INTONATIONAL STRUCTURE OF KUMAMOTO JAPANESE: A PERCEPTUAL VALIDATION

Kikuo Maekawa
The National Language Research Institute
3-9-14, Nishiga'oka, Kita-ku, Tokyo, 115 Japan.
kikuo@tansei.cc.u-tokyo.ac.jp

ABSTRACT

A phonological model of intonation was proposed for Kumamoto *accentless* Japanese and the validity of the model was examined by means of perception experiment. Twenty intonation contours including both well-formed and ill-formed ones were synthesized from the original by means of LPC resynthesis. Naturalness judgments given by 19 native Kumamoto speakers indicated that well-formed contours were more natural than ill-formed ones.

1 INTRODUCTION

Kumamoto Japanese (KJ) is a regional dialect spoken in *Kumamoto* prefecture on the *Kyuushuu* island and is one of the most influential dialects in the *Kyuushuu* area. The dialect is of particular interest from a prosodic phonological point of view, because it is an *accentless* dialect. Following the tradition of Japanese linguistics, I will use the term *accent* to refer to the specification of pitch at the level of the lexicon. An accentless dialect is defined as a dialect which has no lexical specification of tone. In this particular sense, KJ is typologically similar to languages like Korean (Seoul dialect) in that it has neither lexical stress nor lexical tone.

The study of accentless dialects is one of the most underdeveloped areas of Japanese prosody. There is an undocumented but widespread belief that the pitch in some accentless dialects is not systematically controlled. One reason for this belief stems from the fact that traditional descriptive works on Japanese prosody have concentrated on the analysis of word level prosody, i.e. *accent*. Recent developments of intonational phonology combined with various techniques of speech analysis, however, have made it possible to examine the structure of accentless dialects from the phrasal perspective. In this paper, I present a phonological model of the intonation of KJ based on the theory of intonation developed mainly by Pierrehumbert and her colleagues[1-2], and examine the validity of this model by a perception experiment.

2 THE MODEL OF KUMAMOTO INTONATION

The model of KJ intonation described here is based on three assumptions. (1) Intonation is realized as the result of phonetic interpolation and smoothing of the abstract (phonological) tones distributed sparsely in the prosodic structure. (2) There are two types of phonological tone, H and L; the phonetic value of these tones, especially that of H can be modified by the application of phonological and phonetic rules such as downstep or upstep. (3) In accentless dialects, all the phonological tones are associated with one of the nodes in the prosodic structure.

The prosodic structure of KJ consists of two levels, which I will refer to using the labels of υ and α for the upper and the lower levels, respectively. The node υ introduces a couple of tones to the edges of the whole utterance; L_υ to the left and either L_υ or H_υ to the right. α also introduces a couple of tones, H_α and L_α. These tones are autosegmentally linked to some of the syllables of the utterance by the following tonal linking rules.

TL1: *The left hand L_υ is linked to the first syllable of the domain dominated by the node with which the tone is associated.*

TL2: *The right hand L_υ or H_υ is linked to the last syllable of the domain dominated by the node with which the tone is associated.*

TL3: *L_α is linked to the first syllable of the α which immediately follows the α with which the tone is associated. In case there is no following α, the tone is linked to the last syllable of the α with which it is associated.*

TL4: *H_α is linked to one of the syllables dominated by the α node with which the tone is associated.*

In addition, there are two general constraints on the application of the tonal linking rules.

Constraint 1: *Utmost two tones can be linked to the same syllable on condition that the two tones should be associated with different kind of nodes in prosodic structure.*

Constraint 2: *Linking of L_α has priority over linking of H_α at the left edge of an α. In contrast, linking of H_α has priority over that of L_α at the right edge.*

Examples of the prosodic structure of KJ together with the observed F0-contours are given in Figure 1. It is important to note that TL4 has nothing to do with the location of H_α linkage; this tone can be linked to any of the syllables which are dominated by the same α-node with which the tone is associated. There are two consequences of this indeterminacy of the peak tone location, which I will refer to "*wandering* H". First, KJ intonation has a considerably greater degree of freedom of peak location than most Japanese dialects, both accent and accentless. It is likely that this characteristic gave rise to the widespread erroneous impression that intonation in KJ is not linguistically controlled. Second, when the H_α is linked to a syllable which is near the end of an α-phrase which is

relatively long in terms of number of constituency syllables, the resulting intonation rises gradually over the whole phrase. This long rising pattern, as can be seen in panel A of Figure 1 and 2, is one of the most salient phonetic characteristics of the KJ intonation, and is rarely found in other Japanese dialects. Also, KJ has a rule of upstep, which raises the phonetic value [Hz] of the last tone of an utterance, i.e. either L_υ or H_υ. The rule applies whenever the L_α associated with the last α–phrase of an utterance is deleted under the combined effect of the two constraints stated above. Examples of the upstepping of L_υ and H_υ are given in panel A and B of Figure 1, respectively. Details of the KJ intonation and its model are given in reference [3].

3. THE PERCEPTION EXPERIMENT

3.1 STIMULI

A perception experiment was carried out in order to test the validity of the proposed phonological model. The stimuli used in this experiment consisted of twenty utterances of the KJ WH question utterance, *nanno miyu?* [Gloss: *nan*=what, *no*=OBJ, *miyu*=can be seen; What can (you) see?] differing only in the F0 contours. A WH question sentence was selected because in KJ short WH questions are usually realized as a single α-phrase, e.g., $(\text{nanno miyu?})_\alpha$. In contrast, corresponding Yes-No questions of the same length are always realized as two α-phrases such as $(\text{nanka})_\alpha$ $(\text{miyu?})_\alpha$ [=Can (you) see anything?] as shown in Figure 2.

The stimuli were synthesized from a single WH question utterance uttered by a young male KJ speaker using the LPC resynthesis technique. The schematized intonational structures of the stimuli, as shown in Figure 3, can be classified into five different categories. Category A consists of six well-formed contours, whose prosodic structure can be generated by the model stated above. Stimuli A1-A5 share the same prosodic structure disregarding the location of the *wandering* H. In A1, the tone is realized at the left edge of the vowel of the syllable of /nan/; in A2 it is realized at the right edge of the same vowel, and in A3, at the right edge of the syllable. In A4 and A5, the tone is at the right edge of the syllables /no/ and /mi/, respectively. A6 is generated by the deletion of the L_α, which is triggered by the linking of an H_α to the last syllable of its domain, according to the above constraints. In these cases, the deletion triggers the application of upstep.

Members of categories B-E are all ill-formed ones. Category B is concerned with inappropriate tonal deletions. B1 involves the deletion of the utterance initial L_υ, and B2 that of H_α. In B3, the deletion of H_α is followed by the unmotivated application of the upstep. In B4, both H_υ and H_α were deleted. Although members of category C can be generated by the phonological model, they are, from a syntax-phonology mapping point of view, inappropriate as the structure of a WH question sentence because they consist of two separate α-phrases. The peak locations of C1-C3 are the same as in A1-A3. Category D are concerned with inappropriate dislocations of L_α. The tone was realized at the right edges of the syllables /nan/ and /no/ in D1 and D2, respectively. Lastly, category E are inappropriate applications of the upstep rule. Members of this category are identical to the corresponding members of category A except for the height of the H_υ.

These stimuli were recorded randomly 5 to 11 times on an audio tape and were presented to the 29 native speaking subjects of KJ in the city of Kumamoto. The subjects consisted of two different age-groups, 19 high school and college students and 10 middle-aged high school teachers. They were asked to judge whether the utterances they heard were "natural" or "unnatural" WH question of KJ. Because no crucial age-graded difference was found, I will only discuss the student data in this paper.

3.2 RESULTS

Results of the perception experiment are summarized in Figure 4. The naturalness of the members of category A was the highest, thereby testifying the validity of the proposed model. Category E was the second highest. This is a natural consequence since the only difference between categories A and E was the phonetic value of the H_υ. Members of B can be split into two distinct groups; B1-B3 were judged to be more natural than B4 and B5, presumably because the former three were characterized by the final rising contour whereas the last two were not. The lowness of the naturalness rating of category C shows the fundamental importance of the proper mapping between the syntactic and the prosodic structures. The extreme lowness of the naturalness of category D also supports this point.

Difference in the naturalness among the members of Category A reveals two more interesting points. First, comparison among judgments for A1-A3 suggests that the canonical location of H_α is the right edge of a syllable. A similar conclusion can be drawn from a comparison of E1-3, whose peak locations are the same as those of A. Second, and more importantly, the naturalness of A5 and A6 are lower than the other members of the category. Close examination reveals that the relatively low naturalness judgments for these contours relates primarily to inter-individual differences among the subjects. Although 11 out of 19 subjects judged the contour of A6 to be 'natural' more than 5 times out of the total of 6 judgments, there were two subjects who judged A6 to be 'natural' less than two time. This result suggests that there are two types of KJ speakers, those who have the constraint on the deletion of L_α at the right edge of the last α and those who do not. On the other hand, the response pattern for A5 was more complicated and difficult to interpret. Although 5 subjects judged the contour to be 'natural' more than 5 times (out of the total of 6 judgments), most (11 out of 19) of the subjects' responses were distributed in the intermediate range, i.e. between 2 and 4 times. In this case, it is not possible to suppose the existence of individual differences of the constraints as was the case with A6. Rather, it seems that the low naturalness of the contour stemmed from a pragmatic inappropriateness in focus placement. The subjects might have thought that the focus was placed wrongly on the

predicate of /miyu/ in A5 because one semantic function of H_α in KJ is to indicate the location of pragmatic/discourse focus, and focus in an WH question is generally on the WH-word. This interpretation needs to be examined in a separate experiment.

4. CONCLUSION

Results of perception experiment show that Kumamoto Japanese has its own phrase level intonational structure and the native speakers of this dialect can perceive intonational difference in a subtle way. They were especially sensitive to the inappropriateness of the intonational phrasing (of α–phrase) when compared to other problems caused at the phonetic level. The considerable surface pitch variation of KJ caused by the phonological indeterminacy of the *wandering* H tone does not have much effect on naturalness judgments.

REFERENCES

[1] J. Pierrehumbert, *The Phonology and phonetics of English intonation*, MIT diss. 1980.

[2] J.Pierrehumbert and M. Beckman, *Japanese tone structure*, MIT Press, 1988.

[3] K. Maekawa, Akusento to intoneeshon: Akusento no nai chiiki (Accent and intonation in the area of accentless dialects), in *Nihongo onsei no shinkenkyuu*, Vol.2, Sanseido, Tokyo. To appear.

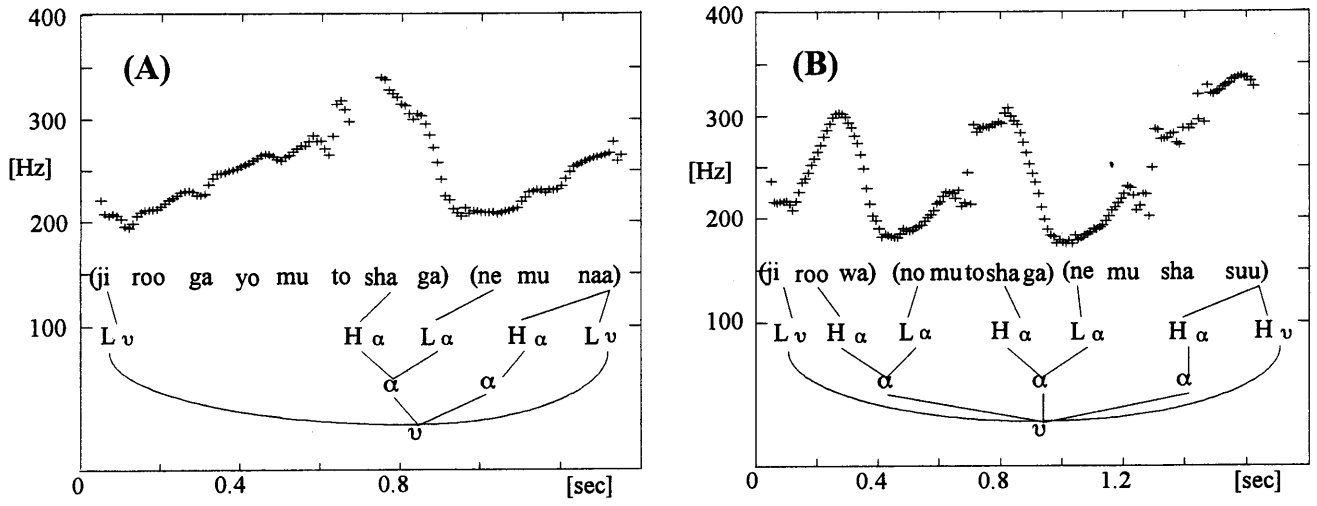

Figure 1. The prosodic structures and the observed F0 contours of two KJ utterances. (A) *jiroo-ga yumu-to-shaga nemu-naa* [When Jiro reads aloud, we become sleepy], and (B) *jiroo-wa nomu-to-shaga nemu-sha-suu* [When he drinks, Jiroo becomes sleepy] as uttered by a young female speaker. Parentheses shows the α-constituency. The last tone of the utterances were upstepped because the L_α of the last α-phrases were deleted.

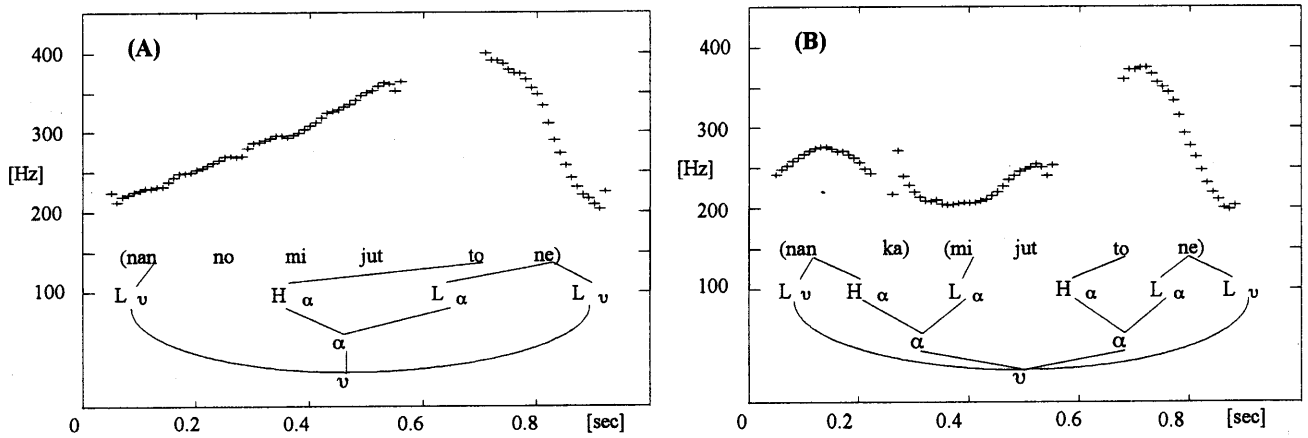

Figure 2. Comparison of α-phrasing in WH and Yes-No questions in KJ. (A) *nan-no miyutto-ne* [What can you see?], and *nanka miyutto-ne* [Can you see anything?]. The contours end with falling pitch because of they end with the emphatic particle of *ne*.

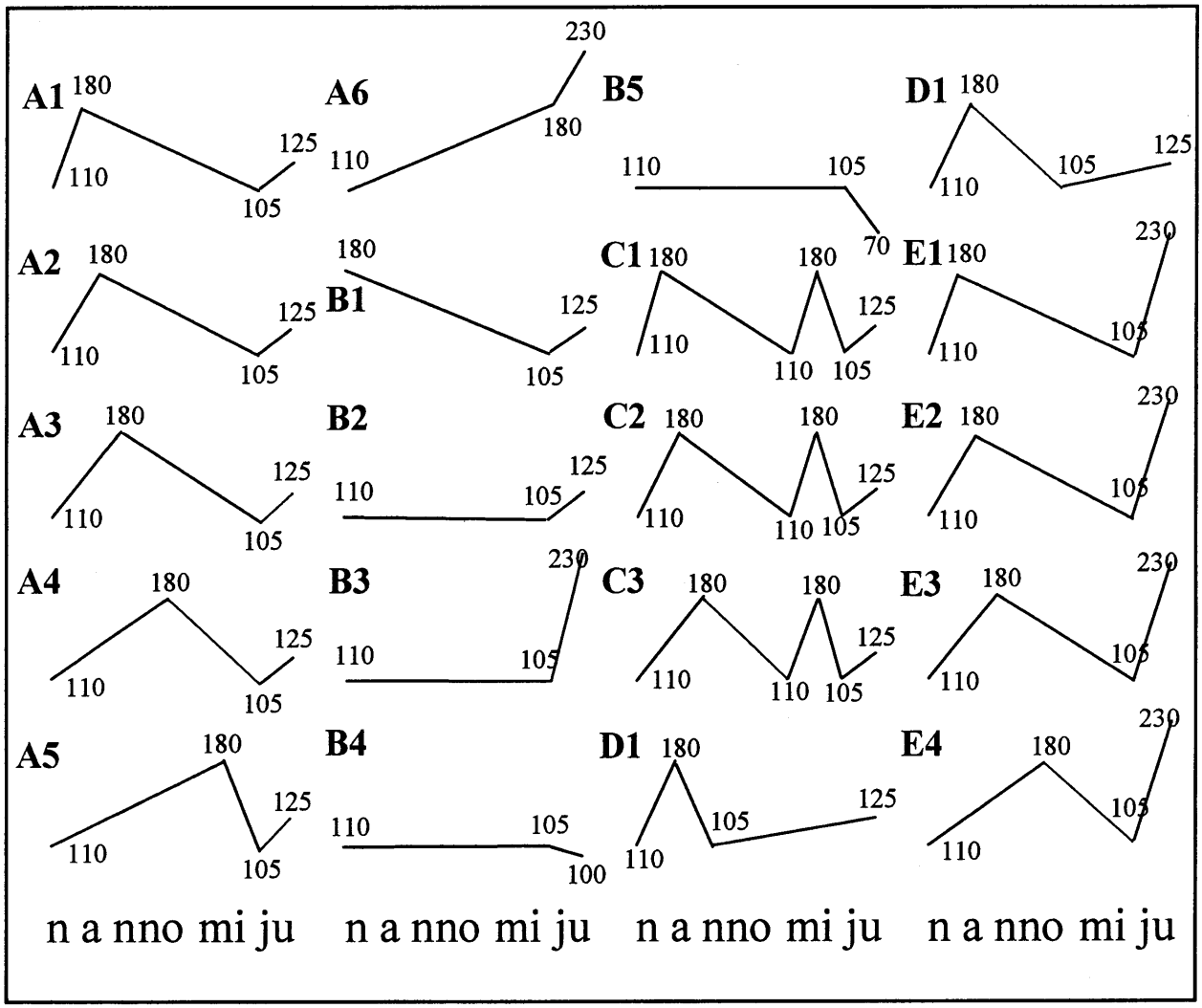

Figure 3. Structure of the twenty synthetic intonation contours used for the perception experiment. Digits in the figure show the F0 values [Hz] of the inflection points in the contours. These points were linearly interpolated.

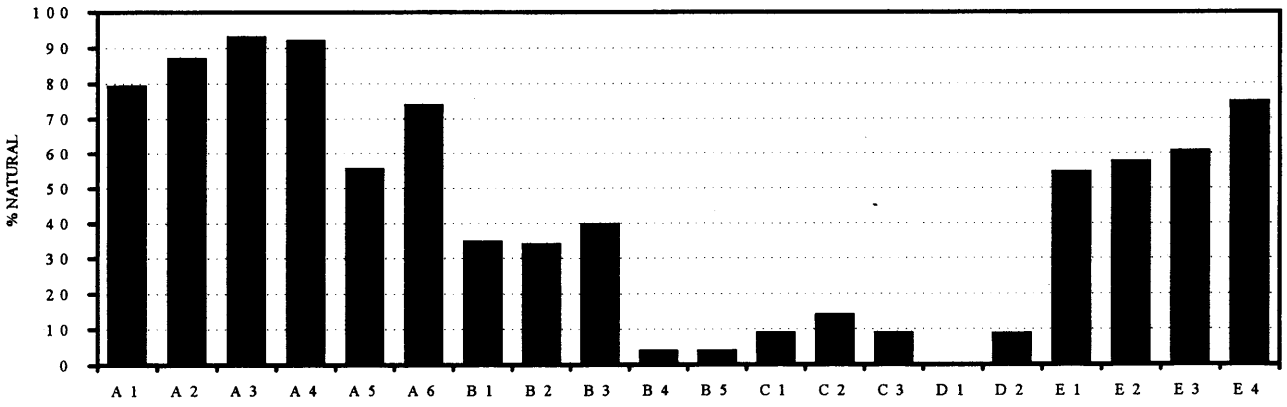

Figure 4. Results of the perception experiment. The abscissa gives the individual F0 contours used in the experiment and the ordinate, the percentage with which a contour was judged to be 'natural.' Pooled data of 19 student subjects.