Durational variations of Japanese long / short vowels in different speaking rates: analysis of a spontaneous speech corpus

Yasuyo Minagawa[†‡], Takayuki Kagomiya[†] and Kikuo Maekawa[†]
†National Institute for Japanese Language, Japan
‡ Japan Society for the Promotion of Science, Japan
E-mail: {minagawa, kagomiya, kikuo}@kokken.go.jp

ABSTRACT

This study examines the temporal variations of Japanese long and short vowels due to differences in accentual conditions, speaking rates and syllable positions by analyzing a large-scale database called the "*Corpus of Spontaneous Japanese"* or CSJ. Analysis revealed a phrase-final lengthening as well as weak but significant effects of accent conditions for long/short vowels and bi-/mono-moraic syllables. Speaking rates consistently influenced temporal features, however these effects were more pronounced in the long vowels than in the short ones, resulting in shortening of the long to short vowel ratios due to an increment of speaking rate. These differences were only significant between the slowest category, Category 1 and the rest of the categories, which means that Japanese durational contrast remained constant in normal, fast and very fast speaking rates, even in spontaneous speech.

1. INTRODUCTION

Japanese is commonly referred to as a mora-timed language in which the duration of each mora is kept roughly constant. Acoustical as well as perceptual evidence of Japanese mora-timing has been explored in many studies from various points of view as reviewed by Warner & Arai (2001a). Some researchers reported that mora-timing was evidenced in the linear relationship between word duration and the number of morae constituting the particular word, while others showed strong tendency towards isochronal moraic unit by compensating the duration of adjacent segments in Japanese. Although perceptual or cognitive approaches of verifying moraic unit are relatively limited, a recent study revealed neurophysiological correlates to a durational contrast based on mora-timed Japanese (Minagawa-Kawai et al. 2002).

The results of these previous studies, particularly acoustical studies have frequently differed or been contradictory of each other, depending on the studies and the definition of mora-timing has varied or sometimes been revised. However, some characteristics regarding timing control in Japanese as compared to other languages with different rhythmic control seem to be clear and undoubted. That is, the extent to which various factors such as prosody, articulatory constraints and grammatical context influence the moraic (syllable) duration, is relatively smaller than in other languages. More specifically, accentual factor (Kaiki et al. 1992), focus (Bradrow et al. 1995), the conditions of syllable position within a word and grammatical word types (content word vs. function word) do not considerably influence the moraic duration. Furthermore, the differences in syllable length due to intrinsic duration of phonemes tend to be smaller in Japanese than in other languages (Minagawa-Kawai 1999). This is in clear contrast to the results of other languages where these factors of different levels crucially influence the syllable duration (Umeda 1975, Crystal & House 1982, Bartcova & Sorin 1987, Hoequist 1989). These relatively invariable traits of Japanese mora, which may be derived from the phonological constraints of durational contrast, could contribute to give rise to the impression of mora-timed rhythm, at least in read-speech.

Not many empirical results have been revealed about the nature of timing control in spontaneous Japanese, except for some reports as in Warner & Arai (2001b) which demonstrated the weaker role of mora as a timing unit. It is very likely that several influential factors regarding duration mentioned above could have different effects, probably stronger effects, on moraic duration in spontaneous speech, though these factors have not been examined thoroughly. Consequently, the present study investigates the durational variations of Japanese long/short vowels and bi-/mono-moraic syllables in spontaneous speech that are due to several phonetic and prosodic factors, i.e. pitch-accent, syllable positions and speaking rates. In order to examine sufficient data on samples which are spoken under natural conditions by speakers with a uniform background, the large-scale speech database called the *Corpus of Spontaneous Japanese* (CSJ) was used for the analysis.

2. METHODS

2.1. The CSJ

The entire CSJ database which comprises about 650 hours of speech and is a compilation of over seven million words, was designed for use in speech recognition as well as for studies on phonetics or linguistics. The main body of the corpus is monologues elicited under two different conditions: academic presentation speech (APS) and simulated public speech (SPS). The speech in the SPS is complied from standard Japanese speakers with a balanced

ISBN 1-876346-48-5

representation of age, and gender. A subset of the corpus, called the *Core*, is provided with segmental labeling and X-JToBI intonation labels (Maekawa et al. 2002) in addition to transcriptions and morphological information that all the corpus have. The segmental labels contain various tags including standard phonetic labeling, filled pauses, pauses, disfluency and other non-verbal voicing information. For a detailed segmentation procedures and criteria, please refer to Maekawa et al. (2000) and Kikuchi et al. (2003).

2.2. The Current Data

Of this large corpus that has been developed, the present study used about five hours of speech from the SPS databank which was fully furnished with labels. The subjects in the data sample this time included seven male and six female speakers all thirty to forty years in age. They were from Tokyo and its environs and spoke so-called Standard Japanese.

Among this data, we excluded some types of segments to make more adequate data in terms of the purpose of this study. Excluded items were words or vowels labeled "(F) filled pauses", "(D) fragmented words", "(L) whispers" "(W) mispronounced words", "<H> prolonged word final vowel that functions as filled pauses" and "<FV> uncertainty of phonetic quality of vowels used as filled pauses". Dipthongs such as /ai/ and /ae/ and devoiced vowels were also excluded.

Syllable positions in an accentual phrase were determined using BI (Break Indices) labels higher than "2" including labels specific to X-JToBI system like "2+p". Accent positions were also determined based on X-JToBI labels. Utterance boundaries in the present corpus were where a pause of longer than 200ms followed. Boundaries where the typical sentence-ending forms of predicates follow a short pause longer than 50ms, were also regarded as utterance boundaries.

Speaking rates were calculated in each utterance unit and the current study designated four speaking-rate categories according to the numbers of mora counted per second: Category1, speaking rates less than 7.5 mora/sec; Category 2, 7.6-8.4 mora/sec; Category 3, 8.5-9.4 mora/sec; Category 4, more than 9.5 mora/sec. In addition to this type of speaking-rate category calculated from all of the data, individual speaking–rate categories were ranked according to the four levels within individual speech rate and they were also considered for the analysis.

3. RESULTS

Since the results analyzed using speaking-rate categories determined from all of the data and those determined from the individual data did not differ greatly, the present study showed the results of the former analysis. In the current study, we employed two types of speech unit for measurements, they are, vowel and mora units. Here, the results given from both types of analysis are demonstrated.

Figure 1 indicates the durations of short vowels (A) and long vowels (B) according to the syllable positions, accentual conditions and the speaking rates. Accented mora

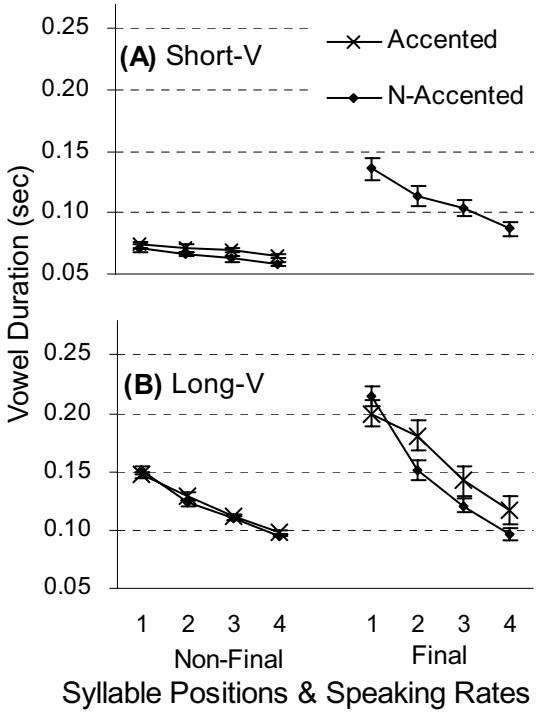

Figure 1 Durations of short (A) and long (B) vowels according to the accent conditions (accented, not-accented), the syllable positions and the speaking rates.

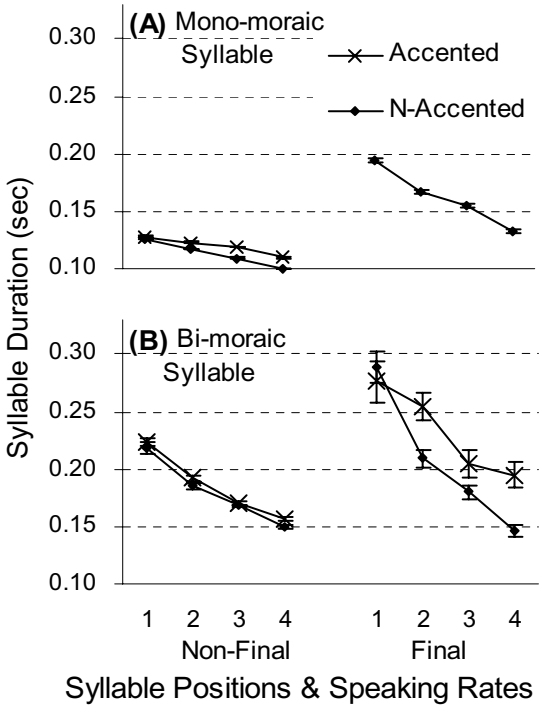

Figure 2 Durations of mono-moraic (A) and bi-moraic (B) syllables according to the accent conditions (accented, not-accented), the syllable positions and the speaking rates.

rarely occurs at phrase-final, thus short vowels in that condition were excluded.

The durations of short vowel were longer in the phrase final position than in the non-final positions. They tended to be shorter as the speaking rates increased; however this tendency was weak in the non-final position. These tendencies were confirmed by ANOVA with accent condition (accented, unaccented) and speaking rate (1, 2, 3, 4) as factors. Main effects were significant for both the accentual conditions (AC) and the speaking rates (SR) (AC, $F(1, 35895) = 266.5$, $p<0.0001$; SR, $F(3, 35895) = 395.1$, $p<0.0001$) and a significant interaction between AC and SR (AC*SR, $p<0.0001$) was found.

The effects of AC and SR on the mono-moraic syllable durations (Fig. 2 A) were relatively comparable to those observed in the vowel durations and the results of ANOVA indicated significant main effects for both of the factors (AC, $F(1, 35895) = 193.5$, $p<0.0001$; SR, $F(3, 35895) = 786.5$, $p<0.0001$) and significant interactions for AC*SR ($p<0.0001$).

Durational characteristics of the long vowel durations (Fig. 1 B) and the bi-moraic syllable durations (Fig. 2 B) seem to be roughly similar to those observed in the analyses of their short counterparts (Fig.1A, 2 A). The difference is the stronger effects of SR in the panels B than in the panels A. In the long vowels at the non-phrase-final position, there are no clear differences in duration between the accented and unaccented vowels, whereas the accented vowels are longer than the unaccented ones in the phrase-final position.

The results of ANOVA conducted for the long vowels with AC, SR and syllable position (SP) as factors revealed significant main effects for SP ($F(1, 4657) = 114.8$, $p<0.0001$) and SR ($F(3, 4657) = 123.3$, $p<0.0001$) with significant interactions for SP*SR ($F(3, 4657) = 12.6$, $p<0.0001$) and AC*SR ($F(3, 4657) = 114.8$, $p<0.001$).

The accent effects were relatively stronger for the durations of bi-moraic syllables (Fig. 2 B) than for those of the long vowels (Fig.1B). In the bi-moraic syllables, main effects were found for AC ($F(1, 4657) = 20.5$, $p<0.0001$) as well as SP ($F(1, 4657) = 112.3$, $p<0.0001$) and SR ($F(3, 4657) = 136.6$, $p<0.0001$). Interactions were significant for SP*AC, SP*SR ($p<0.0001$) and AC*SR ($p<0.001$). A post-hoc test showed that the accented syllables were significantly longer than the unaccented ones in almost all the conditions.

Since inherent duration of vowels could influence the results, z-scores that normalize the durations of all vowel types in the present data were calculated. Using the z-scores as variable, we also performed the same ANOVA as we did for the vowels and virtually the same results as obtained in the vowel analysis were found.

Figure 3 indicates one of the typical subject cases regarding durations of long and short vowels plotted according to speaking rates. The faster the speaking rate becomes, the shorter the durations of long and short vowels are. The slope of regression curve is steeper for the long vowels than for the short vowels, indicating that the effect of speaking rate is larger in the long vowels.

The ratio of long to short vowels and the ratio of bi-moraic to mono-moraic syllables in each subject were calculated and Figure 4 and 5 indicate their averaged ratios as functions of syllable position and speaking rate. For the ratio in the non-phrase-final position, ANOVA using AC and SR as factors was conducted and it was found that

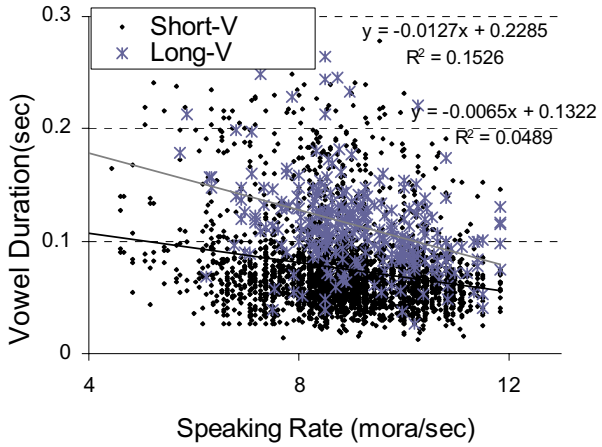

Figure 3 Correlations of durations of long/short vowels and speaking rates in one subject.

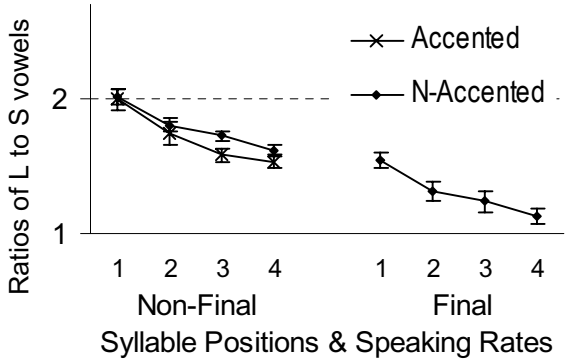

Figure 4 Ratios of long to short vowel durations as functions of syllable position and speaking rate in different accentual conditions

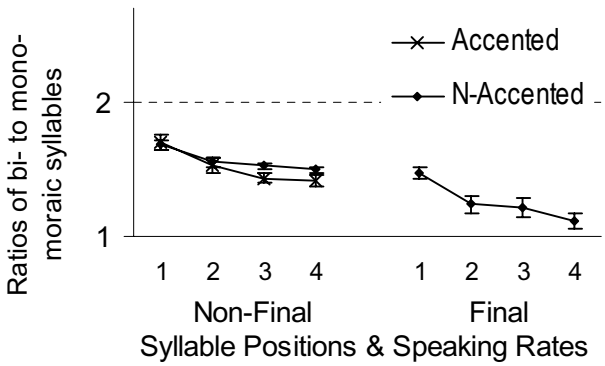

Figure 5 Ratios of bi- to mono-moraic syllable durations as functions of syllable position and speaking rate in different accentual conditions.

ISBN 1-876346-48-5

in both the vowel and syllable analyses, the main effects were significant only for SR (vowel ratio, $F(1, 168) = 64.3$, $p<0.0001$; mora ratio, $F(1, 168) = 12.3$, $p<0.0001$). A post-hoc test revealed that significant differences in syllable ratios were only existed between the SR Category 1 and the rest of the three categories, whereas in the vowel ratio all of the pairs in the four SR categories had significant differences except for the SR Category 3 and SR Category 4. This suggests that the bi-moraic to mono-moraic syllable ratio does not vary in 75% of the current data set, which corresponds to speaking rates faster than 7.5 mora/sec.

4. DISCUSSION

4.1. Effect of syllable position

Similar to the previously reported findings (Weismer & Ingrisano 1979, Takeda et al.1989), phrase-final vowel lengthening was observed in the current data. Compared to the non-phrase-final mono-moraic syllables, the phrase final syllables prolonged from 1.3 to 1.5 times. Moreover these ratios decreased as speaking rate increased. This is essentially opposite to the pattern found in English where phrase-final vowel lengthening become greater with a faster speaking rate (Smith 2002). Such difference in temporal patterns may derive from language-specific factors of timing control due to phonological reasons.

4.2. Accentual effect

As a temporal feature of Japanese, the correlation between accent pattern and vowel duration was reported to be weak in a read-speech analysis (Kaiki et al. 1992). The present data on spontaneous speech revealed that accent had a significant effect on short vowel durations, yet these differences due to accent pattern were small, i.e. differences ranging from 3-14ms in non-final position. Accentual effects may be more emphasized in spontaneous speech than in read speech, particularly for long vowels in phrase-final position.

4.3. Comparisons of long and short vowels

Both averaged data and individual analysis (Fig. 3) on the length of long and short vowels as a function of speaking rate showed that an increased speaking rate had a relatively greater effect on long vowel (bi-moraic syllable) duration. The slope values of regression curves (Fig. 3) were higher for the long vowels than for the short vowels and thus these curves get closer as speaking rate increases. However, these curves do not merge until a point where speaking rate becomes faster than about 14 (mora/sec). Such speaking rates only occur at 0.1% of probability in the much larger CSJ subset containing 884k-word (Maekawa, 2003). This finding indicates that the phonemic length contrast in Japanese is maintained even in a very fast speaking rate.

Because of the greater speaking rate effects of the long vowels, the ratios of long to short vowels (bi-moraic to mono-moraic syllables) have a tendency to decrease as speaking rate increases. However, further detailed analysis regarding the differences among speaking rate categories revealed that significant differences exist only between a SR Category "1" and the rest of the categories. Consequently, it can be concluded that although the ratios of Japanese bi-moraic to mono-moraic syllables differ between slow and normal speaking rates, in general the ratios remain constant in normal, fast, and very fast speaking rates even in spontaneous speech.

5. CONCLUSIONS

The current work showed some temporal features of Japanese long and short vowels in the spontaneous speech corpus. The accentual conditions had a weak but significant effect on durations of vowels and morae. A durational contrast of long/short vowels was revealed to be maintained even in a very fast speaking rate.

ACKNOWLEDGEMENTS This study was supported by Japan Society for Promotion of Science (No.08484).

REFERENCES

[1] A.R. Bradrow, R.F. Port and K.Tajima, *Proc. of ICPhS*, Vol.4, pp.344-347,1995.

[2] B. L. Smith, *Phonetic*a, 59, pp.232-244, 2002.

[3] C. Hoequist, *Phonetica*, pp.203-237,1983

[4] G. Weismer and D. Ingrisano, *J.Speech Lang. Hear. Res*. 40, pp.858-866, 1979.

[5] H. Kikuchi and K. Maekawa, *Proc. of ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, (To Appear) 2003.

[6] K. Bartcova and C.Sorin, *Speech Communication*, 6, pp.245-260,1987.

[7] K. Maekawa, H. Koiso, S. Furui, et al. *Proc. of LREC*, vol.3, pp.947-952, 2000.

[8] K. Maekawa, H. Kikuchi, Y. Igarashi, et al. *Proc. of ICSLP*, vol. 3, pp.1545-1548.2002.

[9] K. Maekawa, *Proc. of ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, (To Appear) 2003.

[10] K. Takeda, Y. Sagisaka and H. Kuwabara, *J. Acoust. Soc. Am.*,86, pp.2081-2087.

[11] N. Kaiki, K. Takeda and Y. Sagisaka, *IECE Trans*, J75-A: 7, pp.467-473, 1992.

[12] N. Warner and T. Arai, *Phonetica*, 58, pp.1-25, 2001a.

[13] N. Warner and T. Arai, *J. Acoust. Soc. Am.*, 109, pp.1144-1156, 2001b.

[14] N. Umeda, *J. Acoust. Soc.Am.*, 58, pp.434-445, 1975.

[15] T.H.Crystal, & A.S. House, *J. Acoust. Soc. Am.*, 83, pp.1574-1585, 1987.

[16] Y. Minagawa-Kawai, *Proc. of ICPhS*, pp.365-368, 1999.

[17] Y. Minagawa-Kawai, K. Mori, I. Furuya, et al. *Neuroreport,*vol.13, pp.581-584, 2002.