Development of perception of second language Japanese geminates: Role of duration, sonority, and segmentation strategy

DEBRA M. HARDISON
Michigan State University

MIKI MOTOHASHI SAIGO
Kansai Gaidai University

Received: December 12, 2007 Accepted for publication: May 4, 2009

ADDRESS FOR CORRESPONDENCE
Debra M. Hardison, Michigan State University, Department of Linguistics and Languages, A-714 Wells Hall, East Lansing, MI 48824. E-mail: hardiso2@msu.edu

ABSTRACT
This study explored the second language perceptual accuracy of Japanese geminates (moraic units) by English-speaking learners at three proficiency levels: beginner (28), low–intermediate (42), and advanced (15). Stimuli included singleton and geminate /t/, /k/, /s/ followed by /a/ or /u/ produced by a native speaker in isolated words and carrier sentences. Results of a forced-choice identification task revealed a significant complex interaction involving all factors. Simple effects tests indicated scores for geminate /s/ followed by /u/ in carrier sentences were significantly lower across proficiency levels. The singleton-geminate contrast for /t/, /k/ was significant for beginning learners, marginal for low–intermediate, and nonsignificant for advanced. Duration differences did not account for all findings. Greater consonant–vowel sonority difference facilitated perception, enhancing mora boundaries for speech segmentation.

One of the basic challenges in second language (L2) speech is the perception of nonnative contrasts. Among temporal phenomena, cross-language differences in the duration parameters of voice onset time (VOT) have been extensively studied in perception and production (e.g., Flege, 1991; Flege & Eefting, 1988; Flege, Schirru, & MacKay, 2003; McClaskey, Pisoni, & Carrell, 1983). Another challenge, also characterized as an issue of duration (e.g., Hayes, 2002), is the perception of geminates in Japanese. In that language, singleton consonants contrast with their longer geminate counterparts. Accurate perception and production of geminates are important in Japanese, and particularly difficult for L2 learners for whom segment duration is not a contrastive feature in their first language (L1; Han, 1992). The question arises as to whether there are contextual factors
in addition to segmental duration that influence L2 perception of geminates and how such influence might change with linguistic experience.

The geminate, as a type of mora, is a unit of timing within the larger context of words. Both morae and syllables can be represented in a single phonological structure in which the mora (µ) takes the place of the rhyme as a constituent of the syllable (σ; e.g., Haraguchi 1996; Hayes, 1989; Kubozono, 1999). As shown in Figure 1, unlike consonant–vowel morae, special morae such as geminates (Figure 1a) and the second half of long vowels (Figure 1c) do not constitute syllables by themselves, and are not accent-bearing units (Kubozono, 1999, Tsubjimura, 1996). The structure in Figure 1a represents the Japanese word kite “postage stamp,” which contains a geminate /t/, three morae, and two syllables. This contrasts with kite “coming” in Figure 1b, which has a singleton /t/, and kiite “listening” in Figure 1c, which has a long vowel /i/. Syllables that include geminate consonants are closed syllables, representing exceptions to the basic Japanese open syllable structure, which consists of an optional onset consonant plus a vowel ([C]V). In the case of geminates (Figure 1a), there is an obstruent coda in the first syllable, followed by what Shibatani (1990) describes as a homorganic consonant onset in the next syllable. Hayes (1989) refers to this as the “consonant melody flopped onto a following vowel-initial syllable” (p. 258).

Acoustic cues to geminate perception for native speakers of Japanese include stop closure duration (e.g., Fukui, 1978) and pitch accent, which is also contrastive. Native speakers require a longer closure duration to perceive a geminate in a low–high ([LH] vs. a high–low [HL]) pitch accent pattern (Ofuka, 2003). In addition, the duration of the preconsonantal vowel is a cue to consonantal length as vowels are longer before geminates than before singletons (e.g., Han, 1994). Hirata (1990a) found listeners perceived a singleton when the ratio of the duration of the preceding vowel and that of the stop closure was low, and a geminate when the ratio was higher. Although postconsonantal vowel duration does not appear to have a significant effect on perceptual accuracy of geminates (Hirato & Watanabe, 1987), misperception of a geminate as a singleton can occur when postconsonantal elements in a carrier sentence are spoken slowly (Hirata, 1990a). In sum, for native speakers, the acoustic cues to perception of the singleton-geminate difference include the duration of the stop closure, duration of the preceding vowel, and rate of speech of the postconsonantal elements of an utterance.

Figure 1. The phonological structures for the geminate, singleton, and long vowel: (a) kitte “postage stamp” (geminate /t/), (b) kite “coming” (singleton /t/), and (c) kiite “listening” (long vowel /ii/). σ, syllable; µ, mora.
Although native speakers tend to show categorical boundaries in the discrimination of singleton versus geminate stops, nonnative data suggest a more continuous pattern (e.g., Hirata, 1990b; Min, 1987). Korean and Chinese speakers learning Japanese in Minagawa and Kiritan’s (1996) study, and Thai speakers in Minagawa’s (1996) study showed effects of pitch accent type on singleton and geminate identification. Learners whose L1 was Korean or Chinese made significantly more errors involving misperceptions of singletons as geminates (vs. geminates as singletons) in HL accent contexts; however, there was no difference in error types in LH contexts. In contrast, L1 English and Spanish learners of Japanese showed no significant difference in errors according to pitch accent pattern. Acoustic measurements revealed the average postconsonantal vowel in the HL contexts was significantly shorter than that in the LH contexts, suggesting the ratio of consonant closure duration to postconsonantal vowel duration may be a possible acoustic influence.

Some findings suggested L1 English learners of Japanese have some perceptual capability with regard to determining the number of morae, but have difficulty attributing duration to the correct segment. When learners were asked to spell English loanwords in Japanese (using Roman letters) produced by a Japanese speaker, their errors generally involved lengthening the vowel in words that native listeners spelled with geminate consonants (Yamagata & Preston, 1999). In fact, perceptual accuracy of long vowels in Japanese by English speakers does not appear to be as problematic as identification of geminates. Motohashi Saigo (2007) found that beginning, intermediate, and advanced learners identified long vowels (e.g., kaaete) in nonwords presented in isolation and in short sentences with 85 to 95% accuracy in a forced-choice task involving minimal triplets (i.e., singleton, geminate, long vowel).

In contrast to most of the research on L2 perception of geminates, Hayes (2002) included the fricative /s/ in her study along with the stops /t/ and /k/. She hypothesized that learners would have less difficulty perceiving the singleton-geminate contrast for /t/ based on the greater durational difference (180 ms) in her stimuli compared to the difference for singleton-geminate /s/ (134 ms) and /k/ (142 ms). Participants in that study were a total of 54 college students in the United States at several levels of Japanese study: first semester, second semester, and a group of combined third and fourth semester students. No details were given regarding their language learning backgrounds or any study abroad participation. Stimuli were 120 pairs of nonwords (60 involving the same token, 60 involving different ones) produced by a native speaker of Japanese involving singleton-geminate pairs for /s/, /t/, and /k/. The task was to determine if the tokens in each pair were the same or different. Results across learner groups indicated that accuracy in discriminating the singleton-geminate contrast was significantly better for /t/ compared to /k/ and to /s/, with no significant difference between /k/ and /s/. The only significant difference between learner groups was found for the more advanced learners compared to a control group.

Although the issue of durational difference has dominated research in this area, other factors related to phonetic context may also be involved. The perceptual distance between the geminate and following vowel may influence perception of the consonant’s acoustic cues, such as its duration, and its identification as
a geminate. This suggests a possible role for the sonority difference between the consonant and following vowel in establishing perceptual distance and mora boundary salience. The greater the distance, the greater the strength of a segment’s cues in perception (Wright, 2004). Sonority is considered an inherent property of a segment, traditionally based on stricture. Generally speaking, the wider the openness of the vocal tract (the smaller the degree of stricture), the greater the energy in the acoustic signal, and the higher the sonority value. This value among vowels is inversely related to height. Of the five vowels in Japanese (i.e., /a, i, e, o, u/), two were selected for the current study: /a/ (low vowel), which has the highest sonority value, and /u/ (high vowel), which has the lowest value among the vowels. In the Tokyo dialect, /a/ is also the longest vowel, and /u/ is the shortest (Yoshida, 2006).

Segments may be classified from highest to lowest in sonority (i.e., a sonority hierarchy) as follows: vowels > glides > liquids > nasals > fricatives > stops. The current study involved singleton and geminate stops /t/ and /k/ and the fricative /s/. A discussion of sonority difference between segments requires quantification; for example, in Selkirk’s (1984) sonority index, /a/ is given a value of 10, /u/ is 8, the fricative /s/ is 4, and the stops are each 0.5. Sonority is involved in the optimal sequencing of segments within a syllable to take advantage of the perceptual robustness of acoustic cues (Wright, 2004). Sonority differences between segments create perceptual distance and affect the salience of boundaries between morae.

Selective attention to the mora is important for perception. The mora plays a role in the temporal organization of speech in production (e.g., Kubozono, 1999), and the segmentation of speech sounds in perception (e.g., Cutler & Otake, 1994; Otake, Hatano, Cutler, & Mehler, 1993). For example, in a monitoring task, native listeners were able to detect speech sound targets more easily when they corresponded to a mora than when they did not (e.g., Cutler & Otake, 1994; Otake et al., 1993). In contrast, native speakers of English use stressed syllables to segment English sound sequences (e.g., Cutler & Norris, 1988; Cutler, Mehler, Norris, & Segui, 1986). Given that in the early stages of L2 acquisition, adult learners rely on those cues pertinent to their L1 (Strange & Shafer, 2008), L1 English learners of L2 Japanese likely focus attention on the syllabic level for segmentation based on its effectiveness in English, especially in connected speech. This pattern of segmentation assigns a consonant to the onset position of a syllable with a corresponding lengthening of the vowel in the preceding open syllable. Such a strategy is compatible with the misperception of the duration of the geminate in Japanese as a feature of the preconsonantal vowel (Yamagata & Preston, 1999).

In the current study, the roles played by duration, sonority, and segmentation strategy were explored, focusing on the influence of consonant type (i.e., singleton and geminate /t/, /k/, /s/), phonetic context (i.e., postconsonantal /a/, /u/), and presentation condition (i.e., word in isolation vs. a brief carrier sentence) on geminate perception by L1 American English learners of Japanese at three levels of proficiency. The research was guided by the following questions:

1. Can differences in duration (singleton vs. geminate) involving medial consonants account for all findings? If not, is there evidence of phonetic context effects?
Specifically, does the sonority difference between a consonant and following vowel influence perception of the consonant’s duration? The sonority difference, and thus, perceptual distance is greater in the case of a stop-vowel (e.g., *sakka* “last summer”) compared to a fricative-vowel sequence (e.g., *sassa* “quickly”). This suggests identification accuracy would be lowest for the sequence of geminate */s/ + */u/**, which has the smallest consonant-vowel sonority difference, and highest for geminate */t/** or */k/** + */a/**, which has the largest difference.

2. Is there evidence of acquisition of the singleton-geminate durational contrast across proficiency levels? We hypothesized the most advanced of the three levels of learners would show greater accuracy in geminate identification because of their greater exposure to, and interaction with, the target language, especially through participation in a study abroad program in Japan. This could contribute to modification of the perceptual system through a shift in attention to the critical cues in the L2 (e.g., Hardison, 2003; Lively, Logan, & Pisoni, 1993).

3. Do learner error patterns suggest a segmentation strategy? L2 learners who fail to perceive the consonant duration accurately may misperceive the geminate as a singleton (see Figure 1b). If duration is perceived, it may be inaccurately attributed to the preceding vowel (see Figure 1c), as reported by Yamagata and Preston (1999). This latter option would be compatible with vowel lengthening in an open syllable in English followed by a consonant singleton onset in the following syllable.

4. Are learners more accurate when the geminate occurs in a word/nonword produced in isolation compared to a carrier sentence? Previous research on L2 geminate perception (e.g., Hayes, 2002; Minagawa, 1996) used words and nonwords produced in isolation; however, perception of morae may be more difficult when the stimuli are embedded in longer stretches of speech that are more challenging to segmentation. Research has shown that lexical status (i.e., word vs. nonword) does not have a significant effect on L2 perception in a forced-choice identification task (e.g., Lively et al., 1993).

The materials and task in the current study were designed to elicit perceptual segmentation based on rhythmic units and/or their boundaries in contrast to segmentation for the purposes of activation and selection of competing word candidates in lexical recognition (Cutler & Otake, 2002). This allowed learners to focus attention on acoustic–phonetic information without engaging higher order cognitive processes. A forced-choice identification task instead of a same-different discrimination task (e.g., Hayes, 2002) was chosen as more ecologically valid to reflect categorization behavior. The between-subjects variable was group (three levels of learner proficiency); within-subjects variables were consonant type (*/s/**, */t/**, */k/**), postconsonantal vowel (*/a/**, */u/**), and stimulus condition (isolated word, carrier sentence).

Method

Participants

Participants were undergraduates at a large university in the United States, native speakers of American English, and ranged in age from 19 to 22. There were no
heritage learners of Japanese. Students from three Japanese courses volunteered to participate. There were 28 students (21 male, 7 female) from the beginner level (first semester of study), 42 (25 male, 17 female) from the low–intermediate (third semester), and 15 (10 male, 5 female) from the advanced level (seventh semester). All courses were taught by native speakers of Japanese. The beginning learners had been studying Japanese for about 3 months prior to data collection. All participants from the advanced level had studied in Japan for one or two semesters, where they were involved in conversational interactions with native Japanese speakers. The courses for all levels included communicative activities. In this program, students are made aware of geminates because of their contrastive role in the language; however, there is no specific instruction in or practice involving the perception or production of geminates. Instructors sometimes correct students’ inaccurate pronunciation, but this is not the focus of the curriculum.

Materials

Materials involved 30 word tokens (see Appendix A): 24 test stimuli and 6 fillers. Of the 24 test stimuli, half began with /a/ (e.g., akka) and half with /sa/ (e.g., sakka). The medial consonants were /s/, /t/, and /k/, followed by /a/ and /u/. These segments were chosen based on their sonority values. Test stimuli included a balanced number of singletons (bisyllabic and bimoraic) and their geminate counterparts (bisyllabic and trimoraic). Most were real words of relatively low pedagogical frequency; for example, itta would be more commonly used to express “went” than satta. Six nonwords were used to complete the geminate-singleton contrasts. Response options for participants were minimal triplets such as sasu (singleton) “stab,” sassu (geminate) “guess,” and saasu (long vowel), a nonword. The use of minimal triplets was based on findings of a pilot study, and those of Yamagata and Preston (1999), indicating learner errors in geminate perception. The filler items were bisyllabic and trimoraic nonwords, each of which contained a long vowel (e.g., kaate), and were used to avoid response bias. Additional stimuli with long vowels were not included in the study because of the amount of time participants were available and concerns for fatigue. The pitch accent pattern for all stimuli was HL(L).

Stimuli were recorded by a female native speaker of Japanese (Tokyo dialect) using a SONY minidisc recorder/player with a SONY unidirectional microphone. In the first recording, each word was produced in isolation. In the second recording, the word was embedded in one of the following carrier sentences: watashi wa _____ toimashita “I said ____” or kore wa _____ desu “This is ____.” For each recording, the speaker was shown the word/sentence on a card, asked to produce it with a natural speaking rate, and to maintain the same tempo throughout the session. One of the authors monitored the consistency of the speech rate by watching a digital metronome (flashing lights on a computer screen). Following the recordings, the waveforms for all test items containing /u/ were examined to ensure that the vowel had not devoiced. In the Tokyo dialect, /u/ can be devoiced when it follows a voiceless consonant and either precedes a voiceless consonant or occurs in word-final position, unless the vowel is in a position to receive an accent (Tsujimura, 1996). Stimuli were tested for auditory intelligibility by native speakers of Japanese, and then were randomized for presentation. Each stimulus
was presented four times in each of the above two conditions. Total stimuli were 240 (4 × 30 × 2 conditions).

**Stimulus durations.** The duration of each of the following segments was calculated for each stimulus: preconsonantal vowel, medial consonant (i.e., stop closure for singleton and geminate /t/, /k/ and period of frication for singleton and geminate /s/), and postconsonantal vowel (see Table 1). Duration was measured in the expanded waveform display using the SUGI Speech Analyzer by Animo, and confirmed with synchronized spectrogram display. Standard segmentation procedures were applied (e.g., Johnson, 1997; Ladefoged, 2003). In measuring the duration of a vowel, the beginning was marked at the onset of complex voicing with higher frequency components, and the end was marked at the offset of these components. The beginning of the stops was marked at the offset of all higher frequency components showing the mouth was closed, and the end was marked at the release burst. The beginning and end of the fricative were marked at the onset and offset of clear frication noise, respectively. Following previous studies, VOT was not included in the duration measurements. The consonants in the present study were in medial position where VOT is generally smaller than in initial position (Han, 1992). Japanese stops are considered weakly aspirated or unaspirated even in initial position. Han (1992) further notes that VOT does not play a significant role in the perception of the geminate-singleton contrast.

As shown in Table 1, the duration of the preconsonantal vowel covaried with that of the following consonant as noted in previous studies (Han, 1994); therefore, vowels were longer preceding geminates compared to singletons. Measurements also confirmed the consistently shorter duration for postconsonantal /u/ versus /a/ (Yoshida, 2006). The duration of all singletons was shorter than that of their geminate counterparts. The singleton-geminate duration ratios in the isolated word condition ranged from 1.7:1.0 for /s/ to 2.3:1.0 for /t/. These fall below the mean ratio reported by Han (1992) of 2.8:1.0 for the stops /t/ and /k/ (no data for fricatives). In Han’s study, the ratios varied from 2.5:1.0 to 3.2:1.0 because of differences across speakers and phonetic contexts.

**Procedure**

Data collection was carried out in sound-attenuated rooms. Stimuli were presented auditorily via centrally located ceiling-mounted speakers. Participants were familiarized with the talker’s voice and the carrier sentences. For each stimulus, there were three options or minimal triplets (e.g., sasu, singleton; sassu, geminate; saasu, long vowel) on response sheets. Participants were instructed to circle the response that matched what they heard (isolated word condition), or what the speaker was introducing (carrier sentence condition). They were told they might not recognize all the target words they would hear, and the sentences served only a carrier function. Interstimulus interval was determined through pilot testing and set at 3 s. A forced-choice versus an open-response task was selected to avoid the problem of determining whether the learners’ spelling was an accurate reflection of their perception. Options were written in Roman letters, with which the learners were familiar. The textbook at the beginning level of Japanese study presents Roman letters along with Japanese characters (hiragana), using two letters for
Table 1. Duration of singletons, geminates, preconsonantal vowels, and postconsonantal vowels (ms)

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>t</th>
<th>a</th>
<th>a</th>
<th>tt</th>
<th>a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isolated word</td>
<td>68</td>
<td>90</td>
<td>144</td>
<td>76</td>
<td>210</td>
<td>139</td>
</tr>
<tr>
<td>Carrier sentence</td>
<td>61</td>
<td>87</td>
<td>142</td>
<td>71</td>
<td>197</td>
<td>137</td>
</tr>
</tbody>
</table>

Singleton-geminate duration difference in isolated word = 120, ratio = 2.3:1.0
Singleton-geminate duration difference in sentence condition = 110, ratio = 2.3:1.0

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>t</th>
<th>u</th>
<th>a</th>
<th>tt</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isolated word</td>
<td>72</td>
<td>100</td>
<td>120</td>
<td>94</td>
<td>208</td>
<td>114</td>
</tr>
<tr>
<td>Carrier sentence</td>
<td>78</td>
<td>109</td>
<td>115</td>
<td>92</td>
<td>204</td>
<td>105</td>
</tr>
</tbody>
</table>

Singleton-geminate duration difference in isolated word = 108, ratio = 2.1:1.0
Singleton-geminate duration difference in sentence condition = 95, ratio = 1.9:1.0

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>k</th>
<th>a</th>
<th>a</th>
<th>kk</th>
<th>a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isolated word</td>
<td>69</td>
<td>94</td>
<td>127</td>
<td>94</td>
<td>196</td>
<td>119</td>
</tr>
<tr>
<td>Carrier sentence</td>
<td>63</td>
<td>93</td>
<td>124</td>
<td>97</td>
<td>198</td>
<td>121</td>
</tr>
</tbody>
</table>

Singleton-geminate duration difference in isolated word = 102, ratio = 2.1:1.0
Singleton-geminate duration difference in sentence condition = 105, ratio = 2.1:1.0

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>k</th>
<th>u</th>
<th>a</th>
<th>kk</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isolated word</td>
<td>75</td>
<td>97</td>
<td>122</td>
<td>96</td>
<td>212</td>
<td>88</td>
</tr>
<tr>
<td>Carrier sentence</td>
<td>68</td>
<td>94</td>
<td>119</td>
<td>96</td>
<td>207</td>
<td>84</td>
</tr>
</tbody>
</table>

Singleton-geminate duration difference in isolated word = 110, ratio = 2.2:1.0
Singleton-geminate duration difference in sentence condition = 113, ratio = 2.2:1.0

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>s</th>
<th>a</th>
<th>a</th>
<th>ss</th>
<th>a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isolated word</td>
<td>78</td>
<td>111</td>
<td>144</td>
<td>102</td>
<td>212</td>
<td>115</td>
</tr>
<tr>
<td>Carrier sentence</td>
<td>72</td>
<td>106</td>
<td>136</td>
<td>99</td>
<td>210</td>
<td>113</td>
</tr>
</tbody>
</table>

Singleton-geminate duration difference in isolated word = 101, ratio = 1.9:1.0
Singleton-geminate duration difference in sentence condition = 104, ratio = 2.0:1.0

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>s</th>
<th>u</th>
<th>a</th>
<th>ss</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isolated word</td>
<td>81</td>
<td>135</td>
<td>125</td>
<td>100</td>
<td>223</td>
<td>87</td>
</tr>
<tr>
<td>Carrier sentence</td>
<td>80</td>
<td>142</td>
<td>116</td>
<td>96</td>
<td>225</td>
<td>84</td>
</tr>
</tbody>
</table>

Singleton-geminate duration difference in isolated word = 88, ratio = 1.7:1.0
Singleton-geminate duration difference in sentence condition = 83, ratio = 1.6:1.0

*Note: Measurements are for stimuli beginning with /sa/ produced in the isolated word and carrier sentence conditions.*
geminates (e.g., sas-su) and long vowels (e.g., saasu). The carrier sentence condition was presented first, followed by the isolated word condition. The tasks were completed on the same day with a break in between.

RESULTS AND DISCUSSION

In this section, the discussion of the results of the perceptual identification task is followed by the analysis of error patterns.

Perceptual identification accuracy

The number of stimuli correctly identified was tabulated for each of the three proficiency levels. To ensure that the lexical status of the stimuli was not a factor, the accuracy data for the real words were compared to the data for the nonwords. A Wilcoxon Signed Ranks Test revealed no significant difference ($Z = -0.854, p > .05$). A comparison was also made between the accuracy data for the stimuli beginning with /sa/ and those beginning with /a/. Results revealed no significant difference; therefore, the data were combined and submitted to a mixed design analysis of variance (ANOVA). The between-subjects factor was group, and the following were within-subjects factors: condition (isolated word, carrier sentence), duration (singleton, geminate), consonant (/t/, /k/, /s/), and vowel (/a/, /u/). All factors showed significant main effects; however, all were also subsumed under a significant complex Group $\times$ Condition $\times$ Duration $\times$ Consonant $\times$ Vowel interaction, $F(4, 164) = 2.87, p < .05, \eta^2_p = .07$. Further in-depth analyses were carried out within each proficiency level. Results are shown in Figures 2–4 for the geminates and singletons with both postconsonantal vowels and in both conditions. Because the stops /t/ and /k/ showed a similar pattern of findings across all groups, their data are combined.

Beginner level. Data for the beginning learners were submitted to a four-factor within-subjects ANOVA. Analysis revealed a significant Condition $\times$ Duration $\times$ Consonant $\times$ Vowel interaction, $F(2, 54) = 5.05, p = .01, \eta^2_p = .16$. As shown in Figure 2a–d, scores were lower for geminates and singletons presented in carrier sentences compared to isolated words; however, the consonant-vowel combinations were differentially affected, especially as geminates. In the word condition (Figure 2a), mean accuracy declined from an average of 77.8% for the geminate stops followed by /a/ (e.g., sakka) to 52.5% for geminate /s/ followed by /u/ (e.g., sassu). In contrast, Figure 2b shows the mean accuracy for the sentence condition, decreasing from an average of 72% for the geminate stops with postconsonantal /a/ to 19.5% for geminate /s/ with /u/. Simple effects tests revealed the duration factor significantly affected identification accuracy for the stops, $F(1, 27) = 4.56, p < .05$ (e.g., /k/). The geminate-singleton contrast was also significant for /s/, with lower scores for the geminate, $F(1, 27) = 11.97, p < .01$. In addition, vowel type influenced scores for /s/, which were lower with postconsonantal /a/ (vs. /a/), $F(1, 27) = 10.86, p < .001$. Stimulus condition affected perceptual accuracy of /s/, which was lower in sentences versus isolated words, $F(1, 27) = 11.46, p < .01$. In sum, although the effect of duration was
significant for all consonants, these findings also point to geminate /s/ followed by /u/, especially in carrier sentences, as a significant area of difficulty for these learners.

**Low–intermediate level.** A four-factor within-subjects ANOVA revealed a significant Condition × Duration × Consonant × Vowel interaction, $F(2, 82) = 3.559$, $p < .05$. Simple effects tests revealed vowel type influenced perceptual accuracy of /s/, with lower scores for postconsonantal /u/, $F(1, 41) = 25.5$, $p < .001$. Scores for /s/ were also lower in the carrier sentence condition, $F(1, 27) = 38.79,$
Figure 3. The identification accuracy by low–intermediate learners: the percentage correct by consonant (/t/ and /k/ combined, and /s/) and vowel (/u/ and /u/) for geminates in (a) isolated words and (b) carrier sentences, and singletons in (c) isolated words and (d) carrier sentences. The error bars enclose ±1.0 SE. [A color version of this figure can be viewed online at journals.cambridge.org/aps]

$p < .001$, as were scores for stimuli with postconsonantal /u/, $F (1, 41) = 4.77$, $p < .05$. Although identification accuracy for geminate /s/ was significantly lower than for its singleton counterpart, $F (1, 41) = 14.05$, $p = .001$, this duration difference was marginally significant for the stops (e.g., /k/), $F (1, 41) = 4.04$, $p = .05$. As shown in Figure 3a–d, there is greater variability in mean identification accuracy for /s/ with the lowest scores in the carrier sentence condition for the geminate followed by /u/ (26.5%). Similar to the beginning learners, these findings indicate geminate /s/ followed by /u/ (e.g., *sassu*) in carrier sentences was a significant area of difficulty for the low–intermediate learners.
Advanced level. A four-factor within-subjects ANOVA revealed a significant Condition × Duration × Consonant × Vowel interaction, $F(2, 28) = 3.33, p < .05$, $\eta^2_p = .19$. As shown in Figure 4a–d, identification accuracy for singletons in isolated words is slightly higher than that for the low–intermediate group, whose accuracy for these stimuli was at least 80%. The difference between the advanced learners and the two lower proficiency levels is most evident in terms of geminate identification. For advanced learners, perceptual accuracy was quite consistent across the consonant–vowel contexts for geminates and singletons in isolated words (see Figure 4a and c), and somewhat lower overall for singletons in carrier sentences shown in Figure 4d. However, there is greater variability for
geminates in the sentence condition as shown in Figure 4b. These observations were supported by results of simple effects tests, which revealed vowel type influenced identification accuracy for /s/, with lower scores for postconsonantal /u/, $F$ (1, 14) = 8.50, $p = .01$. The duration factor also influenced scores for /s/ with lower accuracy for the geminate, $F$ (1, 14) = 6.65, $p < .05$, as did the condition factor, with lower scores in carrier sentences, $F$ (1, 14) = 25.22, $p < .001$. Consistent with the findings for the beginner and low–intermediate groups, results demonstrate that identification accuracy for advanced learners was significantly lower for geminate /s/ followed by /u/ (e.g., sassu) in carrier sentences (34.5%) compared to isolated words (80%). In contrast to the two lower proficiency groups, /s/ was the only consonant for the advanced learners for which scores were significantly affected by the geminate-singleton duration difference.

**Error Patterns**

In the forced-choice task used in this study, when geminates were not accurately identified, misperceptions were either singletons or long vowels. For each level of proficiency, the type of error was tabulated, and converted to a percentage of the total errors for that level. As shown in Figure 5, the relative distribution of error types was consistent across levels of proficiency. The lowest percentage of errors at each level was for misperceptions of geminates as singletons in the isolated word condition (the first bar on the graph per group). The highest percentage of errors at each level was for misperception of geminates as long vowels in the carrier sentence condition (the fourth bar on the graph per group): 41% for the beginner group, 37% for low–intermediate, and 54% for advanced. This is
consistent with the results of Yamagata and Preston (1999) and observations in the classroom. The largest number of misperceptions of geminates as long vowels in both presentation conditions was for geminate /s/ followed by /u/ (e.g., *sassu*) for each learner group. Chi-square analysis explored whether there was a significant difference in the frequency with which a particular geminate–vowel sequence was misperceived as a singleton versus a long vowel. Results did not reach significance ($p = .05$) for any proficiency level or presentation condition.

General Discussion

The current study investigated factors affecting L2 learners’ perceptual development of Japanese geminates. The problems learners face have generally been attributed to the challenge of perceiving the longer duration that characterizes the geminate. For learners whose L1 is English, segment duration is not a contrastive feature. The first two research questions we posed in this study were whether duration alone could account for all the data, and whether differences existed across proficiency levels. Findings indicated the effect of duration was mediated by the type of consonant, postconsonantal vowel, stimulus condition, and learner proficiency. Identification accuracy scores for geminate /s/ followed by /u/ were consistently the lowest especially in carrier sentences, with modest increases in accuracy at each successive proficiency level. A series of simple effects tests revealed the initial significant interaction involving all factors, including group, could be attributed to the changing influence of the duration factor in relation to the others (condition, consonant, and vowel) as proficiency level increased. For beginning learners, the effect of duration was significant for all three consonants. At the low–intermediate level, the effect of duration was significant for /s/ and marginally significant ($p = .05$) for the stops /t/ and /k/. For the more advanced learners, the decline in geminate identification accuracy focused solely on /s/ followed by /u/ (e.g., *sassu*). Overall, findings suggested the need to consider factors in addition to segment duration.

Although geminate /s/ was the longest consonant in the stimuli (see Table 1), its identification accuracy was the lowest, especially when followed by /u/. A comparison of the stimuli *sassu* and *sassa* reveals the preconsonantal vowel durations differed by 2 ms (based on durations from the isolated word condition), the frication durations by 11 ms (longer in *sassu*), and the postconsonantal vowel durations by 28 ms (/u/ is shorter). Therefore, the question arises as to whether postconsonantal vowel duration might be a factor as suggested by previous studies (e.g., Minagawa, 1996). If postconsonantal vowel duration or the consonant–vowel duration ratio were a determining factor, this would make two predictions. First, it would predict a significant difference in perceptual accuracy for geminate /k/ compared to /t/ followed by /u/ (e.g., *sakku* and *sattu*). For this pair, the duration of preconsonantal /a/ differed by 2 ms, the duration of the geminates differed by 4 ms, and the duration of the postconsonantal vowel /u/ differed by 26 ms (longer in *sattu*). However, across learner groups, there was no significant difference in accuracy between geminate /k/ and /t/. There is also no difference between
these words in the sonority values of the consonant and following vowel. Second, the issue of duration would not have predicted the difference in identification accuracy found for stimuli involving geminate /k/ compared to /s/ followed by /u/ (e.g., sakku and sassu). The duration of preconsonantal /a/ differed by only 4 ms, the duration of postconsonantal /u/ differed by 1 ms, and the duration of the geminates differed by 11 ms. Specifically the geminate /s/ (e.g., sassu) was longer by 11 ms, but had the lowest accuracy.

The above findings are compatible with previous research (e.g., Hardison, 2003; Lively et al., 1993) demonstrating contextual variability in L2 perceptual identification accuracy. Consideration of the contextual factors in the current study suggests accuracy may have been influenced by the sonority relationship between the geminate and following vowel. Recall that larger sonority differences create greater perceptual distance and greater salience of acoustic cues, including cues to segment duration. The most marked differences in identification accuracy involved geminate /s/ compared to the stops when followed by /u/ (e.g., sassu vs. sakku), and geminate /s/ when followed by /a/ compared to /u/ (e.g., sassa vs. sassu). The stop-vowel sonority difference is greater than that for the fricative-vowel sequence. In addition, the sonority difference is greater in contexts with /a/ compared to /u/. This difference, representing greater perceptual distance, might facilitate perception of the consonant’s cues to duration and geminate status.

Our third research question concerned the potential role of speech segmentation strategy in addition to duration and sonority relationships to account for learner performance. Evidence of its influence can be seen in the learners’ error patterns (see Figure 5). In the current study, when a geminate was misperceived, response options were either a singleton, or a long vowel followed by a singleton. In the first case, misperception of a geminate (e.g., sakka) as a singleton (e.g., saka), might indicate that learners failed to perceive the longer duration of two elements: the preconsonantal vowel, which serves as an initial cue to the duration of the following consonant with which it covaries, and the geminate itself. Given that these learners were L1 English, they may have followed a segmentation strategy focused on the syllable instead of the subsyllabic mora as the basic rhythmic unit. By choosing the singleton (e.g., saka), the consonant was syllabified as a singleton to the onset position of the second syllable, comparable to the structure in Figure 1c for kite.

The second option was misperception of a geminate as a long vowel (e.g., saaka), which was the predominant error pattern in both stimulus conditions across proficiency levels (see Figure 5). This suggests learners may have detected the longer duration of the preconsonantal vowel and perceived it as two morae instead of one. In a forced-choice task such as the one in this study, this point may have dictated the listeners’ response if they failed to attend to the duration of the following consonant, syllabifying it as a singleton to the second syllable. For L1 English listeners, vowel length is a strong cue, and is compatible with a strategy that syllabifies the consonant as the singleton onset of the next syllable, lengthening the preceding vowel in the resulting open syllable. This also contributes to an explanation for the high accuracy rates found by Motohashi Saigo (2007) for
perception of long vowels by learners at various levels of proficiency. A forced-choice task was also used in that study. One of the stimuli was *kaate* (long vowel), for which the other response options were *kate* (short vowel and singleton) and *katte* (short vowel and geminate). In a situation such as that, the perception of vowel duration is the only challenge for learners. The consonant would be perceived as the singleton onset of the second syllable following a lengthened vowel. As shown in Figure 1c, both morae for the long vowel are captured in one syllable. Perceiving the duration of a vowel may also be easier for learners than perceiving the duration of silence, as in the case of the stop closure period, or the duration of the period of frication.

Our fourth research question dealt with the potential influence of stimulus condition, which contrasted words in isolation with those in short carrier sentences. Although the stimuli in both conditions were produced at the same rate of speech, those embedded within a carrier sentence may have posed a greater challenge to the listeners’ ability to focus attentional resources on detecting the geminate’s duration, and utilizing the mora, or its boundaries, to segment the speech stream. As noted earlier, the task and materials were designed specifically to elicit the learners’ strategy for perceptual segmentation based on rhythmic unit in contrast to segmentation for the purposes of lexical recognition. The stimuli were either nonwords or those with low pedagogical frequency, and the short carrier sentences were provided to the participants. In addition, a forced-choice identification task allowed learners to focus attention on acoustic–phonetic information without engaging higher order cognitive processes. This approach also allowed some comparability with previous studies on L2 geminate perception. Further research is needed to explore learners’ ability to segment speech involving geminates for the purposes of lexical recognition.

In sum, the results of the current study point to a complex interplay of factors in L2 geminate identification accuracy. The challenges for L2 learners of Japanese are (a) to develop an attention-weighting perceptual mechanism for the detection of the mora, (b) to perceive the longer duration of the preconsonantal vowel and identify it as one mora, and (c) to perceive the following consonant’s duration. This latter stage may be maximized by a larger sonority difference between the consonant and vowel. The occurrence of errors, even when sonority difference was maximized, emphasizes the additional important roles of duration and segmentation by moraic units and/or their boundaries.

Future research should also explore the sonority relationship between the consonant and preconsonantal vowel. The current study varied only the postconsonantal vowel to investigate the influence of the sonority difference on the perception of one of the boundaries. Data from more advanced learners might also confirm the direction of perceptual development indicated by the findings of the present study. It is uncertain at what point learner perception of geminates is not significantly affected by contextual variability (e.g., different consonant–vowel sequences) and presentation conditions. In addition, reaction time data in conjunction with accuracy data would be beneficial in assessing the relative degree of automaticity with which learners of different levels of proficiency process this type of stimuli.
APPENDIX A

*Stimuli used in the study (Tokyo dialect)*

<table>
<thead>
<tr>
<th>Singletons</th>
<th>Translation</th>
<th>Geminates</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sasu</td>
<td>Stab</td>
<td>Sassu</td>
<td>Guess</td>
</tr>
<tr>
<td>Asu</td>
<td>NW</td>
<td>Assu</td>
<td>Dominate</td>
</tr>
<tr>
<td>Saku</td>
<td>Tear (verb)</td>
<td>Sakkru</td>
<td>Sack</td>
</tr>
<tr>
<td>Aku</td>
<td>Evil</td>
<td>Akku</td>
<td>Bad mouth (saying something one should not)</td>
</tr>
<tr>
<td>Satu</td>
<td>Volume</td>
<td>Sattu</td>
<td>NW</td>
</tr>
<tr>
<td>Atu</td>
<td>Pressure</td>
<td>Attu</td>
<td>NW</td>
</tr>
<tr>
<td>Sasa</td>
<td>NW</td>
<td>Sassa</td>
<td>Quickly</td>
</tr>
<tr>
<td>Asa</td>
<td>Morning</td>
<td>Assa</td>
<td>NW</td>
</tr>
<tr>
<td>Saka</td>
<td>Refreshments</td>
<td>Sakka</td>
<td>Last summer</td>
</tr>
<tr>
<td>Aka</td>
<td>Red</td>
<td>Akka</td>
<td>Money of poor quality</td>
</tr>
<tr>
<td>Sata</td>
<td>Trouble</td>
<td>Satta</td>
<td>Went</td>
</tr>
<tr>
<td>Ata</td>
<td>NW</td>
<td>Atta</td>
<td>Existed</td>
</tr>
</tbody>
</table>

*Note:* All stimuli have a high–low accent pattern (high–low–low in the case of geminates). NW, nonword.

ACKNOWLEDGMENTS

An earlier version of this paper was presented at the Second Language Research Forum held at the University of Illinois, Urbana–Champaign, in October 2007.

NOTES

1. There are durational contrasts in English across morpheme boundaries; for example, the /p/ in the single-morpheme *topic* has a shorter closure duration than in the dual-morpheme *top pick* (Hayes, 2002). However, there is also a difference in the stress and intonation pattern. Segment duration does not create minimal pairs in English.

2. See Kubozono (1999) for a detailed discussion of the role of the mora in Japanese word-formation processes, adult speech dysfluencies (e.g., stuttering), the *kana* phonetic script, and spontaneous speech errors showing segmentation of words at mora boundaries. He also describes the importance of the syllable in morphological processes in Japanese such as loanword clipping.

3. These findings should be interpreted with some caution as the participants’ spelling may not accurately represent their perception.

4. For the sake of simplicity, and following the practice in the literature, the symbol /u/ is used in this paper to represent the unrounded high back vowel [u] in Japanese.

5. This index differs slightly from Blevins’ (1995) universal sonority scale, which assigned each voiceless stop a value of 1, and from Parker’s (2002) scale in which /s/ is a 3. Although specific sonority values may vary, there is general consensus that vowels (voiced) are the most sonorous and stops are the least.

REFERENCES


