Durational Structure of Japanese Single/Geminate Stops in Three- and Four-Mora Words Spoken at Varied Rates

Yukari Hirata[1], *Shigeaki Amano*[2]

[1] Department of East Asian Languages and Literatures, Colgate University, U.S.A.
[2] Faculty of Human Informatics, Aichi Shukutoku University, Japan

yhirata@colgate.edu, psy@asu.aasa.ac.jp

Abstract

To distinguish Japanese single and geminate stops in two- and three-mora words spoken at varied speaking rates, the ratio of stop closure to the word in native speakers' production was previously found to be a reliable measure. It was not clear, however, whether the stop closure relates more stably (1) to the entire "word" of any length than just (2) to the moras preceding and following the contrasting stops. This study examined this question with three- and four-mora nonsense words in Japanese. Results indicate that the stop closure duration relative to both of the units (1) and (2) were equally useful in accurately classifying single and geminate stops. This implies that the anchor to which the contrasting stop duration normalizes across rates does not have to be the entire "word" although the word is also a stable anchor.

Index Terms: production, Japanese, stop quantity distinction, speaking rate

10.21437/Interspeech.2010-449

1. Introduction

It is known that stop closure duration is the major acoustic correlate for stop quantity distinction in Japanese, e.g., /haka/ 'grave' (two moras) versus /hakːa/ 'mint' (three moras) [1]. However, there is significant overlap of closure duration when these two- and three-mora words are spoken at different speaking rates (Hirata & Whiton, 2005; H&W [2] henceforth). H&W [2] examined disyllables, CV.(C).CV (the contrasting stop consonant underlined), and found that closure-over-word (C/W) duration ratio classified the singleton and geminate categories reliably despite the overlap of absolute closure duration. Specifically, the C/W ratio of 0.35 was found to be the optimal boundary: the produced tokens with greater than 0.35 would have geminates, and those with smaller than 0.35 would have singletons. The results supported relational acoustic invariance [3] in the face of speaker- and rate-variations. What remained unclear in [2] was why it is the word, but not other linguistic units such as adjacent vowels or moras, which was claimed to serve as the stable anchor for varied stop closure duration. It is possible that a word-internal unit smaller than the entire word is sufficient as such an anchor. The present study explored this possibility by analyzing three- and four-mora word pairs.

We set out the following three guiding questions for this study. First, to what extent can the results of two- and three-mora words with singleton and geminate stops (e.g., /haka/ 'grave' vs. /hakːa/ 'mint') [2] be extended to three- and four-mora word pairs (e.g., /hakaku/ 'extraordinary value' vs. /hakːaku/ 'disclosure') in their durational structure? As the first step, we examined this question using three- and four-mora nonsense word pairs where the consonants and vowels were limited to /t k e/.

In examining three- and four-mora words, durational variability is introduced by different types of words. The second question was whether the durational structure of single/geminate distinction is affected by different word types. Specifically, we examine the following three types of words (Table 1): (1) "C," which includes CV only, i.e., CV.(C.)CV.CV (single/geminate consonants underlined), (2) "N," which includes a moraic nasal, i.e., CV.(C.)CV.N, and (3) "V," which includes a long vowel CV.(C.)CV.V. This investigation was not possible when [2] examined two- and three-mora word pairs as in CV.(C).CV. The present study attempts to examine how generalizable H&W's [2] findings are to different types of longer words. For these two questions above, we examined stop closure duration, word duration, and C/W ratio, as in [2].

The third question, which is central to this study, relates to the durational ratio of stop closure to the portion of the three- and four-mora words that excludes the final mora for word types C and N (C/Subword). For example, the "Subword" refers to /tete/-/tetːe/ in pairs such as /tetete/-/tetːete/ and /teten/-/tetːen/. How accurately does the C/Subword ratio classify produced tokens of varied rates into the single and geminate categories? We hypothesized that, if it is some special characteristics of the unit "word" that make the stable durational representation across rates, C/W ratio would classify the singleton and geminate categories with higher accuracy than C/Subword would. This would likely happen if durational isochrony exists in the unit of words [2, 4, 5] but not in individual moras. If, on the other hand, the distinction of single and geminate stops was stably represented in the duration of the stop closure in relation to the preceding and the following moras (i.e., not needing the entire "word"), C/Subword could classify the two categories in the above three- and four-mora words as accurately as C/W ratio would. This scenario would assume that durational isochrony exists in the unit of the mora (see [6] for a review of mora isochrony).

Table 1. *Nonsense words used in the present experiment.*

Word type	Singletons (3-mora words)	Geminates (4-mora words)
C CV.(C.)CV.CV	tetete kekeke	tetːete kekːeke
N CV.(C.)CV.N	teten keken	tetːen kekːen
V CV.(C.)CV.V	teteː kekeː	tetːeː kekːeː

In the traditional mora theory [7] in which moras are assumed to be isochronous, specific predictions can be made about C/W and C/Subword ratios. The singleton stop (C) in CV.CV receives 0.5 mora in the two mora word, and thus, the C/W ratio is predicted to be 0.5/2=0.25. The geminate stop (C.C) in CV.C.CV receives 1.5 moras (one full moraic C and half of a CV mora) in the three mora word, and thus, the C/W ratio is predicted to be 1.5/3=0.5. The actual values were 0.24 (singletons) and 0.48 (geminates) in [2]. For the three- and four-mora word pairs investigated in the present study, the C/W ratio of the singleton word (CV.CV.CV) is predicted to be 0.5/3=0.17, and that of the geminate word (CV.C.CV.CV) is predicted to be 1.5/4=0.38. For C/Subword in the present study, the predicted ratio would be 0.25 and 0.50 as in the case of two- and three-mora words. The traditional mora theory would also assume the word duration ratios of two-, three-, and four-mora words to be 2:3:4, although there is a tendency for the word duration to have smaller increments for additional moras in faster speech [4].

2. Method

2.1. Participants

Four adult native Japanese speakers, two men and two women, participated in recording. They were born and have been living in Tokyo metropolitan areas. They all speak standard Japanese.

2.2. Materials and procedure

Table 1 shows 12 target words used in the present experiment. They were three- and four-mora nonsense words whose segments were limited to /t k e/ and the moraic nasal, and the words had no pitch accent. The contrasting stops /t(:) k(:)/ occurred between the first two syllables. These target words were produced in a carrier sentence /sore wa ____ da to omoimasu/ 'I think that it is ___." spoken at slow, normal, and fast rates. The definition of speaking rates was as follows: "as slowly as possible but still connected across words" for the slow rate, "relaxed and comfortable" for the normal rate, and "as fast as possible without making excessive errors" for the fast rate. Beyond this definition, the rate of speech was self-determined by participants. Each speaker repeated each word three times at the normal rate, followed by the slow and then the fast rates. The order of the target words was randomized for each repetition. There was a total of 432 tokens (4 words x 3 word types x 3 rates x 3 repetitions x 4 speakers) analyzed in the present study.

2.3. Analyses

Digitally recorded materials were labeled using the acoustic analysis software PRAAT. Duration was measured for contrasting stop closure (the underlined (C)C portion in Table 1), the preceding and following vowels, the target word, as well as the entire sentence. Three durational ratios were obtained: the ratio of the stop closure to the preceding vowel (C/V1), the ratio of the stop closure to the word (C/W), and the ratio of the three- and four-mora words with the three-mora word standardized by 3. This last ratio was obtained to examine whether the word duration corresponds with the number of moras in the word: three- and four-mora words as 3:4.

For word types C and N, the duration of the target word excluding the final mora, i.e., "Subword," was measured. For example, "Subword" is the portion such as the first three moras /tet:e/ of the four-mora target word /tet:ete/, and the first two moras /tete/ of /teten/. Since we cannot reliably remove the last mora in word type V (i.e., CV(C)CVV), this word type was not used in the examination of "Subword." The ratio of the stop closure to the "Subword" (C/Subword) was then calculated.

For the above measures, analyses were conducted in three steps. First, an analysis of variance (ANOVA) was conducted for each of the following three duration measures: stop closure, word, and C/W ratio. Quantity (singleton, geminate), speaking rate (slow, normal, fast), and word type (C, N, V) were within-subjects factors. Second, regression analysis was conducted in order to examine whether the durational relationship between the closure and the Subword was linear as in the case of two- and three-mora words in [2]. Finally, in order to examine with which of the four measures (closure, C/V1, C/W, and C/Subword) we could classify the produced tokens into singleton and geminate categories with highest accuracy, optimal boundaries and their classification accuracy were computed, following the method used in [2, 8].

3. Results

3.1. Stop closure duration

Mean stop closure duration was significantly shorter for singletons than geminates (69 vs. 207 ms) [$F(1,3)=190.86$, $p<0.001$], but ranges overlapped between the two categories across rates (singletons: 15-160 ms; geminates: 79-473 ms). Fig. 1 shows that the closure duration of geminates was longer than that of singletons within each speaking rate, but that duration of geminates spoken fast overlapped with that of singletons spoken slowly.

Word types significantly affected closure duration [$F(2,6)=14.53$, $p<0.01$]. It was consistently shorter for word type C (126 ms) than V (148 ms) [$p<0.05$], while there was a marginal difference between word types C and N (139 ms) [$p=0.054$] and no significant difference between V and N [$p=0.092$].

There was a significant interaction of speaking rate and quantity [$F(2,6)=47.15$, $p<0.001$]. The duration difference between singletons and geminates was greater for slower speech (Fig. 1). This durational pattern is comparable to that in [2].

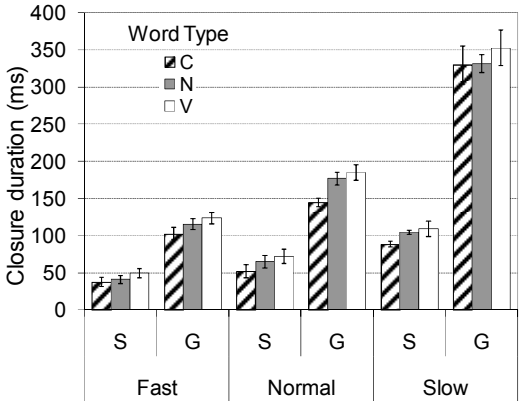

Figure 1: *Closure duration of single [S] and geminate [G] stops by speaking rate and word type [C: CV(C)CVCV; N: CV(C)CVN; V: CV(C)CVV].*

3.2. Word duration

Mean word duration was shorter for faster speech (slow: 827 ms; normal: 418 ms; fast: 306 ms). The mean word duration also differed between the singleton and geminate words (448 vs. 586 ms). The mean ratio of three-mora vs. four-mora words was 3:3.9, extremely close to the theoretical prediction of 3:4. ANOVA indicated that there was a rate by quantity interaction [$F(2,6)=54.45$, $p<0.001$]. The differences between singleton and geminate words were shorter for faster speech. This was reflected in the ratios of singleton and geminate words: 3:4.01 for the slow, 3:3.88 for the normal, and 3:3.77 for the fast rates. These results on word duration were similar to those of [2], except that two- and three-mora words had the durational ratio of 2:3.

The most interesting finding was that word duration was not affected by word type (C: 511; N: 529; V: 511 ms) [$F(2,6)=3.89$, $p=0.082$]. This indicates that word duration reflects the number of moras it contains, regardless of whether the word contains only CV syllables, the moraic nasal, or a long vowel. This also supports the mora theory at the word level [4, 5].

Table 2. *Optimal boundaries for closure duration, C/V1 ratio, C/W ratio, and C/Subword ratio.*

		Optimal boundary***	Classification accuracy (%)
Closure duration	3-4 mora words (Present study)	108 ms	**89.1**
	2-3 mora words (H&W [2])	74 ms	82.2
C/V1 ratio	3-4 mora words (Present study)	1.50	**93.5**
	2-3 mora words (H&W [2])	1.53	92.1
C/W ratio*	3-4 mora words (Present study)	0.25	**97.5**
	2-3 mora words (H&W [2])	0.35	98.0
C/Subword ratio**	3-4 mora words (Present study)	0.35	**97.2**

* The mora theory would predict the C/W boundary for 3- and 4-mora words and 2- and 3-mora words to be 0.27 and 0.38, respectively (see the introduction and [2] for more details).
** "Subword" refers to the entire 3- and 4-mora word minus the last mora for word types C and N.
*** In cases where the optimal boundary was found at two locations, those values were averaged.

3.3. C/W ratio

Mean C/W ratios were significantly greater for the geminate than the singleton words (0.16 vs. 0.35) [$F(1,3)=166.28$, $p<0.001$]. The values were comparable but slightly smaller than the theoretical prediction of 0.17 vs. 0.38. C/W ratios were also affected by speaking rate [$F(2,6)=5.40$, $p<0.05$] and word type [$F(2,6)=31.09$, $p<0.001$], as well as showing a complex interaction of the three factors (quantity, rate, and word type) [$F(4,12)=3.44$, $p<0.05$]. Our main interest, however, was whether C/W ratio can classify the single and geminate tokens accurately across rates and speakers. Table 2 shows optimal boundaries and classification accuracy in the present data (N=432; 4 nonsense words x 3 word types x 4 speakers x 3 rates x 3 repetitions), compared to those in [2]. It shows that, using the C/W ratio of 0.25, 97.5% of the present data was classified accurately into the singleton and geminate categories. Accuracy was higher when we used C/W ratio than using absolute closure duration or C/V1 ratio. This result is very similar to that in [2] (see H&W's values in Table 2).

3.4. C/Subword ratio

Fig. 2 shows the durational relationship between the closure and the Subword, which was the portion of word types C and N that excluded the final mora. Word type V was omitted from this analysis because it is impossible to remove reliably the second mora of a long vowel, e.g., in /tet(:)e:/. The figure shows that the closure duration increased linearly as the Subword duration increased due to the decrease of the speaking rate for both singletons [$y=0.202x+4.820$ (where x = Subword duration; y = closure duration), $r^2=0.694$] and geminates [$y=0.526x-28.262$, $r^2=0.936$]. To provide a point of comparison, Fig. 2 also shows the regression lines obtained from two- and three-mora words (including words such as /heta/ 'unskilled' vs. /het:a/ 'decreased') from [2].

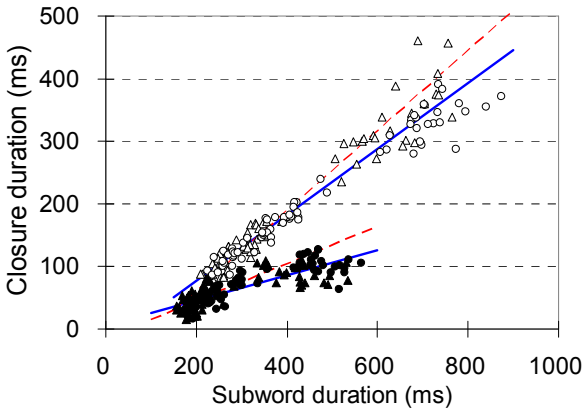

Figure 2: *Closure duration plotted against Subword duration (word minus last mora duration). Filled and unfilled plots represent singletons and geminates, respectively. Triangles and circles represent word types C and N, respectively. Solid lines represent regression lines for the single and geminate tokens of the present data, while dashed lines represent those of two- and three-mora words in [2].*

The next analysis was to examine C/Subword ratio in the present data (Fig. 3) and calculate its classification accuracy. As seen in Fig. 3, the C/Subword ratio values were lower for singleton (mean=0.22) than geminate words (mean=0.45), and the histograms show that the two categories were clearly separated. With the optimal boundary of 0.35, the classification accuracy was found to be 97.2% (Table 2; N=288; 4 nonsense words x 2 word types x 4 speakers x 3 rates x 3 repetitions). This means that C/Subword worked in classifying the singleton and geminate categories as equally well as C/W ratio did (whose classification accuracy was 97.5%). The distribution of C/Subword ratio in Fig. 3 is similar to that in two- and three-mora words examined in [2].

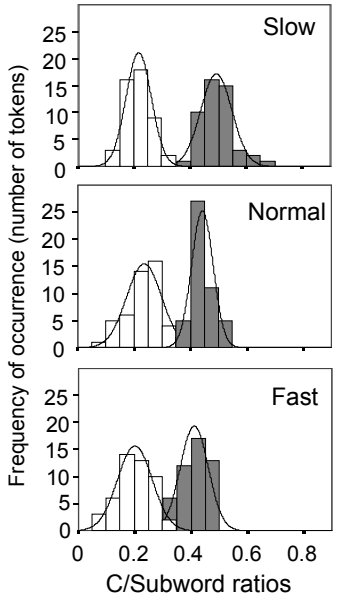

Figure 3: *Distribution of C/Subowrd ratios of singleton (white bars; n=48 at each rate) and geminate (dark bars; n=48 at each rate) tokens at three speaking rates*.

4. Discussion and conclusions

The first question addressed in this study was whether the durational structures of two- and three-mora word pairs and three- and four-mora word pairs are similar with regard to the singleton and geminate stop distinction. In the present investigation of three- and four-mora nonsense word pairs, we found many similarities to the durational structure of two- and three-mora words found in [2]. First, there was a large durational overlap between the singleton and geminate stop closure across three speaking rates. Second, the word duration corresponded well with the number of moras in the words, i.e., 3:4 in three- and four-mora words, and this result is similar to that in [9] which examined real words. The present result extends previous findings about two- and three-mora words being 2:3 in [2], and is consistent with the mora theory at the word level [4, 5]. Third, accuracy in classifying singleton and geminate stops was higher with C/W ratio than with C/V1 ratio and absolute closure duration, and this result was very similar to that in [2].

The second question addressed in this study was whether the durational structure of the single/geminate distinction was affected by different word types. The stop closure duration was shorter in the word type CV.(C)CV.CV than CV.(C)CV.N or CV.(C)CV.V, which is consistent with [9]. However, there was no effect of word type for duration of words, and this result is consistent with the mora theory for the unit of word [2, 4, 5]. Regardless of whether the word includes the CV mora, the moraic nasal, or a long vowel at the end of the word, the word duration indicates the correspondence with the number of moras, as shown in the ratio of 3:4 for singleton versus geminate words.

For the third question, which was central to this study, we examined whether the ratio of stop closure to the Subword (C/Subword) was similar to the C/W ratio in two- and three-mora words as in [2]. The mean C/Subowrd ratios were 0.22 for singleton and 0.45 for geminate words, comparable to the C/W ratios of 0.24 and 0.48 for two- and three-mora words in [2]. Using the C/Subowrd ratio, the produced tokens were classified into the singleton and geminate categories with 97.2% accuracy. This accuracy was as high as that with the C/W ratio in the present experiment (97.5%) and that with the C/W ratio in the nonsense words in [2] (98.0%). The implication of this result is that the distinction of single and geminate stops was stably represented in the duration of stop closure in relation to the preceding and the following moras. In other words, the anchor to which the contrasting stop duration normalizes does not have to be the entire "word" although the unit "word" is also a stable anchor across rates.

Further studies are needed to examine whether the present results also apply to real words, in which variations in segments would create larger durational variability than the present stimuli with controlled segments. Several decades of research have examined production and perception of the Japanese stop quantity distinction as pertaining to theoretical and applied phonetic studies (see [10] for a review). The complete investigation of the production of single and geminate stops spoken at varied rates may provide fundamental information useful for understanding how native Japanese speakers perceive this quantity distinction.

5. Acknowledgments

This study was conducted using facility and service at NTT Communication Science Laboratories, Kyoto, Japan.

6. References

[1] Fujisaki, H., Nakamura, K., and Imoto, T., "Auditory perception of duration of speech and non-speech stimuli", in G. Fant and M. A. A. Tatham [Eds], Auditory Analysis and Perception of Speech, 197-219, Academic Press, 1975.

[2] Hirata, Y. and Whiton, J., "Effects of speaking rate on the single/geminate stop distinction in Japanese", J. of the Acoustical Society of America, 118(3): 1647-1660, 2005.

[3] Pickett, E. R., Blumstein, S. E., and Burton, M. W., "Effects of speaking rate on the singleton/geminate consonant contrast in Italian", Phonetica, 56: 135-157, 1999.

[4] Port, R. F., Dalby, J., and O'Dell, M., "Evidence for mora timing in Japanese", J. of the Acoustical Society of America, 81(5): 1574-1585, 1987.

[5] Han, M., "Acoustic manifestations of mora timing in Japanese", J. of the Acoustical Society of America, 96(1): 73-82, 1994.

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

[7] Han, M., "The feature of duration in Japanese", The Study of Sounds, 10: 65-80, 1962.

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

[9] Hirata, Y., Kato, H., and Tajima, K., "Duration of Japanese singleton and geminate obstruents in two- to four-mora words", Proc. of Acoustics'08Paris, 2351-2356, 2008.

[10] Hirata, Y., "Durational variability and invariance in Japanese stop quantity distinction: Roles of adjacent vowels", J. of the Phonetic Society of Japan, 11(1): 9-22, 2007.