ISCA Archive
http://www.isca-speech.org/archive

6th International Conference on Spoken Language Processing (ICSLP 2000)
Beijing, China
October 16-20, 2000

IDENTIFICATION AND DISCRIMINATION OF SYNTACTICALLY AND PRAGMATICALLY CONTRASTING INTONATION PATTERNS BY NATIVE AND NON-NATIVE SPEAKERS OF STANDARD JAPANESE

Sanae Eda

Ohio State University

ABSTRACT

This paper reports the results of an empirical study which examined the discrimination and identification of prosodic variations by native and non-native speakers of Japanese. The data support the idea that recognizing prosodic cues is easier than associating the correct interpretations with them. This result is in keeping with first language acquisition studies showing that children acquire accurate production of prosodic effects well before they acquire adult like competence in pragmatic use.

1. INTRODUCTION

In recent years, an increasing number of researchers have recognized the importance of prosody in sentence processing. In Tokyo Japanese, several effects have a role. The grouping of phrases into a hierarchy of prosodic constituents is manifested in variation in local pitch range, and provides cues for disambiguating potentially ambiguous surface syntactic structures; the same manipulation can also indicate pragmatic focus of attention [2, 3, 5]. For example, consider the following two utterances; although they are made up of the same string of words, they were produced in response to different questions, and different parts of the utterances (written in bold) are being emphasized. The arrow indicates the beginning of a new phrase, which is cued by the expansion of the overall pitch range; the item that begins the new phrase is perceived as being emphasized.

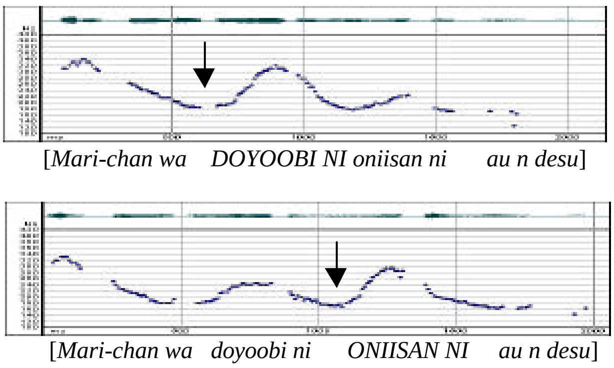

Figure 1: Prosodic grouping by the local pitch range variation.

In addition, [4] and [6] have provided empirical data demonstrating that there are at least five phrase boundary pitch movements (BPMs) in Tokyo Japanese that they consider to be distinct in form and function, and that each BPM can provide cues to indicate certain intentions of the speaker. (See [6] for a detailed discussion on their forms and functions). In this paper, we will limit our investigation to the perception of three BPMs: incredulity question rise (***incredQ***), information question rise (***infoQ***) and insisting rise (***insist***).

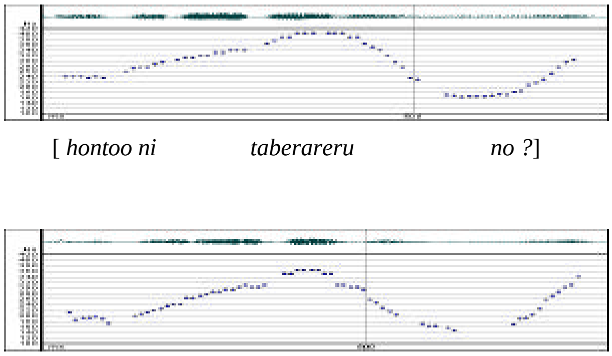

Figure 2: F0 shapes of ***infoQ*** (top) and ***incredQ*** (bottom).

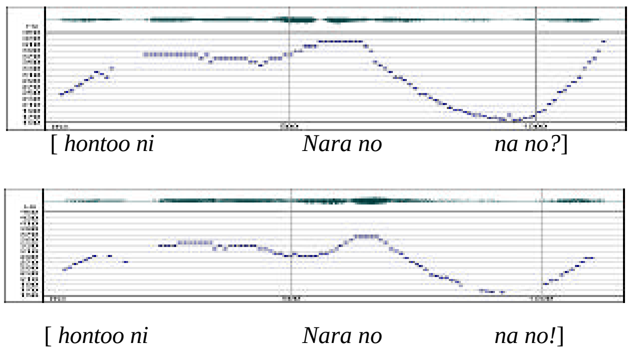

Figure 3: F0 shapes of ***infoQ*** (top) and ***insist*** (bottom). Note the difference in the shapes of the final rise; while ***infoQ*** rises to the top of the speaker's pitch range, insist is characterized by a short rise only to the mid-point.

This paper presents the results of two sets of studies. The first study examined the processing of three types of prosodically marked syntactic and pragmatic contrasts by native and non-native speakers of Japanese. The second study focused more on the non-native speakers' perception of the differences in forms and meanings conveyed by three BPMs.

2. EXPERIMENT 1

2.1. Design

The methodology for this study was derived from [1], which compared the native and non-native perception of intonation by learners of English and Portuguese.

Stimuli. The stimuli used in this study were a total of 36 sentences, consisting of words familiar to all the experimental participants, including the first-year learners of Japanese. A female native speaker of Tokyo Japanese (KY, a graduate student in linguistics) read each sentence three times; twice with the same prosodic patterns, and for the third production, she read it in such a way to make it differ minimally in some aspect of its prosodic structure. This resulted in a total of 108 stimuli. The stimuli were digitally recorded in a sound booth.

Subjects. A total of three groups of subjects participated in the first study: 36 native speakers of English, learning Japanese as a second language (NNS); 21 native speakers of Tokyo Japanese (NS-Tokyo); and 30 native speakers of other Japanese dialects (NS-other).

Tasks. Subjects heard pairs of utterances, and performed two tasks. First task (*discrimination task*) required them to judge whether or not the two utterances in each pair conveyed different meanings. Half of the pairs they heard differed minimally in their intonation patterns, and half of the pairs were two separate productions of the same intonation patterns. Immediately following the discrimination task, they performed the *identification task* in which they chose from among three possible paraphrases. Upon completing the two tasks, the next pair was presented. Subjects continued this procedure until they have heard all 36 pairs of utterances. The study was administered on a computer, and subjects selected their responses by mouse clicks.

2.2. Results

Comparison between native speakers of Tokyo dialect and speakers of other dialects. While there was no significant difference between the two groups in the discrimination task ($Z=-0.71$, $p<0.24$), NS-Tokyo group performed significantly better in the identification task ($Z=-2.91$, $p<0.01$*). This trend can be visually observed in Figure 4; each data point in figure 4 represents the score of each item in the experiment averaged across subjects; the scores of NS-Tokyo and NS-other are plotted on x- and y-axis, respectively. While the data appears equally distributed on both sides of the line in discrimination task (top), the data appears mostly below the line in identification task (bottom) because the denominator is larger than the numerator for most of the items. Due to this difference, we will only use the data from NS-Tokyo in comparing between native and non native performances in the remaining of this paper.

Comparison between native speakers of Tokyo dialect and non-native speakers of Japanese. As expected, there was a statistically significant difference between the performance of native and non-native speakers in both the discrimination ($Z=2.64$, $p<0.01$*) and the interpretation tasks ($Z=-2.55$, $p<0.01$*). This trend can be readily observed in Figure 5.

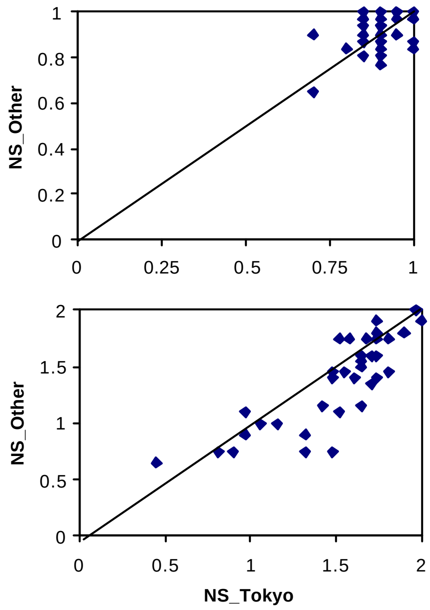

Figure 4: Comparison between NS-Tokyo and NS-other, discrimination task (top) and identification task (bottom).

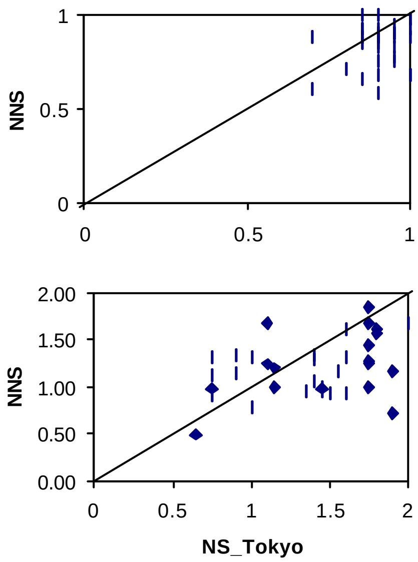

Figure 5: Comparison between NS-Tokyo and NNS, discrimination task (top) and identification task (bottom).

NNS performance. While non-native speakers performed equally well in the discrimination task regardless of their proficiency, their performance in the identification task was strongly predicted by their proficiency ($r=0.447$, $n=36$, $p<0.01$*).

See section 2.2 for more discussion on this issue.

3. EXPERIMENT 2

The second study examined the non-native perception of Japanese intonation, with focus on the variation of BPMs. This study compared three types of rising BPMs (***infoQ, incredQ, insist***) and **falling** BPM. Especially of interest was to explore more fully on the contrast between ***infoQ*** (full rise to the top of the pitch range) and ***insist*** (rise to the mid-point of the pitch range). Assuming that these three BPMs were different types of rising intonation, we hypothesized that they would present differing degrees of difficulties to NNS of Japanese. Specifically, we predicted that ***insistQ*** would be more difficult for native speakers of English since there is no comparable distinction in English, thus requiring them to create a new category for the contrasting intonation patterns without using an analogy to English.

3.1. Design

A native speaker of Tokyo Japanese (HN) recorded the stimuli following a procedure described in Section 2.1. This study involved the same tasks as Experiment 1, but the format of the presentation was modified so that subjects performed the two tasks separately. They heard 36 pairs of utterances for discrimination, and 40 (sentences which were not used in discrimination task) for identification tasks.

A total of 32 native speakers of English participated in this study. All the subjects were participating in a 9-week intensive Japanese program in a total-immersion environment, and the experiment was conducted during the first two weeks of the program. All the participants took the same proficiency test (JSKIT) at the beginning and end of the program, except for the firs year students, who did not take it at the beginning. Although this study was administered at the beginning of the language program, we used the second set of the JSKIT scores because about 1/3 of the subjects were first year students. Since there was a significant correlation between the two sets of test scores ($r=0.947$, $n=22$, $p<0.01$*), we felt justified to use the second set of JSKIT scores.

3.2. Results

The data from this study showed the same trend as the previous study. Figure 6 shows the scatter plots of the subjects' scores in discrimination (top) and identification (bottom) against their JSKIT scores. Figure 6 shows the scatter plots of the NNS scores in discrimination and identification tasks. Different from Figures 4 and 5, each data point in Figure 6 represents a subject's performance, plotted against their JSKIT score. Note the wide variance across the level of their Japanese proficiency in discrimination task (top). Although the regression line in this figure appears to have a slight down-slope, taking into consideration the sizes of variance, intercept and slope, we concluded that there was no learning trend in their discrimination task performance.

As discussed in Section 2.2, while no correlation is found between their accuracy in discrimination and language proficiency ($r=-0.04$, $n=32$, $p>0.10$), their accuracy in associating appropriate interpretations with various BMPs is strongly correlated with their proficiency ($r=0.59$, $n=32$, $p<0.01$*).

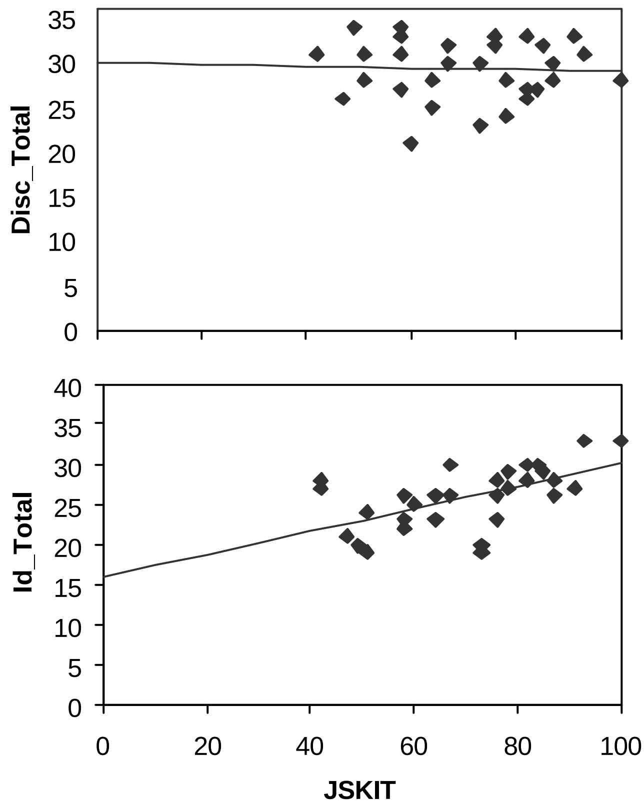

Figure 6. NNS scores in discrimination and identification tasks, plotted against their Japanese proficiency test (JSKT) scores.

Non-native perception of BPMs. Figure 7 shows the percent of correct discrimination of ***falling (Fall), insist rise (CC)*** and ***incredQ rise (QQ)*** in the condition in which the tokens were presented with another token of the same type ("same" condition). We can see that the discrimination is quite good in this condition for all token types, and no significant difference was found between the three token types.

Figure 8 shows the percent of correct discrimination in the condition in which the tokens were presented as intonational minimal pairs. The data is plotted from left to right in the order of difficulty. Interestingly, the most 'difficult' pair turned out to be the contrast between ***falling*** *and* ***insist***. Statistical test (Wilcoxon) shows that, while there is a significant difference between F-CC and CC-Q, there is no significant difference between F-Q and CC-Q conditions. This means that, contrary to the original hypothesis, ***insist rise*** is more difficult to distinguish from, i.e., more easily confused with ***falling*** than ***infoQ rise***.

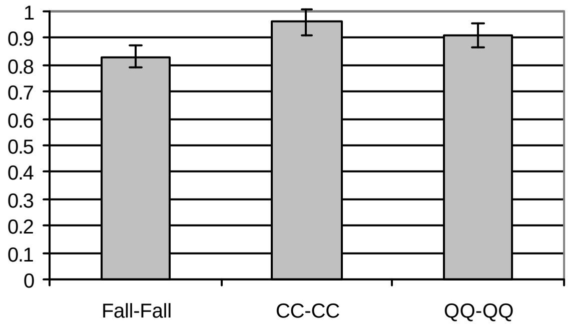

Figure 7. Percent correct discrimination of ***falling, insist rise (CC)*** and ***incredQ rise (QQ)*** in "same" condition.

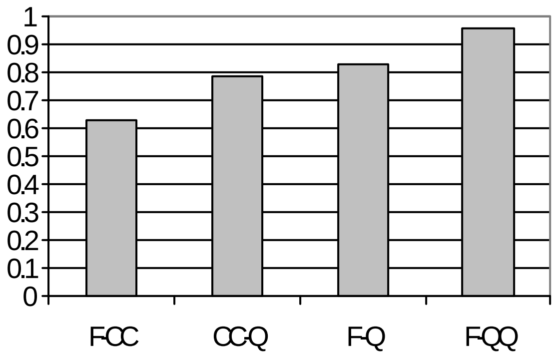

Figure 8. percent of correct discrimination in conditions in which the tokens were presented as intonational minimal pairs.

4. DISCUSSION

This paper reports the results of an empirical study which examined the discrimination and identification of prosodic variations by native and non-native speakers of Japanese.

The data from NNS did not support our hypothesis that the similarity of F0 shapes of the two rising BMPs, ***infoQ*** and ***insist,*** and the lack of comparable distinction in English, would make them difficult to discriminate from each other for native speakers of English. Contrary to our prediction, native speakers perceived ***insist rise*** as more similar to ***falling*** than it is to ***infoQ rise***. There are two possible ways to account for this result. First, it might be due to the fact that ***insist rise*** has similar semantic functions as ***falling*** BPM, which makes them different from ***infoQ***, or, secondly, it might be due to the F0 shape of ***insist rise***, which reaches only to the mid-point of the pitch range. This will be an issue that requires more exploration in the future, and such examination will need to address the question of to what extent the two factors, similarities of F0 shape and functions, would play roles in the perception of intonation patters.

The data obtained from NS and NNS support the idea that recognizing prosodic cues is easier than associating the correct interpretations with them. This result is in keeping with first language acquisition studies showing that children acquire accurate production of prosodic effects well before they acquire adult like competence in pragmatic use.

5. REFERENCES

1. Cruz-Ferreira, M. "A test for non-native comprehension of intonation in English." *IRAL XXVII-1, 1989.*: *pp23-39*.
2. Kubozono, H. "Modeling syntactic effects on downstep in Japanese", In G. J. Docherty and D. R. Ladd, eds. *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. 1992. pp. 368-387.
3. —. *The organization of Japanese prosody*. 1993. Tokyo: Kuroshio Publishers.
4. Maeda, K., and Venditti, J. J. "Phonetic investigation of boundary pitch movements in Japanese." *In International Conference on Spoken Language Processing (ICSLP), Vol. XX, 1998*.
5. Pierrehumbert, Janet and Mary E. Beckman. *Japanese Tone Structure.* 1988. Cambridge, MA: MIT Press.
6. Venditti, J. J., Maeda, K., and van Santen, J. P. H. "Modeling Japanese boundary pitch movements for speech sysnthesis", In *Proceedings of the 3rd ESCA Workshop on Speech Synthesis, 1998*.