VOWEL-ONSET INTERVAL AS A TIMING UNIT FOR SINGLETON/GEMINATE STOP DISTINCTION IN JAPANESE

Yukari Hirata and Connor Forbes

Colgate University, Department of East Asian Languages and Literatures
yhirata@mail.colgate.edu

ABSTRACT

This study examined whether vowel onset intervals in Japanese disyllables can serve as a reliable timing unit in Japanese. Duration of vowel onset intervals was measured in Japanese disyllables containing singleton and geminate stops spoken in a carrier sentence at three speaking rates. Analysis included the vowel onset interval duration divided by the mean mora duration, and this value was predicted to show timing regularity in terms of the mora, according to Brady et al. [1]. Results supported this prediction.

Keywords: Japanese, mora, singleton/geminate stops, V-onset interval, Vmora

1. INTRODUCTION

Consonant length (singleton or geminate) is phonemic in Japanese, e.g., /kako/ (‘past’) vs. /kakːo/ (‘parenthesis’), and there have been numerous studies characterizing its timing structure in terms of acoustic manifestation of duration (as well as intensity and fundamental frequency) [2, 3, 4, 6, 7]. Japanese has been considered as a mora-timed language [4, 11, 13], and this consonant length contrast provides interesting tests for the phonetic and acoustic reality of moras. One long-lasting debate has been whether there is a measurable acoustic unit that would define moras (see [14] for an extensive review).

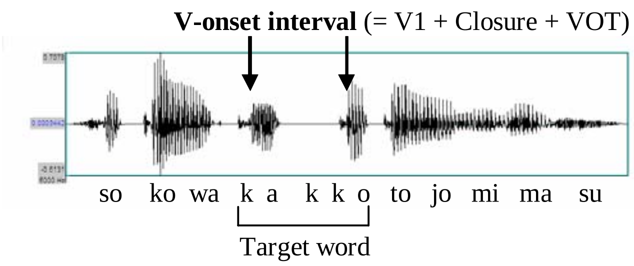
While there is a huge amount of literature addressing this issue with Japanese moras, this paper focuses on one aspect of the durational structures involved in the consonant length distinction in Japanese: Vowel-onset (V-onset) intervals in disyllable words. Examining this durational unit was motivated by Brady, Port, and Nagao [1] and Ofuka, Mori, and Kiritani [9]. Traditionally, duration of a CV mora has been calculated from the beginning of the consonant to the end of the vowel. However, Brady et al. [1] gave relevant perceptual and psychoacoustic rationales for using the interval between two vowel-onsets as a better indicator of a mora. Brady et al. measured V-onset intervals in Japanese sentences with two- to three-mora words as targets, e.g., /toko/ (CV.CV) (2 moras), /toːko/ (CV.V.CV) (3 moras), /tokːo/ (CV.C.CV) (3 moras), and /tomiko/ (CV.CV.CV) (3 moras). V-onset mora (Vmora) was then calculated by dividing each V-onset interval by the mean mora duration which is the mean of several such intervals preceding and following the targets. The results showed a significant degree of correspondence between the phonological weight of moras and the Vmora values.

Arriving at a similar idea from a different experiment, Ofuka et al. [9] noted the crucial role of VC as a perceptual timing unit in Japanese. They conducted a perceptual experiment using an edited [utatane]-[utatːane] continuum, in which the closure duration of the second [t], as well as the duration of vowel [a] preceding and following this [t], was altered. They found that, while native Japanese listeners perceived single and geminate stops categorically based on the closure duration, they tended to give more “geminate” responses when the preceding vowel is longer and when the following vowel is shorter. They argued that these results could not simply be explained if the closure duration was the sole factor, but could be explained if one considers VC as a perceptual unit.

The present study examined whether the V-onset interval and Vmora (=V-onset interval/mean mora duration) show a clear durational structure in Japanese disyllables including singleton and geminate stops. Materials were a subset of those used in Hirata and Whiton [6] and Hirata [5], in which target disyllables were embedded in a carrier sentence spoken at slow, normal, and fast speaking rates by four speakers. Fig. 1 shows an example of the materials with the target word /kakːo/ (‘parenthesis’) embedded in the carrier sentence. Previous findings in [5, 6] that are relevant in the present study are summarized here: (1) the duration of stop closure for singletons and geminates overlapped significantly across three speaking rates so that it was not a good measure to divide the two categories reliably (classification accuracy of 84.1%; Table 1). However, (2) other relational measures, such as durational ratios of closure to word, closure to the preceding vowel, and closure to the following vowel, provided better results to classify the two categories accurately (Table 1). (3) Closure duration for singletons vs. geminates was approximately 1:3, and this was not affected by speaking rate variations. (4) Vowels preceding the contrasting stops were very slightly but significantly longer for the geminate than the singleton context, but the following vowels were shorter after geminates than after singletons. The average duration differences, however, were very small, about 11-16 ms.

If the mora is clearly manifested in the V-onset interval as proposed by Brady et al. [1], we would predict that the V-onset interval duration (underlined below) for disyllables with singletons and geminates, e.g., /kako/ (C**V.C**V = 2 moras) vs. /kakːo/ (C**V.C.C**V = 3 moras), would be 1:2. Furthermore, Vmora, defined as the V-onset interval duration divided by the mean mora duration in the sentence, would also be 1 and 2, respectively. The present study also addressed the question of how accurately the V-onset interval, or the Vmora, can classify the singleton and geminate categories across three speaking rates. This was examined by calculating optimal boundaries (see the analyses section) for distinguishing singletons from geminates in the present data set.

Figure 1: V-onset interval determined by adding the durations of the first vowel, closure, and voice onset time of the target word, e.g., in /kakːo/.



2. METHODS

2.1. Materials

Materials were a subset of the materials (Set 2) used in Hirata and Whiton [6] and Hirata [5]. Set 2 consisted of contrasting singleton and geminate voiceless stops, /t(ː)/ and /k(ː)/ in disyllables using 30 real word pairs. In this data set, all five vowels /i e a o u/ were used as preceding vowels, and /b dʒ t k m n s h ɾ w j/ were used for the first consonants in CV(C)CV. Six out of 30 word pairs began with a vowel, e.g., /ut(ː)a/. All words were embedded into the carrier sentence /soko wa ____ to jomimasu/ ("that is read ____") at three speaking rates: slow (slowest tempo possible), normal (relaxed and comfortable tempo), and fast (fastest tempo possible) [10]. Set 2 totaled 2160 tokens (30 words x 2 lengths x 3 rates x 3 repetitions x 4 speakers). Four native Japanese speakers were used to read the materials (two male and two female), who were originally from Saga, Niigata, and Fukushima prefectures and who spoke standard Japanese. The pitch accent in this data set was deliberately left to each speaker. Pitch accents have small effects on duration [12], but the goal of the present study was to delineate durational invariance in spite of variability associated with different pitch accents, speakers, speaking rates, and segments. For more details on acoustic analysis, recording, and measurement criteria, see [6].

2.2. Analyses

In [6] and [5], the duration of single and geminate stop closures, words, vowels preceding and following the contrasting stops, voice onset time (VOT), and sentences were measured. Nineteen out of a total of 2160 tokens were excluded from the present study due to measurement difficulty in these tokens; thus, a total of 2141 tokens were used in the present analysis.

This study calculated the V-onset intervals for each of 2141 tokens by adding the duration of the preceding vowel, closure, and VOT (Fig. 1). This method is slightly different from that used by Brady et al. [1] which used the "interval between onsets of low-frequency speech energy" derived by an automatic vowel-onset detection algorithm. The present study examined whether we find similar results using these different methods. For the Vmora, the V-onset interval was divided by the mean mora duration. The mean mora duration in the present study was calculated by dividing the sentence duration by the number of moras in it. This method was also different from how Brady et al. determined the mean mora. To be precisely in line with their method, we would need to measure, in the future, the V-onset interval of a larger unit such as from the beginning of /o/ in /so/ to the end of /m/ in /ma/ in the present carrier sentence.

Analysis also included the calculation of “optimal boundaries” as well as classification accuracy with those boundaries, using the V-onset interval values and the Vmora values. Following [8], an optimal boundary is a value that classifies two members of a contrast with highest accuracy. Finding the optimal boundary involved counting the number of misclassified tokens among 2141 tokens, at each of the consecutive boundaries (e.g., 137, 138, 139 ms … for V-onset interval, and 1.30, 1.31, 1.32, … for Vmora). The value with the least number of misclassified tokens is defined as the “optimal” boundary. The goal of examining optimal boundaries was to find how well singleton and geminate categories can be classified when looking at the V-onset intervals and the Vmora.

3. RESULTS

3.1. Duration of V-onset intervals

Fig. 2 shows the distribution of V-onset interval duration for singletons (left panel) and for geminates (right panel), including all three speaking rates for all four speakers. The duration ranged from 66 to 453 ms for singletons, and from 137 to 1039 ms for geminates, and a large overlap (137-453 ms) is observed between the two categories. This pattern of distribution is quite similar to that found for the closure duration in [6]. The mean V-onset interval duration in the present study was 186 ms for singletons and 366 ms for geminates. Their ratio was 1:1.97, extremely close to the hypothetical ratio of 1:2.

3.2. Optimal boundary for V-onset interval duration

In order to examine whether this unit can classify the two categories reliably, the optimal boundary for the V-onset interval duration was calculated. As shown in Table 1, the optimal boundary was 176 ms, which misclassified 525 (447 singletons and 78 geminates) out of a total of 2141 tokens. This yielded a classification accuracy of 75.5%. Table 1 also shows optimal boundaries and classification accuracy based on other durational measures. For example, using closure duration as a criterion, we were able to classify the singleton and geminate categories with 81.4 % accuracy. Comparing all of the measures examined in the table, the V-onset interval duration provided the lowest classification accuracy.

3.3. Correlation between singleton and geminate V-onset intervals

As mentioned in the introduction, the V-onset intervals would theoretically be 1:2 for singleton vs. geminates, and the mean of all tokens supported this prediction in 3.1. To further examine this relationship across three speaking rates, the V-onset intervals for singletons are plotted against the corresponding V-onset intervals for geminates in Fig. 3. Pearson correlation indicated that the relationship is linear ($r = 0.949$, $p < 0.001$), and the regression line ($y = 0.4259x + 30.197$) is very close to the hypothetical line ($y = 0.5x$) (Fig. 3). This indicates that the V-onset intervals for geminates are approximately twice as long as those for singletons.

3.4. Vmora (V-onset interval/mean mora duration)

Next is the examination of the normalized V-onset interval, or Vmora. Vmora in this study is defined as the V-onset interval duration divided by the mean mora duration. Fig. 4 shows the distribution of Vmora values for singletons (white bars) and for geminates (striped bars) across three speaking rates including all four speakers. It shows that the distribution peaks correspond approximately to 1 for singletons and to 2 for geminates. The mean Vmora was 0.96 for singletons, and 1.88 for geminates. Another striking pattern observed in this figure is that there is very little overlap between the singleton and geminate categories.

3.5. Optimal boundary for Vmora

In order to find the exact amount of overlap between the Vmora values of singleton and geminate categories and classification accuracy, we calculated the optimal boundary. As shown in the bottom row of Table 1, the optimal boundary was somewhere between 1.34 and 1.36, and with this boundary, only 0.4% of the entire data set (only 9 out of the total of 2141 tokens) was misclassified. Thus, classification accuracy was 99.6%. Compared with all the other measures examined for optimal boundaries (closure duration, C/V1, C/W, C/V2, and V-onset interval duration), Vmora was the best measure to classify the singleton and geminate categories.

4. DISCUSSION AND CONCLUSIONS

This study examined whether stop length distinction in Japanese disyllables is durationally manifested in the V-onset interval as proposed by Brady et al. [1]. Results indicated that the V-onset interval duration for disyllables (underlined below), e.g., /kako/ (C**V.C**V = 2 moras) vs. /kakːo/ (C**V.C.C**V = 3 moras), was useful in characterizing the phonological weights of moras. The durational ratio was 1:1.97 for singletons vs. geminates, very close to the predicted ratio of 1:2. The result that the ratio was slightly less than 1:2 was consistent with the data in [1]. Furthermore, Vmora, defined as the V-onset interval duration divided by the mean mora duration in the sentence, was 0.96 for singletons, and 1.88 for geminates, and these values are, again, very close to the values of 1 and 2, respectively, predicted from Brady et al.'s V-onset mora theory.

The absolute duration of V-onset intervals was not a very good measure to classify the singleton and geminate categories across three speaking rates, as their classification accuracy was only 75.5% (Table 1). However, the Vmora, which is the normalized V-onset mora interval, was highly effective in classifying the singleton and geminate categories. The Vmora values, which roughly corresponded to the phonological weights of moras, were able to accurately classify the singleton and geminate categories with 99.6 % accuracy (Table 1). This accuracy superseded the accuracy derived from all the other measures (i.e., closure duration, C/V1, C/W, and C/V2 ratios).

Taken together, the present results support the effective use of Vmora in describing the timing structure of Japanese moras, as far as the stop length distinction is concerned. It is worth noting that Vmora not only corresponded to the phonological weights of moras, but also was effective in classifying the singleton and geminate categories, in spite of variabilities associated with different speakers, speaking rates, and segmental compositions of words.

Further studies are needed to examine Vmora in all other parts of the sentence and in different speech styles. We also need to follow up this study by recalculating Vmora based on the corrected "mean mora" duration (see the method section).

5. ACKNOWLEDGMENTS

This study has been supported by research grants from Colgate University.

6. REFERENCES

[1] Brady, M. C., Port, R. F., Nagao, K. 2006. Effects of speaking style on the regularity of mora timing in Japanese. *J. Acoust. Soc. Am* 120 (5), 3208.

[2] Fujisaki, H., Nakamura, K. Imoto, T. 1975. Auditory perception of duration of speech and non-speech stimuli. In: Fant, G., Tatham, M. A. A. (eds), *Auditory Analysis and Perception of Speech.* London: Academic Press, 197-219.

[3] Han, M. 1962. The feature of duration in Japanese. *The Study of Sounds* 10, 65-80.

[4] Han, M. 1994. Acoustic manifestations of mora timing in Japanese. *J. Acoust. Soc. Am* 96 (1), 73-82.

[5] Hirata, Y. 2007. Durational variability and invariance in Japanese stop quantity distinction: Roles of adjacent vowels. *J. Phonetic Society of Japan* 11 (1) (in press).

[6] Hirata, Y., Whiton, J. 2005. Effects of speaking rate on the single/geminate stop distinction in Japanese. *J. Acoust. Soc. Am* 118 (3), 1647-1660.

[7] Idemaru, K. 2005. An Acoustic and Perceptual Investigation of the Geminate and Singleton Stop Contrast in Japanese. Doctoral dissertation, University of Oregon.

[8] Miller, J. L., Green, K. P., Reeves, A. 1986. Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica* 43, 106-115.

[9] Ofuka, E., Mori, Y., Kiritani, S. 2005. Perception of a Japanese geminate stop: the effect of the duration of the preceding/following vowel. *J. Phonetic Society of Japan* 9 (2), 59-65.

[10] Port, R. F. 1977. The Influence of Speaking Tempo on the Duration of Stressed Vowel and Medial Stop in English Trochee Words. Doctoral dissertation, Indiana University Linguistics Club, Indiana University.

[11] Port, R. F., Dalby, J., O'Dell, M. 1987. Evidence for mora timing in Japanese. *J. Acoust. Soc. Am* 81 (5), 1574-1585.

[12] Sugito, M. 2000. *Onsei Hakei wa Kataru (Nihongo Onsei no Kenkyu 4).* Osaka: Izumi Shoin.

[13] Vance, T. J. 1987. *An Introduction to Japanese Phonology*. Albany, NY: State University of New York Press.

[14] Warner, N. and Arai, T. 2001. Japanese mora-timing: A review. *Phonetica* 58, 1-25.

Table 1: Optimal boundaries and their classification accuracy based on V-onset interval duration and Vmora (bottom two rows). Data included 2141 tokens out of a total of 2160 tokens for Set 2 materials of Hirata and Whiton [6] and Hirata [5] (19 tokens unused due to measurement problems). Results of other measures (closure duration, C/V1 ratio, C/W ratio, C/V2 ratio) are from [6] and [5].

	Optimal boundary	Percentage misclassified	Classification accuracy
Closure duration	115 ms	18.6%	81.4%
C/V1 (consonant to preceding vowel duration) ratio	1.53	8.7%	91.3%
C/W (consonant to word duration) ratio	0.35	4.3%	95.7%
C/V2 (consonant to following vowel duration) ratio	1.59-1.76	1.1-1.2%	98.8-98.9%
V-onset interval duration	**176 ms**	**24.5%**	**75.5%**
Vmora (V-onset interval/ mean mora)	**1.34-1.36**	**0.4%**	**99.6%**

Figure 2: Distribution of V-onset interval duration (ms) for singletons (left panel; n = 1068) and for geminates (right panel; n = 1073), including all three speaking rates and four speakers. The dashed line represents optimal boundary computed as 176 ms.

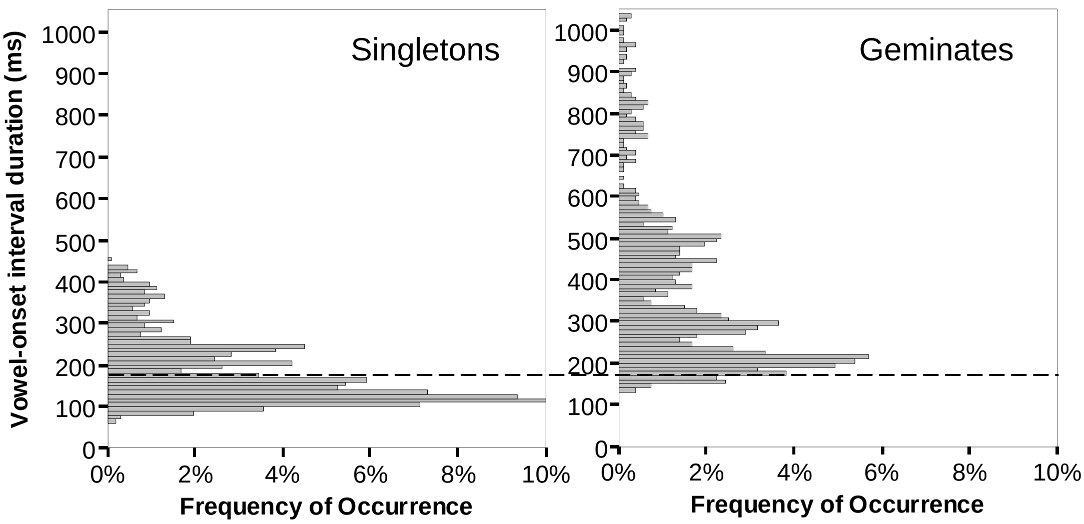


Figure 3: V-onset intervals (ms) for singletons plotted against those for geminates. Each data point represents the average of three repetitions for each of the 30 word pairs, e.g., /kako/ vs. /kak:o/, for each rate for each speaker (n = 360). The regression line (solid; y = 0.4259x + 30.197) is close to the hypothetical line (dashed; y = 0.5x).

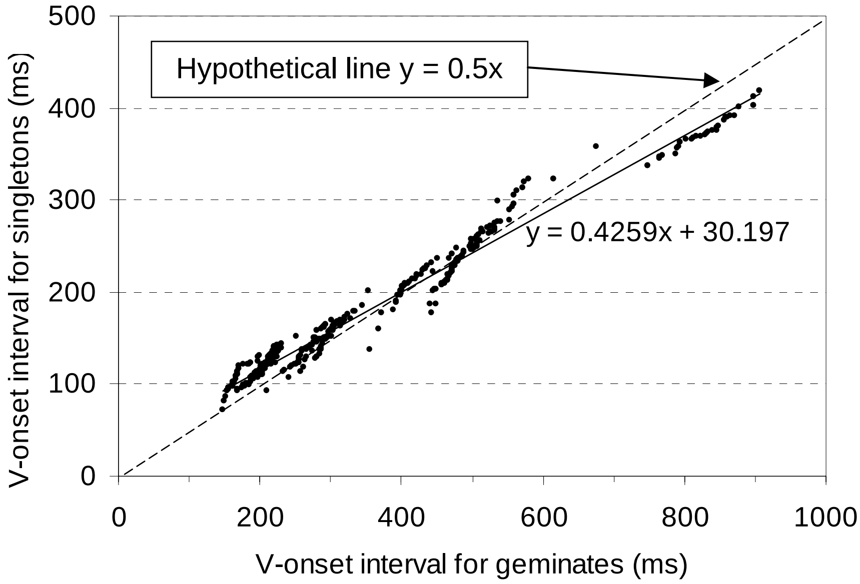


Figure 4: Distribution of Vmora values (V-onset interval normalized by the mean mora duration). The distribution of white bars for singletons and striped bars for geminates indicates their corresponding weights of moras (1 vs. 2).