The Perception of Geminate Stops by L1 English Learners of Japanese

Miyuki Takeuchi
Indiana University

1. Introduction

This study investigates the perception of Japanese geminate stops [pp, tt, kk] by L1 English learners of Japanese in comparison with native Japanese speakers. The investigation attempts to identify the specific perceptual differences between the two groups in their attention to the phonetic features for identifying a singleton/geminate. Ladefoged (2006) defines geminate consonants as “long consonants that can be analyzed as double consonants” (p. 244). When consonants are geminated, the obstruction of airstream is prolonged, that is, the acoustic duration of the stop becomes longer, thus, geminate consonants are pronounced audibly longer than the equivalent singletons. Some languages, such as Italian, Japanese, Arabic, Hebrew, Finnish and Hungarian, have geminate consonants. In these languages the contrast between a single and a geminate consonant is phonemic. The acoustic duration of a geminate consonant relative to a single consonant, as well as the phonemes that can be geminated, vary from one language to another. On the other hand, many languages including English do not have a consonant length contrast morpheme-internally, though English has double consonants that occur across morpheme boundaries such as in the word *bookcase*.

The present study explores whether lack of a consonant length contrast in English affects the way English-speaking learners of Japanese perceive Japanese geminate consonants. The difficulty among adult L2 learners in discriminating novel L2 phonemes that are lacking in their L1 inventory has been reported in previous studies, for example the problem of distinguishing English /r/ and /l/ among Japanese speakers (Takagi, 2002;

Aoyama, Flege, Guion, Akahane-Yamada, and Yamada, 2004). On the other hand, Flege (1995) maintains that "language-specific perceptual patterns are modifiable to some extent" and "the perceived relation of L1 and L2 sounds may change during naturalistic L2 learning" (p. 237). If these arguments are applied to the perception of Japanese geminates by L1-English learners of Japanese, their performance could be improved after some time of exposure to Japanese though it may be difficult for them to attain the same criterion for the identification of a singleton/geminate as that of native Japanese speakers.

Indeed, previous studies on the perception of the Japanese single/geminate consonants by L2 learners (e.g., Hayes, 2002; Hayes, 2005; Hayes-Harb and Masuda, 2008) agree that the phonological distinction between a singleton and a geminate improves with their Japanese proficiency. However, the studies which investigated the production of geminate/singleton consonants by L2 learners reported that their performance, even at the advanced level, was often distinctive from that of native speakers in terms of the rates of the acoustic length between a singleton and a geminate and implementation of the geminate consonants of varied place of articulation. The findings from the previous production studies suggest difficulty in attaining the target norm at a phonetic level among learners. Given that insufficiently developed perceptional criteria affects productions (Flege, 1995), I propose that it should be necessary to investigate learners' perception of geminate/singleton consonants not only at the phonological level but also at the phonetic level, in order to pinpoint the possible perceptual difference between learners and native speakers which may affect learners' production of geminate consonants. Hence, the focus of the present study is on L1 English learners' (ELJ, henceforth) sensitivity to the phonetic features, i.e., the length of the preceding syllable and place of articulation of single/geminate consonants as well as varied lengths of closure duration, in their perception of Japanese stop single/geminate consonants[1]. Learners' performances are compared to that of native Japanese speakers (NJS, henceforth), following the research questions below:

(1) How do the two groups identify the stimuli with varied lengths of closure duration?

(2) Does the length of preceding syllable affect the way the ELJ and NJS identify a singleton/geminate?

(3) Does place of articulation of the stops [p, t, k] affect the way the ELJ and NJS identify a singleton/geminate?

2. The implementation of geminates in Japanese and English

Geminate consonants in native Japanese occur with voiceless stops [p, t, k], fricatives [s, ʃ], and an affricate [tɕ], and they are phonemically distinguished from the equivalent singletons. For example, the segments [t] and [tt] are processed into two separate phoneme categories. Thus, the acoustic stimuli, [oto] and [otto], are categorized separately and are allocated to different meanings, which are 'sound' and 'husband' respectively. When native Japanese speakers produce a minimal pair of a singleton and a geminate, the ratio of acoustic duration of the single and double consonant falls roughly in the range of 2.5-3.2 : 1 according to Han's (1992) data obtained from 10 native Japanese speakers. (In Han's measurement, the voice onset time [VOT] was not included as a part of the stop consonant.) The ratio of geminate vs. singleton approximates 3.0 : 1.

According to Han (1992, 1994), the first part of a geminate consonant (often

[1] Stop consonants are chosen because (a) it is relatively easy to manipulate the length of closure duration and (b) they provide three variables in terms of place of articulation.

assigned to its own phoneme, transcribed /Q/, by Japanese phonologists) fills a component that has a length of a moraic (or syllabic) consonant, and the second part is an equivalent of a single consonant C, which is why the length of a geminate consonant is roughly three times of that of a single consonant. For example, a word [otto] forms three moras, [o. t. to], in which the first part of a geminate [t] lasts as long as the other moras in the word, e.g., [to]. Homma (1981) also reported similar findings to Han's (1992, 1994). Note, however, that the argument above has been controversial since there are other scholars who found the ratio of "geminate vs. singleton" does not reach 3.0 : 1 (e.g., Fukui, 1978; Beckman, 1982)[2].

In English, the consonant length contrast is allophonic. Implementation of the length contrast may vary across individuals or dialects, which does not affect the meaning of a word. For example, whether 'roommate' is pronounced as [rumeɪt] or [rummeɪt], both [m] and [mm] are processed in a single phonemic category, and the whole word, whether it has a single or a double [m], represents the same meaning. However, double (geminate) consonants can occur across morphemes, for example in 'unknown' and 'wholly,' or across words, as in 'this Sunday' (Abercrombie, 1967). The ratios of a singleton to a geminate which are observed across morphemes or words have been reported in several previous studies: Delattre's (1968) spectrographic analysis found that the ratio of /ll/ vs. /l/ in *it will lend* and *it will end* was 1.3 to 1, and that of /nn/ vs. /n/ in *I've seen Nelly* and *I've seen Elly* was 1.5 to 1. Han (1992) analyzed the wideband spectrograms of the production by two native English speakers. Calculation of

[2] Fukui (1978) and Beckman (1982) included the VOT as a part of a stop, and obtained the ratio 2.1-2.4 : 1, and 2.25: 1 respectively. Beckman (1982) reached the ratio 2.79 : 1 with the VOT included in the following vowel. The argument seems to partly involve the issue of measurement.

/tt/ vs. /t/ in *cattail* and *catfish/catnip*, and *get Tom* and *get Mary* showed that the rate of /tt/ to /t/ was 1.9 to 1. Kaye (2005) analyzed the ratio of /mm/ to /m/ in *immature* and *mature* pronounced by 10 native English speakers to find that the ratio was1.82 to 1 on average. Though there was variation in the rate across the phonemes with which germination occurs (1.3-1.9 to 1), these ratios are far from reaching2.5-3.2 to 1 observed in Japanese.

3. Previous studies on perception/production of Japanese geminates by L2 learners

The previous studies on L2 learners' perceptual performance of Japanese geminates mostly agree that the performance improves with experience in Japanese language (Hirata, 1990 b, Hayes, 2002; Mah and Archbald, 2003; Hayes, 2005; Hayes-Harb and Masuda, 2008). For example, Hayes (2002) examined native English speakers with no experience in Japanese and L1 English learners of Japanese in three levels. The informants were asked to determine whether two stimuli, which consist of minimal pairs of a singleton and a geminate or the same two stimuli, are the same or different. The results showed that the participants' sensitivity, or the rate of the correct responses, improved with proficiency. Hayes (2005) examined perceptual performance by English speakers with no experience of studying Japanese, L1 English learners of Japanese and native speakers, using the stimuli consisting of six non-words, in which the duration of stop consonants /t/, /k/, and a fricative /s/ were manipulated into 13 levels. The Japanese native speakers' responses demonstrated that they identified the single/geminate contrast categorically, while monolingual English participants identified the contrast linearly depending on the consonant length. The responses of L1 English learners of Japanese fell in between the native Japanese speakers and English monolinguals, which indicates that the learners had learned the categorical identification of the single/geminate contrast, though not perfectly.

While the studies above reported that learners could approach the target-like categorical distinction between a singleton and a geminate, there are few studies which focused on the specific differences in the perceptual performance between native Japanese speakers and learners. One of such studies is Hirata (1990 a, 1990 b), which reported the effect of length of the preceding syllable on the perception of the singleton/geminate contrast at the word level (although her main concern was rather about the perceptual difference of the singleton/geminate contrast between learners and native Japanese speakers at the sentence level). The study found that the duration of the preceding vowel (which makes a syllable by itself in her study) affected the Japanese participants' judgment: the longer the duration of the preceding vowel was, the more likely the following consonant was judged as a singleton. In contrast, difference in the duration of the preceding vowel did not affect learners' judgment.

Some studies reported that it was difficult for learners of Japanese to attain the target-like production of the singleton/geminate contrast even though their perceptual performance approached the target level. Hayes-Harb and Masuda (2008) demonstrated that their L1 English learners of Japanese, who were able to phonologically distinguished geminates from singletons, did not perform as well in the production tasks. Mah and Archbald's (2003) single informant, who is a native speaker of English with half-a-year experience of studying Japanese, was able to make a distinction between a single and geminate consonant perceptually, whereas her production showed considerable variation across place of articulation and among tokens.

Han (1992) demonstrated specific differences in the production of geminate consonants between L1 English learners (who are fluent in speaking Japanese) and native speakers, more specifically in the rates of the acoustic length between a singleton and a geminate and implementation of the geminate consonants of varied place of articulation. Her informants were asked to produce ten Japanese minimal pairs that contrast single and geminate stops [p/pp, t/tt, k/kk] contained in a carrier sentence. The ratio of geminates to singleton among the native Japanese speakers was 2.72-3.22 : 1, while the learners' ratio was found to be smaller, i.e., 1.67-2.06 : 1. The implementation of geminate consonants across place of articulation by the Japanese speakers showed the following order in terms of the closure duration of the three geminate stops: pp > tt > kk (pp=200.6ms, tt=198.6ms, kk=184.2ms). The order by the learners was: pp > kk > tt (pp=179.2ms, kk=173.8ms, tt=172.0ms). While native speakers' production of [kk] is apparently shorter than the other two, there was not much difference across place of articulation found with learners' production.

It should be noted that "many L2 production errors have a perceptual basis" (Flege, 1995, p. 238). Flege's claim is based on his hypothesis that insufficiently developed L2 phoneme inventories prevent learners from correctly perceiving some L2 sounds: the incorrectly perceived sounds are stored and structured in the long term memory, which causes inaccurate production (p. 236). Assuming the findings on learners' production reported above are the reflection of their perceptual capacity, the difference between learners and native speakers would be recognized in their perceptual performances, too. The smaller closure rate between a geminate and a singleton as well as variability among learners may be shown in learners' perceptual performance as an unclear distinction between a singleton and a geminate, and their non-sensitiveness to the length of the preceding vowel (syllable) and place of articulation might be shown in their lack of the perceptional attention to those features, when experimental stimuli are appropriately controlled for these properties. Hirata (1991a, 1991b), as reported above, examined perception performances by learners and native Japanese speakers, using the stimulus with a variation in the acoustic consonant duration for the target singleton/geminate contrast and with an additional variation in the length of the preceding vowel. However, she conducted the experiment, examining five learner informants of different proficiency levels and with only one type of stimulus. The present study attempts to build on Hirata (1991a, 1991b), examining increased number of participants of the same level with the stimuli controlled for the investigation of both the effects of place of articulation and length of the preceding syllable.

4. The procedure

4.1. Informants

Informants consisted of ten L1 English-speaking learners of Japanese (ELJ) and ten native Japanese speakers (NJS). The average age among the ELJ was 19.3, within the range of 18 to 24, whereas the average age among the NJS was 41.5, within the range of 23 to 60. None of the informants had been diagnosed with any kind of hearing or speech problems.

All the ELJ were students, studying Japanese in a Japanese language course at a Midwestern university in the U.S. They were in the second year of the course, having studied Japanese for a year and two months at the time of the experiment, or admitted in the second year at the beginning of the semester after passing the proficiency examination.

The ELJ informants were in the middle of the third semester of studying Japanese, and can be described as learners at the beginning to intermediate level. Five of the ELJ studied no other foreign language besides Japanese. Three of them studied Spanish, one studied French, and one Bengali in addition to Japanese. Since Spanish, French, and Bengali do not have geminate consonants as a phoneme, the experience in studying these foreign languages was not expected to have an influence on the perception of Japanese geminate consonants.

Ten NJS were recruited as a control. All of them were living in the US at the time of the study. All of them were born and brought up in Japan. All NJS have lived in Japan until the age of 22 or older before coming to the U.S. The length of their stay in the U.S. ranged from three months to 22 years with an average of 6.5 years. All of them spoke English as a second language, though their proficiency in English varied. Four of the NJS were from the area around Tokyo where standard Japanese is spoken. Six of them are from Kansai or Nagoya area where Kansai and Nagoya dialects are spoken. There is little difference in the realization of geminates across the Japanese dialects, and the speakers of Kansai and Nagoya dialects are also exposed to the standard Japanese quite extensively in their daily lives. Therefore, I believe the dialectal variation among the informants most likely did not affect their perception of the single/geminate contrast.

4.2. Stimulus and presentation

The experimental stimulus package consisted basically of nine two-syllable-long non words, *bipa, bita, bika, dapa, data, daka, mupa, muta, and muka*[3]. These words

[3] A few of the singleton or geminate form of these words are found in modern Japanese vocabulary, such as the geminate form of data [datta], which is the past tense of *da* or 'to be,' and the singleton form of bika [bika] ,which means 'beautification,' but the others were not commonly used in everyday modern Japanese.

were created so that each represents one of the single/geminate stops in Japanese [p/pp], [t/tt], [k/kk] preceded by one of the three kinds of vowels [i], [a], and [u] of the five vowels[4] in Japanese, since they represent high, low and back vowels respectively. The vowel [a] was attached at the end, and the first consonant was chosen so that the word and its geminate form were most likely to be non-words. The created stimuli were read by the author (a native Japanese speaker) and were recorded, using *Wavesurfer* on the computer. All of them were pronounced with a pitch accent on the first syllable. The nine non-words were manipulated and arranged for the purpose of the experiment in the following way.

Each two-syllable-long non-word was modified so that the duration of the first syllable represented two lengths (80ms/100ms), and the closure duration of the medial consonants [p], [t], and [k] represented ten different lengths. To take an original non-word *bika,* for example, it has two variations "*bika* 80ms" and "*bika* 100ms" depending on the duration of [bi] (written as *bika*80 and *bika*100 respectively henceforth), and each of them demonstrates ten levels of closure duration with [k]. The 80ms- or 100ms-length of the first syllable was created by adjusting around the middle part of the first vowel of the recorded non-words, using the "cut" function on *Wavesurfer*. The closure duration of the medial consonant was manipulated from 100ms to 370ms by 30ms by adjusting the part between the end of the formants of the first vowel and the start of the burst of the medial consonant, using the "cut" and "insert silence" function on *Wavesurfer*. The Measurement of the stop closure duration should be noted. In the present study, the stop

[4] The five vowels are [a], [i], [u], [e], and [o].

closure duration does not include VOT, or voice onset time[5]. Describing with a phonological representation, the ten variations are on the continuum of [bika] to [bikka].

A practice session was given prior to the experimental session. The practice session consisted of five stimuli that were made from a non-word [upa] with the other accompanying conditions, including intervals, being the same as those set up in the experimental session. For the experimental session, each of the 18 sets, *bika80, bipa80, bita80, daka80, dapa80, data80, muka80, mupa80, muta80, bika100, bipa100, bita100, daka100, dapa100, data100, muka100, mupa100, muta100 was* presented once in this order with about 7-8 seconds interval between the presentation of each set, and then the 18 sets were presented repeatedly in the same order immediately after the first 18 sets were presented. Thus, each informant listened to 36 sets in total. Each set consisted of 10 stimuli with 10 levels of closure duration with 2900-ms interval between each. The order of the presentation of the 10 levels of closure duration in each set was scrambled to create three variations: for example, *bika80-1, bika80-2*, and *bika80-3.* For the presentation of the first 18 sets, one of the three variations was selected randomly, and in the repeated (or the second) presentation of the 18 sets, the variation that was not used for the first presentation was chosen so each participant listened to two variations of each set in terms of the order of the presentation of the 10 levels of closure duration.

Each informant listened to the stimulus package consisted of the practice session and 360 experimental stimuli (18 sets × 2 repetition × 10 levels of closure duration in a set) on the computer[6], and responded to each stimulus by choosing (circling) a singleton or a

[5] Whether to include VOT in the stop closure duration has been discussed in the previous studies: The present study followed Han (1992:113) that observed "most researchers tend to associate VOT with the following vowel" (cited from Port and Rotunno, 1979; Homma, 1981).
[6] No head phones were used.

geminate which was written on a sheet of paper in *hiragana* Japanese sillabary. To take an example of *bika*, the participants were supposed to choose ‘bika’ or ‘bikka’ written in *hiragana* by circling one of the two. *Hiragana* was used because the author assumed the learners, at least in the third semester of studying Japanese, understand the component of the Japanese orthography that represents geminate consonants.

The disadvantages of an off-line task, such as one employed in the present study, have been discussed by some researchers (e.g. Dupoux, Kakehi, Hirose, Pallier and Mehler, 1999). First, they pointed out the lack of response time control. However, the response time was controlled as the timing of the interval between two stimuli (2900ms), during which a participant had to respond by circling one of the two choices. All the participants claimed no problem with the timing after the practice session.

The second potential problem is the effect of orthography. This could be a problem because it is not certain if the participants can correctly match the perceived sound with the corresponding orthography. Therefore the researcher asked each leaner if they understand the orthographical contrast between a single/geminate after the practice session, and nobody claimed any problems.

4.3. Analysis

The responses by each participant were analyzed in terms of the rate of identification with a singleton. For example, if one judges a contrast [bita/bitta] as “bita,” a singleton, the identification rate was 100%, while if one judges it as “bitta,” a geminate, the rate was coded as 0%. These rates were arranged on graphs after they were averaged for each stimulus set by each group. Next, the boundary closure duration was computed for each stimulus set responded by each participant. The boundary closure duration shows the point of division, or the 50 % of the rate between the identification with a singleton and with a geminate. The 50 % boundary point was calculated using logistic regression, which is an appropriate way to find boundary point of this type within the generalized linear model. In order to examine the tendencies within each group, the following values were computed: (a) the average boundary values by each group; (b) the average boundary values for each of the three preceding vowels [i, a, u] by each group; (c) the average boundary values for each of the two lengths of preceding syllable (80/100ms) by each group; (d) the average boundary values for each of the three places of articulation of the stops [p, t, k] by each group.

5. Results

5.1. 50 % boundary point and statistic analysis

Table 1 shows 50 % boundary closure duration as an average of the relevant variables among the ELJ, namely the types of preceding vowels [i, a, u], place of articulation of the stops [p, t, k], the lengths of the preceding syllable. In the table, "Category" means a set with the same preceding vowel and the same length of preceding syllable. For example, "i 80" stands for the set with the preceding vowel [i] and 80ms-long preceding syllable, i.e., *bipa80, bita80, and bika80*. The "category" is further classified across place of articulation; i.e., p, t, and k under the column "place of articulation." For example, the column "p" (under "Place of articulation") of the category "i 80" in Table 1 shows the average 50% boundary closure duration with the stimulus *bipa80* among the ELJ. Table 2 shows 50 % boundary closure duration among the NJS.

Table 1. 50 % boundary closure duration by ELJ (ms)

Category	Place of articulation			Category average
	p	t	k	
i 80	188.3	193.0	174.4	185.2
a 80	187.0	174.8	194.2	185.3
u 80	185.4	189.9	181.9	185.7
i80/a80/u80 total	186.9	185.9	183.5	185.4
i 100	193.1	191.8	175.7	186.9
a 100	182.3	184.1	197.5	188.0
u 100	176.6	182.4	185.3	181.4
i100/a100/u100 total	184.0	186.1	186.2	185.4
ELJ total	185.5	186.0	184.8	185.4

Table 2. 50 % boundary closure duration by NJS (ms)

Category	Place of articulation			Category average
	p	t	k	
i 80	215.6	206.6	194.6	205.6
a 80	214.1	202.1	200.6	205.6
u 80	208.1	215.6	206.7	210.1
i80/a80/u80 total	212.6	208.1	200.6	207.1
i 100	227.6	224.5	208.1	220.1
a 100	214.1	210.8	211.0	212.0
u 100	215.6	229.1	209.6	218.1
i100/a100/u100 total	219.1	221.5	209.6	216.7
NJS total	215.9	214.8	205.1	211.9

The values of the individual 50 % boundary values were submitted to a four-way 3×2×3×2 repeated measures analysis of variance (ANOVA) with the within-subject factors being consonants with different place of articulation (p, t, k), lengths of preceding syllable (80, 100ms), and preceding vowels (i, a, u), and with the between-subject factor being listener groups (ELJ, NJS), in order to find whether there is an effect of each variable and interaction of a combination of the variables in the identification of a singleton/geminate. ANOVA showed that there is a main effect of listener group (ELJ vs. NJS) [$F(1, 18) = 10.422$, $p = 0.005$], a main effect of consonants (p, t, k) [$F(2, 36) = 4.354$, $p = 0.020$], and a significant interaction of preceding vowels and consonants [$F(4, 72) = 7.582$, $p = 0.000$][7]. No other significant main effects or interactions are found. These mean that the 50 % boundary points are different between ELJ and NJS, and across the consonant conditions; and the listeners' reactions to the three preceding vowels differ across consonants. Since the effect of between-subject factor (group) was found, the boundary values for each group were submitted separately to a three-way 3×2×3 repeated measures of ANOVA with the within-subject factors being consonants, lengths of preceding syllable, and preceding vowels. This revealed that there is a main effect of lengths of preceding syllable with the NJS [$F(1, 9) = 8.667$, $p = 0.016$], but not with the ELJ [$F(1, 9) = 0.000$, $p = 0.997$]; and there is a main effect of consonant conditions with the NJS [$F(2, 18) = 8.229$, $p = 0.003$], but not with the ELJ [$F(2, 18) = 0.057$, $p = 0.945$]. Post hoc pairwise comparisons with the Bonferroni correction among the NJS across consonant conditions showed that there is a significant difference between [p] and [k] (M_difference = 11.250, $p = 0.017$), and [t] and [k] (M_difference = 10.217, $p = 0.023$), but not between [p] and [t] (M_difference = 1.033, $p = 1.000$). These show that the NJS

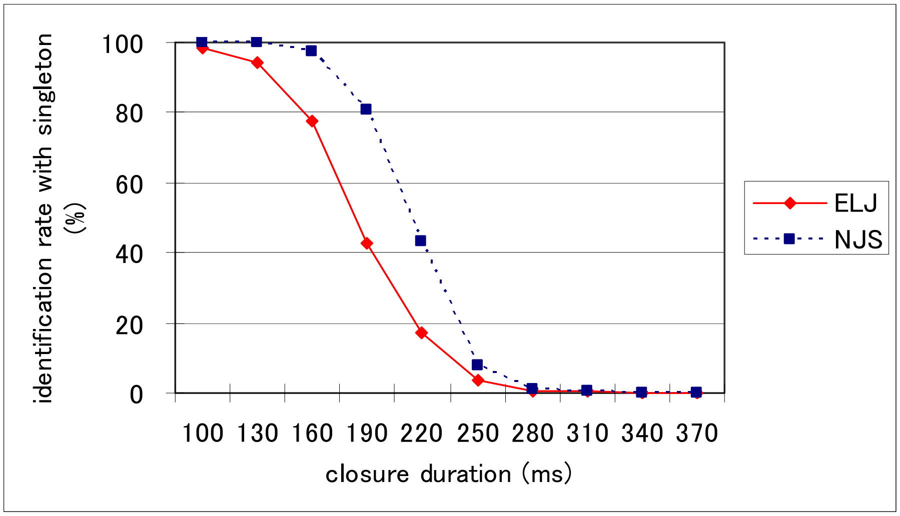
[7] The p value for which the null hypotheses is rejected was set at $p < .05$.

differentiates their identification of the stimuli between 80ms- and 100ms-preceding syllables, and across consonant conditions (place of articulation), specifically between the stimulus with [k] and those with [p] and [t]; but the ELJ does not react to either condition. A significant effect of interaction between consonants and preceding vowels was found only with the ELJ [$F(4, 36)= 7.927$, $p = 0.000$], which means that the ELJ's reactions to the stimuli with different place of articulation differ depending on the types of preceding vowel. No other significant main effects or interactions were found.

5.2. Comparison between ELJ and JNS

The timing of the identification has been found statistically different between the ELJ and NJS. Such a gap can be observed in Figures 1, which shows the identification rate with a singleton by the ELJ and NJS. To obtain each value on the graphs, at the point of the 190ms closure duration (on the x-axis) for example, all the identification rates with the 190ms closure duration were averaged for each group, and the values were plotted on the y-axis. The solid and dotted lines represent the values by the ELJ and NJS respectively. The gap between the two groups is clear. For example, at 190 ms-closure duration, about 40 % of the ELJ identified the stimulus as a singleton, while 80 % of the NJS did so. Although both ELJ's and JNS's identification rates form typical sigmoid function, the ELJ's identification rates start declining earlier (at a shorter closure duration point) with less sharper inclination than the NJS's rates.

Figure 1. Identification rate with a singleton / ELJ vs. NJS



5.3. Reactions to the stimuli with 80ms- and 100ms- preceding syllables

NJS's identification rates to the stimuli with 80ms-preceding syllable have been found to be statistically different from those with 100ms-preceding syllable, while ELJ's identification rates have not shown a gap. The difference in the reactions to the stimuli with two different lengths of the preceding syllable between the two groups is observed in Figures 2 and 3, which show the identification rate with a singleton of the stimuli with 80ms- and 100ms-preceding syllables by the ELJ and the NJS respectively. The solid and dotted lines represent the values for 80ms- and 100ms-preceding syllables.

In Figure 2, two graphs almost overlap entirely, which visually shows that the ELJ respond the stimuli with 80ms- and 100ms-preceding syllables in the same manner. Figure 3 shows a gap between the two graphs. The NJS respond to the stimuli with 80ms- and 100ms- preceding syllable in a slightly different manner. Between 160ms- and 250ms- closure duration, the NJS's point of identification with a geminate is delayed with the stimuli of 100ms-preceding syllable compared to those of 80ms: the gap ranges from 2 to 9 %. Though the gap is not so large, this phenomenon is observed consistently between 160ms- and 250ms-closure duration (see also Table 2 for the comparison).

Figure 2. Identification rate with a singleton by ELJ (80- /100ms-preceding syllables)

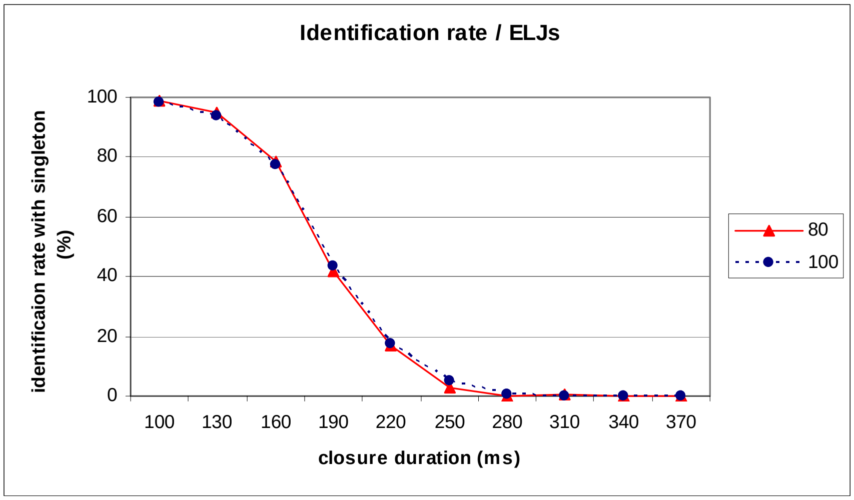
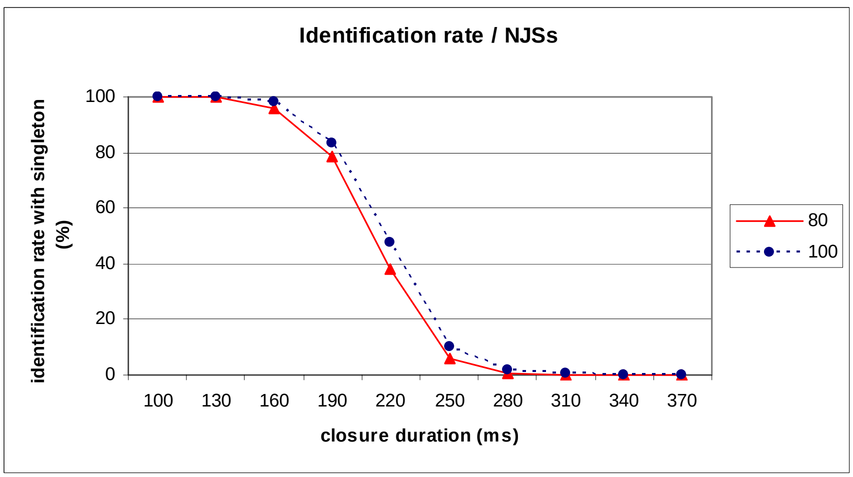


Figure 3. Identification rate with a singleton by NJS (80- /100ms-preceding syllables)



5.4. Reactions to the stimuli across places of articulation

NJS's identification rates of the stimuli across consonant conditions (places of articulation) have been found to be statistically different, while ELJ's identification rates have not. The NJS's reactions to [k] are different from those to [p] and [t]. The difference in the reactions to the stimuli with varied consonant conditions between the two groups is observed in Figures 4 and 5, which show the identification rate with a singleton by the ELJ and by the NJS across places of articulation. The narrower solid, dotted, and wider solid lines represent the values for stimulus with [p], [t], and [k] respectively.

Figure 4. Identification rate with a singleton by ELJ (across places of articulation)

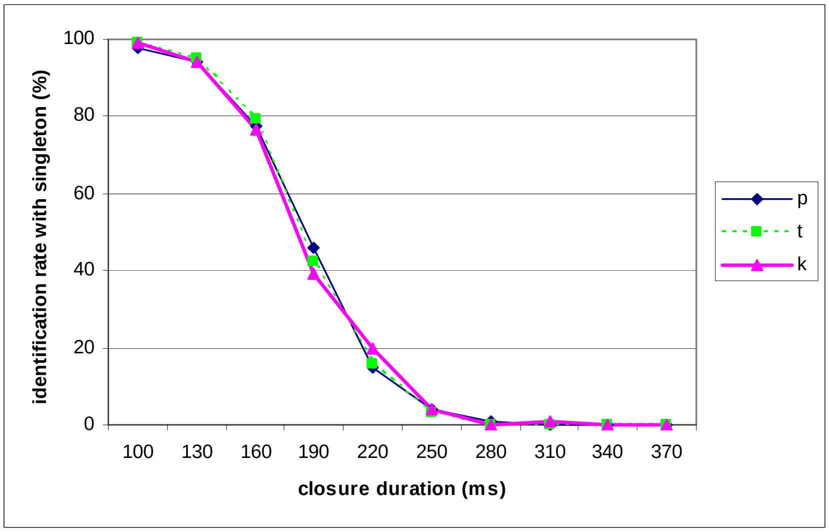
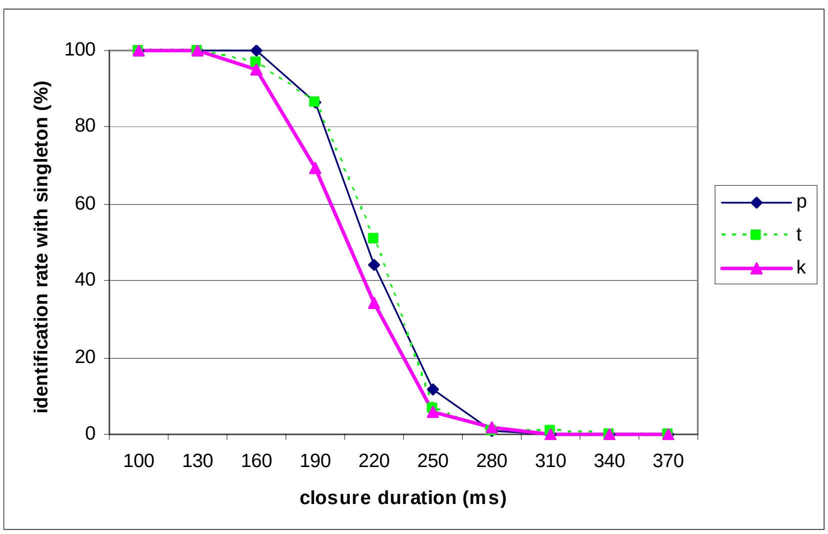


Figure 5. Identification rate with a singleton by NJS (across places of articulation)



In Figure 4, three graphs almost overlap, which visually shows that the ELJ respond the stimuli across places of articulation in the same manner. Figure 5 shows a gap between [k] and the other two, [p] and [t]. The NJS respond to the stimuli with [k] in a different manner. Between 160ms- and 220ms-closure duration, the NJS identifies the stimuli with [k] more likely as a geminate than those with [p] and [k]. In other words, in

NJS's perception, the stimuli with [k] can be identified as a geminate with shorter closure duration.

6. Discussion

The results of the experiment demonstrated that the timing of the identification with a singleton/geminate was different between the ELJ and NJS, though both groups basically showed a typical sigmoid function in accordance with the acoustic duration of the consonants. The ELJ identified certain stimulus more likely as a geminate than the NJS at certain closure duration of between 130ms and 220ms. Moreover, the duration of the preceding syllable and place of articulation of a singleton/geminate were found to affect NJS's identification, but not ELJ's. The NJS identified the stimulus with shorter preceding syllable more likely as a geminate, and the stimulus with [k] more likely as a geminate than those with [p] and [t] at certain closure duration. The types of the preceding vowel were found to have no effect on the identification by either group.

The fact that the ELJ's identification function revealed a sigmoid function similar to that of the NJS suggests that the ELJ has established two phonological categories, i.e., singleton and geminate, though not perfectly. However, in contrast to Hayes and Masuda's (2008) perception study, which claimed no significant difference between learners and native Japanese speakers in their phonological distinction of a geminate from a singleton, the present study demonstrated that, between 130 and 220ms closure duration, the ELJ identified certain stimuli with a geminate at a systematically higher percent-frequency than the NJS did. Hirata (1990 b) and Hayes (2005) investigated how learners identified the stimuli of various closure duration (13 – 17 levels) in comparison with native Japanese speakers, and claimed that learners' function approached the native speakers' level without discussing the difference between the two groups (at least at a word level). They did not point out learners' identification with a geminate at shorter closure duration than native speakers, which is one of the findings from the present study. I will propose three possibilities below to account for this phenomenon among the ELJ.

The first possibility is that the ELJ, who know that a Japanese geminate has longer stop duration than a singleton, have not yet established a well-defined criterion for the categorization. Therefore they may start being confused as soon as the closure duration gets somewhat longer than the stimulus that is clearly perceived as a singleton. At the 160ms-closure duration, for example, the NJS's identification rates with a singleton were 96 and 98 % for the stimuli with 80ms- and 100ms- preceding syllables, while the ELJ's rates were 78 and 77 % respectively. The NJS were almost sure that the stimulus with 160 ms closure duration was a singleton, which indicates that the NJS's threshold between a singleton and a geminate is clearly over 160ms. By contrast, the ELJ's criterion may not be firmly established or it may be wider to cover a lager range, which caused some stimuli with 160ms closure duration to be judged as a geminate.

Another possibility is the influence of the timing of double consonants realized morpheme-externally in English (the learners' L1). Han (1992) suggested the L1 influence to explain underdifferentiation in the durational contrast between single and double consonants in some of her L1 English informants' L2 Japanese production. The L1 influence could impact learners' perception, too. The previous studies (Delattre, 1968; Han, 1992; Kaye, 2005) reported that the rate of closure duration of double to single consonants in English was lower than the rate with Japanese (see page 4-5). The stop closure duration of English double consonants does not reach a double of an equivalent singleton, and this timing of English may have influenced the new criterion for the identification of Japanese geminates that is being established among the ELJ. If this is true, their boundary closure duration leans towards shorter. The influence of L1 (English) timing on the perception of Japanese geminate stops accounts specifically for the shorter boundary closure duration that has been demonstrated in the present study. The possibility of learners' confusion does not specifically explain the shortness of the boundary closure among the learners. In the present study, there is no case in which the ELJ, as a group, identified certain stimulus with a geminate at longer closure duration than the NJS. In this sense, the influence of the L1 timing could be more plausible than learners' confusion as a cause of the phenomenon. For stronger argument for the L1 influence, English data from the ELJ group, such as one examined in Han's (1992) study is necessary.

The third possibility is that both factors argued above, i.e. the confusion and L1 influence, may be interacting in the ELJ's judgment. The stimulus with the closure duration of 100ms is exclusively judged as a singleton, but the ELJ may be confused by some of the stimuli with the closure duration of 130ms or more. Given that the ratio of the acoustic duration between a double and a single consonant in English is 1.3-1.9 to 1, the closure duration of 130-190ms is the range in which English speakers produce double consonants when they occur morpheme-externally. Therefore, the timing of learners' confusion and the threshold of English single/double consonants coincides, which may have tempted them to perceive some uncertain stimuli in this range as a geminate.

The present study found that the duration of the preceding syllable and place of articulation of a singleton/geminate had an effect on NJS's identification, but not on

ELJ's. The result concerning the duration of preceding syllable confirms Hirata (1990a). The present finding, as well as Hirata (1990a) (see page 6 for details), suggests that the Japanese speakers' identification of a singleton/geminate counts not only on the closure duration but also on the relative perceivable length of the target to the preceding segment. This makes sense if Japanese speakers perceive or produce the first part of a geminate consonant /Q/, which is timed as one mora, in accordance with the mora-timing of the preceding syllable. In other words, when the preceding part is longer, the duration of the following geminate should be longer, which is why it is necessary to have longer closure duration for a stimulus with a longer preceding syllable in order to be judged as a geminate.

Ofuka, Mori, and Kiritani (2005), on the other hand, made a conflicting finding. They showed that their native Japanese speaker informants judged the part /(Q)t/ in the phrase /uta(Q)tane/ more likely as geminate when the preceding vowel /a/ is longer. They propose that the VC is the base unit for the perception of a geminate. On their hypothesis, when the preceding vowel /a/ becomes longer, the VC unit, /a(Q)t/, that includes the /(Q)t/ becomes longer, and so /(Q)t/ is judged more likely as a geminate. What should be noted is that the experimental condition with Ofuka et al's study is different from that of the present study as well as Hirata (1990a). The present study used two syllable (non-) words as the stimuli and presented them in isolation, while Ofuka et al used four-syllable carrier phrase that contains the singleton/geminate consonant in the third syllable. Ofuka et al argues that, under the condition where a two syllable word is presented in isolation, one judges the speech timing based on the first vowel because the length of the second vowel is unstable under such condition. Because the studies that used two syllable word as a stimulus, including the present study, Hirata (1990a), and Hirato and Watanabe (1987), showed a consistent finding, it is likely that native Japanese speakers depend on the duration of the first vowel when two syllable words are presented in isolation in order to time the overall speech rate, as Ofuka et al holds.

Hence, at least under the condition where two syllable words are used as stimuli, the NJSs showed their sensitivity to the duration of preceding syllable for their identification of the following consonant length, while the ELJs did not. The finding suggests the acquisition of the criterion based on closure duration is not enough for learners to attain a native-like management of a singleton/geminate contrast. It also points to the difficulty in acquiring the prosodic feature of Japanese, including mora-timing, for English-speaking learners.

Regarding place of articulation of a singleton/geminate, the NJS were sensitive to the difference, while the ELJ were not. The NJS identified the stimulus with [k] more likely as a geminate than those with [p] and [t] at certain closure duration. Han's (1992) production data, which showed [kk] was produced shorter than [pp], and than [tt], suggests that there may be constraints on the production of stop geminates across place of articulation among native speakers. A possibility of such constraints could be the timing difference based on inherent duration of phonetic segments.

Inherent duration of phonetic segments, including voiceless stops, was discussed by Umeda (1977) and Klatt (1976). Umeda (1977) holds that labial phonemes are the longest, dentals are shorter, and velars are the shortest in the word-initial stressed condition as inherent characteristics (p. 848). Klatt (1976) argues similarly that bilabial stops are longer than alveolars and velars (p. 1213). In addition, Klatt (1976) claims "duration serves as a cue to the phonetic identity of many segment types" (p. 1217). Though Umeda (1977) and Klatt's (1976) arguments are based on their English data, the constraints across place of articulation should be applicable to other languages because the constraints are related to the features of the oral articulators. Assuming the NJS are under the control of the inherent duration of the segments (stops), it is probable that the production of each stop consonant serves as a basis of the NJS's internal representation. Given a velar stop [k] is perceived at a shorter closure duration than a labial [p] and a dental [t], a velar geminate stop [kk] is perceived at relatively shorter closure duration than [pp] and [tt] under the same constraints as a singleton. The NJS's identification of the stimuli with [k] at shorter closure duration suggests that the NJS's criterion for the identification is so finely tuned that it is sensitive to the slight difference in the inherent duration of the stop phonemes caused by dynamics of articulation. In the present experiment, the assumption was substantiated with the shorter boundary closure duration of [k], but longer closure duration with [p] was not statistically confirmed. In contrast to the NJS, the ELJ did not react to the difference across place of articulation although they should be under the control of the inherent duration of the stops. What this indicates would be the timing of stop duration of a geminate, as a rate to a singleton or the timing of a geminate itself, may have not been established in the learners' internal representation at the same level as native Japanese speakers.

Regarding the participants' reactions across preceding vowels, ANOVA revealed no main effect of types of preceding vowels, but it showed significant interaction between consonant conditions and preceding vowels among the ELJs. It means that neither NJS nor ELJ differentiated their judgment across types of preceding vowels, and the

ELJ's reactions to consonant conditions vary across types of preceding vowel. It should be noted, however, that the present study does not make it clear which factors, preceding vowel, preceding syllable, or even an entire non-word, contributed to this result since difference in the preceding vowels, [i], [a], and [u], concurs with a difference in the preceding syllable, [bi], [da], and [mu], and a difference in an entire (non-)word with an identical medial consonant. I propose that the effect of interaction of consonant conditions and preceding vowels in ELJ's perception may indicate ELJ's non-sensitiveness to stimuli across place of articulation: their reactions to place of articulation are not fixed, varying across the conditions of preceding vowels, preceding syllable, or the entire word.

What could be the factors that affect the ELJ's judgment, then? A possibility was suggested in the comments by three of the ELJ in an interview with the researcher. They claimed that context helps in their distinction of a singleton from a geminate during ordinary conversation in Japanese. In this case, their attention is not directed to the specific phonetic features of a word, but to the meaning. Therefore unfamiliarity of a word makes it challenging for them. Without robust phonetic criterion, the ELJ would find it hard to identify a singleton/geminate under the conditions of the present study in which non-words were used as stimuli without contexts. It is also assumed that when trying to perceive non-words, the ELJ's judgment may be affected depending on whether each stimulus is close to real words with or without a geminate consonant, or how close it is to some real words that they have actually heard as a singleton or a geminate.

There are some issues that are left for future investigation. The present study has suggested a possibility of the learners' L1 influence in their identification of geminate consonants. The examination of the ELJ's English production data of the double consonants that occur beyond a word boundary would help inquiries into the possibility of L1 influence on the L1 English learners' perception of Japanese geminates. In addition, the stimuli used in the experiment should be refined. Although the participants' reaction to the type of preceding vowel was not a main focus of the study, the effect of the type of preceding vowel remained unclear in the present study. In order to analyze the effect, it is necessary to control the stimuli so that the variable of a preceding vowel can be isolated. For the examination of the effect of the duration of preceding syllable, more variation or more range in the lengths of preceding syllable would be helpful to make the finding more confirmative. Lastly, though the present study looked only at the perception, Japanese production data from the same participants would allow stronger arguments by showing the interaction between learner's perception and production.

7. Conclusion

The findings from the present study demonstrated differences in the perception of singleton/geminate contrast at a phonetic level between learners of Japanese and native Japanese speakers. The findings support the arguments in the previous studies (e.g., Han 1992, Mah and Archibald, 2003), which found a phonetic difference in production between learners of Japanese and native Japanese speakers, by demonstrating the difference in the perception data of the two groups. Though Hayes-Harb and Masuda (2008) reported no significant difference in the phonological distinction of a singleton/geminate contrast between learners and Japanese native speakers, the present study demonstrated differences between the two groups at a phonetic level. The learners in the present study identified a certain stimulus with a singleton/geminate based on the lengths of closure duration, but they did not react to the accompanying conditions that affected native speakers' identification, namely a difference across place of articulation of the stops and duration of preceding syllable. This suggests that the learners have not established the phonetic criterion well enough to consistently react to each stimulus with varied features. The findings also suggest that the learners have not reached native-like timing management of a singleton/geminate contrast in accordance with the acoustic duration of the preceding syllable.

L1 (English) influence could explain why the ELJs in the present study consistently identified a certain stimulus as a geminate at shorter closure duration than the NJS. Another possibility could be that the ELJs' knowledge about geminate consonants in Japanese with their insufficiently established criterion causes confusion when they are presented the stimuli which can not be judged clearly as a singleton or a geminate. Or both learners' L1 influence and confusion may be interacting in their performance.

Refinement of the stimuli and additional examination into learner's English and Japanese productions would be necessary for investigations that could further illuminate learners' systems in handling Japanese geminate consonants.

REFERENCES

Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Aoyama, K., Flege, J.E., Guion, S.G., Akahane-Yamada, R. and Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics 32*, 233-50.

Beckman, M. (1982). Segment duration and the 'Mora' in Japanese. *Phonetica, 39*, 113-135.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels In Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1568-1578.

Delattre, P. (1968). *The General Phonetic Characteristics of Languages* (final report, 1967-1968). Santa Barbara: US Office of Education and the University of California, Santa Barbara.

Flege, J. E. (1995). Second-language speech leaning: Theory, findings, and problems. In W. Strange (Ed.), *Phonological development: Models, research, implications* (pp. 233 -277). Timonium, MD: York Press.

Fukui, S. (1978). Nihongo no heisaon no enchoo tanshuku niyoru sokuon hisokuon toshiteno chooshu. [Perception of the Japanese stop consonants with reduced and extended durations]. The *Bulletin, Journal of Phonetic Society of Japan, 159.* 9-12.

Han, M. (1992). The timing control of geminate and single stop consonants in Japanese: A challenge for nonnative speakers. *Phonetica*, 49, 102-127.

Han, M. S. (1994). Acoustic manifestations of mora timing in Japanese. *Acoustical Society of America, 96 (1).* 73-82.

Hayes, R. L. (2002). The perception of novel phoneme contrasts in a second language: A developmental study of native speakers of English learning Japanese singleton and

Geminate consonant contrasts. *Coyote Papers, 12*, 28-41.

Hayes-Harb, R. (2005). Optimal L2 speech perception: Native speakers of English and Japanese consonant length contrasts. *Journal of Language and Linguistics*. 1-29.

Hayes-Harb, R. & Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research 24*, pp. 5- 33.

Hirata, Y. (1990a). Tangoreberu bunreberu niokeru nihonjin no sokuon no kikitori. [Perception of geminated stops in Japanese word and sentence levels]. *Onsei Gakkai Kaiho [The Bulletin], 194*, 23-28.

Hirata, Y. (1990b). Tangoreberu bunreberu niokeru nihonjin no sokuon no kikitori – eigo o bogo tosuru nihongo gakushuusha no baai. [Perception of geminated stops in Japanese word and sentence levels by English-speaking learners of Japanese language]. *Onsei Gakkai Kaiho [The Bulletin], 195*, 4-10.

Hirato, N. & Watanabe, S. (1987). Sokuon no chikaku to koozokuboin no jizokujikan tono kankee. [The relation between the perceptual boundary of voiceless plosives and their moraic counterparts and the length of vowels following closure]. *Onse Gengo II*, 99-106.

Homma, Y. (1981). Durational relationship between Japanese stops and vowel. *Journal of Phonetics 9*, 273-281.

Kaye, A. S. (2005). Gemination in English. *English Today 82, vol.21*. 43-55.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of Acoustical Society of America, 59,* 1208-1221.

Ladefoged, P. (2006). *A Course in Phonetics, Fifth Edition*. Boston, MA. Thomson Higher Education.

Mah, J. and Archibald J. (2003). Acquisition of L2 length contrasts. *Proceedings of the 6th Venerative Approaches to Second Language Acquisition Conference,* ed. Juanna M. L. et al., 208-212.

Ofuka, E., Mori, Y. & Kiritani, S. (2005). Sokuon no chikaku ni taisuru senkoo koozoku boinnchoo no eikyoo. [Perception of Japanese geminate stop: the effect of the duration of the preceding / following vowels]. *Journal of the phonetic society of Japan*, 9-2, 59-65.

Port, R., Rotunno, R. (1979). Relation between voice-onset time and vowel duration. *Journal of the Acoustical Society of America, 66*. 654-662.

Takagi, N. (2002). The limits of training Japanese listeners to identify English /r/ and /l/: eight case studies. *Journal of the Acoustical Society of America 111*, 2887-96.

Umeda, N. (1977). Consonant duration in American English. *Journal of Acoustical Society of America, 61,* 846-858.