Mispronunciation of Japanese singleton and geminate stops by Korean and Taiwanese Mandarin speakers

Kimiko Yamakawa¹*, Shigeaki Amano²† and Mariko Kondo³‡

¹Faculty of Contemporary Culture, Shokei University,
2–8–1 Musahigaoka, Kikuyo, Kikuchi-gun, Kumamoto, 861–8538 Japan
²Faculty of Human Informatics, Aichi Shukutoku University,
2–9 Katahira, Nagakute, 480–1197 Japan
³School of International Liberal Studies, Waseda University,
1–6–1 Nishiwaseda, Shinjuku-ku, Tokyo, 169–8050 Japan

(Received 13 April 2020, Accepted for publication 23 September 2020)

Abstract: Non-native Japanese speakers often mispronounce Japanese singleton and geminate stops. Previous studies have pointed out that non-native speakers’ mispronunciations are caused by an inadequate closure duration, which is the primary acoustic cue to distinguish singleton and geminate stops. However, the durations of preceding and following segments of the closure have not been fully investigated. In this study, the durations of the closure and the preceding and following consonant-vowel segments were analyzed to clarify the characteristics of Japanese singleton and geminate stops mispronounced by Korean and Taiwanese Mandarin speakers. The results revealed that the non-native speakers pronounced singleton stops with a longer closure and a shorter preceding consonant-vowel segment than the native Japanese speakers. In contrast, they pronounced geminate stops with a shorter closure and a longer following consonant-vowel segment than the native Japanese speakers. These results indicate that non-native speakers’ mispronunciations of Japanese singleton and geminate stops are caused by both inadequate closure duration and anteroposterior segment durations. It is likely that the reason for the mispronunciation is the difference in the rhythmic unit between Japanese and the first languages of the non-native speakers.

Keywords: Mispronunciation, Singleton and geminate stop, Duration ratio, Non-native speaker

1. INTRODUCTION

Geminate consonants are long consonants; phonetically, a long constriction of a consonant. Geminate consonants occur in many languages such as Arabic, Berber, Danish, Finnish, Italian, Malayalam, and Tamil. They also occur in Japanese, usually with voiceless obstruents and nasals. This study focuses on geminate stops in Japanese. The Japanese language phonologically distinguishes singleton and geminate stops. For example, [kako] (past) and [kak:o] (parentheses) are distinguished and have different meanings in Japanese. The primary acoustic cue to distinguish singleton and geminate stops is the closure duration [1,2]. That is, a singleton stop has a short closure duration, whereas a geminate stop has a long closure duration. This difference means that controlling the timing of speech segments is critical to pronouncing these stops correctly.

However, the closure duration covaries with speaking rate. That is, native Japanese speakers’ closure duration is longer when the speaking rate is low, whereas it is shorter when the speaking rate is high [e.g., 3,4]. This covariation occurs for both singleton and geminate stops. Therefore, in some cases, the closure duration of a geminate stop at a high speaking rate is shorter than that of a singleton stop at a low speaking rate [e.g., 3,4]. This means that native Japanese speakers do not rely on the absolute duration of the closure to distinguish the stops. Instead, they consider the durational cues relative to the speaking rate and correctly distinguish the stops.

As for the durational cues, Amano and Hirata [5,6] found that the most important cues for separating singleton and geminate stops at various speaking rates were the durations of closure and subword. The subword (see Fig. 1) is defined as a portion from the beginning of the preceding consonant-vowel (CV) segment to the end of the following...
CV segment, and here a ‘segment’ means a CV speech unit. Amano and Hirata [5,6] showed that the category boundary of the stops is represented by a linear equation of the two durational cues regardless of speaking rate, and native Japanese speakers use this category boundary to distinguish singleton and geminate stops.

In this study, and in the previous studies by Amano and Hirata [5,6], the consonant is defined as a segment with nonzero intensity. That is, it is defined as a consonant such as a plosive or affricate that does not include a silent part preceding the consonant burst. On the basis of this definition, the plosive or affricate starts at the beginning of the burst, but not at the beginning of a closure, whereas different consonants such as fricatives and nasals start at the beginning of the frication and nasalization, respectively. In all cases, a consonant starts with a nonzero intensity part.

This definition has merit in that a consistent acoustic feature (i.e., the beginning of nonzero intensity) can be applied to the start point of all kinds of consonants, regardless of whether the consonant is located at the utterance initial or medial of a word or a sentence. Some previous studies [e.g., 7] included a silent part in a plosive and affricate. However, in such cases, the consonant would start with a silent part when in a medial position of an isolated word, whereas it would start with a burst part when in an initial position. Moreover, the consonant in the initial position of a word would start with a silent part when the word is embedded in the medial position of a sentence, whereas it would start with a burst part when the word is embedded in the initial position of a sentence. This inconsistency is not desirable when trying to identify the general relationship between an acoustic feature and a phoneme/word unit.

On the basis of this consonant definition, the symbol C (also C1 and C2) in this study and the previous studies by Amano and Hirata [5,6] indicates a consonant excluding a silent part (i.e., a closure part). Also, a subword [5,6] is defined as the following sequence: a consonant-vowel segment preceding a closure (hereafter, the preceding CV segment), the closure itself, and the consonant-vowel segment following it (hereafter, the following CV segment). That is, the subword is defined as a C1V1_c1C2V2 sequence, where C1 and C2 are consonants excluding a closure, V1 and V2 are vowels, and cl is a closure (Fig. 1). Note that the subword duration was measured as the time between the beginning of C1V1 and the end of C2V2. When V2 and its following vowel are consecutive, for example, [setai] (household) and [set:ai] (reception), the end of V2 is defined as the midpoint time of the formant transition between V2 and its following vowel.

Although geminate consonants are common in many languages, studies on second language (L2) phonological acquisition of Japanese have shown that learners have various problems relating to both the production and perception of geminate consonants [e.g., 8–10]; it is one of the most common errors in L2 Japanese phonology [11]. Their mispronunciations are probably caused by the fact that they are not good at controlling the timing of speech segments. Their pronunciation of singleton stop closures tends to be too long, whereas geminate stop closures are too short [e.g., 12].

Amano and Hirata’s [5,6] findings about Japanese singleton and geminate stops by native Japanese speakers introduced the idea that a mispronunciation of those Japanese stops by non-native Japanese speakers could be identified by the category boundary derived from a native speaker’s correct pronunciation (Fig. 2). In Fig. 2, the dashed line indicates the native Japanese speakers’ category boundary for singleton and geminate stops based on subword and closure durations. Native speakers’ pronunciations of singleton and geminate stops are distributed below and above the category boundary, respectively (not shown in Fig. 2). Non-native speakers’ pronunciations of
singleton and geminate stops are assumed to be distributed within the dotted and solid ovals, respectively. Both types of pronunciation are assumed to cross the native speakers’ category boundary. Non-native speakers’ productions in the area above the native speakers’ category boundary are categorized as a geminate stop, whereas productions in the area below the native speakers’ category boundary are categorized as a singleton stop.

If a non-native speaker’s pronunciation of a singleton stop falls above the category boundary, it means that the singleton stop is mispronounced as a geminate stop. This type of mispronunciation is hereafter referred to as S → G mispronunciation. In contrast, if a non-native speaker’s pronunciation of a geminate stop falls below the category boundary, it means that the geminate stop is mispronounced as a singleton stop. This type of mispronunciation is hereafter referred to as G → S mispronunciation.

By this method based on the findings of Amano and Hirata [5,6], non-native Japanese speakers’ mispronunciations can be clearly categorized into two subcategories, namely, S → G and G → S mispronunciations. Another important point is that non-native speakers’ pronunciation of Japanese is affected by their native language [8,10]; consequently, the ratio of the S → G and G → S mispronunciations might differ between different languages. Therefore, the first aim of this study is to examine the effects of non-native speakers’ native language on the ratio of S → G and G → S mispronunciations specified by the category boundary.

When non-native Japanese speakers do the S → G mispronunciation, their closure duration is expected to be longer than that of native Japanese speakers, whereas for the G → S mispronunciation, it is expected to be shorter than that of native Japanese speakers. However, what about the durations of the preceding and following CV segments? This question is the basis of the second aim of this study. Are the durations of anteroposterior CV segments in non-native speakers’ mispronunciations consistent with those in native speakers’ pronunciations, or do they deviate? Also, are there certain biases caused by a non-native speaker’s first language when they deviate from native speakers’ pronunciations? If the durations of anteroposterior CV segments in mispronunciations systematically deviate from those in native speakers’ pronunciations, this might be another cause of non-native speakers’ mispronunciations. If this systematic deviation occurs regardless of the non-native speaker’s first language, it might be another piece of evidence for the universal tendency in the mispronunciation of Japanese singleton and geminate stops.

However, it is inadequate to simply compare the raw segment durations pronounced by native and non-native speakers because the variation in speaking rate will affect the segment duration and render the results unclear. Therefore, to control the effects of speaking rate, it would be better to compare the ratios of segment duration to subword duration. A subword was used to obtain the ratio because it is a necessary and sufficient unit to normalize the raw durations of the closure, preceding CV, and following CV segments. Therefore, in this study, we analyzed the duration ratios of these singleton and geminate stops to examine how non-native speakers’ mispronunciations of singleton and geminate stop segments differ from native speakers’ pronunciations.

### 2. SPEECH RECORDING

#### 2.1. Speakers

Korean and Taiwanese Mandarin speakers (hereafter, TW Mandarin) were recruited as non-native Japanese speakers because the Korean and TW Mandarin languages have prosodic characteristics that are different from Japanese (Table 1). Ten Korean speakers (3 males and 7 females), 10 TW Mandarin speakers (4 males and 6 females), and 10 native Japanese speakers (4 males and 6 females) participated in the speech recording. The average ages were 24.1 years (Min. = 20, Max. = 30, SD = 3.4) for the native Japanese speakers, 20.7 years (Min. = 19, Max. = 23, SD = 1.5) for the Korean speakers, and 22.7 years (Min. = 21, Max. = 26, SD = 1.1) for the TW Mandarin speakers. The native Japanese speakers spoke standard Japanese, and came from the Tokyo metropolitan area and its suburbs. The Korean speakers were from Seoul and its suburbs. The TW Mandarin speakers were from Taipei and its suburbs.

The Korean and TW Mandarin speakers had N3 or higher certificates of the Japanese-language proficiency test (JLPT). The JLPT has five levels: N1, N2, N3, N4, and N5. N1 indicates “Very Proficient” and requires the ability to understand Japanese used in a variety of circumstances, whereas N5 indicates “Very Basic” and requires the ability to understand basic Japanese. N3 corresponds to the ability to understand Japanese used in everyday situations to a certain degree. Therefore, the Korean and TW Mandarin speakers in this study can be regarded as having more than intermediate skills of Japanese.
2.2. Word Items for Recording

A 343-word set was constructed with minimal pairs of words derived from several studies [8,10,13,14]. The minimal pairs consisted of either contrasting singleton and geminate consonants, short and long vowels, with and without a moraic nasal, or with and without the glide /j/ following the consonant in the syllable onset. The total number of words was less than double the number of minimal pairs, because some minimal pairs shared the same word as a counterpart [e.g., the [kok|i] (secretary) - [kok;i] (tableware) pair and the [kok|i] (secretary) - [co:k;i] (sanity) pair].

In the 343-word set, all geminate words had a geminate stop at the second mora except one geminate word having it at the third mora. This deviation of geminate stop position reflects the tendency in Japanese language for a geminate stop to occur frequently at the second mora but not at the third or subsequent mora.

The accent patterns in the 343-word set of minimal pairs were not controlled. Therefore, some minimal pairs had the same accent pattern, but some did not. This was because it was difficult to find minimal pairs with the same accent pattern, and it is difficult for many non-native speakers to speak Japanese with a correct accent pattern.

In addition, although some previous studies showed that accent patterns had some effect on geminate perception [15,16], they are not a major cue for the identification of singleton and geminate stops, which mainly depends on the duration of speech segments [1,2]. Therefore, accent patterns were not controlled in the minimal pair test words in this study.

All 343 words were used for recording, but only 30 minimal pairs with contrasting singleton and geminate stops (see Appendix) were used for the analysis. All geminate words in the 30 minimal pairs had a geminate stop at the second mora. Other basic Japanese words, such as [hoci] (star) and [sakana] (fish), which were not part of the word set, were used for practice recordings before the real recording session.

2.3. Procedure

The Japanese speakers were recorded in Tokyo, Japan, the Korean speakers in Seoul, Korea, and the TW Mandarin speakers in Taipei, Taiwan. The recordings were all conducted in a soundproof or quiet room using a microphone (Sony, ECM-999), with a pop filter (Stedman, Proscreen XL), A/D converter (Roland, UA25-EX), and a personal computer (Toshiba, Dynabook SS RX2).

In each trial, the speaker pressed the space key on the computer keyboard to start the recording. A word item was presented on the computer screen in Hiragana, which was embedded in the carrier sentence [korewa ____ dato omoi masu] (“I suppose that this is __”), and the speaker pronounced the sentence at a normal speaking rate. The pronunciation was digitally recorded with a 16-bit quantization and 48-kHz sampling frequency, and was stored as a digital audio file in the computer. When the speaker finished saying the sentence, she/he pressed the space key again to stop the recording.

After 10 practice trials, the speakers pronounced the 343 words in carrier sentences, with the recordings divided into 9 blocks of 35 words and 1 block of 28 words. Each speaker took a break of a few minutes after every two blocks. The order of the words was randomized for each speaker. The accent patterns of the words pronounced by the native Japanese speakers were checked, and incorrect accent patterns were regarded as mistakes and the relevant words were rerecorded at the end of that session. The accent patterns of the non-native Japanese speakers were not checked because of the difficulty that non-native speakers have in correctly producing Japanese accent patterns. However, if a speaker hesitated to pronounce a word or made a mistake, the sentence was recorded again at the end of that block.

3. ANALYSIS

3.1. Word Items for Analysis

Thirty minimal pairs of Japanese words with singleton and geminate stops (see Appendix) spoken by the native and non-native Japanese speakers were used for the analysis. From the initial total of 1,800 word items, 7 items were not used in the analysis because of misrecording, giving a final total of 599 items pronounced by the native Japanese speakers, 599 items pronounced by the Korean speakers, and 595 items pronounced by the TW Mandarin speakers.

3.2. Segment Duration Measurement

Professional labelers used the waveform and spectrogram of each recorded item to mark the start and end times of each phoneme segment. The durations of the closure and the preceding and following CV segments were calculated using the time information for both singleton and geminate stops. Subword durations were obtained as the sum of the durations of the preceding CV segment, the closure, and the following CV segment.

3.3. Category Boundary of Native Japanese Speakers

Using the 599 items pronounced by the native Japanese speakers, we obtained the category boundary of singleton and geminate stops by discriminant analysis as Eq. (1):

$$0.0889x - y - 61.4 = 0,$$

where $x$ is the subword duration and $y$ is the closure duration in milliseconds.
3.4. Mispronunciation by Non-native Japanese Speakers

The mispronunciations of the non-native Japanese speakers were identified using Eq. (1). When the left side of Eq. (1) < 0, it indicates that the non-native speakers intended to pronounce a singleton but that their pronunciation was categorized as a geminate (i.e., $S \rightarrow G$ mispronunciation). In contrast, when the left side of Eq. (1) > 0, it indicates that the non-native speakers intended to pronounce a geminate but that their pronunciation was categorized as a singleton (i.e., $G \rightarrow S$ mispronunciation).

The mispronunciation ratios of the non-native Japanese speakers are shown in Table 2. For both the Korean and TW Mandarin speakers, the $S \rightarrow G$ mispronunciation ratio was more than 20% whereas the $G \rightarrow S$ mispronunciation ratio was less than 10%, indicating that the $S \rightarrow G$ mispronunciation was more frequent than the $G \rightarrow S$ mispronunciation.

3.5. Segment Duration

To clarify the segment characteristics of singleton and geminate stops in the native Japanese speaker’s pronunciations and non-native Japanese speaker’s mispronunciations, the mean durations were calculated for the preceding CV, closure, following CV, and subword (Table 3).

For the native Japanese speakers, the $t$-test was conducted to examine the differences in duration between the stop types. As expected, the closure duration of a geminate stop was significantly longer than that of a singleton stop [$t(597) = 38.6, p = 3.0 \times 10^{-164}$]. Similarly, the duration of the preceding CV segment of the geminate stop was significantly longer than that of the singleton stop [$t(597) = 6.85, p = 1.9 \times 10^{-11}$]. In contrast, the duration of the following CV segment of the geminate stop was significantly shorter than that of the singleton stop [$t(597) = 4.35, p = 1.6 \times 10^{-5}$]. These findings agree with those of previous studies about native Japanese speakers [e.g., 17,18]. Therefore, it can be said that the native Japanese speakers in this study pronounced the items normally.

The $t$-test was also conducted for the non-native Japanese speakers. The closure duration of a designated geminate stop in $G \rightarrow S$ mispronunciations was significantly shorter than that of a designated singleton stop in $S \rightarrow G$ mispronunciations for the Korean speakers [$t(83) = 4.95, p = 3.8 \times 10^{-6}$] and the TW Mandarin speakers [$t(104) = 9.75, p = 2.5 \times 10^{-16}$], which is opposite to the results of the native Japanese speakers. Furthermore, the duration of the preceding CV segment did not differ between the designated singleton and

### Table 2: Ratio (%) of non-native Japanese speakers’ mispronunciation identified using native Japanese speakers’ category boundary of singleton and geminate stops (Eq. (1)).

| Mispronunciation type | First language of non-native speaker |
|-----------------------|--------------------------------------|
| $S \rightarrow G$     | Korean (24.4 (299)) TW Mandarin (26.4 (295)) |
| $G \rightarrow S$     | Korean (4.0 (300)) TW Mandarin (9.3 (300)) |

**Note:** The number of items in a population is shown in parentheses.

### Table 3: Segment duration (ms) for singleton and geminate stops in pronunciation by native Japanese speakers and in mispronunciation by non-native Japanese speakers with first language of Korean or TW Mandarin. Ratio of preceding CV duration to following CV duration is shown in the rightmost column.

| Speakers’ first language | Designated stop type | Mispronunciation type | $n$ | Preceding CV | Closure | Following CV | Subword | Preceding CV to Following CV |
|-------------------------|----------------------|-----------------------|-----|--------------|---------|-------------|--------|-----------------------------|
|                         |                      |                       |     | $M$ | $SD$ | $M$ | $SD$ | $M$ | $SD$ | $M$ | $SD$ | $M$ | $SD$ |
| Japanese                | Singleton            | —                     | 299 | 131.3 | 38.9 | 51.3 | 18.6 | 97.3 | 21.8 | 280.0 | 57.1 | 1.35 |
|                         | Geminate             | —                     | 300 | 154.5 | 43.9 | 131.3 | 30.7 | 89.4 | 22.8 | 375.2 | 74.1 | 1.73 |
| Korean                  | Singleton            | $S \rightarrow G$    | 73  | 112.3 | 32.2 | 104.7 | 16.8 | 127.0 | 30.2 | 344.0 | 59.0 | 0.88 |
|                         | Geminate             | $G \rightarrow S$    | 12  | 128.2 | 50.7 | 79.3 | 13.8 | 111.4 | 30.0 | 318.9 | 70.6 | 1.15 |
| TW Mandarin             | Singleton            | $S \rightarrow G$    | 78  | 135.1 | 40.5 | 123.0 | 21.9 | 155.4 | 34.8 | 413.5 | 58.9 | 0.87 |
|                         | Geminate             | $G \rightarrow S$    | 28  | 151.1 | 51.1 | 79.4 | 14.6 | 134.6 | 29.6 | 365.1 | 66.4 | 1.12 |
Table 4  Segment ratio of singleton and geminate stops to subword duration in pronunciation by native Japanese speakers and in mispronunciation by non-native Japanese speakers with first language of Korean or TW Mandarin.

| Speakers’ first language | Designated stop type | Mispronunciation type | Segment       |          |          |          |
|-------------------------|----------------------|-----------------------|---------------|----------|----------|----------|
|                         |                      |                       | Preceding CV  | Closure  | Following CV |
|                         |                      |                       | M   | SD   | M   | SD   | M   | SD   |
| Japanese                | Singleton            | —                     | 0.465 | 0.081 | 0.184 | 0.060 | 0.351 | 0.060 |
|                         | Geminate             | —                     | 0.409 | 0.073 | 0.352 | 0.063 | 0.238 | 0.039 |
| Korean                  | Singleton            | S → G                 | 0.324 | 0.062 | 0.308 | 0.041 | 0.368 | 0.054 |
|                         | Geminate             | G → S                 | 0.392 | 0.082 | 0.256 | 0.057 | 0.351 | 0.059 |
| TW Mandarin             | Singleton            | S → G                 | 0.324 | 0.071 | 0.300 | 0.049 | 0.376 | 0.066 |
|                         | Geminate             | G → S                 | 0.405 | 0.081 | 0.221 | 0.043 | 0.373 | 0.077 |

geminate stops for the Korean speakers \((t(83) = 1.45, p = 0.15)\) or the TW Mandarin speakers \((t(104) = 1.66, p = 0.99)\). Again, this is different from the results of the native Japanese speakers, for which the preceding CV segment of the geminate stop was longer than that of the singleton stop. In contrast, the results of the durations of the following CV segments were inconsistent between the Korean and the TW Mandarin speakers. Similarly to the native Japanese speakers, the duration of the following CV segment of the designated geminate stop was significantly shorter than that of the designated singleton stop for the TW Mandarin speakers \((t(104) = 2.81, p = 5.9 \times 10^{-3})\) but not for the Korean speakers \((t(83) = 1.66, p = 0.10)\).

However, the ratio of the duration of the preceding CV to the following CV (Table 3) suggests that the native and non-native Japanese speakers had different durations in these CV segments. For singleton stops, Korean and TW Mandarin speakers had ratios of 0.88 and 0.87, respectively, compared with a much larger ratio of 1.35 for the native Japanese speakers. Similarly, for geminate stops, even though the non-native Japanese speakers’ ratios (1.15 for Korean and 1.12 for TW Mandarin) were more than one, they were still much smaller than the ratio of 1.73 for native Japanese speakers. These findings indicate that segment durations of the preceding and following CV differ between the native and non-native Japanese speakers.

3.6. Segment Ratio to Subword Duration

The subword durations seem to be different between the native and non-native Japanese speakers for both stop types (Table 3). To examine the difference in the subword durations between the speakers’ first languages, analysis of variance was conducted for a singleton stop and a geminate stop. There was a significant main effect of first language for a singleton stop \(F(2, 447) = 177.8, p = 1.6 \times 10^{-57}\) and a geminate stop \(F(2, 337) = 3.55, p = 0.030\). These results indicate that subword durations differ between Japanese, Korean, and TW Mandarin speakers regardless of the stop type. Namely, speaking rates are different between these speakers.

When the speaking rate is significantly different, it is not reasonable to compare raw segment durations between Japanese, Korean, and TW Mandarin speakers. To compensate for the differences in speaking rate, we used the ratio of each segment duration to the subword duration. Specifically, the mean ratio of the duration of each segment in each subword was obtained for the singleton and geminate stops pronounced by the Japanese, Korean, and TW Mandarin speakers (Table 4). The segment ratios were different between the native and non-native Japanese speakers. For example, an inverted order of ratio between the native and non-native speakers was observed in the closure and preceding CV segments.

3.7. Segment Ratio Difference from Native Japanese Speakers

To clarify the difference in segment ratio between native and non-native Japanese speakers, the mean segment ratio of the native Japanese speakers was used as standard segment ratio for each word item. That is, the ratio difference was obtained by subtracting the native speakers’ segment ratio from the corresponding non-native speakers’ segment ratio for each word. The ratio differences for the preceding CV segment, closure, and following CV segment for \(S → G\) mispronunciation and \(G → S\) mispronunciation are respectively shown in Figs. 3 and 4. Note that, the ratio difference was calculated from native Japanese speakers’ singleton pronunciations for \(S → G\) mispronunciation and \(G → S\) mispronunciation in Fig. 3, whereas it was calculated from Japanese speakers’ geminate pronunciations for \(G → S\) mispronunciation in Fig. 4. A statistical test was performed to examine whether the segment ratio difference is equal to zero (Table 5). Significant differences revealed by the test are indicated with an asterisk in Figs. 3 and 4.

A consistent tendency was observed in both the Korean and TW Mandarin speakers. That is, for \(S → G\) mispronunciations, the segment ratio difference compared with
the native Japanese speakers was negative for preceding CV segments, but positive for closures (Fig. 3). In contrast, for G → S mispronunciations, the segment ratio difference was negative for closures, but positive for following CV segments (Fig. 4). These ratio differences indicate that, compared with the native Japanese speakers’ pronunciation, the closure was too long and the preceding CV segment was too short in non-native speakers’ S → G mispronunciations, whereas in G → S mispronunciations, the closure was too short and the following CV segment was too long.

4. DISCUSSION

In this study, non-native Japanese speakers’ mispronunciations were identified using a category boundary of singleton and geminate stops of native Japanese speakers. Then, segment durations and ratios of closure, preceding CV, and following CV were analyzed and compared between the native and non-native speakers.

The results of segment duration ratios showed that, compared with the pronunciation by native Japanese speakers, the non-native speakers produced a longer closure in S → G mispronunciations (Fig. 3) but a shorter closure in G → S mispronunciations (Fig. 4). These results were expected because, in the S → G mispronunciations, singleton stops that should have a shorter closure were categorized as geminate stops that have a longer closure, whereas in the G → S mispronunciations, geminate stops that should have a longer closure were categorized as singleton stops that have a shorter closure.

A notable new finding in this study is that, when the non-native speakers made an S → G mispronunciation, they produced a shorter preceding CV segment than the native speakers (Fig. 3), but when they made a G → S mispronunciation, they produced a longer following CV segment than the native speakers (Fig. 4). In addition, for S → G mispronunciations by the non-native speakers, preceding CV segment shortening occurred together with closure lengthening, whereas for G → S mispronunciations, following CV segment lengthening occurred together with closure shortening. Therefore, we can say that the mispronunciations of singleton and geminate stops by the non-native speakers in this study were caused by not only inadequate closure duration but also inadequate anteroposterior segment durations.

Moreover, these significant effects were observed consistently in both the Korean and TW Mandarin speakers. Previous studies [e.g., 8,10] have pointed out that
each native language (L1) has a different effect on the pronunciation of a non-native language (L2). Therefore, in this study, we could have expected that the deviations in segment duration ratios in the $S \rightarrow G$ and $G \rightarrow S$ mispronunciations would differ between non-native speaker’s L1s (i.e., Korean and TW Mandarin). However, the Korean and TW Mandarin speakers showed a similar tendency regarding the durations of closure and preceding and following CV segments in each type of mispronunciation. Therefore, L1 did not have any different effects on the pronunciation of Japanese in this study.

As shown in Table 1, Korean has phrase accent prominence whereas TW Mandarin has tone prominence. In addition, Korean and TW Mandarin have different syllable structures. Nevertheless, both groups of non-native speakers showed very similar results. Therefore, the prominence and syllable structure cannot be related to the cause of mispronunciation. A rhythmic unit that is a syllable for both Korean and TW Mandarin but a mora for the cause of mispronunciation. Therefore, the non-native speakers showed very similar results. Therefore, the syllable structures. Nevertheless, both groups of non-native speakers showed very similar results. Therefore, the prominence and syllable structure cannot be related to the cause of mispronunciation. A rhythmic unit that is a syllable for both Korean and TW Mandarin but a mora for Japanese would be the cause of mispronunciation.

The shorter duration of the preceding CV in the $S \rightarrow G$ mispronunciations by the non-native speakers, as shown in Fig. 3, may be because they have different rhythmic units and syllabification of Japanese sound sequences compared with the native Japanese speakers. As explained in the introduction, geminate stops are phonetically a stop with a long closure [CVC.V]. However, phonologically, they are considered as a doubling of a stop with a syllable boundary between the double stops /CVC.CV/. Closed syllables are rare in Japanese which mainly has light open syllables, but they are much more common in the non-native speaker’s L1. Syllables are an important phonological unit in these languages. Therefore, when the non-native speakers mispronounced singleton stops as geminate stops, they probably applied their L1 syllable structure to a singleton stop in /CV.CV/ and mispronounced it as a geminate stop in /CVC.CV/. In contrast, the native Japanese speakers would have used the mora rather than the syllable to produce the sequence as /CV-CV/ (here, the hyphen “-” indicates a mora boundary). Assuming that the native and non-native speakers produce at the same tempo, then the /CVC/ syllable and the /CV/ mora would have the same duration. However, since the /CVC/ syllable contains more segments than the /CV/ mora, the non-native speakers would need to shorten the /CV/ compared to the native Japanese speakers, as shown in Fig. 3.

In the case of the mispronunciation of geminate stops as singleton stops by the non-native speakers (Fig. 4), the difference in duration from native speakers could be explained by non-native speakers’ inadequate compensation. When the non-native speakers mispronounced a geminate stop in /CVC.CV/ as a singleton stop in /CV.CV/, they tended to produce the geminate stops with insufficient closure duration while also increasing the following CV duration. The reason for the excessively long duration of the following /CV/ syllable may have been because the non-native speakers knew that a geminate consonant should be longer than a singleton consonant, and so when they mistakenly produced a “geminate” consonant with a short closure, they then over-compensated by extending the duration of the following /CV/ syllable.

The non-native speakers made fewer $G \rightarrow S$ mispronunciations than $S \rightarrow G$ mispronunciations (Table 2). The reason for the fewer $G \rightarrow S$ mispronunciations may have been a hypercorrection of the geminate stop. That is, the non-native speakers were aware of their low skill in distinguishing between singleton and geminate stops and they knew that a geminate stop has a long closure, so they intentionally pronounced the closure of a geminate stop longer than usual. Moreover, in the experiment, the geminate words were presented with the Hiragana “っ” which explicitly indicates a geminate stop. This explicit indication would make the non-native speakers pay attention to the geminate stop and may have promoted their intentional pronunciation of a longer closure. As a result, they succeeded in pronouncing geminate stops with fewer errors.

Whatever the cause, the results of this study suggest that in order to reduce mispronunciations of singleton stops, non-native speakers should try to produce a shorter closure duration and extend the preceding CV segment. In addition, to produce a correct singleton stop, as shown for the native Japanese speakers in Table 3, the duration of the preceding CV segment (131.3 ms) should be longer than that of the following CV segment (97.3 ms). To reduce mispronunciations of geminate stops, the results also suggest that non-native speakers should produce not only a long closure but also a short following CV segment. In addition, to produce a correct geminate stop (Table 3), the duration of the following CV segment (89.4 ms) should be shorter than that of the preceding CV segment (154.5 ms).

Word items in this study had a geminate stop only at the second mora. However, the most important cues for distinguishing single and geminate stops are closure duration and subword duration, so mora position should not greatly affect the results. Therefore, it should also be possible to use this method for geminate stops at the third or subsequent mora because a subword can be defined as any two successive CV segments in a word. Even so, it would be useful to compare words with geminate stops at different mora positions to clarify the effects of rhythmic units on non-native Japanese speakers’ mispronunciations.

Another point that should be noted is the difference in mispronunciation frequency between individuals. Depend-
ing on their Japanese ability, some non-native Japanese speakers may make more mispronunciations than others. Therefore, in extreme cases, some non-native speakers may make very frequent mispronunciations, whereas other speakers may not make any mispronunciations. In this study, all the non-native speakers made some S → G and/or G → S mispronunciations, but none of the speakers excessively made mispronunciations. It is likely that most non-native speakers will make some mispronunciations, but it would be useful to conduct a further study of individual differences in mispronunciations.

In this study, the non-native speakers’ mispronunciations were defined as a discriminant error, with the criterion of a category boundary of singleton and geminate stops by native speakers. However, in the strictest sense, the criterion should be the perception boundary of native speakers, because the mispronunciation is always identified by a listener in speech communication. Therefore, in future studies, it should be examined whether the mispronounced singleton and geminate stops in this study would be perceived as such by native Japanese speakers. Although this has not been confirmed, it is highly probable that S → G mispronunciations would be perceived as a geminate stop whereas G → S mispronunciations would be perceived as a singleton stop, because the perception and production boundaries of singleton and geminate stops almost coincide [5,6], and therefore, the two boundaries would make almost the same discriminant errors.

This study focused only on geminate stops and not on other types of geminates in Japanese. An example of another type is a geminate fricative (e.g., [hasːeː] (pronunciation)). The primary cue for a geminate fricative is frication duration, and it is affected by speaking rate, much like a geminate stop is. It is unclear whether the boundary of singleton and geminate fricatives can be represented by a linear function, as in Amano and Hirata’s study [5,6], nor whether non-native speakers have the same duration deviation in anteroposterior segments of a geminate fricative, as was found in this study. These questions should be clarified for geminate fricatives and for various other types of geminates in future studies.

ACKNOWLEDGMENTS

This study was supported by JSPS KAKENHI Grant Numbers JP22320081, JP24652087, JP25284080, JP26370464, and JP17K02705. We would like to thank Professor Kyung-Ja Ryoo and Dr. Kazuya Iihoshi of Sangmyung University; Professor Chiu-yu Tseng of the Institute of Linguistics, Academia Sinica; Professors Chang-Ho Lin and Meng-Ling Hsu of Ming Chuan University; Professor Rong-Kuan Shen of Shih Hsin University; Professor Yaw-Huei Maa of Tamkang University for their assistance in the utterance recordings.

REFERENCES

[1] H. Fujisaki and M. Sugito, “Physical properties of speech sounds,” in Iwanami Japanese series 5: Phonology, S. Ohno and T. Shibata, Eds. (Iwanami, Tokyo, 1977), pp. 63–106 (in Japanese).
[2] S. Kawahara, “The phonetics of obstruent geminates, sokuon,” in The Mouton Handbook of Japanese Language and Linguistics, H. Kubozono, Ed. (De Gruyter Mouton, Berlin, 2015), pp. 43–77.
[3] Y. Hirata and J. Whiton, “Effects of speaking rate on the single/geminate stop distinction in Japanese,” J. Acoust. Soc. Am., 118, 1647–1660 (2005).
[4] K. Idemaru and S. Guion-Anderson, “Relational timing in the production and perception of Japanese singleton and geminate stops,” Phonetics, 67, 25–46 (2010).
[5] S. Amano and Y. Hirata, “Perception and production boundaries between single and geminate stops in Japanese,” J. Acoust. Soc. Am., 128, 2049–2058 (2010).
[6] S. Amano and Y. Hirata, “Perception and production of singleton and geminate stops in Japanese: Implications for the theory of acoustic invariance,” Phonetics, 72, 43–60 (2015).
[7] M. Han, “Acoustic manifestations of mora timing in Japanese,” J. Acoust. Soc. Am., 96, 73–82 (1994).
[8] Y. Sukegawa, “Utterance tendency of non-native Japanese speakers: Results of questionnaire survey,” Japanese Speech and Education, Research Report of Grant-in-Aid for Scientific Research on Priorities Areas by Ministry of Education, Science and Culture, pp. 187–222 (1993) (in Japanese).
[9] T. Toda, “Acquisition of special morae in Japanese as a second language,” J. Phon. Soc. Jpn., 7(2), pp. 70–83 (2003) (in Japanese).
[10] M. Kondo, “Language-specific and language-universal L1 interference on phonological acquisition of Japanese learners: Results of questionnaires to Japanese teachers,” Bulletin of the Graduate Division of Letters, Arts and Sciences of Waseda University, III 57, 21–34 (2012) (in Japanese).
[11] Y. Hirata, “L2 phonetics and phonology,” in The Handbook of Japanese Language and Linguistics, H. Kubozono, Ed. (De Gruyter Mouton, Berlin, 2015), pp. 719–762.
[12] K. Masuda and R. Hayes-Harb, “Native English speakers’ strategies of producing moraic obstruents,” in Linguistics and Education of Japanese IV, M. Minami, Ed. (Kurosio, Tokyo, 2005), pp. 259–270 (in Japanese).
[13] N. Minematsu, K. Nishina and S. Nakagawa, “Read speech database for foreign language learning,” J. Acoust. Soc. Jpn. (J), 59, 345–350 (2003) (in Japanese).
[14] K. Yamakawa, S. Amano and M. Kondo, “Development of Japanese read-word database for non-native speakers of Japanese,” Proc. 17th Oriental Chapter of the International Committee for Co-ordination and Standardization of Speech Databases and Assessment Techniques (COCOSDA), pp. 65–70 (2014).
[15] E. Ofuka, “Perception of a Japanese geminate stop /tː/: The effect of pitch type and acoustic characteristics of preceding/following vowels,” J. Phon. Soc. Jpn., 7(1), pp. 70–76 (2003) (in Japanese).
[16] H. Hung, “Perceptions of Japanese geminate stops among Taiwanese learners of Japanese,” J. Phon. Soc. Jpn., 16(2), pp. 15–27 (2012) (in Japanese).
[17] N. Campbell, “A study of Japanese speech timing from the syllable perspective,” J. Phon. Soc. Jpn., 3(2), pp. 29–39 (1999).
[18] K. Idemaru and S. Guion, “Acoustic covariants of length contrast in Japanese stops,” J. Int. Phon. Assoc., 38, 167–186 (2008).
### APPENDIX: MINIMAL PAIRS OF JAPANESE WORDS

| Orthography | IPA transcription | Meaning | Orthography | IPA transcription | Meaning |
|-------------|-------------------|---------|-------------|-------------------|---------|
| 部下        | [bu:ka]           | subordinate | 物価        | [bu:ka]           | prices  |
| 画家        | [gaka]            | painter   | 学科        | [gaka]            | department |
| ガキ        | [gak:i]           | brat      | 楽器        | [gak:i]           | musical instrument |
| 墓          | [haka]            | grave     | 発火        | [haka]            | combustion |
| 事件        | [dziken]          | affair    | 実験        | [dziken]          | experiment |
| 変えて       | [kaete]           | changed   | 帰って       | [kaete]           | returned |
| 過去        | [kako]            | past      | 括弧        | [kako]            | parenthesis |
| 麦          | [kate]            | pabulum   | 勝手        | [kate]            | selfishness |
| 来た        | [kita]            | have come | 切った       | [kita]            | have cut |
| 来て        | [kite]            | coming    | 切手        | [kite]            | stamp |
| 帰途        | [kito]            | on the way home | きっと       | [kito]            | surely |
| 子機        | [kok:i]           | cordless handset | 国旗        | [kok:i]           | national flag |
| 町          | [mat:i]           | town      | マッチ       | [mat:i]           | match |
| 枕          | [makura]          | pillow    | 真っ暗       | [makura]          | pitch darkness |
| 坂          | [saka]            | slope     | 作家        | [saka]            | writer |
| 先          | [saki]            | tip       | 殺氣        | [saki]            | menacing atmosphere |
| 世界        | [sekai]           | world     | 石灰        | [sekai]           | caustic lime |
| 世間        | [seken]           | society   | 石鹸        | [seken]           | soap |
| 咳          | [sek:i]           | cough     | 石器        | [sek:i]           | stone ware |
| 世帯        | [setai]           | household | 接待        | [setai]           | reception |
| 瀬戸        | [seto]            | strait    | セット       | [seto]            | set |
| 式          | [siki]            | equation  | 漆器        | [siki]            | lacquer ware |
| して        | [siete]           | done      | 知って       | [siete]           | known |
| 便途        | [sito]            | use       | 嫌姫       | [sito]            | jealousy |
| スパイ       | [supai]           | spy       | 酸っぱい     | [supai]           | sour |
| 所見        | [soken]           | observation | 食券        | [soken]           | food ticket |
| 書記        | [sok:i]           | secretary | 食器        | [sok:i]           | tableware |
| 盾          | [tate]            | shield    | 立って       | [tate]            | stood |
| 敵          | [tek:i]           | enemy     | 鉄器        | [tek:i]           | ironware |
| 時          | [tok:i]           | time      | 突起        | [tok:i]           | prong |

*Note: The pitch patterns of some word pairs are different.*