbers of stimuli (from the lowest to the highest). Responses of Russian subjects are shown by solid lines, of Japanese subjects – by dashed lines, and of British subjects – by dotted lines.