Dept. for Speech, Music and Hearing
Quarterly Progress and Status Report

The location of the F0 turning point as a cue to mora boundary

Nagano-Madsen, Y. and Eriksson, L.

journal: STL-QPSR
volume: 30
number: 1
year: 1989
pages: 041-045

THE LOCATION OF THE FO TURNING POINT AS A CUE TO MORA BOUNDARY

Yasuko Nagano-Madsen and Lars Eriksson
Dept. of Linguistics and Phonetics, Lund University

Abstract

The perceptual significance of the Fo maxima and minima in relation to the mora boundary in Japanese was tested by shifting the original Fo turning points systematically both forward and backward by using the ILS program. The synthesized versions in which the points were shifted greatly in either direction were rejected but the shift to the following consonant allowed wider range of acceptance. The reasons for rejection may be categorized as tonal, intonational, and temporal change. Fall stimuli and rise stimuli had different sensitivity in perceptual response and words with moraic nasal showed almost no change in perception.

INTRODUCTION

This paper reports the results of a pilot study to test the perceptual significance of the "location of the Fo (fundamental frequency) turning point" in Japanese. The location of the turning point is one of the core concepts in the Lund Model of Intonation (Gårding, 1983) and has been found to be useful for all the languages that have been described within the framework.

Our observation is that the occurrence of the turning points in a language is most typically found either (1) on a vowel or (2) at and around the mora- and syllable-boundary depending on whether it is a mora or a syllable that is the unit of tone assignment in the language. Japanese (Fujisaki, 1983; Nagano-Madsen, 1987) and Eskimo (Nagano-Madsen, 1988) are the examples of mora type. For the Eskimo data, it has been suggested that the turning point may be used as a segmentation criteria for dividing a sequence of vowels into two moras instead of dividing it at mid point. For Hausa, Lindau (1986) remarked that the points may be regarded as a phonetic correlate of syllable boundary because of their regular fixation. Some several Kwa languages of the Niger-Congo family, including Yoruba and Igbo (Nagano-Madsen, forthcoming), are further examples of the syllable type. All these languages belonging to the second group seem to share the perceptual similarity that their accent or tone can be stated in terms of level tone as against contour tone.

The main questions we have asked in this study were:

1. how much freedom is granted for the turning point to move away from the mora boundary in order to be perceived as an original utterance and
2. what happens afterwards.

EXPERIMENT

Some bimoraic and trimoraic Japanese words with different vowel types were recorded by one of the authors (a speaker of the Kochi dialect). The utterance was meant as an answer to "what did you say?" It was spoken slowly and clearly, and also slightly emphatically. As the different vowel type and the number of mora did not show any obvious influence for the purpose of the present experiment, only the following bimoraic words were used for the actual perception test. For simplicity, we have used H and L description.

ama (HL) "heaven" / ama (LH) "a woman diver"
mama (HL) "mother"/ mama (LH) "food"
an (HL) "a plan"/ hen (LH) "strange"

The durations of the segment ranged as follows: the first vowel from 123 ms to 160 ms, the intervocalic [m] from 110 ms to 125 ms, and the second vowel from 160 ms to 210 ms. The duration of the moraic nasals was 185 ms and 340 ms, respectively. The original utterance was then manipulated to shift the Fo maxima and minima every 25 ms (5 frames) from the original mora boundary at four steps both forward and backward by using ILS editing programs developed at the department (cf. Fig. 1). Our basic principles of synthesis were based on the production data (Nagano-Madsen, 1987) We have chosen the constant elements as follows: (1) initial rise and the original fall for the HL pattern and (2) the original rise for the LH pattern. The rest was adjusted by extension or cutting of horizontal lines. In addition, a small bump or valley that appeared toward the end of an utterance was cut if it was at a high frequency region because when this portion was shifted to the central part of the vowel, it resulted in shaky voice quality. Synthesis made in this way sounded better than the versions in which the entire pitch contour was moved forward and backward.

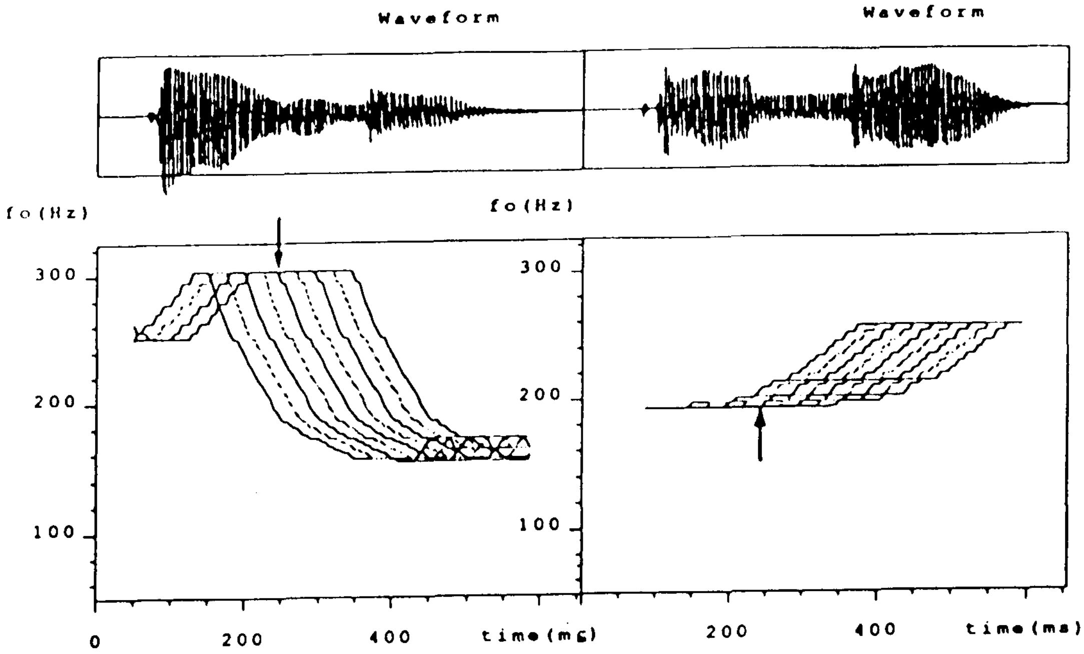

Fig. 1. Wave form and fundamental frequency contour of /ámal and /amá/. The arrow indicates the original version.

We obtained eight synthesized versions plus original, a total of nine versions for each word. A panel of five native Japanese listeners was given information as to the purpose of the experiment and how the stimuli were made. They listened to the original utterances and were asked to say the corresponding words themselves. Then some of the synthesized versions were presented as examples, and they were instructed to concentrate on the tonal and temporal impression rather than the quality of synthesis. They then listened to the randomized nine versions and evaluated them by four steps, e.g., (1) sounds very much like the original or it is perfectly acceptable, (2) it is o.k., (3) slightly different, and (4) definitely different. A response for (1) counted as three points, (2) as two points, (3) as one point, and (4) as zero point. A short interview was held after the experiment session in order to summarize the listeners' impression about the rejected versions.

RESULTS

The results are shown graphically in Fig. 2. The horizontal axis shows the nine stimuli numbered from the left (cf. Fig. 1), no. 5 being the original. The vertical axis shows the averaged responses calculated as described above. If all the five listeners chose (1), its score will be 15, while if all of them chose (4), the score will be zero. The higher the score, the higher the level of acceptance. The responses to *ama* and *mama* were very similar, therefore the results are pooled and shown by white squares. On the other hand, the responses to the words with moraic nasal were clearly different and shown separately by the black squares.

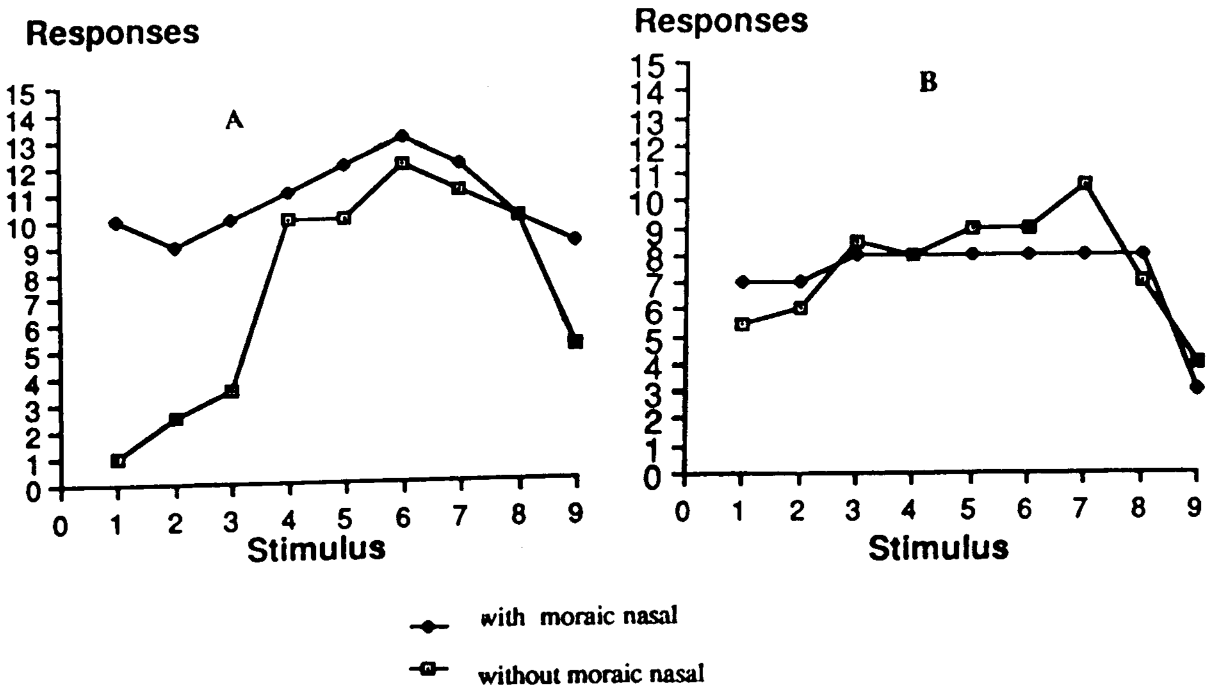

Fig. 2. Averaged responses for the fall (A) and rise (B) stimuli.

In general, we observe that the graph is not symmetrical but slightly right weighted. This means that the shift of the turning points to the right to the following consonant was more acceptable than the shift to the preceding vowel.

All the no 9 (rightmost) versions were rejected except for the *an* stimulus. The location of the turning point for the no 9 stimuli was shifted to the end point of the intervocalic [m] for both fall and rise stimuli. The leftmost stimulus had slightly different results depending on whether it was a fall or a rise. Nos 1, 2, and 3 were clearly rejected for the fall stimuli. The turning point for the no 3 version was 50 ms from the original position. As for the rise stimuli, nos 1 and 2 were rejected with less confidence. In other words the rise stimuli allowed slightly wider range of shift to the preceding vowel than the fall stimuli did. It looks like the listeners were more sensitive to the fall stimuli both at the level of confidence and at the point of perceptual change. Words with moraic nasal showed much less change in perception. Except for no 9 in *hen*, none of the versions were clearly rejected. When it had a rise stimuli, the response was at least as sensitive as indicated by almost a level line from no 1 to no 8.

DISCUSSION

The present experiment has shown that the great deviation of the turning point either to the preceding vowel or to the following consonant had an effect of making the original utterance sound different. Some of the perceptual features appear to be in good agreement with the production data. The shift of the points to the following consonants is often observed when an utterance is spoken emphatically. This may be the reason why nos 6 and 7 sometimes scored even higher than the original (no 5).

That the shift of the points for the rise stimuli earlier to the vowel than the fall stimuli is also in good agreement with the production data. The point of rise was more susceptible to change by tempo and tonal context (Nagano-Madsen, 1987).

Another perceptual correspondence to the production data is the case with moraic nasal. Although we have chosen the best sample in which the turning points appeared at V-N (moraic nasal) boundary, the acoustic data for these words were not always consistent, e.g., part of a moraic nasal was nasalized, etc. Perceptually, all except for the no 9 version of *hen*, sounded "just the same" to the listeners. These observations suggest that the words with moraic nasal have different tonal manifestation from the words without moraic nasals. Since the moraic nasal occurs only at syllable final position preceded by a vowel both in Japanese and in Eskimo, it may be the case that the vowel plays some role together with the following moraic nasal in tonal manifestation. In this respect, it may be interesting to compare the behaviour of moraic nasals with that of syllabic nasals in some Kwa languages as they only occur word initially followed by a consonant.

The rejected versions appeared to have different reasons. Based on the interview with the listeners and on our own analysis, we may classify them into three types:

1. Temporal impression was different, e.g., the original [a] sounded as bimoraic [aa] or prolonged [a:].
2. The tonal or intonational category was different.
3. Synthesized versions were not good enough as a natural speech, e.g., voice quality was funny.

The authors heard the changes in the above categories (1) and (2) quite systematic. Nos 1 and 2 fall stimulus had an effect to change the original H to HL tone and the original [a] to [aa]. Likewise the no 8 rise stimuli changed the original L to LH and the [a] to [a:]. The listeners described the change in no 8 as that of from statement to question. We think the reason why the no 8 stimulus was not rejected badly was that it sounded perfectly all right as a natural utterance and some listeners gave a good score for that.

Since a change in temporal impression is generally accompanied by a change in tonal impression, it is not easy to point out what exactly is responsible for the perceptual change. However, from the available data (Ainsworth, 1986), as well as a small indication in our data, we feel that temporal impression may be affected by clear pitch change as it was shown by the fall stimuli in the present experiment. The rise stimulus, on the other hand, was very gradual and it is doubtful if listeners can perceive such a gradual change at lower frequency regions clearly as a point of pitch change as such. This also reminds us of the different context in which Fo minima can occur. In a sequence of HLH, the minima is generally preceded by a steep fall and the point of rise is more sharply marked (cf. Hausa and Eskimo data). No 9 was heard more like LL and not LH (Rise).

The present data is in good agreement with the argument in House (1988) who considers for languages like Swedish and Chinese which use contour movement, it is important to time the Fo with C-V boundary so that the right movement will be on the vowel. The Japanese data suggest that it avoids having the movement in the vowel. House as well as Bruce (1977) consider the relative timing of fall in relation to the vowel rather than the location of the turning point itself as a relevant tonal event. Our data are also in good agreement with this view except that the location of the pitch change in fall stimuli (and presumably rise stimuli as well if preceded by a sharp fall) may be responsible for changing temporal impression. And, if this is the general phenomenon, languages in which quantity distinction is crucial as it is in Japanese and in Eskimo, the location of the turning point can not be arbitrarily assigned.

Since our data are only word citation clear speech, we are not sure if the same results will be obtained in more natural speech in which the duration of the segments are much reduced. In order to confirm some of the points made in this paper, further study is necessary.

Acknowledgement

Thank you Gösta (Bruce) and David (House) for much useful advice.

References

Ainsworth, W.A. (1986): "Pitch change as a cue to syllabification," J. of Phonetics **14**, pp. 257-264.

Bruce, G. (1977): *Swedish Word Accents in Sentence Perspective,* Travaux de l'Institut de Linguistique de Lund XII. Gleerups, Lund.

Fujisaki, H. (1983): "Dynamic characteristics of voice fundamental frequency in speech and singing," pp. 39-55 in (P.F. MacNeilage, ed.) *The Production of Speech*, Springer-Verlag, New York.

Gårding, E. (1983): "A generative model of intonation," pp. 11-25 in (A. Cutler & R. Ladd, eds.) *Prosody: Models and Measurements*, Springer, Heidelberg.

House, D. (1988): "Perceptual constraints and tonal features," paper presented at the Sixth International Phonology Meeting, Krems, Austria, 1-4 July.

Lindau, M. (1986): "Testing a model of intonation in a tone language," J.Acoust.Soc.Am. **80**:3, pp. 757-764.

Nagano-Madsen, Y. (1987): "Effect of tempo and tonal context on fundamental frequency contours in Japanese," *Working Papers 31* (Lund University), pp. 103-115.

Nagano-Madsen, Y. (1988): "Phonetic reality of the mora in Eskimo," *Working Papers 34* (Lund University), pp. 79-82.