PERCEPTION OF INTONATIONAL CHARACTERISTICS OF WH AND NON-WH QUESTIONS IN TOKYO JAPANESE

Kikuo Maekawa

National Language Research Institute, Tokyo, Japan.

ABSTRACT

The intonational difference between wh and non-wh questions in Tokyo Japanese was examined. Perception experiments involving synthetic intonation revealed that the most important cue for the discrimination between the two types is the lack of salience of intonation boundary after the wh-word, rather than the prominence of the focused wh-word per se.

1. INTRODUCTION

That syntactic behavior of wh and non-wh questions differ is well recognized by grammarians. It seems to be less recognized by those who are working with Japanese prosody that the two question types differs significantly in their prosodic domains as well. As a matter of fact,the difference does not consist in a mere difference of final rise but rather concerns the overall intonation shapes.

2. MATERIAL

Wh-questions are marked with wh-words like dare (who), doko (where), nani (what) etc. Incidentally, there are a class of words which are not wh-words but morphologically very similar to them: dareka (someone), dokoka (somewhere), nanika (something)etc. Those words are semantically marked, given their indefinite-pronoun-like meaning. As the result of their morphological similarity, we can construct pairs of wh and non-wh-questions like (1) and (2), where syntactic and accentual configurations are exactly the same across two sentences. (Apostrophes denote accent locations.)

(1)[na'ni-ga]$_{NP}$[mi-e'-ru]$_{VP}$
what-Nom. see-Pot.-Pres.
= What can (you) see?

(2)[na'nika]$_{NP}$ [mi-e'-ru]$_{VP}$
something see-Pot.-Prs.
= Can (you) see anything?

Fig.1 shows typical examples of the F0 contours of(1) and (2) uttered by a male speaker of Tokyo Japanese (TJ). Their intonational difference can be expressed in terms of their focus placement. Roughly speaking, the focus of a wh-question like (1) is on the wh-word, while the focus of a non-wh question like (2) is on its predicate. Usually the difference in focus placement is reflected in the prosodic structures of these sentences. According to the theory proposed by Pierrehumbert & Beckman [1], the difference can be represented in terms of the difference of the 'intermediate phrase' defined as the domain of 'catathesis.' While the whole utterance makes up an intermediate phrase in (3), the utterance is divided into two different intermediate phrases in (4). (It is interesting that the same prosodic difference can be observed in two 'accentless' Japanese dialects[2].)

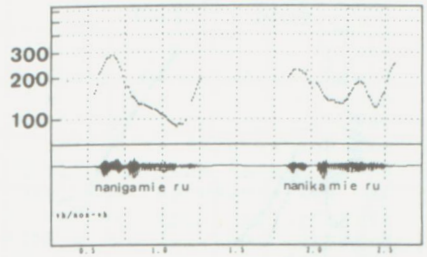

Fig.1 The F0 contours of wh question /naniga mieru / (left) and non-wh question /nanika mieru/ (right) as uttered by a male Tokyo Japanese speaker. The frequency scale is logalithmic.

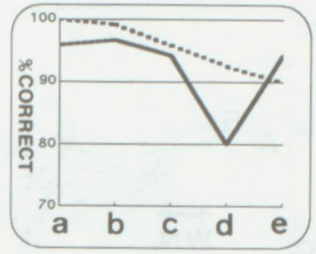

Fig.2 % correct identifications of wh question (real line) and non-wh question (dotted line). The abscissa represents the masking types indicated in the text.

(3) [na'niga mie'ru]
(4) [na'nika] [mie'ru]

In Fig.1, the peak F0 value of naniga is clearly higher than its nanika counterpart, and testifies to the presence of focus in the wh-word. This kind of focus-driven prominence in the wh-word is realized consistently, but it is by no means the only characteristic of wh-intonation. Rather, what makes the intonation shape of(1) visually distinct from that of (2) is the lack of salience of the prosodic boundary between NP and VP(a quick rise at the beginning of mieru). In short, there are two possible phonetic cues to the difference between (1) and (2): prominent F0 peak of the wh-word (Pw) and the salience of the prosodic boundary(Sb).

3. EXPERIMENT 1

The aim of the first experiment was to examine if native speakers of TJ can in fact discriminate the two question types solely by means of intonation.The difference of (1)and (2) consists in the /k/-/g/ consonantal contrast as far as the segmental tier is concerned. So it was expected that subjects would be forced to rely on prosodic cues if we erased these consonants and then filled the resulting silence with white noise. On this reasoning,the following ten stimuli were prepared. The underlines show the time stretch replaced with noise.

(1a) nanigamieru
(1b) nanigamieru
(1c) nanigamieru
(1d) nanigamieru
(1e) nanigamieru
(2a) nanikamieru
(2b) nanikamieru
(2c) nanikamieru
(2d) nanikamieru
(2e) nanikamieru

In erasing sequences of segments, care was taken to rid the effect of coarticulation as much as possible. Consequently,the white noise penetrates more or less into the final part of preceding segment and the beginning of following segment in all cases. All manipulation of original utterances, which were sampled in 10KHz/16bits

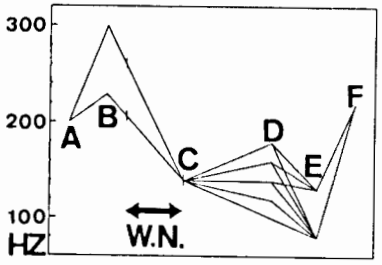

Fig. 3 Schematic structure of the synthetic stimuli. Control points A-F were linearly interpolated as a gross approximation to natural intonations. The thick arrow indicates the time stretch masked with white noise.

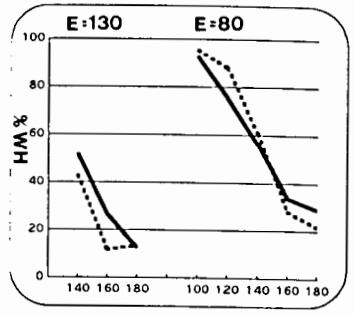

Fig. 4 % wh-judgments of sixteen synthetic stimuli as the function of the D values (abscissa). Real lines stand for the stimuli with B=300Hz (prominent wh), and dotted lines stand for those with B=230Hz(not prominent).

condition, was made on a computer. These stimuli were presented to eleven speakers of TJ in random order in a quiet listening condition. The subjects were requested to identify whether the utterance they heard was (1) or (2). No notice concerning the relevance of prosody was given. Fig.2 summarizes the result of the first experiment. Real and dotted lines show respectively the percentages of correct identification of wh and non-wh-question types. The overall average correct identification rate is quite high (92.2% for wh's and 95.5% for non-wh's), showing that natural utterances are full of prosodic cues. However, Fig. 2 provides us with little information about the relative importance of Pw and Sb. Both of these would seem to have equal importance in the identification task. (And it cannot be denied that cues other than the F0 shapes made certain contribution.)

4. EXPERIMENT 2

The aim of the second experiment was to examine the relative importance of Pw and Sb by using synthesized speech in which both cues were controlled. Fig.3 shows the schematic structure of the stimuli synthesized. A-F of Fig. 3 denote the points where the contour is controlled. Point A is the beginning of the utterance and is fixed at 200Hz. Point B is concerned with the cue Pw; its F0 value is either 300Hz or 230Hz. Point C stands for the beginning of the predicate mieru and is fixed at 140Hz. Point D is taken as representative of the cue Sb and is 180, 160, 140, 120 or 100Hz. Point E is the beginning of the sentence final rise and is either 130 or 80Hz. Point F is the target of the rise and is fixed at 220Hz. Of all the twenty combinations of the F0 values of B, D and E, the four combinations in which the E value is higher than the D value were eliminated because these give rise to intonational configurations which are impossible in TJ. The remaining sixteen intonation contours were synthesized by PARCOR method, using the PANASYNS program developed by Hiroshi Imagawa and Shigeru Kiritani. The stimuli were presented to the same listeners in the same manner as in the previous experiment. Fig.4 shows the percentages with which each stimulus was perceived as wh-question.The abscissa of the figure is a composite representation of D values for the stimuli with E=130 Hz (the leftward three values) and for the stimuli with E=80Hz (the rest). The real and dotted lines stand respectively for the stimuli with B=300Hz and B=230Hz. This figure shows clearly that the contribution of the D value is greater by far than that of the B value. Although a raised B value(300Hz) makes some contribution to subjects' judgment of wh-question,this effect is observed only when D is relatively high (180Hz or 160 Hz). Once D is set to relatively low values (120Hz or 100Hz), the stimuli were perceived mostly as wh-question irrespective of the B values.

5. DISCUSSION AND CONCLUSION

The two experiments reported here lead us to reconsider the phonetic nature of focus in TJ, stressing the importance of the salience of the prosodic boundary. In this respect, it is noteworthy that Fujisaki & Kawai[3] and Kori[4] have independently pointed out that focus not only increases the prominence of the focused constituent but also reduces the prominence of the following constituents. Kori also suggests that prominence of the final constituent of an utterance is more reduced than that of the other constituents. This analysis, which is based on production data, seems to be congruent with my perception data. Fig. 4 indicates that in order for a stimulus to be identified as a wh-question with 90% accuracy, it is necessary that the D value be lower than 120Hz i.e. lower than the right edge of the preceding NP. The data presented here and that of Kori and that of Fujisaki & Kawai suggest that any theory of phonetics that assumes that the effect of focus is limited only to the constituent marked as focused is inappropriate and to be revised. Finally, it should be pointed out that one important problem was left untouched: whether the difference of intonation examined in this study is specific to the pair of wh and non-wh questions. The line of reasoning that I followed in this study predicts that the difference is not a specific one. It is expected that the same intonational difference is observed in any pair of sentences having the same difference of focus placement as the one observed between (3) and (4).

6. REFERENCES

[1]Pierrehumbert,J.&M.Beckman (1988),Japanese Tone Structure. The MIT Press.
[2]Maëkawa,K.(1990), 'Muakusentohoogenno intoneeshon,' in Onseigengo, 4, 87-110.
[3]Fujisaki,H & H.Kawai (1988), 'Realization of linguistic information in the voice fundamental frequency contour of the spoken Japanese,'Ann.Bull.RILP,22,183-191.
[4]Kori,S.(1989), 'Kyoochooto intoneeshon,' in Sugito. M.ed. Kooza nihongoto nihongokyooiku,Vol.2. Tokyo, Meijishoin.

ACKNOWLEDGMENT

I am very grateful to Osamu Mizutani of NLRI for his comments on an earlier draft of this paper.