Durational Variability and Invariance in Japanese Stop Quantity Distinction: Roles of Adjacent Vowels

Yukari HIRATA*

1. Introduction


Acoustic, psychoacoustic, and perceptual studies in the mechanism of quantity distinctions in Japanese consonants and vowels trace a long history (Han 1962, Fujisaki et al. 1975, Fukui 1978, Homma 1981, Watanabe and Hirato 1985, Hirato and Watanabe 1987, Port et al. 1987, Hirata 1990a, Aoyama 2001, Kinoshita et al. 2002, Minagawa et al. 2003, Hirata 2004a, Hirata and Whiton 2005, Ofuka et al. 2005). One fundamental issue is that, although duration is the primary acoustic correlate and perceptual cue to quantity distinction (Han 1962, Otsubo 1981), various factors such as speaking rate affect duration of target and surrounding segments (Hirata 2004a). This makes it difficult to pinpoint what acoustic parameter is invariant and responsible for native speakers’ consistent perception of “short” and “long” phonemes. The questions of whether there exists a single invariant acoustic parameter for a
phonological feature in spite of all kinds of acoustic variabilities, and whether that invariant parameter corresponds to the invariant phonological percept are central issues in phonetics (Picket et al. 1999). This many-to-one mapping is the core problem that various speech perception theories attempt to account for (Liberman et al. 1967, Lahiri et al. 1984, Liberman and Mattingly 1985, Fowler 1986, Diehl and Kluender 1989).

The present study is an endeavor delineating variability and invariance in the production of quantity distinction in Japanese voiceless stops. It follows up on a recent acoustic study, Hirata and Whiton (2005), in which duration of various parts of Japanese disyllables was examined when they were spoken in a carrier sentence at three speaking rates. Hirata and Whiton’s (2005) three major findings relevant to the present study were as follows. First, there was no overlap between the duration of singleton and geminate closures at each speaking rate, but there was significant overlap in the duration of the singletons spoken at a slow rate and the geminates spoken at a fast rate. Second, the durational ratio of geminate and singleton closures was approximately 3:1, and this ratio was not affected by speaking rate variations. Third, ratio of stop closure to the preceding vowel and ratio of closure to word were affected by speaking rates, but these ratios classified the singleton and geminate categories across rates better than absolute closure duration did. Most notably, closure to word duration ratio, at the boundary ratio of 0.35, was found to be the best classifier among the three parameters examined. In other words, if closure duration was less than 35% of the disyllable word duration, the word almost certainly included a singleton stop, and if closure duration was greater than 35% of the word duration, it almost certainly included a geminate stop.

What was not clear in Hirata and Whiton (2005), however, was the role of vowels adjacent to the contrasting stops. A question remained as to whether these vowels show systematic differences that are reliable indicators of stop quantity when speaking rate varies. Three goals are set out in the present study, regarding (1) duration of vowels preceding single and geminate stops, (2) duration of vowels following those contrasting stops, and (3) durational ratios of stop closure to the preceding vowel (C/V1) and to the following vowel (C/V2). Below, specific questions and previous studies that motivated the present study are described.

With regard to the first goal, previous studies found a tendency for vowels to be longer when they precede geminates than singletons (Fukui 1978, Aizawa 1985, Takada 1985, Campbell 1999, Idemaru 2005). Vowels preceding voiceless stops are also known to play a crucial role in the perception of stop quantity distinction in Japanese (Watanabe and Hirato 1985, Hirata 1990a). Specifically, perceptual boundaries between single and geminate stops depend on the duration of closure relative to the duration of the preceding vowel. Thus, preceding vowels play a role as an “anchor.” There is also some indication that longer preceding vowels associated with geminates than singletons might be used as a secondary cue to stop quantity distinction (Fukui 1978, Uchida 1993).

The present study addressed the question of whether this tendency is stable in the face of speaking rate variations. Does this durational relationship hold when speaking rate varies or at any individual speaking rate? If so, what is the magnitude of effects in terms of absolute duration? Absolute duration might not play a large role in native speakers’ perception, as they almost automatically normalize speaking rate and adapt to given contexts (Fujisaki et al. 1975). However, absolute duration matters greatly for non-native speakers. For example, native English speakers are unable to adapt to speaking rates and tend to misperceive a phonemic “short” vowel spoken slowly as a phonemic “long” vowel, and a phonemic “long” vowel spoken quickly as “short” (Tsurutani 2003, Hirata 2005, Tajima 2006). Thus, finding the magnitude of vowel duration differences and examining their stability might help non-native speakers to understand the mechanism of Japanese stop quantity distinction.

With regard to the second goal of this study, previous production studies found that vowels are shorter when they follow geminates than singletons (Fukui 1978, Aizawa 1985, Takada 1985, Campbell 1999, Idemaru 2005). The perceptual role of vowels following single and geminate stops, however, is inconsistent so far. Hirato and Watanabe (1987) found that the following vowel in /ap(α)a/ presented in isolation did not affect the identification of single and geminate stops. Idemaru (2005) also found similar results with edited /set(α)ә/ in isolation. However, Ofuka et al. (2005) examined perception of /utat(ә)әne/ embedded in a carrier sentence, and found that there were more geminate responses (i.e., categorical boundary values were smaller) when the following vowel was shorter. While the present study does not directly address this issue in perception, the present examination of distributional characteristics of following vowels in natural, read speech might give
new ideas for future perception experiments. The questions addressed in this study include whether vowels are consistently shorter when they follow geminates than singletons in materials that vary in speaking rate. What is the magnitude of this effect in absolute duration? Is the difference in absolute vowel duration consistent across rates? How strong is the effect of stop quantity relative to the effect of speaking rate, and how might these two effects interact?

For the third goal of the present study, the ratio of the stop closure to the preceding and following vowel (C/V1 and C/V2) in disyllables was examined. As noted earlier, the C/V1 ratio is important in the perception of stop quantity: how long the closure duration needs to be for singletons or geminates depends on the preceding vowel. Indeed, Hirata and Whiton’s (2005) production data showed that the C/V1 ratio alone can be used to classify the two stop categories with 91–92% accuracy. However, the relative importance of the C/V2 ratio is unclear. Traditional mora theory (Han 1962, Kimura 1985) would seem to predict no difference between the C/V1 and C/V2 ratios. However, if vowels are consistently shorter following geminates than singletons, and consistently longer preceding geminates than singletons (e.g., Campbell 1999), then C/V2 would differ more between geminate and singleton categories than C/V1 would. That is, the following “shorter” vowels enhance the long duration of geminates, and the following “longer” vowels enhance the short duration of singletons. The following vowels having an inverse relationship with the stop quantity (as opposed to the preceding vowels being in direct proportion to the stop quantity) might play an important role in perception, according to Diehl and Klunder’s (1989) Auditory Enhancement Hypothesis. The present production data might shed light on the stability and relative importance of these two ratios as indicators of stop quantity distinction. The third goal of this study, therefore, was to compare mean ratios of C/V1 and C/V2, and to compare accuracy in dividing the two stop categories using C/V1 and C/V2 ratios.

Finally, it is important to clarify what was not examined in the present study. First, this study only examined production data, and no perception experiment was conducted. Thus, terms used in this study, such as “good indicators of stop quantity distinction,” solely refer to the degree of stability and systematicity to which native Japanese speakers produced materials related to stop quantity distinction. Two sets of production materials in the present study, one with controlled segments and the other with more varied segments, were used to characterize duration, and invariance in the face of speaking rate and segmental variabilities. Acoustic examination goes hand in hand with perceptual examination, and each gives insight into the other. Second, this study only examined the acoustic parameter, duration. It has been pointed out that other acoustic parameters such as fundamental frequency and intensity are associated with Japanese stop quantity distinction (Ofuka 2003, Idemaru 2005), and that phonological pitch accents and fundamental frequency contours have effects on perception and production of duration (Mori 2001, Hirata 1990a, Kinoshita et al. 2002, Minagawa et al. 2003). Although the present study focused on duration, the interplay between duration and these non-durational factors should continue to be examined in the future.

2. Method

2.1 Materials

Materials used in this study were the identical sets of words and a carrier sentence used in Hirata and Whiton (2005) (see the Appendix). Set 1 had six pairs of nonsense disyllables, contrasting in single and geminate voiceless stops, /t(ː)/ and /k(ː)/, with non-high vowels (/a/, /e/, and /o/). Set 2 had 30 real-word pairs contrasting in the same stop quantity in disyllables. As in the Appendix, preceding vowels in Set 2 included all five vowels /i e a o u/, and the first consonants in CV(C)CV varied in /b dʒ t k m n s h r w j/, while some words began with a vowel. These words in the two sets were spoken in a 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) (Port 1977). Set 1 had a total of 720 tokens (6 words×2 quantities×3 rates×5 repetitions×4 speakers), and Set 2 had 2160 tokens (30 words×2 quantities×3 rates×3 repetitions×4 speakers). Materials were read by four native Japanese speakers (two male and two female), who spoke standard Japanese but were originally from Saga, Niigata, and Fukushima prefectures. The pitch accent of words was on the first mora for Set 1, but was deliberately left up to each speaker for Set 2. Pitch accents have small effects on duration (Sugito 2000, Mori 2001, Hirata 1990a), but the goal of the present study and Hirata and Whiton (2005) was to delineate durational invariance in spite of variability associated with different pitch accents, segments, speakers, and speaking
rates. For more details of recording, acoustic analysis, and measurement criteria, see Hirata and Whiton (2005).

2.2 Analyses

In Hirata and Whiton, duration of single and geminate stop closures, voice onset time (VOT), vowels preceding the contrasting stops, words, and sentences was measured. In the present study, vowels following the contrasting stops were measured, and compared with the preceding vowels. Analyses of Variance (ANOVAs) were conducted separately for preceding and following vowels and for Sets 1 and 2, with vowel duration as a dependent variable. For Set 1, rate (slow, normal, fast), consonant quantity (single, geminate), vowel type (/a/, /e/, /o/), and consonant type (/t/, /k/) were within-subjects factors. For Set 2, rate (slow, normal, fast) and consonant quantity (single, geminate) were within-subjects factors. For post-hoc analyses, multiple comparisons were made with Bonferroni correction.

In addition to analysis of vowel duration, ratios of stop closure to the preceding and following vowels (C/V1 and C/V2, respectively) were calculated, and “optimal boundaries,” as well as classification accuracy with those boundaries, were computed. Following Miller, Green, and Reeves (1986) and Pind (1995), an optimal boundary is a value that classifies two members of a contrast with highest accuracy (see Hirata and Whiton for more details). The procedure involves counting the number of misclassified tokens among, for example, 720 tokens in Set 1, at each of the consecutive boundary ratios (e.g., 1.00, 1.01, 1.02, …). The value that had the least number of misclassified tokens was said to be the “optimal” boundary. The goal of examining optimal boundaries was to find how accurately C/V2, C/V1, and other parameters can classify the singleton and geminate categories.

3. Results

3.1 Duration of preceding vowels

Figure 1 shows durational distribution of vowels preceding single and geminate stops for Set 2, including all three speaking rates. It shows that the range of duration is similar between the singleton context (18–250 ms; left panel) and the geminate context (13–255 ms; right panel), but that the peaks of the distribution curves differ between the two categories. The overall difference between the two distribution curves appears to be quite small. This is to be expected, however, as it indicates a sub-phonemic difference within the phonemic “short vowel” category, unlike a large between-category difference for phonemic “short” versus “long” vowels as in Hirata (2004a). The short vowels in the present study cannot stretch more for the slower speaking rate or for enhancing the consonant quantity distinction because that might cause confusion between phonemic short and long vowels. In fact, the present short vowel duration range is exactly the same range found for the short vowels spoken at three speaking rates in Hirata (2004a). In sum, these vowels alone would not help to distinguish the following single and geminate stop distinction.

ANOVA for Set 1 showed that speaking rate had a significant effect on vowel duration \( F(2, 6) = 34.10, \)
Durational Variability and Invariance in Japanese Stop Quantity Distinction: Roles of Adjacent Vowels

Vowel duration was significantly longer for the slow than normal (122 vs. 68 ms), and for the normal than fast rate (52 ms). Consonant type also had a significant effect \[ F(1, 3)=11.66, p=0.042 \]: vowels were significantly longer in words with /t/ than /k/ (85 vs. 77 ms). As the main issue of this section, the main effect for consonant quantity was found to be significant \[ F(1, 3)=27.10, p=0.014 \]. Vowels were significantly longer, on average, when they preceded geminates than singletons (89 vs. 73 ms), although the difference is only 16 ms. There was no other significant effect or interaction in Set 1. The left panel of Figure 2 shows the duration of vowels preceding single versus geminate stops for each rate (Set 1). The vowel duration difference between the singleton and geminate contexts was significant for the slow and normal rates, but the difference did not reach significance for the fast rate.

For Set 2, in which segmental composition of words varied more extensively, results were consistent with those of Set 1, in which segments were controlled. As in the Appendix, the preceding vowels in Set 2 included all five vowels /i e a o u/, and the first consonants in CV(C)CV also varied in /b d g t k m n s h r w j/, while some words even began with a vowel. In spite of these segmental differences, which are known to increase durational variability, an effect of speaking rate \[ F(2, 6)=25.04, p=0.001 \], and, more notably, an effect of consonant quantity \[ F(1, 3)=11.35, p=0.043 \] were found, and there was no significant interaction. The right panel of Figure 2 shows these two effects (Set 2). For the normal and fast rates, vowels preceding geminates were significantly longer than those preceding singletons, but the difference did not reach significance for the slow rate. Results of Sets 1 and 2 generally support previous studies showing a small duration difference between vowels preceding singletons and geminates (Fukui 1978, Campbell 1999, Ofuka 2003), but the difference may not always be significant for different speaking rates.

3.2 Duration of following vowels

Figure 3 shows durational distribution of vowels following single and geminate stops for Set 2, including all three speaking rates. It shows that the range of duration for the singleton context (left panel) stretched more to the longer values (11–264 ms) than for the geminate context (right panel) (22–205 ms), although the peaks of the distribution curves were similar between the two categories.

For Set 2 data, all four main effects of rate (slow, normal, fast), consonant type (/t/, /k/), vowel type (/a/, /e/, and /o/), and consonant quantity (single, geminate) were significant. Vowel duration was longer for slower rates (129 vs. 63 vs. 46 ms) \[ F(2, 6)=27.68, p=0.001 \], and longer after /t/ than /k/ (81 vs. 78 ms) \[ F(1, 3)=10.13, p=0.05 \]. Effects of vowel type \[ F(2, 6)=19.96, p=0.002 \] indicated that duration of /a/ (82 ms) was significantly longer, on average, than duration of /o/ (76 ms), but these two vowels did not significantly differ from /e/ (80 ms). As the main interest of this section, there was a significant effect of consonant quantity \[ F(1, 3)=15.08, p=0.03 \]. As in the previous studies, vowels were shorter when they followed geminates than singletons (74 vs. 85 ms), and these following vowels showed a durational pattern opposite to that of preceding vowels. A rate by consonant quantity interac-

Fig. 2 Duration of vowels preceding single and geminate stops spoken at three speaking rates. The error bars represent one standard error from the mean. **p<0.01; *p<0.05; n.s. p>0.05
Vowels following singletons

Vowels following geminates

Fig. 3 Distribution of duration of vowels following single and geminate stops spoken by four speakers across three rates (Set 2).

Fig. 4 Duration of vowels following single and geminate stops spoken at three speaking rates. The error bars represent one standard error from the mean. **p=<0.01; *p=<0.05; n.s. p>0.05

In summary, as far as absolute duration of vowels is concerned, the present results are consistent with previous studies (Fukui 1978, Campbell 1999, Ofuka 2003) in that there is a small vowel duration difference between the singleton and geminate contexts. However, this difference does not seem to be a necessary and sufficient feature associated with the consonant quantity distinction.

3.3 Durational ratios of stop closure to preceding and following vowels (C/V1 and C/V2)

In Hirata and Whiton (2005), mean C/V1 ratios across three rates were 1.00 [SD=0.36] (Set 1) and 1.06 [SD=0.40] (Set 2) for singletons, and 2.38 [SD=0.81] (Set 1) and 2.61 [SD=1.03] (Set 2) for geminates. In the present study, mean C/V2 ratios were calculated to be
0.86 [SD=0.26] (Set 1) and 0.98 [SD=0.27] (Set 2) for singletons, and 2.93 [SD=0.93] (Set 1) and 3.25 [SD=0.87] (Set 2) for geminates. Figure 5 plots these ratios (circles for singletons and triangles for geminates) as well as their standard deviations (horizontal bars). The comparisons between the results of C/V1 and C/V2 suggest that the mean ratios were further apart between the singleton and geminate categories for C/V2 (the mean differences of 2.07–2.27) than for C/V1 (the mean differences of 1.38–1.55). The standard deviations also separated the two categories more clearly for C/V2 than C/V1.

One of the main goals of this study was to examine how accurately C/V2 ratios can classify single and geminate stop categories across three speaking rates. For this, optimal boundaries that classified the two categories with highest accuracy were calculated. As shown in the bottom rows of Table 1, the C/V2 optimal boundary in Set 1 was either 1.50 or 1.51, similar to the optimal boundary found for C/V1 ratio (1.52 or 1.53), and both of the values, 1.50 and 1.51, classified the singleton and geminate categories with 98.9% accuracy. This means that, if C/V2 ratio is smaller than 1.50 (or 1.51), the disyllable word almost always includes a single stop, and if it is greater than 1.50 (or 1.51), the word almost always includes a geminate stop. In addition, the values anywhere between 1.36 and 1.74 were found to classify the two stop categories with over 98.0% accuracy. This classification accuracy was higher than the accuracy with the C/V1 optimal boundary ratio (92.1%) in Set 1 (Table 1).

For Set 2, the optimal boundaries of any of 1.59–1.76 classified the singleton and geminate categories with 98.8–98.9% accuracy. Thus, when C/V2 ratio is smaller than 1.59, the word almost always includes a single stop, and when it is greater than 1.76, the word almost always includes a geminate stop. In addition, the values anywhere between 1.51 and 1.76 were found to yield classification accuracy over 98%.

Combining the present results with those of Hirata and Whiton (2005), C/V2 was found to classify the singleton and geminate categories with highest accuracy among the parameters examined so far, closure duration, C/V1 ratio, C/W ratio, and C/V2 ratio (Table 1). However, in the case of Set 2, this does not mean that C/V2 ratio would, in general, be a better correlate of consonant quantity distinction than C/V1 ratio because of the nature of data set 2. C1V1 in Set 2 had a wider range of consonants and vowels, while C2V2 were limited to /ta te to ka ke ko/ (See the Appendix). Thus, the C/V2 ratio being more reliable than C/V1 in Set 2 might simply be a reflection of the segmental
Table 1  Optimal boundaries for closure duration, C/V1 ratio, C/W ratio, and C/V2 ratio. Set 1 includes 720 tokens, and Set 2 includes 2160 tokens, both spoken by all four speakers at three speaking rates.

<table>
<thead>
<tr>
<th></th>
<th>Optimal boundary</th>
<th>Number of misclassified tokens</th>
<th>Percentage misclassified</th>
<th>Classification accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Closure duration</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Set 1</td>
<td>73 or 74 ms</td>
<td>122 single 6 geminate</td>
<td>17.8%</td>
<td>82.2%</td>
</tr>
<tr>
<td>Set 2</td>
<td>115 ms</td>
<td>235 single 167 geminate</td>
<td>18.6%</td>
<td>81.4%</td>
</tr>
<tr>
<td><strong>C/V1 ratio</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Set 1</td>
<td>1.52 or 1.53</td>
<td>25 single 32 geminate</td>
<td>7.9%</td>
<td>92.1%</td>
</tr>
<tr>
<td>Set 2</td>
<td>1.53</td>
<td>114 single 74 geminate</td>
<td>8.7%</td>
<td>91.3%</td>
</tr>
<tr>
<td><strong>C/W ratio</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Set 1</td>
<td>0.35</td>
<td>8 single 7 geminate</td>
<td>2%</td>
<td>98%</td>
</tr>
<tr>
<td>Set 2</td>
<td>0.35</td>
<td>46 single 47 geminate</td>
<td>4.3%</td>
<td>95.7%</td>
</tr>
<tr>
<td><strong>C/V2 ratio</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Set 1</td>
<td>1.50 or 1.51</td>
<td>6 single 2 geminate</td>
<td>1.1%</td>
<td>99.8%</td>
</tr>
<tr>
<td>Set 2**</td>
<td>1.59–1.76</td>
<td>10–20 single 5–13 geminate</td>
<td>1.1–1.2%</td>
<td>98.8–98.9%</td>
</tr>
</tbody>
</table>

*The boundary ratios of 1.36–1.74 all yielded high classification accuracy of 98.1–98.9% with 1–13 singleton and 0–13 geminate tokens misclassified.

**The boundary ratios of 1.51–1.58 also yielded high classification accuracy of 98.0–98.7% with 22–38 singleton and 4–5 geminate tokens misclassified.

[Results of closure duration, C/V1 ratio, and C/W ratio are from Hirata and Whiton (2005).]

compositions of C2V2 and C1V1. However, we can make a strong conclusion based on data set 1. Set 1 had only the six word types, /ta(t:)a/, /te(t:)e/, /to(t:)o/, /ka(t:)a/, /ke(t:)e/, /ko(t:)o/, and thus, both sides of the contrasting stops were the identical vowels, which should yield the same amount of durational variability. In this data set, the C/V2 ratio served better in classifying the consonant quantity (with 98.9% accuracy) than the C/V1 ratio (with 92.1% accuracy). This result serves as evidence that the mutually enhancing durational relationship between the contrasting stops and the following vowels helps to classify the two stop categories with higher accuracy, at least in terms of the present production data.

4. Discussion and conclusions

The present study investigated whether duration of vowels preceding and following single and geminate stops in Japanese disyllables varied systematically across speaking rates, and whether those vowels were stable indicators for the stop quantity. The first question addressed in this study was whether the previous findings regarding the preceding vowels, i.e., longer vowels preceding geminates than singletons, held true when speaking rate varied. The distribution of vowel duration in Figure 1 suggested that the duration range was similar between the singleton and geminate contexts. However, there were significant effects of stop quantity, as well as of speaking rate, on the preceding vowel duration in the overall data of Sets 1 and 2. When each rate was examined separately (Figure 2), two of the three rates in both sets showed that vowels were significantly longer before geminates than singletons. Even though the duration differences did not reach significance for the fast rate of Set 1 and the slow rate of Set 2, they showed a tendency for the expected direction: vowels tended to be longer before geminates than singletons. The differences between the two quantity contexts were quite small, ranging from 11 to 18 ms on average, and this effect was much smaller than effects of speaking rates, which yielded a difference of about 70 ms between the fast and slow rates. However, the effect of stop quantity on vowel duration was clearly observed even when there was a great segmental variation within words as in Set 2. These results altogether are consistent with the previous findings.

The second question addressed in this study was whether vowels following the geminate stops were shorter than those following the singleton stops. The distribution of vowel duration in Figure 3 indicated that the duration range was smaller for the geminate than singleton context, but the medians did not appear to differ very much between the two contexts. However, there were small but significant effects of stop quantity, as well as effects of speaking rates, on vowel duration.
in the overall data of Sets 1 and 2. Looking at the individual rates in Figure 4, vowels tended to be shorter after geminates than singletons, as reported in the previous studies. However, these differences between the singleton and geminate contexts were not robust. Only for the slow rate of Set 1, vowels were significantly shorter after geminates than singletons, with a durational difference of 22 ms. Effects of speaking rate were greater, yielding the differences of about 70–90 ms between the fast and slow rates. If one compares effects of stop quantity on the preceding versus following vowel duration, effects seem to be slightly greater for the preceding vowels.

The third question of this study relates to the durational ratios of C/V1 and C/V2. Specifically, the study addressed the question of how accurately C/V2 alone can classify the singleton and geminate categories, and compared its relative classification accuracy with that of C/V1. As mentioned in the introduction, the traditional mora theory would predict no difference between the C/V1 and C/V2 ratios. However, given the previous finding on the vowel shortening after geminates and the vowel lengthening before geminates, the present study predicted that C/V2 ratios would exhibit a greater difference between the two stop categories. The results supported this prediction (Figure 5). The mean C/V2 ratios were 0.92 and 3.05 for singletons and geminates, while the mean C/V1 ratios were 1.03 and 2.50, respectively (averaged across two data sets). Thus, the distance between the mean ratios of the singleton and geminate categories was greater for C/V2 than C/V1.

Results of optimal boundaries for both sets of data showed that C/V2 ratio classified the singleton and geminate categories with higher accuracy (98.9%) than C/V1 ratio did (91–92%). In particular, the results of Set 1 provide evidence that C/V2 ratio can generally classify the two stop categories with higher accuracy, since the identical segmental variability before and after the contrasting stops (e.g., /tatt(:)a/, /tet(:)e/, and /kok(:)o/) enables the strict C/V1 versus C/V2 comparison. These results mean that, just by considering the closure duration relative to the following vowel alone, we are able to tell with 98.9% accuracy whether a disyllable includes a single or geminate stop, regardless of speaking rates, and regardless of the different segmental composition of words.

How do the present results fit in the extant literature of Japanese stop quantity distinction? Watanabe and Hirato (1985) and Hirata (1990a) found that the primary perceptual cue to the stop quantity distinction was the closure duration relative to the preceding vowel. There is a discrepancy with regard to the perceptual role of the following vowels: Hirato and Watanabe (1987) did not find the significant perceptual role of the following vowels, but Ofuka et al. (2005) did. Ofuka et al. noted that this discrepancy might be due to the fact that Hirato and Watanabe used isolated disyllables in which the vowel following the contrasting stops was in an utterance final position, whereas Ofuka et al. used a four-syllable word embedded in a carrier sentence. There is greater durational variability in phrase/utterance final positions (e.g., Mori 2004), and perceptual sensitivity decreases in those positions (Kato et al. 2003). Thus, the following vowel in a phrase-final position did not play a perceptual role in Hirato and Watanabe, but that in a phrase-medial position did in Ofuka et al.

In the present study, vowels following the contrasting stops were not in a phrase-final position, as the target disyllables were embedded in a carrier sentence /soko wa ___ to jomimasu/. The present production data showed a well-structured durational relationship between the contrasting stops and the following vowels (as well as the preceding vowels). Although absolute duration of these vowels alone could not be an independent correlate for the singleton/geminate distinction, C/V2 ratio served as an extremely good classifier for the stop quantity distinction across rates. It only makes sense that such well-structured relationships in the acoustic signals are used by native listeners, and that perceptual effects of vowels following the contrasting stops are present in phrase-medial positions. Thus, the present production data are compatible with the perceptual data of Ofuka et al. (2005). It is interesting to note that C/V2 ratio, if not in a phrase-final position, might possibly be a mutually enhancing perceptual cue, as shown in the inverse durational relationship between the closure and the following vowel. The use of C/V2 is well in line with Diehl and Kluender’s (1989) Auditory Enhancement Hypothesis: shorter vowels after geminates auditorily make the length of geminates sound longer, and longer vowels after singletons make the length of singletons sound shorter.

What might be the relative importance of C/V1 and C/V2 in the perception of stop quantity distinction? Although the present production data showed C/V2 as a better classifier than C/V1, we might need to consider the fact that these speech sounds come into listeners in sequence in on-line processing. It is possible that listeners are normally able to identify the stop quantity dis-
tinction with V1 and the target stop before complete information about V2 reaches to their ears, particularly when a preceding phrase gives listeners reliable speaking rate information. By the time V2 arrives in the listeners, C/V2 might play a role in confirming the prior decision. It would be interesting to conduct gating experiments (Grosjean 1980) in which one can examine on-line processing of Japanese words contrasting in quantity, and can find out up to which portion of words are sufficient for accurate recognition of those words.

How about other units that are proposed to be good candidates for reliable stop quantity distinction? How are all those pieces of information related to each other? One parameter, duration ratio of stop closure to word (C/W), was found to classify the two stop categories with high accuracy (95.7–98.0%) across rates in Hirata and Whiton’s (2005) production data (see Table 1). According to the traditional mora timing theory Hirata and Whiton credited, C/W ratio can accommodate for the difference in the number of moras in words. For example, C/W ratios in two/three-mora words (CV.CV versus C.V.CV) are expected to be 0.25 (=0.5/2 moras) and 0.50 (=1.5/3 moras), on average, with the boundary ratio of 0.375, but in the case of three/four-mora words (CV.CV.CV versus C.V.CV.CV or CV.CV.C.CV), ratios would be 0.17 (=0.5/3 moras) and 0.38 (1.5/4 moras) with the boundary ratio of 0.275. This C/W ratio was proposed to take advantage of the durational stability of word (Port et al. 1987), even when durational compensation within each mora is not perfect. One can expect, therefore, that the usefulness of C/W ratio might appear when examining a set of words with extremely varied consonants and vowels. Other possible parameters are VC as a perceptual unit, as proposed by Ofuka et al. (2005), and ratio of closure to the preceding CV mora (Idemaru 2005). We currently do not have a complete agreement in the relative importance of all of these parameters. Further studies in the interplay among all of these parameters would be necessary for our complete understanding of the stop quantity distinction in Japanese.

All of the above parameters and multiple pieces of information normally might not come in conflict with one another in natural speech, and native listeners might use them altogether for a coherent percept of a single or geminate stop. As mentioned in the introduction, nondurational cues such as fundamental frequency and phonological pitch accents can affect the stop quantity distinction as well. Acoustic redundancy associated with the stop quantity distinction is clear just by looking at the four parameters examined in Table 1, in which any of the parameters alone can classify the two stop categories over 80%, and this redundancy is probably to help listeners to perceive words under various circumstances (e.g., in noisy backgrounds). The question to be asked would, then, be not so much about which is and is not the definite primary perceptual cue, but rather about how these various parameters are used in different speech contexts.

For example, Hirata (1990a) examined native Japanese speakers’ perception of /it(:)a/ in two different contexts, one in which these words were presented in isolation, and the other in which the words were followed by a carrier sentence of varied rates. The results showed that perceptual cues listeners used were different between these two contexts. Similarly, Hirata and Lambacher (2004) examined whether native Japanese listeners could identify vowel length accurately when disyllables containing short and long vowels, originally spoken in a carrier sentence of different speaking rates, were excised out of the sentence context. Results showed that the perceptual accuracy significantly deteriorated in this excised context, even though disyllables themselves contained clear durationual information. This indicates that native listeners use preceding and/or following phrases for the accurate identification of vowel length, and that those disyllables alone might not contain totally sufficient cues to vowel length, especially when speaking rate varies. These two studies might suggest that native listeners’ perception is flexible and uses whatever available acoustic information in given contexts. If so, then, strict invariants might not be necessary in speech perception for the purpose of meaningful communication (Lindblom 1991).

In conclusion, this study showed that, in spite of speaking rate variations, vowels were fairly consistently longer when they preceded geminate than single stops, and that vowels tended to be shorter when they followed geminate than single stops. Effects were greater for the preceding than following vowels in terms of absolute duration. However, the study also found that the closure to following vowel ratio classified the two stop quantity categories with higher accuracy than the closure to preceding vowel ratio did. Further studies are necessary to reach the complete understanding of the mechanism of consonant quantity distinction in Japanese. Investigating words longer than disyllables in both production and perception will provide additional insight into this issue. It will also be extremely useful to investigate various contexts in
which the stop quantity distinction is made, such as isolated words versus words in sentence-initial, sentence-medial, and sentence-final positions. Finally, investigating the perceptual role of, and the acoustic analysis of, one or two moras outside the target words will be helpful. For example, the present target words formed a phrase with the following particle /to/ in a carrier sentence, and it is possible that this particular context made V2 and C/V2 ratio stable. It will be of interest to examine whether the same target words followed by different particles such as /to/, /ga/, /ni/, and /wa/ yield the same C/V2 stability.

Acknowledgments

Ideas of this study emerged shortly after the workshop, Kyoto Sokuon Kenkyukai, through Phonological Association in Kansai (PAIK) and Kyoto Sangyo University, Kyoto, Japan, in March of 2006. I thank Itsue Kawagoe and Masako Ariai for organizing this wonderful workshop. I also thank Osamu Fujimura and the other workshop participants for the inspiring discussion, and Yoko Mori, Etsuko Ofuka, and Jon Bernard for their insightful comments on the earlier version of this manuscript. Thanks also go to my students, Jake Whiton and Connor Forbes, for helping measurement of the present production data. This study was supported by Colgate University’s Research Council.

[Notes]

1) In the present study, “C/V1” refers to the ratio of either single or geminate stop closure (C1 or C2,C2) to the preceding vowel (V1) in disyllables (C1)V1(C2,C2)V2. Likewise, “C/V2” refers to the ratio of either single or geminate stop closure (C1 or C2,C2) to the following vowel (V2).

2) This paper followed International Phonetic Association transcription (Pullum and Ladusaw 1986) for a geminate obstruent, as in /ita/ (‘went’), which is equivalent to /iQta/ in Japanese phonology (Vance 1987).

3) The duration ratio of singleton and geminate stop closures is approximately 1:3 (in the range of 1:2.6–3.3, Han 1962, Homma 1981, Han 1992); this result is stable across speaking rates (Hirata and Whiton 2005). What would mora theory predict the weight of the stops themselves to be? Given that a CV receives one mora and a CCV two moras, and that a geminate has three times the length of a singleton, then a singleton would receive 0.5 moras and a geminate 1.5 moras. From this it follows that the vowel in any CV would also receive 0.5 moras. Assuming that CV moras are structurally identical regardless of their position in the word, then each segment of a CV mora will always receive half a mora. Although the C and V in a CV are clearly not of equal duration acoustically (Warner and Ariai 2001, Campbell and Sagisaka 1990), this implicit phonological weighting provides a point of comparison in the process of interpreting various durational values and ratios (see Hirata and Whiton 2005, for more discussion).

4) The central claim of this hypothesis is that multiple gestural components for producing a segment covary not because of physiological constraints, but because these components altogether yield redundancy in acoustic signals. These multiple, redundant acoustic correlates, in turn, create auditory effects that are mutually enhancing. One example they give is that perception of stop voicing, whose major acoustic correlate is VOT, is enhanced by duration of preceding vowels.

5) Lindblom (1991) argues, “… phonetic invariance is not necessary at all for adequate lexical access, because successful speech understanding presupposes gestures that are sufficiently contrastive but not necessarily physically constant” (p. 21).

References


Durational Variability and Invariance in Japanese Stop Quantity Distinction: Roles of Adjacent Vowels

ence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian,” *Journal of Phonetics* 30: 2, 229–258.


Mor, Y. 本屋順子（2001）「アクセントの最終 lengtheningの影響」『音声研究』7: 1, 70–76.


Otsobo, K. 大塚一夫（1981）「日本人的促音の有無の判別能力について」『名古屋大学総合言語センターレ」語学文化集』3: 1, 39–47.


Sugito, M. 杉藤美代子（2000）「音声波形は音をる」『日本音声学研究』4大阪：和泉書院.


Toda, T. 户田貴子（2003）「外国人学習者の日本語特殊拍の習得」『音声研究』7: 2, 70–83.


(Accepted February, 2007)
Appendix  [Materials from Hirata and Whiten (2005)]

Set 1: Nonword stimuli

<table>
<thead>
<tr>
<th>k</th>
<th>/kaka/</th>
<th>/kaka/</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/keke/</td>
<td>/keke/</td>
</tr>
<tr>
<td></td>
<td>/kok/</td>
<td>/kok/</td>
</tr>
<tr>
<td>t</td>
<td>/tata/</td>
<td>/tata/</td>
</tr>
<tr>
<td></td>
<td>/tte/</td>
<td>/tte/</td>
</tr>
<tr>
<td></td>
<td>/toto/</td>
<td>/toto/</td>
</tr>
</tbody>
</table>

Set 2: Real word stimuli

<table>
<thead>
<tr>
<th>ka</th>
<th>/buka/</th>
<th>‘subordinate’</th>
<th>/buka/</th>
<th>‘price’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/joka/</td>
<td>‘free time’</td>
<td>/joka/</td>
<td>‘four days’</td>
</tr>
<tr>
<td></td>
<td>/dkka/</td>
<td>‘market price’</td>
<td>/dkka/</td>
<td>‘parents’ house’</td>
</tr>
<tr>
<td></td>
<td>/rika/</td>
<td>‘science’</td>
<td>/rika/</td>
<td>‘early summer’</td>
</tr>
<tr>
<td></td>
<td>/nika/</td>
<td>‘second chapter’</td>
<td>/nika/</td>
<td>‘daily schedule’</td>
</tr>
<tr>
<td></td>
<td>/haka/</td>
<td>‘gravestone’</td>
<td>/haka/</td>
<td>‘mint’</td>
</tr>
<tr>
<td></td>
<td>/skka/</td>
<td>‘hill’</td>
<td>/skka/</td>
<td>‘writer’</td>
</tr>
<tr>
<td></td>
<td>/ika/</td>
<td>‘less than’</td>
<td>/ika/</td>
<td>‘first chapter’</td>
</tr>
<tr>
<td></td>
<td>/aka/</td>
<td>‘red’</td>
<td>/aka/</td>
<td>‘worsen’</td>
</tr>
<tr>
<td>ke</td>
<td>/jake/</td>
<td>‘desperation’</td>
<td>/jake/</td>
<td>‘windbreaker’</td>
</tr>
<tr>
<td></td>
<td>/ake/</td>
<td>‘open’</td>
<td>/ake/</td>
<td>‘disbelief’</td>
</tr>
<tr>
<td></td>
<td>/kke/</td>
<td>‘bet’</td>
<td>/kke/</td>
<td>‘beriberi’</td>
</tr>
<tr>
<td>ko</td>
<td>/hako/</td>
<td>‘box’</td>
<td>/hako/</td>
<td>‘eight (counter)’</td>
</tr>
<tr>
<td></td>
<td>/nekko/</td>
<td>‘cat’</td>
<td>/nekko/</td>
<td>‘root’</td>
</tr>
<tr>
<td></td>
<td>/nakko/</td>
<td>‘past’</td>
<td>/nakko/</td>
<td>‘parenthesis’</td>
</tr>
<tr>
<td>ta</td>
<td>/heta/</td>
<td>‘unskilled’</td>
<td>/heta/</td>
<td>‘decreased’</td>
</tr>
<tr>
<td></td>
<td>/buta/</td>
<td>‘pig’</td>
<td>/buta/</td>
<td>‘hit (past tense)’</td>
</tr>
<tr>
<td></td>
<td>/wata/</td>
<td>‘cotton’</td>
<td>/wata/</td>
<td>‘divided’</td>
</tr>
<tr>
<td></td>
<td>/nita/</td>
<td>‘boiled’</td>
<td>/nita/</td>
<td>(surname)</td>
</tr>
<tr>
<td></td>
<td>/uta/</td>
<td>‘song’</td>
<td>/uta/</td>
<td>‘sold’</td>
</tr>
<tr>
<td>te</td>
<td>/mate/</td>
<td>‘wait (imperative)’</td>
<td>/mate/</td>
<td>‘wait’</td>
</tr>
<tr>
<td></td>
<td>/sate/</td>
<td>‘let me see…’</td>
<td>/sate/</td>
<td>‘leave’</td>
</tr>
<tr>
<td></td>
<td>/ute/</td>
<td>‘hit (imperative)’</td>
<td>/ute/</td>
<td>‘sell’</td>
</tr>
<tr>
<td></td>
<td>/kte/</td>
<td>‘win (imperative)’</td>
<td>/kte/</td>
<td>‘selfish’</td>
</tr>
<tr>
<td></td>
<td>/tate/</td>
<td>‘vertical’</td>
<td>/tate/</td>
<td>‘stand’</td>
</tr>
<tr>
<td>to</td>
<td>/soto/</td>
<td>‘outside’</td>
<td>/soto/</td>
<td>‘softly’</td>
</tr>
<tr>
<td></td>
<td>/mito/</td>
<td>(place name)</td>
<td>/mito/</td>
<td>‘catcher’s mitt’</td>
</tr>
<tr>
<td></td>
<td>/moto/</td>
<td>‘formerly’</td>
<td>/moto/</td>
<td>‘more’</td>
</tr>
<tr>
<td></td>
<td>/oto/</td>
<td>‘sound’</td>
<td>/oto/</td>
<td>‘husband’</td>
</tr>
<tr>
<td></td>
<td>/kato/</td>
<td>‘transition’</td>
<td>/kato/</td>
<td>‘cut’</td>
</tr>
</tbody>
</table>

(All words in Set 1 were presented in *katakana* orthography, and speakers were told that they were all nonsense words, although some words such as /tata/ are real words when they are written in appropriate *kanji* and *hiragana* orthography. Speakers were asked to read all words in Set 1 so as to have a pitch accent on the first mora, but the location of pitch accents for real words in Set 2 was left up to the speakers.)