THE COMBINED EFFECTS OF PROSODIC VARIATION ON JAPANESE MORA TIMING

Ann R. Bradlow, Robert F. Port, and Keiichi Tajima
Indiana University, Bloomington, Indiana, U.S.A.

ABSTRACT

Previous research has shown that the mora functions as a consistent timing unit in Japanese, such that there is a linear relationship between total word duration and the number of moras in the word. The present study extends the examination of Japanese as a mora-timed language by investigating the combined effects of variation in overall speaking rate, sentence-level focus, as well as number of moras on total word duration. The data indicate that Japanese timing is consistently constrained by the mora-timing principle. These findings have implications regarding the function of this articulatory-acoustic regularity from the listener's point of view.

INTRODUCTION

In both traditional and contemporary work, Japanese is described as a mora-timed language. As evidence of mora-timing in Japanese, traditional accounts point to the mora-based metrical structure of traditional Japanese poetry, to the *kana* orthographic system in which each symbol represents a mora-sized unit, and to the general native-speaker awareness of the mora ("onsetsu" or "haku") as a unit of equal timing. More recent instrumental analyses have focused on finding acoustic evidence of Japanese mora timing. Although Beckman [1] found that moras of varying segmental content did not exhibit the expected durational consistency, Port et al. [2] and Han [3] showed a linear relationship between the number of moras in a word and the total word duration. Data from these latter studies indicated that the mora-based regularity of Japanese timing is expressed at the level of word-sized units: words of a given number of moras all achieve a constant duration which is directly proportional to the number of moras in the word. Both of these studies revealed durational compensation at the "sub-moraic" level in order to achieve a target total word duration that is consistent with this mora-timing principle. The present study was designed to explore the effect on mora-timing of prosodic variation at the "supra-moraic" level. Specifically, we investigated the combined effects of variation in overall speech rate and sentence-level focus on the duration of words with varying numbers of moras. Our expectation was that the linear relationship between total word duration and number of moras in the word would persist even under conditions that might otherwise be expected to perturb this regularity in the time dimension. Such a result would provide further evidence regarding the extent to which Japanese mora timing is regular and immutable, and thus likely to form the basis of linguistic segmentation in both the production and perception domains.

METHODS

The "supra-moraic" prosodic factors of interest in this study were overall speaking rate and sentence-level focus. In order to manipulate these parameters, a set of nine paired sentences was compiled. In each pair, one sentence placed the target word in a position of sentence-level focus (contrastive focus), whereas the other sentence placed the target word in a neutral position (broad focus). Contrastive focus was achieved by constructing sentences that contrasted the target word with another word in the sentence. An example of a sentence pair is given below in Table 1.

Table 1. Example of a sentence pair with the target word in bold.

Contrastive focus:
Tanaka san wa umibe ***ga*** suki na no de wa naku, umibe ***mo*** suki na no de su.
'It's not that Mr. Tanaka *only* likes the seaside, it's that he ***also*** likes the seaside.'
Broad focus:
Tanaka san wa umibe ***mo*** suki na no de su.
'Mr. Tanaka ***also*** likes the seaside.'

The target words consisted of one, three, or five moras, with three tokens for each mora length. Thus, there were three tokens for each of three mora lengths for each of two focus conditions giving a total of 3x3x2=18 sentences. Two separate lists were prepared: one for the "contrastive focus" sentences and one for the "broad focus" sentences.

Four native Japanese speakers, two males (both age 32 years) and two females (age 32 and 29 years), served as subjects. All speakers were employees at ATR Research Laboratories in Kyoto, Japan, and did the recordings during regular work hours. Although all of the subjects were currently living in the Kyoto prefecture, they were all originally from the Tokyo region and were considered speakers of the Tokyo dialect.

The speakers were recorded in an anechoic chamber at ATR Human Information Processing Research Laboratories in Kyoto, Japan. Each speaker read the two randomized lists of sentences three times: first at a "normal" speaking rate, then at a fast rate, and finally at a slow rate. No attempt was made to ensure that all speakers spoke at the same rates in an absolute sense; rather, the focus of our attention was on achieving three different rates for each speaker in a relative sense. Different randomization orders were used for each subject and for each rate. The order of reading the "contrastive focus" and the "broad focus" lists was consistent across speaking rates for each subject, but was counter-balanced across subjects.

The recordings were digitized and analyzed using the Entropic Signal Processing System on a SUN SPARCstation 5 in the Speech Research Laboratory at Indiana University. The target word in each sentence was marked by time cursors in the waveform. Spectrographic displays were used in conjunction with the waveform displays to determine the onset and offset of the target word. Total word durations and peak fundamental frequencies of the target words were extracted from these labeled portions of the speech files.

RESULTS

Duration

Figure 1 shows the effect of speaking rate and sentence focus on total word duration (in milliseconds) as a function of the number of moras in the word. This plot shows the data for all four speakers pooled; however, the general pattern is the same for each individual speaker. As seen in this plot, at each of the three speaking rates there is a linear relationship between number of moras and total word duration. This plot also indicates that this relationship is not affected by whether the target word is embedded in a sentence with contrastive focus or not. A four-factor ANOVA with Speaker (four levels), Number of Moras (one, three, five), Focus (contrastive, broad), and Speaking Rate (fast, medium, slow) as factors was performed. This analysis revealed a main effect of Number of Mora ($F(2,144)=1352.8$, $p<.01$), of Speaking Rate ($F(2,144)=399.0$, $p<.01$), and of Speaker ($F(3,144)=64.0$, $p<.01$); however there is no main effect of Focus ($F(1,144)=.269$, $p=.60$). The main effect of Speaker is due to individual differences in overall speaking rate, since absolute speaking rate was not controlled across speakers and speakers differed in their "normal" speaking rate. Speakers also differed in the extent to which the fast and slow speaking rates differed from the medium rate; however, for each of the four speakers we find the same significant pattern of results that we find for the pooled data. In all cases, there is a consistent linear relationship between total word duration and number of moras in the word, and, this linearity remains unperturbed by variation in speaking rate and by variation in focus condition.

This observed linearity confirms that in Japanese the time dimension is constrained by a strict principle of mora timing. In light of the finding that variation in sentence-level focus is not reflected in the acoustic signal by durational differences, we performed a comparison of peak fundamental frequency for the target word across the two focus and three rate conditions. Our expectation was that the contrast between the target words in the contrastive-focus versus broad-focus sentences would be reflected in the acoustic signal by a difference in fundamental frequency peak (in Hertz). Such a finding would confirm that the sentence pairs we used in this study were effective in eliciting a

difference between contrastive- versus broad-focus sentences in the acoustic-phonetic domain despite the lack of a difference in duration. This would in turn validate our interpretation of the consistency of mora timing in Japanese, even under conditions which might be expected to affect the time dimension.

Fundamental frequency

Figure 2 shows the peak fundamental frequency (in Hertz) for the target words in the contrastive- and broad-focus sentences as a function of number of moras (top panel) and as a function of speaking rate (bottom panel). These plots show the data for all four subjects pooled; however, the general pattern is the same for each individual speaker. As seen in both panels of Figure 2, the target word F0 peak in the contrastive-focus condition is consistently higher than in the broad-focus condition. Furthermore, the F0 peak is not affected by overall speaking rate (bottom panel). In an ANOVA with Speaker, Focus, Number of Mora, and Rate as factors, there is a main effect of Speaker ($F(3,144)=360.5$, $p<.01$), reflecting the individual differences in fundamental frequency range, a main effect of Focus ($F(1,144)=9.9$, $p<.01$), and a main effect of Number of Moras ($F(2,144)=23.5$, $p<.01$) possibly reflecting the effect of the different segmental structures of the particular tokens. There is no main effect of Rate ($F(2,144)=1201.8$, $p=.12$). Thus, whereas total word durations remain unaffected by variation in sentence-level focus, fundamental frequency peak does vary according to this factor.

SUMMARY AND DISCUSSION

The data in this study provide additional evidence that, in Japanese, phonetic variation in the time dimension is severely constrained by the principle of mora-timing. As shown in earlier work (e.g. [2] and [3]) segment durations at the sub-moraic level exhibit compensatory lengthening and shorting in order to achieve a target total word duration. The present study extends this work by providing evidence of the consistency of mora timing in Japanese in the face of supra-moraic prosodic variation. This constrained acoustic-phonetic structure of Japanese timing is in contrast to a language such as English where the time dimension simultaneously reflects various prosodic features [4]. Additionally, this timing regularity in the acoustic domain raises questions about its usefulness for Japanese listeners. In a series of perception experiments with English and Japanese subjects, Cutler and Otake [5] investigated whether the Japanese listeners' exhibited a sensitivity to mora boundaries in a phoneme monitoring task. Their results showed different listening strategies for the Japanese and English subjects that could be traced to the importance of moras in Japanese versus the importance of syllables in English. This result, in conjunction with the acoustic data, suggests that the Japanese mora-timing principle is an example of a general linguistic timing principle that is under the speaker's control, and to which the listener is sensitive.

ACKNOWLEDGMENTS

We are grateful to Chiemi Kuroki for help with the recordings, to Reiko A. Yamada for help with the sentence lists, and to the speakers for their time. All are affiliated with ATR Human Information Processing Research Laboratories in Kyoto, Japan. This work was supported by NIDCD Training Grant DC-00012, by NIDCD Research Grant DC-00111, and by ONR N0001491J-1261 to Indiana University.

REFERENCES

[1] Beckman, M. (1982), "Segment duration and the 'mora' in Japanese," *Phonetica* , vol. 39, pp. 113-135.

[2] Port, R., Dalby, J., & O'Dell, M. (1987), "Evidence for mora timing in Japanese," *J. Acoust. Soc. Am.*, vol. 81, pp. 1574-1585.

[3] Han, M. (1994), "Acoustic manifestations of mora timing in Japanese," *Journal of the Acoustical Society of America,* vol. 96, pp. 73-82.

[4] Behne, D. and Nygaard, L. (1992) "Concurrent effects on duration I: Vowels," *Research on Speech Perception Progress Report No. 17.* Bloomington, IN: Speech Research Lab., Indiana Univ.

[5] Cutler, A. and Otake, T. (1994), "Mora or phoneme? Further evidence for language-specific listening," *Journal of Memory and Language,* vol. 33, pp. 824-844.

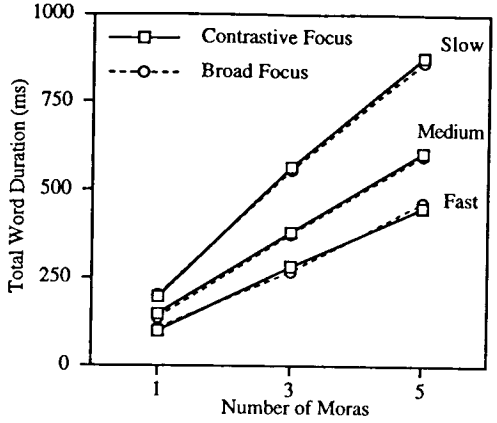

Figure 1. Total word duration as a function of number of moras for words in sentences with contrastive and broad focus at three speaking rates.

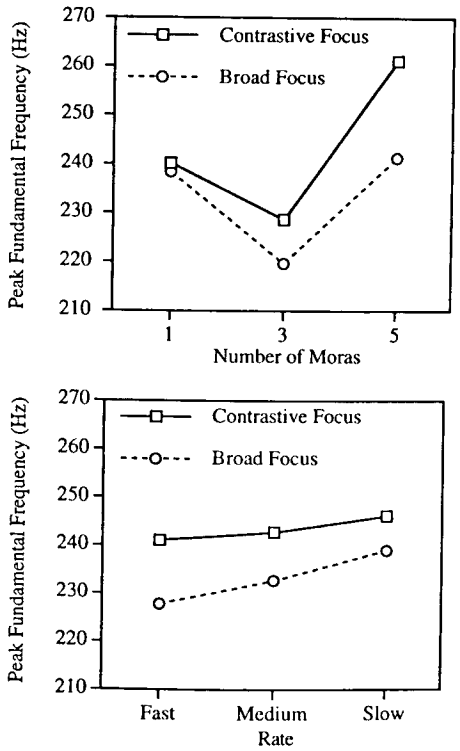

Figure 2. Peak fundamental frequency as a function of number of moras (top panel) and speaking rate (bottom panel) for words in sentences with contrastive and broad focus.