Vietnamese speakers’ mispronunciation of Japanese singleton and geminate stops

Kimiko Yamakawa1,*, Shigeaki Amano2,† and Mariko Kondo3,‡

1Faculty of Contemporary Culture, Shokei University, 2–6–78 Kuhonji, Chu-ku, Kumamoto, 862–8678 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 1 December 2021, Accepted for publication 17 April 2022)

Abstract: Vietnamese speakers’ mispronunciations of Japanese singleton and geminate stops were identified using the category boundary of the stops pronounced by native Japanese speakers. To clarify the characteristics of the Vietnamese speakers’ mispronunciations, their speech segment durations were analyzed. In comparison with native Japanese speakers’ correct pronunciations, Vietnamese speakers mispronounced a singleton stop with a longer closure and a shorter preceding consonant-vowel segment, whereas they mispronounced a geminate stop with a shorter closure and a longer following consonant-vowel segment. These results were consistent with the findings of Korean, Taiwanese Mandarin, and Thai speakers in previous studies, suggesting that non-native speakers may have a common tendency to have inadequate durations of closure and anteroposterior consonant-vowel segments in mispronunciations of Japanese singleton and geminate stops.

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

1. INTRODUCTION

The Japanese language distinguishes singleton and geminate stops. Phonologically, a geminate stop is assumed to take one mora long. A mora is a processing unit of rhythm in the Japanese language (Table 1) [1,2], and the length of a Japanese word can be represented by the number of morae. There are normal and special types of mora. Special mora consists of a geminate consonant, moraic nasal consonant, and long vowel. A normal mora has a consonant-vowel or vowel-only structure. In contrast, a special mora has a consonant-only structure for geminate and moraic nasal consonants, and a vowel-only structure for a long vowel. Within these three special morae, this study focused on a geminate stop consonant.

A stop consonant consists of the closure and burst segments. The closure duration preceding a burst segment is short in a singleton stop whereas it is long in a geminate stop. More precisely, the primary acoustic variable used to distinguish these stops is the closure duration [e.g., 3,4]. However, the closure duration depends on the speaking rate: it is short at a fast speaking rate and long at a slow speaking rate [e.g., 5]. Because of this speaking-rate dependency, the duration of a singleton stop at a slow speaking rate is longer than the duration of a geminate stop at a fast speaking rate in some cases. As shown in this example, the closure duration is not an invariant cue for distinguishing the stops.

Even though the closure duration depends on the speaking rate, native Japanese speakers can pronounce singleton and geminate stops with a clear distinction. This fact indicates that the closure duration alone cannot explain the distinction between the singleton and geminate stops. To address this problem, Amano and Hirata [6,7] introduced a subword duration as a secondary acoustic variable. The subword (Fig. 1) was defined as a portion from the beginning of the consonant-vowel segment preceding the closure (hereafter, the preceding CV segment) to the end of the consonant-vowel segment following the closure (hereafter, the following CV segment), where the CV segment was defined as a non-zero amplitude segment so that it did not include a closure. Amano and Hirata [6,7] demonstrated that a linear equation of closure duration (primary acoustic variable) and subword duration (secondary acous-

* e-mail: jin@shokei-gakuen.ac.jp
† e-mail: psy@asu.aasa.ac.jp
‡ e-mail: mkondo@waseda.jp
[doi:10.1250/ast.43.241]}
tic variable) could distinguish singleton and geminate stops in various speaking rates. More specifically, the linear equation represents a category boundary to separate the Japanese singleton and geminate stops at various speaking rates.

Probably using this category boundary, native Japanese speakers can easily distinguish singleton and geminate stops. In contrast, non-native Japanese speakers are inaccurate at controlling the durational properties of these stops [e.g., 8, 9]. For example, Vietnamese speakers often confuse the Japanese singleton and geminate stops. They mispronounce a singleton stop as a geminate stop (e.g., [bˈko] ‘subordinate’) and a geminate stop as a singleton stop (e.g., [sˈiti] ‘switch’) [10].

Yamakawa et al. [11] pointed out that, in addition to an inappropriate closure duration, the durations of preceding and following CV segments are also related to the non-native speakers’ mispronunciations. Using the procedure of Amano and Hirata [6,7], they obtained a linear equation of closure and subword durations as the category boundary of singleton and geminate stops from production data of native Japanese speakers with a small discriminant error (4.5%). Subsequently, using the category boundary as a criterion of classification, they identified mispronounced singleton and geminate stops by non-native Japanese speakers: Korean and Taiwanese Mandarin speakers.

There are two types of mispronunciations (Fig. 2). One of them is “S → G mispronunciation” which occurs to non-native speakers when they attempt to pronounce a singleton stop, but their pronunciation is classified as a geminate stop. The other one is “G → S mispronunciation” which occurs to non-native speakers when they attempt to pronounce a geminate stop, but their pronunciation is classified as a singleton stop.

Yamakawa et al. [11] analyzed the data of Korean and Taiwanese Mandarin speakers and found that the S → G mispronunciation has a longer closure and a shorter preceding CV segment than those segments in a singleton stop correctly pronounced by native Japanese speakers. In contrast, the G → S mispronunciation has a shorter closure and a longer following CV segment than those segments in a geminate stop correctly pronounced by native Japanese speakers. These results indicated that mispronunciations of singleton and geminate stops are related to inadequate preceding and following CV segment durations as well as an inadequate closure duration.

Yamakawa and Amano [12] analyzed Thai speakers’ mispronunciations of Japanese singleton and geminate stops using the same procedure as Yamakawa et al. [11].
They found the same tendency in segment durations in Thai speakers’ mispronunciation: a longer closure and a shorter preceding CV segment in the S → G mispronunciation, and a shorter closure and a longer following CV segment in the G → S mispronunciation.

The results of these previous studies suggest that non-native speakers may have a common tendency of inadequate segment durations of closure, preceding CV, and following CV in mispronunciations of Japanese singleton and geminate stops. Further research with non-native speakers in other languages is needed to confirm these findings. Based on this background, this study selected Vietnamese speakers and analyzed their mispronunciations of Japanese singleton and geminate stops.

Vietnamese is a tone language. It has three major dialects spoken in the northern, central, or southern regions of Vietnam [13]. The northern dialect that is spoken in Hanoi and its vicinity has six tones [14]. The central dialect spoken in the Hue area and the southern dialect spoken in Ho Chi Minh and its vicinity have five tones [15,16], whose pitch patterns differ from the northern dialect [17].

In terms of prominence, Vietnamese is similar to Taiwanese Mandarin with four-tone patterns [18], and Thai with five-tone patterns [19,20]. Nguyen [14] listed 23 consonants and 12 vowels in Vietnamese. A rhythmic unit of Vietnamese is a syllable with a structure of (C)(C)V(C) (Table 1). Voiceless stops such as /p/, /t/, and /k/ are unexploded at a syllable-final position [14,21]. The vowel duration is longer at a word-final position than at other positions [22]. It is also longer in an open syllable than in a closed syllable [23].

Vietnamese has the same syllable structure [(C)(C)V(C)] as Japanese (Table 1). However, it has a different prominence and a rhythmic unit: Vietnamese has a tone prominence and a syllable rhythmic unit, whereas Japanese has pitch accent prominence and a mora rhythmic unit. Moreover, Vietnamese differs from Korean, Taiwanese Mandarin, and Thai in the prominence, rhythmic unit, and syllable structure (Table 1).

Therefore, if data on Vietnamese speakers’ mispronunciations are obtained, they will add new information about the prosodic characteristics related to non-native speakers’ mispronunciations that were investigated in other languages in previous studies [11,12]. Specifically, if the Vietnamese speakers have a similar mispronunciation tendency as the Korean, Taiwanese Mandarin, and Thai speakers, it would support the assumption that non-native speakers have a different general tendency to pronounce Japanese singleton and geminate stops.

Based on this notion, this study recorded Vietnamese speakers’ pronunciations of Japanese singleton and geminate stops and identified their mispronunciations using the same procedure as Yamakawa et al. [11]. Subsequently, it analyzed the segment durations of preceding CV, closure, and following CV segments to obtain further information about the characteristics of non-native speakers’ mispronunciations.

2. SPEECH RECORDING

2.1. Speakers

Twenty-four Vietnamese speakers (11 males and 13 females) participated in the speech recording. Their average age was 21.4 years (Min. = 20, Max. = 23, SD = 0.8). Fifteen of them (10 males and 5 females) spoke the northern dialect of Vietnamese. They were from Ha Noi and its suburbs (14 speakers), and Nghe An (one male speaker). Their average age was 21.7 years (Min. = 21, Max. = 23, SD = 0.7). The remaining 9 speakers (1 male and 8 females) spoke the southern dialect of Vietnamese. They were from Ho Chi Minh and its suburbs, and Lam Dong (one female speaker). Their average age was 21.0 years (Min. = 22, Max. = 20, SD = 0.7). All speakers could read Hiragana and they have never lived in Japan. They had at least an N3 certificate of the Japanese-language proficiency test (JLPT), which has five levels from highly proficient (N1) to basic (N5). Therefore, the Vietnamese speakers in this study can be regarded as having an intermediate or higher ability in Japanese.

2.2. Word Items for Recording

Word items for recording were the 343 words used by Yamakawa et al. [11]. The 343-word set was constructed with minimal pairs of words. The minimal pairs consisted of either a) contrasting singleton and geminate consonants, b) short and long vowels, c) with and without a moraic nasal, or d) 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. 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. 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, the accent patterns are not a major clue for the identification of singleton and geminate stops, which mainly depends on the duration of speech segments [3,4].

2.3. Procedure

The 343-word items used by Yamakawa et al. [11] were recorded in a soundproof room at the Vietnam Academy of Science and Technology in Hanoi, and in a
recording studio (Yeu Ca Hat) in Ho Chi Minh, Vietnam. The recording procedure was the same as the one taken by Yamakawa et al. [11]. The word items were recorded using a microphone (Sony, ECM-999), a pop filter (Stedman, Proscreen XL), an A/D converter (Roland, UA25-EX), and a personal computer (Toshiba, Dynabook SS RX2). In each recording trial, the speaker was asked to push the space key on the computer keyboard to start the recording. A word item embedded in a carrier sentence, /korewa _ dato omoi masu/ (“I suppose that this is .”), was presented at the center of the computer screen in Hiragana. The speaker pronounced the sentence at a normal speaking rate. The pronunciation was digitally recorded with 16-bit quantization and 48-kHz sampling frequency, and it was stored as a digital audio file on the computer. After the pronunciation was finished, the speaker pushed the space key to stop the recording. The speaker first performed 10 practice trials. Next, the speaker pronounced 343 words in carrier sentences that were divided into nine blocks of 35 words and one block of 28 words. The speaker took a break for several minutes after every two blocks. The word order was randomized for each speaker. 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

Yamakawa et al. [11] analyzed 30 minimal pair words (Appendix) selected from the 343-word items. Following their analysis procedure, the same 30 minimal pair words pronounced by each Vietnamese speaker were selected from the 343 recorded items. The minimal pair words contrasted singleton and geminate stops. They were 2 or 3 morae long for singleton words and 3 or 4 morae long for geminate words. As a control condition of analysis, the data of the same 30 minimal pair words pronounced by Japanese speakers were obtained from Yamakawa et al. [11]. By excluding the misrecorded items, there were 1,434 and 599 items by Vietnamese and Japanese speakers, respectively.

3.2. Segment Duration Measurement

Paid professional labelers measured the start and end times of each phoneme segment of the selected items. Using the measured start and end times, the duration of the closure of the singleton and geminate stops was calculated. Durations were also calculated for the preceding and following CV segments of the closure. Subword durations were calculated as the sum of the duration of the preceding CV segment, closure, and following CV segment.

Figure 3 is a scattergram of singleton and geminate stops on the coordinate plane of closure and subword durations. In addition to Vietnamese speakers’ data, Fig. 3 Scattergram of singleton and geminate stops on the coordinate plane of closure and subword durations pronounced by (a) Japanese speakers, (b) Korean speakers, (c) Taiwanese Mandarin speakers, (d) Thai speakers, and (e) Vietnamese speakers. A solid line represents category boundary of the stops of Japanese speakers. Japanese, Korean, and Taiwanese Mandarin speakers’ data were obtained from Yamakawa et al. [11]. Thai speakers’ data were obtained from Yamakawa and Amano [12].
Japanese, Korean, and Taiwanese Mandarin speakers’ data [11] and Thai speakers’ data [12] are plotted. The solid line is a category boundary (Eq. (1) in Sect. 3.3) of singleton and geminate stops pronounced by Japanese speakers.

### 3.3. Mispronunciation Identification

The mispronunciations of Vietnamese speakers were identified using the native Japanese speaker’s category boundary of singleton and geminate stops in Yamakawa et al. [11]. The category boundary is represented by Eq. (1):

\[ 0.0889x - y + 61.4 = 0, \]  
\[ (1) \]

where \( x \) is the subword duration and \( y \) is the closure duration in milliseconds. This category boundary is reliable because the discriminant error of singleton and geminate stops was only 4.5% for the 599 items pronounced by Japanese speakers [11].

When the left side of Eq. (1) \( < 0 \), it indicates that the Vietnamese speakers were designated 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 Vietnamese speakers were designated to pronounce a geminate but that their pronunciation was categorized as a singleton (i.e., \( G \rightarrow S \) mispronunciation). The \( S \rightarrow G \) mispronunciation ratio was 35.4% for 714 singleton items whereas the \( G \rightarrow S \) mispronunciation ratio was 1.8% for 719 geminate items. That is, the Vietnamese speakers made \( S \rightarrow G \) mispronunciations more frequently than \( G \rightarrow S \) mispronunciations \( (z = 16.37, p < 0.001) \).

This mispronunciation tendency is consistent with that of Korean and Taiwanese Mandarin speakers (Table 2). That is, Korean speakers made more frequent \( S \rightarrow G \) mispronunciations (24.4%) than \( G \rightarrow S \) mispronunciations (4.0%) \( (z = 7.16, p < 0.001) \), and Taiwanese Mandarin speakers also made more frequent \( S \rightarrow G \) mispronunciations (26.4%) than \( G \rightarrow S \) mispronunciations (9.3%) \( (z = 5.45, p < 0.001) \). However, Vietnamese speakers’ tendency differs from Thai speakers who made fewer \( S \rightarrow G \) mispronunciations (18.8%) than \( G \rightarrow S \) mispronunciations (26.5%) \( (z = 2.22, p < 0.05) \).

### 3.4. Segment Duration

Table 3 shows segment durations of the preceding CV, closure, following CV, and subword for singleton and geminate stops in mispronunciation by Vietnamese speakers in this study, and in pronunciation by Japanese speakers in Yamakawa et al. [11]. The Vietnamese speakers’ mispronunciation had a significantly shorter closure duration in a geminate stop than in a singleton stop \( ( t(264) = 5.33, p = 2.1 \times 10^{-7}) \). The opposite relation was found in Japanese speakers. They had a longer closure duration in a geminate stop than that in a singleton stop [11]. The Vietnamese speakers’ duration of the preceding CV segment did not differ between singleton and geminate stops \( ( t(264) = 1.01, p = 0.32) \). However, this is different from the Japanese speakers’ duration of the preceding CV segment that is longer in a geminate stop than in a singleton stop.

#### Table 3 Segment duration (ms) for singleton and geminate stops in pronunciation by native Japanese speakers and in mispronunciation by Vietnamese speakers. The ratio of preceding CV duration to following CV duration is shown in the rightmost column.

<table>
<thead>
<tr>
<th>Speakers’ first language</th>
<th>Designated stop type</th>
<th>Mispronunciation type</th>
<th>( n )</th>
<th>Preceding CV</th>
<th>Closure</th>
<th>Following CV</th>
<th>Subword</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( M )</td>
<td>( SD )</td>
<td>( M )</td>
<td>( SD )</td>
</tr>
<tr>
<td>Japanese</td>
<td>Singleton</td>
<td>—</td>
<td>299</td>
<td>131.3</td>
<td>38.9</td>
<td>51.3</td>
<td>18.6</td>
</tr>
<tr>
<td></td>
<td>Geminate</td>
<td>—</td>
<td>300</td>
<td>154.5</td>
<td>43.9</td>
<td>131.3</td>
<td>30.7</td>
</tr>
<tr>
<td>Vietnamese</td>
<td>Singleton</td>
<td>( S \rightarrow G )</td>
<td>253</td>
<td>172.1</td>
<td>65.2</td>
<td>132.3</td>
<td>32.6</td>
</tr>
<tr>
<td></td>
<td>Geminate</td>
<td>( G \rightarrow S )</td>
<td>13</td>
<td>190.5</td>
<td>39.8</td>
<td>83.8</td>
<td>14.3</td>
</tr>
</tbody>
</table>

Note. Japanese data were obtained from Yamakawa et al. [11].
Table 4  Segment ratio to a subword duration for singleton and geminate stops in pronunciation by native Japanese speakers and in mispronunciation by Vietnamese speakers.

<table>
<thead>
<tr>
<th>Speakers’ first language</th>
<th>Designated stop type</th>
<th>Mispronunciation type</th>
<th>Segment</th>
<th></th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Preceding CV</td>
<td>Closure</td>
<td>Following CV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Japanese</td>
<td>Singleton</td>
<td>—</td>
<td>0.465</td>
<td>0.081</td>
<td>0.184</td>
<td>0.060</td>
<td>0.351</td>
<td>0.060</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Geminate</td>
<td>—</td>
<td>0.409</td>
<td>0.073</td>
<td>0.352</td>
<td>0.063</td>
<td>0.238</td>
<td>0.039</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vietnamese</td>
<td>Singleton</td>
<td>S → G</td>
<td>0.360</td>
<td>0.077</td>
<td>0.281</td>
<td>0.056</td>
<td>0.359</td>
<td>0.058</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Geminate</td>
<td>G → S</td>
<td>0.409</td>
<td>0.085</td>
<td>0.202</td>
<td>0.040</td>
<td>0.330</td>
<td>0.066</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Japanese data were obtained from Yamakawa et al. [11].

The Vietnamese speakers’ duration of the following CV segment was significantly shorter in a geminate stop than in a singleton stop \( t(264) = 3.10, p = 2.1 \times 10^{-3} \), which is consistent with the Japanese speakers’ duration of the following CV segment [11]. These results indicate that the Vietnamese speakers’ mispronunciations have different characteristics from the Japanese speakers’ correct pronunciations not only in a closure segment but also in a preceding CV segment.

The duration ratio of the preceding CV to the following CV in Table 3 indicates that Japanese and Vietnamese speakers have a different relationship between these CV segments. For a singleton stop, the Vietnamese speakers’ ratio (0.95) was lower than the Japanese speakers’ ratio (1.35). For a geminate stop, the Vietnamese speaker’s ratio (1.37) was lower than the Japanese speaker’s ratio (1.73). These results indicate that, compared to Japanese speakers, Vietnamese speakers produce either a shorter preceding CV segment or a longer following CV segment.

A statistical test was performed for each singleton and geminate stop to examine the difference in subword durations between Japanese and Vietnamese speakers. No difference was observed in a geminate stop \( t(311) = 1.84, p = 0.103 \), but a significant difference existed for a singleton stop \( t(550) = 29.3, p = 2.9 \times 10^{-19} \). These results indicate that speaking rates were different between these speakers at least for a singleton stop.

3.5. Segment Ratio

As speaking rates of Japanese and Vietnamese speakers are significantly different at least concerning a singleton stop, it is not reasonable to compare their raw segment durations because they vary with the speaking rate. To compensate for the duration differences caused by the speaking rate, the ratio of each segment duration to the subword duration was used for the analysis. Specifically, the mean ratio of each segment duration to the corresponding subword duration was obtained for singleton and geminate stops (Table 4). The segment ratios in Table 4 seem to be different between the Japanese and Vietnamese speakers. For example in the preceding CV segment, the Japanese speakers had a higher ratio in a singleton stop than in a geminate stop, whereas the Vietnamese speakers had a lower ratio in a singleton stop than in a geminate stop.

To clarify the difference between Japanese and Vietnamese speakers, the ratio differences were analyzed. First, the ratios of the preceding CV, closure, and following CV segment in each word were averaged for Japanese speakers. Next, the ratio difference in these segments was obtained by subtracting the average segment ratio from the Vietnamese speakers’ segment ratio for each word. Figures 4 and 5 show the ratio differences in the preceding CV, closure, and following CV segments for S → G and G → S mispronunciations, respectively. Note that for S → G mispronunciations, the ratio difference was calculated from native Japanese speakers’ singleton pronunciations, whereas for G → S mispronunciations, it was calculated from Japanese speakers’ geminate pronunciations.
Fig. 5 Ratio difference of speech segments in non-native speakers’ G \(\rightarrow\) S mispronunciations from the native Japanese speakers’ geminate pronunciations. Error bars represent standard deviation. Korean and Taiwanese Mandarin speakers’ data and Thai speakers’ data were obtained from Yamakawa et al. [11] and Yamakawa and Amano [12], respectively.

A statistical test was performed to examine whether the segment ratio difference was equal to zero for Vietnamese speakers. The test results revealed that the segment ratio difference for the S \(\rightarrow\) G mispronunciation (Fig. 4) was significantly lower than 0 in the preceding CV segment \(t(252) = 24.5, p = 1.6 \times 10^{-68}\) but was higher than 0 in the closure \(t(252) = 24.9, p = 4.6 \times 10^{-70}\) and the following CV \(t(252) = 4.89, p = 1.8 \times 10^{-6}\). In contrast, for the G \(\rightarrow\) S mispronunciation (Fig. 5), the segment ratio difference was significantly lower than 0 in a closure \(t(12) = 7.56, p = 6.7 \times 10^{-6}\) but was higher than 0 in a following CV segment \(t(12) = 7.07, p = 1.3 \times 10^{-5}\). Although the ratio difference of the following CV segment in the S \(\rightarrow\) G mispronunciation is significant, the difference is very small and almost zero (Fig. 4). Therefore, it is probably appropriate to regard the ratio as being substantially not different from the Japanese speakers’ segment ratio. These results are summarized that, in comparison to the Japanese speakers’ correct pronunciation, Vietnamese speakers have too long closure and too short preceding CV segment in S \(\rightarrow\) G mispronunciations, whereas for G \(\rightarrow\) S mispronunciations, they have too short closure segment and too long following CV segment.

4. DISCUSSION

This study investigated the characteristics of Vietnamese speakers’ mispronunciations of Japanese singleton and geminate stops. Their mispronunciations were identified using the category boundary of the stops pronounced by native Japanese speakers. The results revealed that the Vietnamese speakers made more frequent S \(\rightarrow\) G mispronunciations than G \(\rightarrow\) S mispronunciations, which is consistent with the mispronunciation characteristics of Korean and Taiwanese Mandarin speakers in Yamakawa et al. [11].

As Yamakawa et al. [11] pointed out, the fewer G \(\rightarrow\) S mispronunciations than S \(\rightarrow\) G mispronunciations would be caused by non-native speakers’ hypercorrection of a Japanese geminate stop. Because Vietnamese speakers are probably aware that a geminate stop has a long closure, it is likely that they intentionally pronounce the closure of the geminate stop longer than usual. This resulted in a successful geminate pronunciation and achieved a smaller ratio of the G \(\rightarrow\) S mispronunciation. Furthermore, because Vietnamese speakers probably have an ambiguous category boundary of singleton and geminate stops, they cannot pronounce a singleton stop with a sufficiently short closure. As a result, they more frequently made S \(\rightarrow\) G mispronunciations than G \(\rightarrow\) S mispronunciations as shown in Table 2.

The results of the segment duration analysis (Table 3) revealed that Vietnamese speakers produced a longer closure in S \(\rightarrow\) G mispronunciations but a shorter closure in G \(\rightarrow\) S mispronunciations. It is reasonable that S \(\rightarrow\) G mispronunciations had a longer closure because they were categorized as geminate stops. It is also reasonable that G \(\rightarrow\) S mispronunciations had a shorter closure because they were categorized as singleton stops. These results are expected and are not surprising.

Notable findings of this study pertain to preceding and following CV segments. Vietnamese speakers produced a shorter preceding CV segment in S \(\rightarrow\) G mispronunciations (Fig. 4) and a longer following CV segment in G \(\rightarrow\) S mispronunciations (Fig. 5) than native Japanese speakers. Furthermore, the segment characteristics of Vietnamese speakers were consistent with those of Korean and Taiwanese Mandarin speakers [11], and Thai speakers [12]. However, because the number of G \(\rightarrow\) S mispronunciations of Vietnamese speakers was very small \((n = 13, Table 2)\), their results of G \(\rightarrow\) S mispronunciations might be less reliable than those in other languages.

The above findings suggest that non-native speakers in any language may have a common tendency to have an inadequate duration of the closure, preceding, and following CV segments in mispronunciation of Japanese singleton and geminate stops. However, this should be examined in a future study using non-native speakers in other languages such as English and French.

What can be the cause of the segment characteristics of Vietnamese speakers’ mispronunciations? The syllable structure cannot be the cause because the Vietnamese language has the same structure \((C)(C)V(C)\) as the Japanese language (Table 1). Because the Vietnamese language has a tone prominence whereas the Japanese language has a pitch accent prominence (Table 1), the
pronunciation may be the cause. However, the same characteristics of mispronunciation as Vietnamese speakers were observed in Korean speakers [11] even though the Korean language has a phrase accent prominence that differs from the tone prominence in the Vietnamese language. Therefore, the prominence is not the cause.

The most probable cause is a rhythmic unit. The Japanese language has a mora rhythmic unit, whereas the Vietnamese language, as well as the Korean, Taiwanese Mandarin, and Thai languages, have a syllable rhythmic unit (Table 1). Because non-native Japanese speakers in these languages commonly showed the same tendency of segment duration in mispronunciation, a rhythmic unit is likely to be the cause of the characteristics of mispronunciation.

According to the study by Yamakawa et al. [11] on mispronunciations by Korean and Taiwanese Mandarin speakers, the rhythmic-unit difference between mora and syllable causes the deviation of the segment durations. This can be applied to explain the deviation of segment durations by Vietnamese speakers (Figs. 4 and 5).

Namely, when the Vietnamese speakers made S → G mispronunciations, they probably applied a syllable structure and mispronounced a singleton sequence /CV.CV/ as /CVC.CV/ (here, “—” indicates a syllable boundary, and “C” indicates a consonant with a closure). The /CVC/ can be an unreleased stop that is legal for /p/, /t/, and /k/ at a syllable-final position in Vietnamese [21]. Since the Vietnamese speakers pronounced a longer closure segment in S → G mispronunciations (Fig. 4) and /CVC/ contains more segments than /CV/, the speakers would need to shorten the duration of the preceding /CV/ in /CVC/ to achieve almost the same duration as /CV/ in the original singleton sequence /CV.CV/.

On the other hand, when the Vietnamese speakers made G → S mispronunciations, they applied a syllable structure to a geminate sequence /CV.C-CV/ (here, “—” indicates a mora boundary) and mispronounced it as /CV.CV/. Since the Vietnamese speakers pronounced a shorter closure segment in G → S mispronunciations (Fig. 5), and since /CV/ contains fewer segments than /C-CV/, they would need to lengthen the duration of /CV/ to achieve the duration that they assume for a “geminate” segment. While these explanations seem reasonable, further research is necessary to confirm them.

Previous studies investigated a rhythmic unit in the context of categorizing language types. They suggested that the rhythmic unit is, for example, a stressed syllable for English and Dutch, a syllable for French and Italian, and a mora for Japanese and Tamil [24]. To represent the language types in terms of a rhythmic unit, several kinds of indexes were proposed. The indexes were, for example, the percentage of vocalic duration (%V) [24], the standard deviation of vocalic (ΔV) or consonantal durations (ΔC) [24], the variability index (VI) of syllable duration [25], and the pairwise variability index (PVI) of vocalic or consonantal durations [26].

Although some studies successfully classified several languages by the indexes [e.g., 24, 26], one index or a set of indexes could not identify one language [27], suggesting that the indexes do not properly capture the rhythmic characteristics of languages. However, these results were obtained using the speech data pronounced by native speakers. For the speech data pronounced by non-native speakers like in this study, the indexes might be able to appropriately represent non-native speakers’ rhythmic characteristics. Relationships between these indexes and the current results about non-native speakers might provide new insights into the effects of a rhythmic unit on non-native learners’ difficulties in learning a foreign language. Future studies should investigate this point.

This study used a geminate stop consisting of plosives and affricates. However, the Japanese geminate consonant can be a fricative. In this study, the geminate consonant was voiceless. However, it can be a voiced consonant such as /b/ and /z/, although voiced geminate consonants are not frequent. This study used a part of geminate consonants. In future research, the findings of this study should be examined using other types of consonants described above.

Only a normal speaking rate was used in this study. As the distinction between singleton and geminate stop consonants is duration sensitive, the speaking rate might have effects on non-native speakers’ mispronunciation. For example, pronunciation at a fast speaking rate is more difficult than at a normal or slow speaking rate. Therefore, more mispronunciations are expected and there might be a different pattern of segment duration. This possibility should be examined in future studies.

The accent pattern and position of singleton and geminate stops were not controlled in this study. Although these properties of word items may have effects on the production of the stops, they would be not strong because they do not directly relate to the closure duration which is the main feature for the distinction between singleton and geminate stops. Speaking rate which directly relates to the closure duration probably has more powerful effects than the accent pattern and stop position. However, future studies should control the accent pattern and stop position to obtain precise results.

The nonnative speakers’ mispronunciations were identified with a category boundary of singleton and geminate stops produced by native speakers in this study. In the strictest sense, the criterion of mispronunciation should be the native listener’s perception boundary. However, because Amano and Hirata [6,7] have demonstrated that the perception and production category boundaries of singleton
and geminate stops in native speakers coincide across a wide range of speaking rates, the production boundary in this study is probably the same as the perception boundary. However, it would be still worthwhile to conduct a perception experiment to confirm their consistency in future studies.

This study did not analyze the relationship between speakers’ Japanese proficiency and their mispronunciation characteristics. It is not easy to precisely measure Japanese proficiency because no appropriate method has been developed for measuring non-native speakers’ ability to speak, listen, and understand spoken Japanese. Even the JLPT levels do not correctly reflect this ability. However, the relationship is an interesting topic and should be investigated in the future.

Although this study collected data from Vietnamese speakers of northern and southern dialects, the difference between the dialects was not analyzed because there were not sufficient data for analysis. For example, only one the G → S mispronunciation was found in 270 pronunciations of the nine southern-dialect Vietnamese speakers, while only 11 G → S mispronunciations were found in 360 pronunciations of the 15 northern-dialect Vietnamese speakers. Future studies should increase the number of data and determine the difference in mispronunciations between the dialects.

In summary, this study investigated Vietnamese speakers’ mispronunciations of Japanese singleton and geminate stops. It was found that the segment durations of the preceding CV, closure, and following CV have the same tendency as those in mispronunciation by Korean, Taiwanese Mandarin, and Thai speakers revealed in previous studies [11,12]. The results of this study provide further support to the assumption that non-native speakers have a common tendency of segment durations in the mispronunciation of Japanese singleton and geminate stops. However, further research is necessary to confirm this hypothesis.

ACKNOWLEDGMENTS

This study was supported by JSPS KAKENHI Grant Numbers JP22320081, JP24652087, JP25284080, JP26370464, and JP17K02705. We would like to thank Professor Chi Mai Luong of the Vietnam Academy of Science and Technology in Hanoi and Professors Nguyen Thu Huong and Nguyen Tien Luc of Vietnam National University in Ho Chi Minh City for their assistance in the utterance recordings.

REFERENCES

Appendix: Minimal pairs of Japanese words

<table>
<thead>
<tr>
<th>Orthography</th>
<th>Singleton word</th>
<th>IPA transcription</th>
<th>Meaning</th>
<th>Orthography</th>
<th>Geminate word</th>
<th>IPA transcription</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>に下</td>
<td>[buka]</td>
<td>subordinate</td>
<td></td>
<td>物価</td>
<td>[bukka]</td>
<td>prices</td>
<td></td>
</tr>
<tr>
<td>画家</td>
<td>[gaka]</td>
<td>painter</td>
<td></td>
<td>学科</td>
<td>[gaka]</td>
<td>department</td>
<td></td>
</tr>
<tr>
<td>ガキ</td>
<td>[gak'i]</td>
<td>brat</td>
<td></td>
<td>良器</td>
<td>[gaki]</td>
<td>musical instrument</td>
<td></td>
</tr>
<tr>
<td>腹戸</td>
<td>[haka]</td>
<td>grave</td>
<td></td>
<td>発火</td>
<td>[haka]</td>
<td>combustion</td>
<td></td>
</tr>
<tr>
<td>便戦</td>
<td>[dzieken]</td>
<td>affair</td>
<td></td>
<td>実験</td>
<td>[dzieken]</td>
<td>experiment</td>
<td></td>
</tr>
<tr>
<td>離えて</td>
<td>[kaete]</td>
<td>changed</td>
<td></td>
<td>帰って</td>
<td>[kaete]</td>
<td>returned</td>
<td></td>
</tr>
<tr>
<td>過去</td>
<td>[kako]</td>
<td>past</td>
<td></td>
<td>括弧</td>
<td>[kako]</td>
<td>parenthesis</td>
<td></td>
</tr>
<tr>
<td>核戦</td>
<td>[kite]</td>
<td>pabulum</td>
<td></td>
<td>勝手</td>
<td>[kate]</td>
<td>selfishness</td>
<td></td>
</tr>
<tr>
<td>来た</td>
<td>[kita]</td>
<td>have come</td>
<td></td>
<td>切っ株</td>
<td>[kita]</td>
<td>have cut</td>
<td></td>
</tr>
<tr>
<td>来て</td>
<td>[kite]</td>
<td>coming</td>
<td></td>
<td>切手</td>
<td>[kite]</td>
<td>stamp</td>
<td></td>
</tr>
<tr>
<td>帰途</td>
<td>[kito]</td>
<td>on the way home</td>
<td></td>
<td>きっと</td>
<td>[kito]</td>
<td>surely</td>
<td></td>
</tr>
<tr>
<td>子機</td>
<td>[kok'i]</td>
<td>cordless handset</td>
<td></td>
<td>国旗</td>
<td>[kok'i]</td>
<td>national flag</td>
<td></td>
</tr>
<tr>
<td>町</td>
<td>[matchi]</td>
<td>town</td>
<td></td>
<td>マッチ</td>
<td>[matchi]</td>
<td>match</td>
<td></td>
</tr>
<tr>
<td>枕</td>
<td>[makura]</td>
<td>pillow</td>
<td></td>
<td>真っ暗</td>
<td>[makura]</td>
<td>pitch darkness</td>
<td></td>
</tr>
<tr>
<td>坡</td>
<td>[saka]</td>
<td>slope</td>
<td></td>
<td>作家</td>
<td>[saka]</td>
<td>writer</td>
<td></td>
</tr>
<tr>
<td>先</td>
<td>[sak'i]</td>
<td>tip</td>
<td></td>
<td>殺気</td>
<td>[saki]</td>
<td>menacing atmosphere</td>
<td></td>
</tr>
<tr>
<td>世界</td>
<td>[seki]</td>
<td>world</td>
<td></td>
<td>石灰</td>
<td>[seki]</td>
<td>caustic lime</td>
<td></td>
</tr>
<tr>
<td>世界間</td>
<td>[sekken]</td>
<td>society</td>
<td></td>
<td>石鹸</td>
<td>[sekken]</td>
<td>soap</td>
<td></td>
</tr>
<tr>
<td>病</td>
<td>[seki]</td>
<td>cough</td>
<td></td>
<td>石器</td>
<td>[seki]</td>
<td>stone ware</td>
<td></td>
</tr>
<tr>
<td>世帯</td>
<td>[setai]</td>
<td>household</td>
<td></td>
<td>接待</td>
<td>[setai]</td>
<td>reception</td>
<td></td>
</tr>
<tr>
<td>源戸</td>
<td>[seto]</td>
<td>strait</td>
<td></td>
<td>セット</td>
<td>[seto]</td>
<td>set</td>
<td></td>
</tr>
<tr>
<td>式</td>
<td>[siki]</td>
<td>equation</td>
<td></td>
<td>漆器</td>
<td>[siki]</td>
<td>lacquer ware</td>
<td></td>
</tr>
<tr>
<td>して</td>
<td>[sute]</td>
<td>done</td>
<td></td>
<td>知って</td>
<td>[sute]</td>
<td>known</td>
<td></td>
</tr>
<tr>
<td>帯tein</td>
<td>[sito]</td>
<td>use</td>
<td></td>
<td>嫉妬</td>
<td>[sito]</td>
<td>jealousy</td>
<td></td>
</tr>
<tr>
<td>スパイ</td>
<td>[supai]</td>
<td>spy</td>
<td></td>
<td>醞っぱい</td>
<td>[supai]</td>
<td>sour</td>
<td></td>
</tr>
<tr>
<td>所見</td>
<td>[sokeki]</td>
<td>observation</td>
<td></td>
<td>食券</td>
<td>[sokeki]</td>
<td>food ticket</td>
<td></td>
</tr>
<tr>
<td>書記</td>
<td>[sokeki]</td>
<td>secretary</td>
<td></td>
<td>食器</td>
<td>[sokeki]</td>
<td>tableware</td>
<td></td>
</tr>
<tr>
<td>布</td>
<td>[tate]</td>
<td>shield</td>
<td></td>
<td>立って</td>
<td>[tate]</td>
<td>stood</td>
<td></td>
</tr>
<tr>
<td>備</td>
<td>[tek'i]</td>
<td>enemy</td>
<td></td>
<td>鉄器</td>
<td>[tek'i]</td>
<td>ironware</td>
<td></td>
</tr>
<tr>
<td>時</td>
<td>[tok'i]</td>
<td>time</td>
<td></td>
<td>突起</td>
<td>[tok'i]</td>
<td>prong</td>
<td></td>
</tr>
</tbody>
</table>

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