Perceptual boundary and discrimination sensitivity of Japanese singleton and geminate stops in Japanese and Taiwanese Mandarin speakers

Kimiko Yamakawa\textsuperscript{1,*}, Shigeaki Amano\textsuperscript{2,†} and Mariko Kondo\textsuperscript{3,‡}

\textsuperscript{1}Faculty of Contemporary Culture, Shokei University, 2–6–78 Kuhonji, Chu-o-ku, Kumamoto, 862–8678 Japan
\textsuperscript{2}Faculty of Human Informatics, Aichi Shukutoku University, 2–9 Katahira, Nagakute, 480–1197 Japan
\textsuperscript{3}School of International Liberal Studies, Waseda University, 1–6–1 Nishi-waseda, Shinjuku-ku, Tokyo, 169–8050 Japan

(Received 12 July 2021, Accepted for publication 6 December 2021)

Abstract: To clarify non-native speakers’ perception of Japanese singleton and geminate stops, the perceptual boundary and discrimination sensitivity of these stops were investigated with native speakers of Japanese and Taiwanese Mandarin. The perceptual boundary of Japanese speakers varied with speaking rate, but Taiwanese Mandarin speakers had an almost constant perceptual boundary regardless of speaking rates. Taiwanese Mandarin speakers also had a much lower discrimination sensitivity, only about 15% that of native Japanese speakers. These results indicate that Taiwanese Mandarin speakers’ perception of Japanese singleton and geminate stops is ambiguous due to their inadequate perceptual boundary and low discrimination sensitivity. Meanwhile, vowel devoicing preceding or following the stops decreased the discrimination sensitivity in both Japanese and Taiwanese Mandarin speakers, suggesting an effect of language-independent hearing processing on discrimination sensitivity instead of an effect of language-dependent speech processing. Our results suggest that multiple factors play a role in non-native speakers’ perception of Japanese singleton and geminate stops.

Keywords: Geminate stop, Vowel devoicing, Non-native speaker, Perceptual boundary, Discrimination sensitivity, Speaking rate

1. INTRODUCTION

The Japanese language has several geminate consonants (e.g., [ki] in [kako] ‘parenthesis’) that are the counterpart of singleton consonants (e.g., the second [k] in [kako] ‘past’). Geminate consonants include, for example, voiceless stops ([p], [t], and [k]), voiceless affricates ([ts], and [ts]), voiceless fricatives ([s], and [z]). It also consists of voiced consonants such as [b], [d], [g], [dz], and [dz] although these geminate consonants occur much less frequently, and primarily in recent loan words [1].

Regarding a geminate stop, the closure duration preceding a burst segment is a primary acoustic cue [e.g., 1,2,3]. Namely, geminate and singleton stops can be distinguished with the closure duration, which is longer for a geminate stop than a singleton stop. However, the closure duration varies with the speaking rate. For both geminate and singleton stops, the closure duration becomes shorter as the speaking rate increases and longer as the speaking rate decreases [4,5]. In some cases, the duration of a singleton stop at a slow speaking rate is longer than that of a geminate stop at a fast speaking rate. Therefore, the absolute closure duration cannot be an invariant acoustic cue for singleton and geminate stops.

Nevertheless, native Japanese speakers can easily distinguish the two stops at various speaking rates, indicating the existence of a secondary acoustic cue to account for the variation of closure duration with speaking rate. This secondary acoustic cue might be the duration of such segments as a preceding vowel [6–8], a preceding mora [8], a following vowel [7,8], a subword [4,5], or a word [4,5]. Each of these secondary acoustic cues may contribute in some degree to the distinction between geminate and singleton stops because all these cues can provide information on the speaking rate.

Meanwhile, non-native speakers have difficulties in distinguishing Japanese singleton and geminate stops [9–11]. For example, non-native speakers such as English
Korean speakers had a smaller sensitivity than Japanese speakers, possibly because their first languages have the rhythm unit and phoneme system that differ from those of the Japanese language.

Previous studies examined the characteristics of non-native speakers’ perceptual boundary and discrimination sensitivity. As an example of a perceptual boundary study, Oba et al. [16] conducted a perception experiment to obtain the perceptual boundary of singleton and geminate consonants. In their experiment, Japanese and German speakers made a two-alternative-forced-choice of short (i.e., singleton) or long (i.e., geminate) for items in a stimulus continuum of Japanese singleton and geminate consonants. They found that German speakers had a different perceptual boundary from Japanese speakers. Namely, their perceptual boundary shifted toward longer (i.e., shift to a geminate consonant) than the Japanese speaker’s boundary.

However, the shift direction is not consistent among previous studies. For example, the perceptual boundary shifted to a shorter duration in English speakers in Hirata [12] and Taiwanese Mandarin speakers in Hung [19]. While in Kingston et al. [14], the perceptual boundary of English, Italian, and Norwegian speakers shifted to a longer duration. In Uchida [18], the perceptual boundary of Mandarin speakers with high Japanese proficiency shifted to a shorter duration in a descending series of stimulus presentations, whereas it shifted to longer in an ascending series of stimulus presentations. Although these discrepancies may in some cases be due to the different characteristics of the speakers’ first languages, previous studies only agree that the perceptual boundaries of non-native speakers differ from that of Japanese speakers.

Besides a perceptual boundary, some studies investigated non-native speaker’s discrimination sensitivity for singleton and geminate consonants. They used the d-prime ($d'$) of the signal detection theory as an index of the sensitivity. For example, Sonu et al. [17] showed that Korean speakers had a smaller $d'$ than Japanese speakers, indicating that they are less sensitive than Japanese speakers to the difference between singleton and geminate consonants. The smaller $d'$ was also obtained for Italian and Norwegian speakers [14] as well as English speakers [13,14]. Previous studies have consistently shown the low discrimination sensitivity of non-native speakers.

However, there are several problems with previous studies. Firstly, previous studies have only roughly estimated the perceptual boundary by simply taking the value at which the response ratio crosses 50%. It is well known that such estimation in a narrow range of stimuli near 50% response ratio is easily affected by random noise in the response data of perception experiments. The perceptual boundary should be obtained through regression analysis using a psychophysical equation such as a logistic function and a cumulative distribution function of the standard normal distribution. Regression analysis can reasonably and accurately estimate the perceptual boundary because it uses response data for the whole range of stimuli.

Secondly, previous studies have examined discrimination sensitivity qualitatively rather than quantitatively. They only described non-native speakers as having lower discrimination sensitivity than Japanese speakers simply because the response curve of non-native speakers appeared to have a gentler slope. Discrimination sensitivity should be measured as the gradient of the regression function at the 50%-response-ratio point that corresponds to the perceptual boundary. This quantitative measurement of sensitivity provides a clearer understanding of the perception characteristics of non-native speakers. In this research direction, Hung [19] analyzed the gradient of the logistic functions for Japanese and Taiwanese Mandarin speakers. However, he did not describe at which point of the stimulus continuum the gradient was calculated. Therefore, it is unclear whether the gradient corresponds to the sensitivity at the perceptual boundary.

Also, although previous studies have measured discrimination sensitivity between singleton and geminate stops with $d'$, this is different from the sensitivity at the perceptual boundary. Since previous studies used typical singleton and geminate items as stimuli to measure $d'$, it represents the sensitivity in distinguishing certain singleton and geminate items not at the perceptual boundary. Moreover, the size of $d'$ depends on how far apart the singleton and geminate items are on the stimulus continuum. Therefore, $d'$ does not represent discrimination sensitivity between the two categories of singleton and geminate consonants in the strict sense.

Furthermore, previous studies have not examined the effects of speaking rate. Although some studies mimicked the variation in speaking rates by extending or diminishing the duration of vowels preceding and following the singleton and geminate consonants, such local manipulation might not exactly reflect the speaking rate information that is spread throughout the sentence. To explore speaking rate effects, it would be desirable to use stimuli of singleton and geminate items embedded in a carrier sentence spoken at various speaking rates.

Finally, no previous studies have examined the effects of devoiced vowels. In Japanese, the vowels [i] and [u] are frequently devoiced when occurring between voiceless consonants [20], while devoicing occurs less frequently for the other vowels ([a], [e], and [o]) and at slower speaking
rates. A recent corpus-based study [21] revealed that the occurrence of vowel devoicing is probabilistic and not obligatory. Since a vowel is a high-intensity segment with a lasting duration, it probably provides some time information for speech perception. Vowel devoicing means dissipation of this time information. Since the perception of singleton and geminate consonants is sensitive to time information, it may be affected by nearby vowel devoicing.

Accordingly, this study aimed to clarify the characteristics of non-native speakers’ perception of Japanese singleton and geminate stops considering the factors of speaking rate and vowel devoicing. Specifically, their perceptual boundary and discrimination sensitivity were precisely estimated by regression analysis with a logistic function in comparison with those of Japanese speakers. This study focused on Taiwanese Mandarin speakers as non-native speakers because the Taiwanese Mandarin language differs from the Japanese language in many respects. For example, the Taiwanese Mandarin language has a syllable rhythm unit and tone prominence, whereas the Japanese language has a mora rhythm unit [22,23] and pitch accent prominence [24]. More importantly, the Taiwanese Mandarin language does not have a geminate stop, whereas the Japanese language does. These differences may affect Taiwanese Mandarin speakers’ perception of singleton and geminate stops.

2. EXPERIMENT

2.1. Participants

The participants were paid for participating in the experiment. There were 25 native Japanese speakers (2 males and 23 females) and 28 native Taiwanese Mandarin speakers (5 males and 23 females). Nineteen of the Japanese speakers (2 males and 17 females) were undergraduate students at Aichi Shukutoku University in Aichi, Japan. They were from Nagoya city and its suburbs. The other six female Japanese speakers were undergraduate students at Shokei University in Kumamoto, Japan. They were from Kumamoto city and its suburbs. The average age of all the Japanese speakers was 20.6 (SD = 0.8).

All the Taiwanese Mandarin speakers were undergraduate students learning the Japanese language in the Department of Oriental Language and Literature at Tzu Chi University in Hualien, Taiwan. They were from Hualien city and its suburbs. Their average age was 20.5 (SD = 1.1). One Taiwanese speaker held the N1 certificate of JLPT (Japanese-Language Proficiency Test), six speakers N2, eight speakers N3, and one speaker N5, where N1 is the highest level and N5 is the lowest level of Japanese proficiency. The other 12 speakers had no JLPT certificates. All the Taiwanese Mandarin speakers could read Japanese written in hiragana.

2.2. Stimulus Recording and Preparation

2.2.1. Speech item recording

Speech items were recorded to obtain candidates of original speech items for the stimulus continuum. The speech items comprised 12 minimal pairs of onomatopoeic words such as [piτapita] and [piτapita] that contrast singleton and geminate stops at the underlined part in the second mora. The minimal pair words satisfied one of the conditions inducing devoicing at 1) the first vowel (V1), 2) the second vowel (V2), 3) both the first and second vowels (V1 & V2), and 4) none of the vowels.

Onomatopoeic words were used for the minimal pairs because a geminate onomatopoeia is an emphatic form of a singleton onomatopoeia, which means they have the same meaning but different degrees of emphasis. Moreover, these onomatopoeic words are free of lexical bias because they are almost equally familiar to native Japanese speakers. Furthermore, the onomatopoeic words do not violate the Japanese phonotactic rules. These characteristics are desirable for recording the natural pronunciations of speech items.

Each of the 12 minimal pairs was embedded in the carrier sentence [koreṣu ḏa tomo omoi masu] (“I suppose that this is _”) following the recording procedure in Hirata and Whiton [25], 19 native Japanese speakers from the Tokyo metropolitan area and its suburbs were asked to pronounce the 12 minimal pairs four times each at fast, normal, and slow speaking rates as non-accented words (i.e., a low-high-high-high pitch for singleton onomatopoeia and a low-high-high-high-high pitch for geminate onomatopoeia) in a soundproof room at Waseda University in Tokyo. Note that because devoicing is probabilistic and does not always occur, there were both voiced and devoiced V1 and V2 pronunciations in the devoicing conditions. The speakers’ pronunciations were digitally recorded in a linear PCM format with 16-bit quantization and 48-kHz sampling frequency.

2.2.2. Speech item selection

The recorded speech items at the fast and normal speaking rates were used as candidates for selecting original speech items for constructing the stimulus continuum for the perception experiment, but the items at the slow speaking rate were not, because there were few instances of devoicing at the slow speaking rate. Moreover, only geminate items were used for the selection because Amano and Hirata [4] showed that stimuli derived from singleton and geminate stops result in an almost equal perceptual boundary at fast and normal speaking rates.

Original speech items for stimulus continua were selected under the condition that the same speaker pronounced the same geminate word with voiced and devoiced V1 or V2 at both fast and normal speaking rates. Vowels without clear periodicity were categorized as
devoiced vowels. The following geminate items produced by several speakers were selected as original speech items: [pi:tapita] and [pitapita] by a 21-year-old male speaker and [pi:kapika] and [pi:kapika] by a 24-year-old male speaker in the V1 devoicing condition, [kat:ikat:ei] and [kat:ikatei] by an 18-year-old male speaker and [taputaputu] and [taputapu] by a 19-year-old male speaker in the V2 devoicing condition. Although a single speaker was desirable for these items, such a selection was not available from the recorded speech items. No vowel devoicing other than V1 and V2 occurred in these items. There were 16 original speech items in total (2 words × 2 devoicing positions × 2 voicing/devoicing occurrences × 2 speaking rates). These items were used as the original speech items when preparing a stimulus continuum for the experiment.

Other items, [pi:k:upuku] and [pi:k:upuk], in the “V1 & V2” devoicing condition were also extracted for the stimulus continuum in the perception experiment. However, the results of the experiment showed that the response ratio of Japanese speakers did not reach close enough to 0.0 or 1.0 at either endpoint of the stimulus continuum derived from these items, indicating that the stimulus continuum was inadequately prepared and the obtained data were not reliable. Therefore, the data in the “V1 & V2” devoicing condition were excluded from the analysis.

2.2.3. Stimulus preparation

After processing all original geminate items to have equal intensity, stimulus continua from singleton to geminate stops in minimal pair words (Table 1) were prepared by modifying the closure duration at the second mora of the original geminate item using an originally developed computer program. The closure duration of each stimulus continuum is shown in Table 2. To achieve the geminate response ratios being approximately between 0% and 100% at least for native Japanese speakers, the minimum and maximum closure duration and the step size were decided by a preliminary listening investigation by the authors.

### Table 1

| Vowel devoicing | Minimal pair words | Geminate |
|-----------------|-------------------|----------|
| V1              | [pi:tapita]       | [pi:tapita] |
|                 | [pitapita]       | [pitapita] |
| V2              | [kat:ikat:ei]     | [kat:ikatei] |
|                 | [kat:ikatei]     | [kat:ikat:ei] |
| V1              | [pi:kapika]       | [pi:kapika] |
|                 | [pi:kapika]       | [pi:kapika] |
| V2              | [taputaputu]      | [taputapu] |
|                 | [taputapu]       | [taputaputu] |

### 2.3. Procedure

The 25 Japanese speakers participated in the experiment in a soundproof room at Aichi Shukutoku University, or in a quiet room at Shokei University. Twenty-six of the Taiwanese Mandarin speakers participated in the experiment in a quiet room at Tzu Chi University, while the other two speakers participated at Shokei University where they were visiting as exchange students.

In each trial, the stimulus that was stored in a computer (Dynabook SS RX2, Toshiba) was diotically presented only once to the participant in a constant and comfortable sound level through an audio interface (UA-22, Roland) and headphones (MDR-Z900HD, Sony). After the stimulus was presented, two response buttons, one for singleton and the other for geminate, were displayed side by side in the center of the computer screen. A singleton or geminate word was shown in hiragana on each button. To avoid bias, the positions of the buttons were interchanged for half of the Japanese and Taiwanese Mandarin participants. Participants were asked to listen to the stimulus and select the button that matched what they perceived.

Each stimulus was presented 10 times in a random order for each participant. After 20 practice trials with the stimuli at the endpoints of the stimulus continuum, participants performed 1,600 trials at their own pace, taking a one-minute break at every 160 trials and a 10-minute break after 800 trials. The experiment took about 90 minutes including practice and breaks.

### 2.4. Results

#### 2.4.1. Response ratio

The ratio of geminate responses was calculated by dividing the number of geminate responses by the total number of trials for each stimulus in the stimulus continuum. Figure 1 shows the geminate response ratio as a function of closure duration for Japanese and Taiwanese Mandarin speakers. Figure 1 also shows the response curve estimated by logistic regression, where the dependent variable was the response ratio and the independent variable was the closure duration. The curve (i.e., logistic function) was well fitted to the data, as seen in Fig. 1.

#### 2.4.2. Perceptual boundary

The perceptual boundary of singleton and geminate stops for each stimulus continuum was obtained as the closure duration yielding a 50% response rate on the fitted logistic function. Since ANOVA of the closure duration indicated no significant differences between the V1 and V2 devoiced items at both fast and normal speaking rates, the
data of these two type items were pooled as devoiced items and entered in the following analysis.

Figure 2 shows the closure duration at the perceptual boundary. The perceptual boundary of Japanese speakers tends to be shorter at the fast speaking rate than the normal speaking rate, and it tends to be shorter for voiced than devoiced items. However, no such clear tendencies appear for Taiwanese Mandarin speakers.

For the closure duration at the perceptual boundary, a three-way ANOVA was performed with the factors of ‘native language,’ ‘speaking rate,’ and ‘voicing/devoicing.’ The results showed that the three-way interaction was not significant. The two-way interaction between ‘native language’ and ‘voicing/devoicing’ was significant \(F(1,24) = 4.31, p < 0.05\), but the other two-way interactions were not.

Following the significant interaction, two-way ANOVA was performed with the factors of “speaking rate” and “voicing/devoicing” in each native language. In Japanese speakers, the interaction was not significant and the “voicing/devoicing” factor was marginally significant \(F(1,12) = 3.35, p = 0.092\). The “speaking rate” factor was significant \(F(1,12) = 5.29, p < 0.05\), indicating that the closure duration at the perceptual boundary is significantly shorter at the fast speaking rate than the normal speaking rate.

Meanwhile, neither the interaction nor the two factors were significant in Taiwanese Mandarin speakers. These results indicate that the closure duration at the perceptual boundary is almost constant in Taiwanese Mandarin speakers regardless of speaking rate and devoicing.

A two-way ANOVA for each speaking rate was performed with the factors of “native language” and “voicing/devoicing.” At the fast speaking rate, the interaction was not significant. Only the factor of “native language” was significant \(F(1,12) = 27.8, p < 0.001\), indicating that the closure duration at the perceptual boundary is shorter for Japanese speakers than Taiwanese Mandarin speakers. At the normal speaking rate, no interaction and factors were significant, indicating that the closure duration at the perceptual boundary does not differ between native languages or between voiced and devoiced conditions.

### 2.4.3. Discrimination sensitivity

The discrimination sensitivity of singleton and geminate stops was obtained as the gradient of the fitted logistic function at the perceptual boundary. That is, the differential coefficient of the function was calculated at the closure duration that gave a 50% response rate. Figure 3 shows the gradient at the perceptual boundary. Taiwanese Mandarin speakers have a much lower gradient than Japanese speakers. As shown in Fig. 3, the ratios of the Taiwanese

### Table 2  Closure duration of the original speech item and stimulus continuum between singleton and geminate stops.

| Speaking rate | Vowel devoicing | Original speech item | Closure duration (ms) | Original speech item | Stimulus continuum | Step size |
|---------------|-----------------|----------------------|-----------------------|----------------------|-------------------|----------|
|               |                 |                      | Min.                  | Max.                 | Step size         |          |
| Fast          | V1              | [pitapita]           | 121                   | 25                   | 95                | 10       |
|               | —               | [pitapita]           | 130                   | 25                   | 95                | 10       |
| Normal        | V1              | [pitapita]           | 189                   | 40                   | 110               | 10       |
|               | —               | [pitapita]           | 227                   | 30                   | 100               | 10       |
| Fast          | V1              | [pikapika]           | 107                   | 15                   | 85                | 10       |
|               | —               | [pikapika]           | 93                    | 10                   | 80                | 10       |
| Normal        | V1              | [pikapika]           | 158                   | 35                   | 105               | 10       |
|               | —               | [pikapika]           | 131                   | 10                   | 94                | 12       |
| Fast          | V2              | [katekatei]          | 83                    | 10                   | 94                | 12       |
|               | —               | [katekatei]          | 98                    | 1                    | 71                | 10       |
| Normal        | V2              | [katekatei]          | 101                   | 10                   | 94                | 12       |
|               | —               | [katekatei]          | 101                   | 1                    | 78                | 11       |
| Fast          | V2              | [taptuntapun]        | 82                    | 30                   | 100               | 10       |
|               | —               | [taptuntapun]        | 75                    | 25                   | 95                | 10       |
| Normal        | V2              | [taptuntapun]        | 149                   | 35                   | 105               | 10       |
|               | —               | [taptuntapun]        | 150                   | 35                   | 105               | 10       |
Mandarin speakers’ gradient to the Japanese speakers’ gradient were in the range 0.13–0.18 under the voiced and devoiced conditions at fast and normal speaking rates.

For the gradient at the perceptual boundary, a three-way ANOVA was performed with the factors of “native language,” “speaking rate,” and “voicing/devoicing,” showing that the three-way and two-way interactions were not significant. Although the “speaking rate” factor was not significant, the “native language” factor $[F(1, 24) = 610.8, p < 0.001]$ and the “voicing/devoicing” factor $[F(1, 24) = 4.54, p < 0.05]$ were significant. Therefore, the gradient is significantly lower in Taiwanese Mandarin.

**Fig. 1** Geminate response ratio as a function of closure duration for each singleton-geminate stimulus continuum with devoiced and voiced items at fast and normal speaking rates. The eight panels in the first and second rows are for the V1 devoicing condition, while the eight panels in the third and fourth rows are for the V2 devoicing condition. The closed circles (●) and open circles (○) respectively correspond to Japanese and Taiwanese Mandarin speakers’ data. Solid and broken curves respectively represent the fitted logistic function for Japanese and Taiwanese Mandarin speakers’ data.
speakers than Japanese speakers, and significantly lower under the devoicing condition than the voicing condition. These results indicate that Taiwanese Mandarin speakers have lower discrimination sensitivity than Japanese speakers and that the devoiced items afford lower discrimination sensitivity than the voiced items.

3. DISCUSSION

The perceptual boundary and the discrimination sensitivity of singleton and geminate stops were investigated in Japanese and Taiwanese Mandarin speakers in this study. Regarding the perceptual boundary, Japanese speakers had a shorter closure duration at the fast speaking rate than at the normal speaking rate. This result suggests that Japanese speakers adjust the perceptual boundary according to the speaking rate to correctly distinguish singleton and geminate stops. This result agrees with previous studies [2,4,5] that reported speaking rate dependency for singleton and geminate perception.

Conversely, Taiwanese Mandarin speakers did not show such rate dependency. The results of ANOVA indicate that their closure durations did not differ between fast and normal speaking rates, which suggests that they have an almost constant perceptual boundary across speaking rates. Sonu et al. [17] showed that Korean speakers also used a constant criterion across speaking rates in the perception of singleton and geminate stops. Therefore, there might be a general tendency for non-native speakers to have a constant perceptual boundary regardless of speaking rate. If Taiwanese Mandarin speakers had a constant perceptual boundary, it would have caused problems at the fast speaking rate because their boundary would have been significantly different from that of Japanese speakers; i.e., the Taiwanese Mandarin speakers’ perceptual boundary was near 90 ms at the fast speaking rate, while the Japanese speakers’ perceptual boundary was about 50 ms (Fig. 2). Therefore, there is a shift of about 40 ms between the Taiwanese and Japanese speakers’ boundaries. With such a large shift range, Taiwanese Mandarin speakers may have misperceived geminate stops as singleton stops, whereas Japanese speakers are likely to have perceived the geminate stops properly.

This estimate would be correct if Taiwanese Mandarin speakers had the same high discrimination sensitivity as Japanese speakers. However, they had much lower discrimination sensitivity than Japanese speakers under all conditions (Fig. 3). Their sensitivity was only about 15% of the Japanese speakers’ sensitivity. This fact makes their misperception even worse. With such low discrimination sensitivity, misperception would have occurred over a wider range of closure duration than 50–90 ms. Consequently, Taiwanese Mandarin speakers may have misperceived geminate stops as singleton stops even at closure durations greater than 90 ms. In addition, Taiwanese Mandarin speakers may have misperceived singleton stops as geminate stops when closure durations were shorter than 50 ms, whereas Japanese speakers could still perceive correctly as a singleton stop.

The low discrimination sensitivity also causes a problem at a normal speaking rate. Although the results of ANOVA indicate that Taiwanese Mandarin speakers have almost the same perceptual boundary as Japanese speakers, the results do not mean that they have no problems in perceiving singleton and geminate stops at a normal speaking rate. Their low discrimination sensitivity causes misperception of a singleton stop as a geminate stop at shorter closure durations than the perceptual
boundary and also causes misperception of a geminate stop as a singleton stop at longer closure durations than the perceptual boundary.

For example, for [pitapita] in the panel in the most upper row and the rightmost column panel in Fig. 1, Japanese speakers perceived the geminate stop with a ratio of almost 1.0 at a 100-ms closure duration, whereas Taiwanese Mandarin speakers perceived the geminate stop with ratios of only about 0.7. That is, Taiwanese Mandarin speakers misperceived a geminate stop as a singleton stop with about a 30% error rate. At 30-ms closure duration, Japanese speakers perceived the geminate stop at almost a 0.0 ratio (i.e., they perceived singleton stop at almost a 1.0 ratio), whereas Taiwanese Mandarin speakers perceived the geminate stop at ratios of about 0.4. Namely, Taiwanese Mandarin speakers misperceived a singleton stop as a geminate stop with about a 40% error rate.

Taken together, we can say that Taiwanese Mandarin speakers’ perception of singleton and geminate stops is affected by two factors: the shift of the perceptual boundary and the low discrimination sensitivity. Some previous studies [e.g., 16, 19] suggested that these two factors affect non-native speakers’ perception, which agrees with the results of this study. Moreover, this study revealed the effects of the two factors more clearly and quantitatively than previous studies.

The difference between voiced and devoiced items did not affect the perceptual boundary in this study. However, the ANOVA results indicated a weak tendency for the devoiced items to have a longer closure duration than the voiced items. Therefore, we cannot completely deny the effect of vowel devoicing on the perceptual boundary. Further research is necessary on this point.

Discrimination sensitivity was significantly lower for the devoiced items than the voiced items. A possible cause of the lower sensitivity is that, since the devoiced vowel preceding or following the singleton and geminate stops cannot provide time information, the duration of the stops is more ambiguously perceived with devoiced vowels than voiced vowels that can provide clear time information. This ambiguous perception leads to lower sensitivity in the distinction of the stops.

Note that the results of ANOVA for discrimination sensitivity indicated no interactions between “native language” and “voicing/devoicing” factors, and the results also indicated that the main factor of “voicing/devoicing” was significant. Namely, a lower discrimination sensitivity in devoiced items than voiced items was found in both Japanese and Taiwanese Mandarin speakers. Therefore, the lower discrimination sensitivity in devoiced items is language independent. This result suggests that the lower discrimination sensitivity with devoiced items is related to the sensory processing of hearing information but not to the cognitive processing of language information. Further investigation is necessary to confirm this hypothesis.

This study did not treat individual differences in the Japanese proficiency of Taiwanese Mandarin speakers. It was hard to analyze individual effects in this study because the number of participants was not controlled between proficiency levels. However, it is expected that individuals with high Japanese proficiency may have a better ability than those with low Japanese proficiency in perceiving singleton and geminate stops. Previous studies supported this expectation. For example, using English speakers at the beginner, intermediate, and advanced levels of Japanese, Hardison and Saigo [26] found that the identification ratio of the singleton and geminate consonants improves with their Japanese proficiency level. Namely, the higher their proficiency is, the more accurately they can perceive singleton and geminate consonants.

The same tendency is expected for Taiwanese Mandarin speakers. Individuals with high Japanese proficiency may be good at discriminating singleton and geminate stops. However, what remains unclear is the process whereby speakers’ perception ability improves. In the explanation of the improvement process, both abilities for the perceptual boundary and discrimination sensitivity should be considered. Which ability is more crucial for improvement? Are these abilities acquired similarly, or is one acquired earlier and the other later? These points should be examined in a future study.

Only voiceless stops ([p], [t], and [k]) were used for singleton and geminates consonants in this study. However, singleton and geminates consonants can be contrasted with other phonemes such as affricates ([ts] and [tc]) and fricatives ([s] and [ʃ]) as well as with voiced consonants such as [b], [d], [g], [ʣ], and [ʣ], although the gemination of these voiced consonants is not frequent. These other types of consonants should be examined in a future study.

The perception of singleton and geminate consonants is difficult not only for Taiwanese Mandarin speakers but also for other non-native speakers such as English, French, Korean, Thai, and Vietnamese speakers. It would be worthwhile to examine whether the perception characteristics obtained in this study are specific to Taiwanese Mandarin speakers or are generally found in other non-native speakers. If the lower discrimination sensitivity in devoicing items relates to the sensory processing of hearing information, no language dependency is expected. In addition, if the perceptual boundary relates to the cognitive processing of language information, there might be language-dependent effects. These hypotheses should be verified in the experiments using participants in various languages.

In summary, this study clarified significant characteristics of Taiwanese Mandarin speakers’ perceptual boun-
dary and discrimination sensitivity of Japanese singleton and geminate stops. 1) Their perceptual boundary was almost constant at fast and normal speaking rates, suggesting that they cannot adapt the boundary to variations in speaking rate. This constant perceptual boundary is one of the factors for their misperception of singleton and geminate stops. 2) Their discrimination sensitivity was only about 15% of Japanese speakers’ sensitivity. This low sensitivity is another factor for their misperception of singleton and geminate stops. 3) Discrimination sensitivity decreased for vowel-devoiced items in both Japanese and Taiwanese Mandarin speakers, suggesting that the effect of vowel devoicing might reflect language-independent hearing processing rather than language-dependent speech processing. These results suggest that multiple factors play a role in non-native speakers’ perception of singleton and geminate stops.

There are many Japanese learners in Taiwan. In 2018, 170,159 Taiwan Mandarin speakers learned the Japanese language [27]. For these Japanese learners, this study may provide beneficial knowledge about their tendency in perceiving Japanese singleton and geminate stops, which would help them to learn to speak the Japanese language.

ACKNOWLEDGMENTS

This study was supported by JSPS KAKENHI Grant Numbers JP22320081, JP24652087, JP25284080, JP26370464, and JP17K02705. We would like to thank Professors Lai Jhih Wei and Akiyoshi Yaotani of Tzu Chi University in Hualien City for their assistance in the experiment.

REFERENCES

[1] I. Kawagoe, “The phonology of sokuon, or geminate obstruents,” in *Handbook of Japanese Phonetics and Phonology*, H. Kubozono, Ed. (De Gruyter Mouton, Boston, 2015), pp. 79–119.

[2] H. Fujisaki and M. Sugito, “Physical properties of speech sounds (in Japanese),” in *Iwanami Japanese series 5: Phonology*, S. Ohno and T. Shibata, Eds. (Iwanami, Tokyo, 1977), pp. 63–106.

[3] S. Kawahara, “The phonetics of obstruct geminates, sokuon,” in *The Mouton Handbook of Japanese Language and Linguistics*, H. Kubozono, Ed. (De Gruyter Mouton, Berlin, 2015), pp. 43–77.

[4] S. Amano and Y. Hirata, “Perception and production boundaries between single and geminate stops in Japanese,” *J. Acoust. Soc. Am.*, 128, 2049–2058 (2010).

[5] S. Amano and Y. Hirata, “Perception and production of singleton and geminate stops in Japanese: Implications for the theory of acoustic invariance,” *Phoneticia*, 72, 43–60 (2015).

[6] Y. Hirata, “Perception of geminated stops in Japanese word and sentence levels (in Japanese),” *J. Phon. Soc. Jpn.*, 194, 23–28 (1990).

[7] E. Ofuka, Y. Morii and S. Kiritani, “Perception of a Japanese geminate stop: The effect of the duration of the preceding/following vowel (in Japanese),” *J. Phon. Soc. Jpn.*, 9, 70–76 (2005).

[8] K. Idemaru and S. Guion-Anderson “Relational timing in the production and perception of Japanese singleton and geminate stops,” *Phonetica*, 67, 25–46 (2010).

[9] Y. Sukekawa, “Utterance tendency of non-native Japanese speakers: Results of questionnaire survey (in Japanese),” *Japanese Speech and Education, Research Report of Grant-in-Aid for Scientific Research on Priority Areas by Ministry of Education, Science and Culture*, pp. 187–222 (1993).

[10] T. Toda, “Acquisition of special morae in Japanese as a second language (in Japanese),” *J. Phon. Soc. Jpn.*, 7, 70–83 (2003).

[11] M. Kondo, “Language-specific and language-universal L1 interference on a phonological acquisition of Japanese learners: Results of questionnaires to Japanese teachers (in Japanese),” *Bull. Grad. Div. Lett. Arts Sci. Waseda Univ.*, III 57, pp. 21–34 (2012).

[12] Y. Hirata, “Perception of geminated stops in Japanese word and sentence levels by English-speaking learners of Japanese language (in Japanese),” *J. Phon. Soc. Jpn.*, 195, 4–10 (1990).

[13] R. Hayes-Harb and K. Masuda, “Development of the ability to lexically encode novel second language phonemic contrasts,” *Second Lang. Res.*, 24, 5–33 (2008).

[14] J. Kingdon, S. Kawahara, D. Chambless, D. Mash and E. Brenner-Also, “Contextual effects on the perception of duration,” *J. Phon.*, 37, 297–320 (2009).

[15] K. Tsukada, F. Cox, J. Hajek and Y. Hirata, “Non-native Japanese learners’ perception of consonant length in Japanese and Italian,” *Second Lang. Res.*, 34, 179–200 (2018).

[16] R. Oba, A. Braun and J. Handke, “The perception of Japanese geminates by native and non-native listeners,” *Phonetica*, 92, 9–29 (2005).

[17] M. Sonu, H. Kato, K. Tajima, R. Akahane-Yamada and Y. Sagisaka, “Non-native perception and learning of the phonemic length contrast in spoken Japanese: Training Korean listeners using words with geminate and singleton phonemes,” *J. East Asian Linguist.*, 22, 373–398 (2013).

[18] T. Uchida, “Characteristics of auditory cognition of long vowels and double consonants for Chinese students in learning Japanese language (in Japanese),” *Jpn. J. Educ. Psychol.*, 41, 414–423 (1993).

[19] H. Hung, “Perceptions of Japanese geminate stops among Taiwanese learners of Japanese (in Japanese),” *J. Phon. Soc. Jpn.*, 16, 15–27 (2012).

[20] M. Fujimoto, “Vowel devoicing,” in *Handbook of Japanese Phonetics and Phonology*, H. Kubozono, Ed. (De Gruyter Mouton, Boston, 2015), pp. 167–214.

[21] K. Maekawa and H. Kikuchi, “Corpus-based analysis of vowel devoicing in spontaneous Japanese: An interim report,” in *Voicing in Japanese*, J. Weijer, K. Nanjo and T. Nishihara, Eds. (De Gruyter Mouton, Berlin, 2008), pp. 205–228.

[22] M. Han, “Acoustic manifestations of mora timing in Japanese,” *J. Acoust. Soc. Am.*, 96, 73–82 (1994).

[23] T. J. Vance, *The Sounds of Japanese* (Cambridge University Press, Cambridge, 2008).

[24] S. Haraguchi, “Pitch accent and intonation in Japanese,” in *Autosegmental Studies on Pitch Accent*, H. Hulst and N. Smith, Eds. (De Gruyter Mouton, Berlin, 2010), pp. 123–150.

[25] 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).

[26] D. M. Hardison and M. M. Saigo, “Development of perception of second language Japanese geminates: Role of duration, sonority, and segmentation strategy,” *Appl. Psycholinguist.*, 31, 81–99 (2010).

[27] The Japan Foundation, *Survey report on Japanese-language education abroad 2018* (2020).