Language Learning & Technology
http://llt.msu.edu/vol13num2/motohashisaigohardison.pdf

June 2009, Volume 13, Number 2
pp. 29-47

ACQUISITION OF L2 JAPANESE GEMINATES: TRAINING WITH WAVEFORM DISPLAYS[1]

Miki Motohashi-Saigo, Kansai Gaidai University

Debra M. Hardison, Michigan State University

The value of waveform displays as visual feedback was explored in a training study involving perception and production of L2 Japanese by beginning-level L1 English learners. A pretest-posttest design compared auditory-visual (AV) and auditory-only (A-only) Web-based training. Stimuli were singleton and geminate /t,k,s/ followed by /a, u/ in two conditions (isolated words, carrier sentences). Fillers with long vowels were included. Participants completed a forced-choice identification task involving minimal triplets: singletons, geminates, long vowels (e.g., *sasu*, *sassu*, *saasu*). Results revealed a) significant improvement in geminate identification following training, especially for AV; b) significant effect of geminate (lowest scores for /s/); c) no significant effect of condition; and d) no significant improvement for the control group. Most errors were misperceptions of geminates as long vowels. Test of generalization revealed 5% decline in accuracy for AV and 14% for A-only. Geminate production improved significantly (especially for AV) based on rater judgments; improvement was greatest for /k/ and smallest for /s/. Most production errors involved substitution of a singleton for a geminate. Post-study interviews produced positive comments on Web-based training. Waveforms increased awareness of durational differences. Results support the effectiveness of auditory-visual input in L2 perception training with transfer to novel stimuli and improved production.

INTRODUCTION

The duration of vowels and consonants is a contrastive feature in Japanese unlike English. In Japanese, singleton consonants contrast with their longer geminate counterparts, which are a special type of mora. Accurate perception and production of morae are important for comprehension; however, they can be problematic elements for second-language (L2) learners. Training studies to date have not addressed this particular issue.

Briefly, a mora is a unit of timing, which plays a role in the temporal organization of speech in production (e.g., Kubozono, 1999). Neighboring moraic units tend to show equal duration (Port, Dalby, & O'Dell, 1987). As a unit of rhythm, the mora also plays a role in the segmentation of speech sounds for lexical recognition (e.g., Cutler & Otake, 2002). Special morae, such as geminates, do not constitute syllables by themselves as shown in the following examples, where the moraic units are separated by a dot and syllables by a hyphen: a) moraic obstruent: /*ki.t.te*/, /*kit-te/* "postage stamp," and b) the second half of a long vowel: /*ki.i.te*/, /*kii-te*/ "listening."[2]

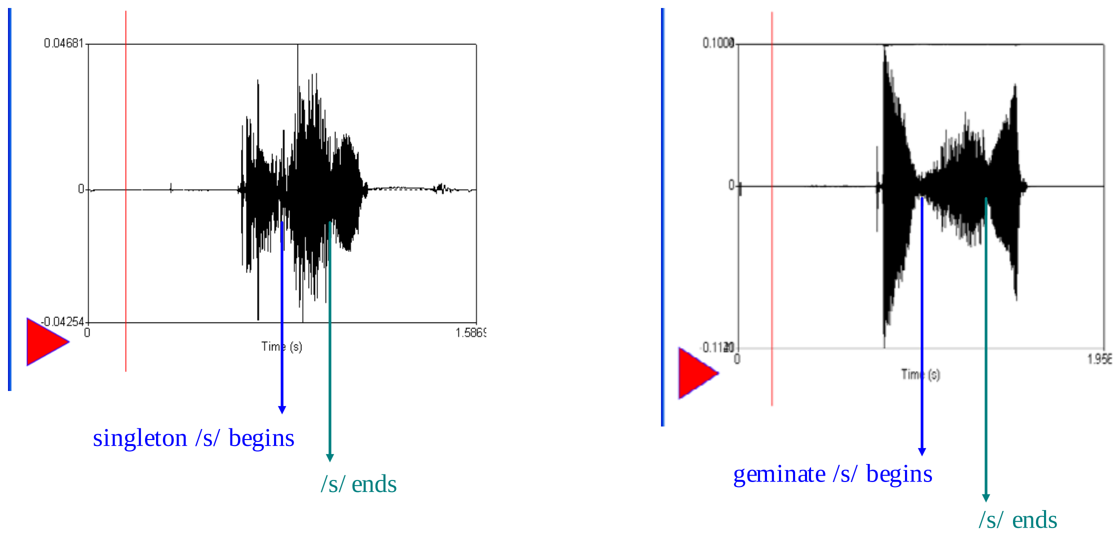
An important acoustic cue for geminate obstruent perception by native speakers of Japanese is stop closure duration (e.g., for /p/, /t/, /k/) or period of frication (e.g., for /s/) (e.g., Fukui, 1978). However, perception is influenced by other factors such as pitch accent, which is also contrastive in Japanese (Ofuka, 2003). For example, *kata* produced with a H(igh)-L(ow) pattern means "shoulder," while *katta* (HLL) with a geminate stop means "won." In contrast, *kata* with a LH pattern means "form" and *katta* (LHH) means "bought." Using synthesized stimuli to alter the closure duration of /t/ along a *kata* – *katta* continuum for both pitch accent patterns, Ofuka found native speakers required a longer closure duration to perceive a geminate in the LHH pattern in contrast to the HLL. In addition, Hirata (1990) found that for stimuli embedded in carrier sentences, a singleton was perceived as a geminate if postconsonantal elements were spoken quickly, and geminates were perceived as singletons if these elements were spoken more slowly. The duration of the vowel preceding the consonant also influenced native perception of a geminate (Hirata, 1990). Vowels are longer before geminates than before singletons (Han, 1994).

Several studies have suggested some acoustic cues L2 learners may be using in geminate perception. A study by R. Hayes (2002) suggested consonant duration may play a role for learners as well as native speakers. She found significantly better singleton-geminate discrimination accuracy for /t/ versus /k/, and /t/ versus /s/ by L1 English learners in the U.S. at different levels of proficiency. The duration difference between singleton and geminate /t/ in her stimuli was the greatest (180 ms) compared to that for /k/ (142 ms) and /s/ (134 ms). Korean and Chinese speakers learning Japanese in Minagawa and Kiritani's (1996) study showed effects of pitch accent type on singleton and geminate identification. Learners 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. The average postconsonantal vowel was significantly shorter in the HL context compared to that in the LH context. In contrast, L1 English and Spanish speakers showed no significant difference in errors according to pitch pattern.

Some research has suggested that L1 English learners of Japanese are able to detect moraic weight but may not attribute it to the correct segment. When learners were asked to write (using Roman letters) English loan words in Japanese that were produced by a Japanese speaker, their errors generally involved lengthening the vowel in words that native listeners spelled with geminates (Yamagata & Preston, 1999). In a recent study, Hardison and Motohashi-Saigo (in press) found that L1 English learners in the first and third semesters of Japanese study in the US also misperceived geminates more often as long vowels. More advanced learners who had studied for one to two semesters in Japan made significantly fewer errors overall. Participants in that study were presented with an identification task involving minimal triplets: singleton (e.g., *sasu*), geminate (e.g., *sassu*), and long vowel (e.g., *saasu*). Results revealed significantly higher accuracy for targets in isolated words compared to carrier sentences, for geminate /k/ versus /s/, and contexts where the postconsonantal vowel was /a/ versus /u/.[3] The most difficult context was geminate /s/ followed by /u/ in a word embedded in a carrier sentence. Duration differences alone did not account for all findings. Results suggested the sonority difference between consonant and vowel might be a factor in establishing perceptual distance between segments.

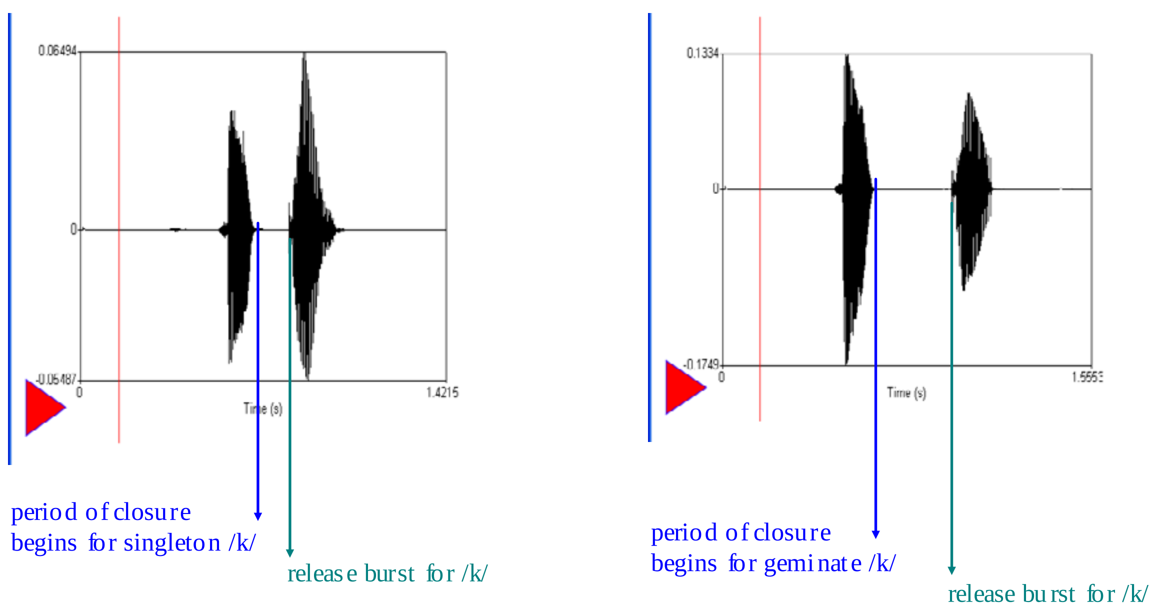
Sonority is an inherent property of a segment, which reflects its degree of acoustic energy and its robustness as an acoustic cue (e.g., Wright, 2004). Values generally range from 0 to 10; for example, in Selkirk's (1984) sonority index, stops (e.g., /t/, /k/) are the lowest in value at 0.5, followed by fricatives (e.g., /s/ at 4), nasals, liquids, and then vowels. High vowels are less sonorous than low ones; therefore, /a/ has greater sonority (value of 10) than /u/ (8).[4] The sonority difference between geminate /s/ and /u/ (e.g., *sassu*), the problematic context noted above, is smaller than the difference between /s/ and /a/, or in a stop-vowel sequence. Results of the study pointed to a complex interplay of factors in L2 geminate identification accuracy. The identification process may be facilitated by a larger sonority difference and thus perceptual distance between consonant and vowel. The findings suggested several challenges for learners: a) to develop an attention-weighting perceptual mechanism for the detection of moraic boundaries as the units of rhythm for speech segmentation (e.g., Cutler, 1997); b) to perceive the longer duration of the preconsonantal vowel as a cue to the following geminate, and identify it as one mora; and c) to perceive the following consonant's duration.

In the present study, we sought a user-friendly Web-based training program that could enhance perception of geminates for L2 learners by providing a visual display of the duration of segments in the speech stream. Other types of visual displays, such as those for pitch, have been effective tools in speech training for L2 learners (e.g., Chun, 2002; Chun, Hardison, & Pennington, 2008; Hardison, 2004, 2005; Levis & Pickering, 2004). To emphasize segmental duration in the current study, speech waveform displays were chosen. The waveform is a representation of rapid variations in air pressure caused by actions of a speaker's vocal organs superimposed on the outgoing airflow. Sample waveforms from the current study are shown in Figures 1 and 2.



(a) *asu* (nonword) (b) *assu* 'dominate'

Figure 1. Waveform displays for singleton and geminate /s/



(a) *aka* 'red' (b) *akka* 'money of poor quality'

Figure 2. Waveform displays for singleton and geminate /k/

The vertical lines in the waveform represent pulses produced by vibrating vocal cords. Figure 1 shows the singleton-geminate contrast for the fricative /s/, where duration is indicated by the period of frication. Figure 2 shows this contrast for the stop /k/, where duration is indicated by the period of closure prior to the release burst for the stop. Arrows are drawn to show approximately where the singleton and geminate consonants begin and end.

Previous studies demonstrating successful L2 perception training at the segmental level have shown the advantages of auditory-visual over auditory-only input (Hardison, 2003), and been characterized by findings that support the ability of learners to generalize improved perceptual abilities to novel stimuli and new talkers, with transfer to improved production (e.g., Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Hardison, 2003; Lively, Logan, & Pisoni, 1993). The nature of the relationship between perception and production continues to be somewhat elusive. In an intensive auditory training study involving L1 Japanese learners of English /r/ and /l/, Bradlow et al. (1997) found no significant Spearman rank-order correlation between rates of improvement for perception and production. However, analysis did reveal a significant positive rank-order correlation between the learners' pretest perception accuracy scores and the effectiveness of the perception training for them. The pretest performance was considered to be a fairly good predictor of training effectiveness although there was considerable variation across individual learners.

To evaluate the potential of waveform displays as a training tool, the current study was guided by the following research questions:

1. Is auditory-visual (AV) perception training using a waveform display better than auditory-only (A-only) training for the improvement of geminate perception? As noted in previous studies, input from both modalities (i.e., auditory and visual) is more beneficial for learners; therefore, we hypothesized that AV training involving waveform displays would produce greater accuracy than A-only training.
2. Are there dominant perception and production error patterns? Based on previous research (Hardison & Motohashi-Saigo, in press; Yamagata & Preston, 1999), more misperceptions of geminates as long vowels were anticipated; however, this pattern may not hold for incorrect production of geminates.
3. Is accuracy better for stimuli in words presented in isolation compared to those in short carrier sentences? We hypothesized that identification accuracy might be greater for the isolated word condition. Stimuli embedded in carrier sentences require the listener to segment a longer stretch of speech, creating a greater challenge in the detection of moraic units.
4. Does training generalize to novel stimuli, and transfer to improved production of geminates? Generalization and transfer have emerged over the past 15 years as two of the hallmarks of successful perception training. Both suggest the potential for improved perceptual capabilities to extend beyond the training session to the natural language environment, which is the ultimate learning objective.
5. What is the relationship between rates of improvement for perception and production? Consistent with previous research, learners' production of geminates was expected to improve as their perception improved with training; however, variation in rates of improvement in these two domains was also expected.

The current study employed a pretest-posttest design with a control group to compare the effects of AV versus A-only training in L2 Japanese geminate perception, and investigate the possible generalization to novel stimuli and transfer to production improvement. Data collection followed this sequence: production pretest, perception pretest, perception training, production posttest, perception posttest, test of generalization of training, and interviews with individual participants. The between-subjects variable was group (AV, A-only, control); within-subjects variables were time (pretest, posttest), condition (isolated words, carrier sentences), consonant (/t/, /k/, /s/), and vowel (/a/, /u/). The dependent variable was identification accuracy scores.

METHOD

Participants

A total of 40 learners participated in this study: 15 received AV training, 15 received A-only training, and 10 formed the control group, which participated in the pre- and posttest production and perception testing but did not receive any training. All participants were native speakers of American English; there were no heritage learners of Japanese. All were enrolled in an elementary Japanese language course at a university in Japan as part of a study abroad program. They had been learning Japanese for about 2 months when this study began. They had not received any prior Japanese instruction. The course met 5 days per week for 50 minutes each day. It emphasized oral communication skills; however, there was no specific focus on segmental perception or production. The instructor was a native speaker of Japanese (Tokyo dialect). The students also took lecture courses in their major areas of study (e.g., economics) taught in English by native speakers of English. They resided in an international student dormitory where communication among students was generally in English.

Materials

Production Testing Stimuli

There were 30 tokens: 24 involving medial consonants (12 singletons, 12 geminate counterparts) and 6 fillers with long vowels /a, e, u, i/ (see Appendix A). Of the 24 tokens involving consonants, half began with /a/ (e.g., *akka* 'money of poor quality'), and half with /sa/ (e.g., *sakka* 'last summer'). Most were real words of relatively low pedagogical frequency, which were not likely to have been known to beginning learners. For example, *itta* is more commonly taught than *satta* to express 'went'. Some nonwords were used to complete the singleton-geminate contrasts.[5] These stimuli involved a balanced number of singletons (bisyllabic and bimoraic) and their geminate counterparts (bisyllabic and trimoraic). The tokens with long vowels were also bisyllabic and trimoraic, and were included to reduce the biasing of learner attention toward the consonants. The consonants were /t/, /k/, /s/, followed by /a/ or /u/, and were chosen for their sonority values. Each token was produced in isolation and then in one of two possible short carrier sentences. In the sentence condition, half of the tokens were produced in one sentence and half in the other. The matching of token to carrier sentence was counterbalanced across participants. The two carrier sentences were written out for the participants as follows: *watashi wa _____ toiimashita* 'I said _____'; *kore wa ______ desu* 'This is ______.' The total number of productions by each participant was 120 (2 x 30 tokens x 2 conditions).

Perception Testing Stimuli

Materials for perception testing were the same as those for production. The stimuli with long vowels were included in the perception tasks to avoid response bias. In addition, previous studies indicated that geminates are often misperceived as long vowels (Yamagata & Preston, 1999).

Stimuli were recorded by a female native speaker of Japanese (Tokyo dialect) using a SONY digital minidisc recorder/player (44.1 kHz sampling rate) with a SONY microphone. Each word was recorded twice in a quiet room in the university's language lab using a HL accent pattern (HLL for geminates): once in isolation and then in one of the two possible carrier sentences noted above. The talker was instructed to use a natural and consistent rate of speech, which was monitored by the researcher using a digital metronome (flashing lights on the computer screen). Waveforms, generated in Praat[6], were examined to ensure that /u/ 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). No devoicing was noted. Stimuli were tested with native speakers of Japanese for auditory intelligibility, and then randomized for auditory presentation to participants. The auditory-only presentation of testing stimuli provided a basis for comparison of the scores from the AV and A-only training groups as the auditory modality was the one they shared. Each stimulus was presented three times in each of the above two conditions (isolated word, carrier sentence). Total testing stimuli were 180 (3 presentations x 30 tokens x 2 conditions).

Perception Training Stimuli.

Due to a limited number of possible tokens (i.e., in order to control the preconsonantal phonetic material), the pre- and posttest stimuli were also used for perception training. Recording details were the same as described above. The waveforms for the AV group in training were created in Praat and converted to SWF files using Macromedia Flash MX 2004 for Web presentation. This allowed the learners to listen to the sound and view the entire waveform display as the cursor moved through the word. The A-only group had access to auditory input only. For both groups, the order of presentation of stimuli was randomized and changed daily although all stimuli were presented with equal frequency every day for a total of 120 (2 presentations x 30 tokens x 2 conditions).

Test of Generalization Stimuli

Following the posttest, a test of generalization (TG) was conducted to determine whether the results from the perception training could be generalized to unfamiliar tokens produced by a new talker, who was also a female native speaker of the Tokyo dialect.[7] Stimuli included singleton and geminate /s/, /t/, /k/ from training, and /p/, which had not been presented before in this study. Postconsonantal vowels were /i/ and /e/, which can occur in postconsonantal position with these consonants, but constituted untrained contexts. The vowel /i/, a high vowel, has the same sonority value as /u/ (8), and /e/ has the value of 9. There were 16 stimuli beginning with /sa/ that included either a singleton or a geminate (See Appendix B), and 6 fillers (nonwords) with long vowels. These were recorded in a quiet room in the language lab following the same procedure used for the preparation of the testing and training materials. This talker was also instructed to use a natural and consistent rate of speech, which was monitored by the researcher using a digital metronome. As with the other phases of the study, these stimuli were checked for auditory intelligibility, and were randomized for auditory presentation within two conditions: words in isolation and carrier sentences. Each stimulus was presented three times in each condition. Total stimuli were 132 (3 presentations x 22 tokens x 2 conditions).

Duration Measurements

Duration measurements for the stimuli used in the perception testing and training are given in Table 1. Duration was measured in the expanded waveform display using the SUGI Speech Analyzer by Animo[8], and confirmed with synchronized spectrogram display. Standard segmentation procedures were applied (e.g., Johnson, 1997; Ladefoged, 2003). To measure the duration of a vowel, the beginning was marked at the onset of complex voicing with higher frequency components, and the end at the offset of these components. The beginning of the stops was marked at the offset of all higher frequency elements 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. Following previous studies (e.g., Han, 1994), VOT (voice onset time) was not included in the duration measurements. The consonants of interest 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 in this study covaried with that of the following consonant; therefore, as expected, the vowels were longer preceding geminates than singletons (Han, 1994). Measurements also confirmed the consistently shorter duration of postconsonantal /u/ versus /a/ (Yoshida, 2006). The duration of all singletons was shorter than that of the corresponding geminates.

Table 1. Perception Testing and Training Stimuli: Duration of Singletons, Geminates, Pre- and Post-consonantal Vowels (in milliseconds)

Singleton			Geminate		
a	t	a	a	tt	a
68	90	144	76	210	139
Singleton-Geminate Duration Difference: 120; Ratio of 2.3:1.0					
a	t	u	a	tt	u
72	100	120	94	208	114
Singleton-Geminate Duration Difference: 108; Ratio of 2.1:1.0					
a	k	a	a	kk	a
69	94	127	94	196	119
Singleton-Geminate Duration Difference: 102; Ratio of 2.1:1.0					
a	k	u	a	kk	u
75	97	122	96	212	88
Singleton-Geminate Duration Difference: 110; Ratio of 2.1:1.0					
a	s	a	a	ss	a
78	111	144	102	212	115
Singleton-Geminate Duration Difference: 101; Ratio of 1.9:1.0					
a	s	u	a	ss	u
81	135	125	100	223	87
Singleton-Geminate Duration Difference: 88; Ratio of 1.7:1.0					

Note. Measurements are from stimuli beginning with /sa/ produced in the isolated word condition. Comparisons with the sentence condition stimuli revealed differences of 1-13 ms. The duration of the preconsonantal vowel in stimuli beginning with /a/ differed by 0-12 ms from those shown above. Measurements were comparable for the stimuli used in the test of generalization.

Procedure (chronological order)

Production Testing

All productions were recorded in a quiet room in the university's language lab with the same recording equipment used in the preparation of the perception testing and training stimuli. Participants were given a word list using Roman letters so they could look over the words before recording them. They were instructed to use a HL(L) accent pattern, with which they were familiar. Each word was produced in isolation and then in one of the two possible short carrier sentences, which were written out for them as follows: *watashi wa* ____ *toiimashita* 'I said ____'; *kore wa* ____ *desu* 'This is ____.' Data were stored in WAV format on a CD-ROM, and copies were made for the native-speaker raters.

Perception Testing

The perception pretest and posttest were administered in a sound-attenuated classroom prior to and immediately following the two-week training period. Stimuli were presented via centrally-located ceiling-mounted speakers. Participants completed a three-alternative forced-choice identification task for both the carrier sentence and isolated word conditions, which were presented in that order. Response options were minimal triplets: singleton, geminate, and long vowel (e.g., *sasu*, *sassu*, *saasu*). Participants were instructed to circle the word on their response sheets that represented what they heard. Interstimulus interval for all testing was determined through pilot testing and was set at 3 seconds. No feedback was given during testing.

Perception Training

Assignment of participants to the AV training, A-only training, or control group was made based on their pretest performance to create comparable groups prior to training. Comparison of pretest scores across groups using a one-way ANOVA revealed no statistically significant difference in the pretest scores for these groups, $F(2,42) = .464$, $p > .05$; therefore, learners in each group represented a range of abilities. Both training groups (AV and A-only) participated in separate trial sessions with one of the researchers to ensure participants understood how to access the materials on the Web site for training, respond, and receive feedback. The content of this session differed for the two training groups. The AV group received information about geminates, and a demonstration of speech waveforms comparing singletons and geminates although none of the stimuli from the study was used in the demonstration. The A-only group received the same information about geminates, but was not given a waveform demonstration.

During training, after each stimulus (AV or A-only) was presented, participants could proceed to the next screen where they were shown three response options: singleton (e.g., *sasu*), geminate (e.g., *sassu*), and long vowel (e.g., *saasu*).[9] These options appeared in a dialog box created with a Web quiz creating tool.[10] The task for each stimulus was "Click the word that you heard." After selecting an option, feedback was provided. If the response was incorrect, participants had a second chance to listen to the stimulus and respond before the correct answer was shown.

The students participated in the training sessions outside of class time. They were instructed to complete 10 training sessions, one per day. These sessions were self-paced. Although students were instructed to report any problems they encountered during training, no reports were received.

Test of Generalization

Following the posttest, both training groups took a perception test of generalization in a classroom setting. This was also a forced-choice three-alternative task in which participants were instructed to circle the word on their response sheets that matched what they heard. Interstimulus interval was 3 seconds. As with the other perception tests, response options involved a singleton, geminate, and long vowel. No feedback was provided.

Post-Study Interviews

A follow-up interview was conducted with each of the participants. Each interview lasted about 10 minutes. Participants were asked to comment on areas of difficulty for them in the perception and production of Japanese, and the effectiveness of the training with regard to geminates. They were also asked for overall comments on the training procedure, especially the Web-delivery format.

RESULTS AND DISCUSSION

The pretest and posttest scores were tabulated for each of the three groups (AV, A-only, control). Recall that a one-way ANOVA revealed no significant difference in the pretest scores for these groups. A paired t-test was conducted to compare the overall pretest and posttest perception scores for the control group. This revealed no significant difference, $t(14) = -.807$, $p > .05$; therefore, this group's data were excluded from further analyses. Comparison of data from the training groups for the stimuli beginning with /sa/ and those beginning with /a/ revealed no significant difference; therefore, these data were combined for further analyses.

The remainder of the discussion of results follows this order: a) overall effectiveness of perception training on geminate identification accuracy and comparison of training type (AV vs. A-only); b) in-depth analysis within each training group, including the effects of geminate and vowel type, and comments on singleton and long vowel identification accuracy; c) analysis of perception errors; d) test of generalization; e) effect of perception training on production; f) analysis of production errors; g) investigation of the perception-production relationship; and h) a brief summary of the post-study interview comments. An alpha level of .05 was used for all statistical tests.

Effectiveness of Perception Training

Participants completed the perception training sessions as instructed with an average duration of about 25 minutes per session. To evaluate the effectiveness of this training on geminate identification and compare training types, accuracy scores were submitted to a mixed ANOVA. The between-subjects variable was group (AV, A-only). The within-subjects variables were time (pretest, posttest) and presentation condition (isolated word, carrier sentence). There was a significant effect of time, $F(1,28) = 50.114$, $p < .001$; posttest scores were higher than those of the pretest. There was a significant effect of group, $F(1,28) = 6.986$, $p < .05$; and a significant Time x Group interaction, $F(1,28) = 9.380$, $p = .005$. The group receiving AV training showed greater improvement. Although there was no statistically significant effect of condition, $F(1,28) = 1.388$, $p > .05$, Figure 3 shows the pre- and posttest scores for both training groups, for both the isolated words and words in carrier sentences. As shown in the graph, mean scores improved the most for the AV group (from 65% to 88% for the word condition; 54% to 87% for the sentence condition), and the least for the A-only group (58% to 70% for the word condition; 53% to 63% for the sentence condition). Given the lack of a significant effect of condition or interactions involving this factor, its levels were collapsed for subsequent analyses within each training group.

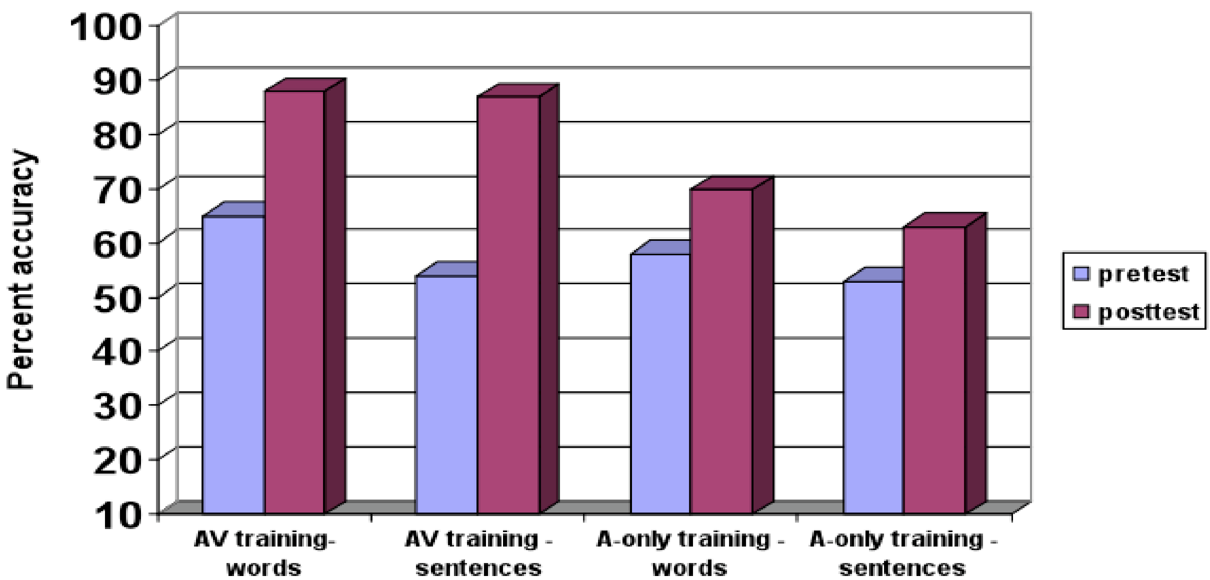


Figure 3. Effectiveness of perception training on identification accuracy of L2 Japanese geminates: Comparison of training type (AV vs. A-only) and condition (isolated words, carrier sentences)

AV Training Group

For a more in-depth analysis, a three-way within-subjects ANOVA was conducted on the scores from the AV training group. Variables were time (pretest, posttest), geminate type (/t/, /k/, /s/), and vowel type (/a/, /u/). The time factor was included again in this analysis in order to explore possible interactions with geminate and/or vowel type. There was a significant effect of time, $F(1,14) = 28.00$, $p < .001$, and a significant effect of geminate type, $F(2,28) = 10.858$, $p < .001$. Pairwise comparisons (with Sidak adjustment for multiple comparisons) revealed the significant differences were between geminate /t/ and /s/, and geminate /k/ and /s/, but not between the two geminate stops. There was a significant Time x Geminate interaction, $F(2,28) = 3.512$, $p < .05$. Although perception of all geminates improved in both vowel contexts, /s/ showed more improvement than the stops for which the pretest scores were higher.

Identification accuracy for singletons in the pretest was high at an average (across singleton-vowel contexts) of 92% for the isolated word condition and 89% for the carrier sentence condition. Posttest accuracy increased to an average of 96% in both conditions. Pretest accuracy for the long vowels in the filler items was 88% in isolated words and 87% in carrier sentences. This rose to 92% in the posttest for both conditions. Because of the high pretest scores and minimal variation across contexts, no further analyses were warranted for these data.

A-only Training Group

A three-way within-subjects ANOVA was conducted on the geminate identification accuracy scores from the A-only training group. Variables were time (pretest, posttest), geminate type (/t/, /k/, /s/), and vowel type (/a/, /u/). Again, there was a significant effect of time, $F(1,14) = 34.017$, $p < .001$. There was also a significant effect of geminate type for this group, $F(2,28) = 5.888$, $p < .01$. Pairwise comparisons (with Sidak adjustment) revealed the significant differences were between geminate /t/ and /s/, and /k/ and /s/. A significant Time x Geminate interaction, $F(2,28) = 7.304$, $p < .05$ indicated a difference in the rate of improvement per geminate, with /s/ showing the greatest amount of improvement. There was also a significant Time x Geminate x Vowel interaction, $F(2,28) = 12.157$, $p < .01$. Pretest scores were lowest for geminate /s/ followed by /u/ (e.g., *sassu*); over the course of training, this geminate-vowel sequence showed the greatest improvement.

Perceptual accuracy of singletons in the pretest, averaged across singleton-vowel contexts, was 92% in both conditions and rose to 95% in the posttest. Pretest identification accuracy of the long vowels in the filler items was 86% in the isolated word condition, and 85% in the carrier sentences. Posttest accuracy was 90% and 88% respectively. As with the AV training group, these data did not warrant further analysis.

Analysis of Perception Errors

Both the AV and A-only training groups showed a similar pattern of perception errors when trying to identify geminates. Chi-square analysis was used to compare the frequency with which each error type occurred (i.e., misperception of a geminate as a singleton or as a long vowel). In the perception pretest, there was a significantly greater number of misperceptions of a geminate as a long vowel (e.g., $\chi^2 = 12.24$, $df = 1$, $p < .001$ for the AV group). This pattern held across all geminate-vowel sequences. Far fewer errors occurred in the posttest; however, their distribution followed the same pattern, favoring misperception of the geminate as a long vowel.

Test of Generalization

A mixed ANOVA compared the scores of both training groups for geminate identification accuracy from the posttest and the test of generalization (TG). The between-subjects variable was group (AV, A-only); the within-subject variables were test (posttest, TG), and condition (words, sentences). There was a significant main effect of group, $F(1,28) = 48.91$, $p < .001$. As shown in Figure 4, the AV group's mean scores for both conditions in the posttest and TG were higher than those for the A-only group. There was no significant effect of or interactions involving condition. There was a significant effect of test, $F(1,28) = 7.255$, $p < .05$; however, there was also a significant Test x Group interaction, $F(1,28) = 48.910$, $p < .001$. The decline in accuracy between the posttest and TG averaged 5% across both conditions for the AV group and 14% for the A-only group. Some of the decline may be attributable to the challenging TG stimulus set. Although it included /s, t, k/ from the testing and training materials there was a new consonant /p/, and new postconsonantal vowels /i, e/. In addition, the TG stimuli were produced by an unfamiliar voice, requiring generalization ability for both voice and segment type simultaneously. Given the role of the postconsonantal vowel and its relationship to the consonant in geminate perception (Hardison & Motohashi-Saigo, in press), the TG stimulus set would likely have been a challenging one.

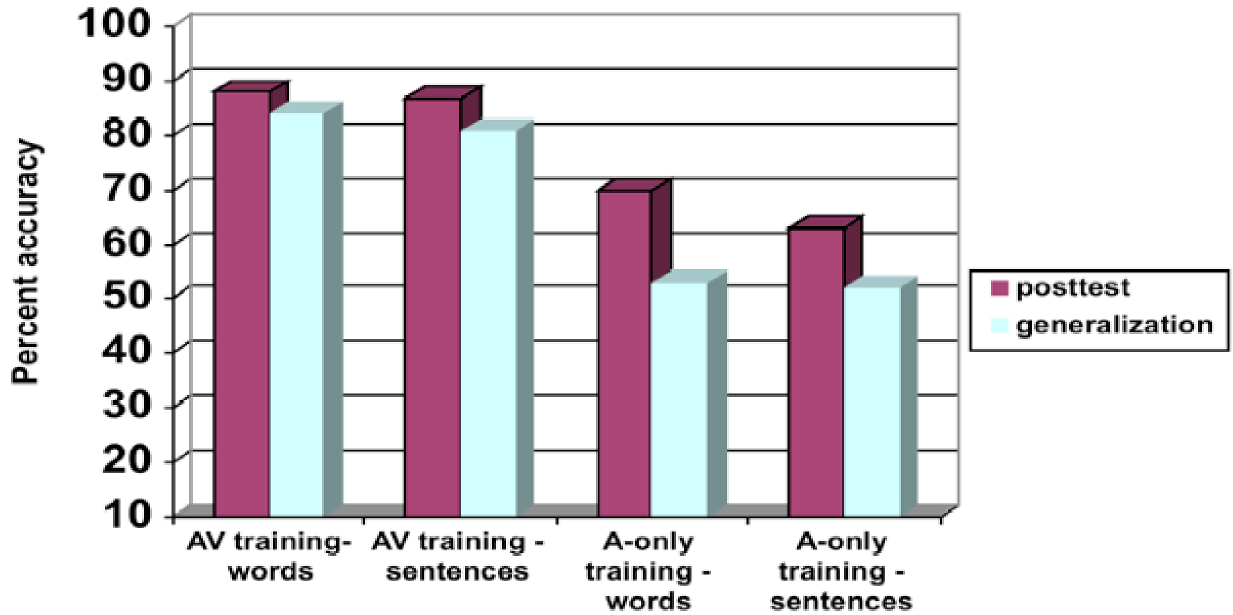


Figure 4. Generalization of perception training: Comparison of geminate identification accuracy in the posttest and test of generalization for the AV and A-only training groups in both conditions (isolated words, carrier sentences)

To investigate the effects of geminate and vowel type, accuracy data were submitted to a two-factor within-subjects ANOVA for each training group. For the AV group, there was no significant effect of geminate type, $F(3,42) = .574$, $p > .05$, or vowel type, $F(1,14) = 1.909$, $p > .05$, and no significant interaction between these factors, $F(3,42) = .412$, $p > .05$. Despite the lack of statistical significance, the trend in identification accuracy shown by the TG data was similar to that shown by the pretest-posttest analysis. The greatest accuracy was noted for all geminates when followed by /e/ versus /i/, the highest score being for /te/ (*satte*) (93.5 %) followed by /ke/ (*sakke*) (90%). For contexts with /i/, the highest accuracy was for /ti/ (*satti*) and /pi/ (*sappi*) (each 80%) and the lowest was for /si/ (*sassi*) (73.5%). As a high vowel, /i/ has the same sonority value (8) as /u/. The lower accuracy in the generalization test for geminate /s/ followed by /i/ parallels the lower accuracy for geminate /s/ followed by /u/ in the analysis of the pretest-posttest data.

A two-factor within-subjects ANOVA was also carried out for the geminate accuracy data from the A-only group. There was a significant effect of geminate type, $F(3,42) = 6.357$, $p = .001$; pairwise comparisons (with Sidak adjustment) revealed the difference was between perception of geminate /k/ and /s/, and geminate /p/ and /s/. However, there was no significant effect of vowel type or interaction between geminate and vowel type. As with the AV training group, the A-only group's data revealed that the highest scores were for contexts where the geminate was followed by /e/ versus /i/. The highest score was for /te/ (*satte*) at 63.5% and the lowest for /si/ (*sassi*) at 20%.

Effect of Perception Training on Production

Learner productions were rated by 5 native speakers of Japanese (Tokyo dialect) who reported no sustained experience with English speakers. Ratings were carried out in a quiet room in the university's language lab where each rater was given a headset. Productions were blocked by learner. For the carrier sentence condition, raters were given the sentences on a piece of paper with a blank for the target word. For both conditions, their task was to "write the word you think you heard in standard Japanese orthography." For each correct identification by a rater, 1 point was assigned. Inter-rater reliability of .84 (Pearson r) was considered satisfactory.

Scores were tabulated for each training group and the control group. As with the perception results, a comparison of scores for tokens beginning with /sa/ and those beginning with /a/ showed no significant difference in terms of production accuracy; therefore, these data were combined for analysis. A paired *t*-test comparing the pretest and posttest production scores for the control group revealed no significant improvement, $t(-.713)$, $df = 14$, $p > .05$. Scores for the training groups were submitted to a mixed ANOVA. The between-subjects variable was group (AV, A-only); the within-subject variables were time (pretest, posttest) and condition (isolated word, carrier sentence). There was a significant effect of time, $F(1,28) = 91.260$, $p < .001$. A significant Time x Group interaction, $F(1,28) = 9.554$, $p < .01$, indicated that although both groups had improved as a result of perception training, the AV training group improved more. There was no significant effect of or interactions involving the condition factor; therefore, this factor's levels were collapsed. Further analyses explored the effect of geminate and vowel type within each training group.

AV Training Group

A three-factor within-subjects ANOVA was carried out on the production scores for geminates for the AV training group. Variables were time (pretest, posttest), geminate type (/t, /k, /s/), and vowel type (/a/, /u/). As with the perception analyses, the time factor was included here in order to detect any interactions. There was a significant effect of time, $F(1,14) = 47.888$, $p < .001$, geminate type $F(2,28) = 22.423$, $p < .001$, and vowel type $F(1,14) = 8.267$, $p = .01$. There were no significant interactions. As shown in Figure 5, pretest accuracy for both conditions improved with training to a comparable degree (31% to 67% for isolated words, 31% to 66% for carrier sentences). Mean scores for geminates were highest for /k/ and lowest for /s/. Pairwise comparisons (with Sidak adjustment) of geminate productions revealed the significant differences were between geminate /t/ and /s/, and between /k/ and /s/. Scores were higher for stimuli involving postconsonantal /a/ versus /u/.

Accuracy of singleton production in the pretest was an average of 88% across singleton-vowel contexts in the isolated word condition, and 86% for the carrier sentences. These figures rose to 93% and 94% respectively in the posttest. As with the scores for singleton perception, those for singleton production showed very little variation across contexts; therefore, no further analyses were conducted.

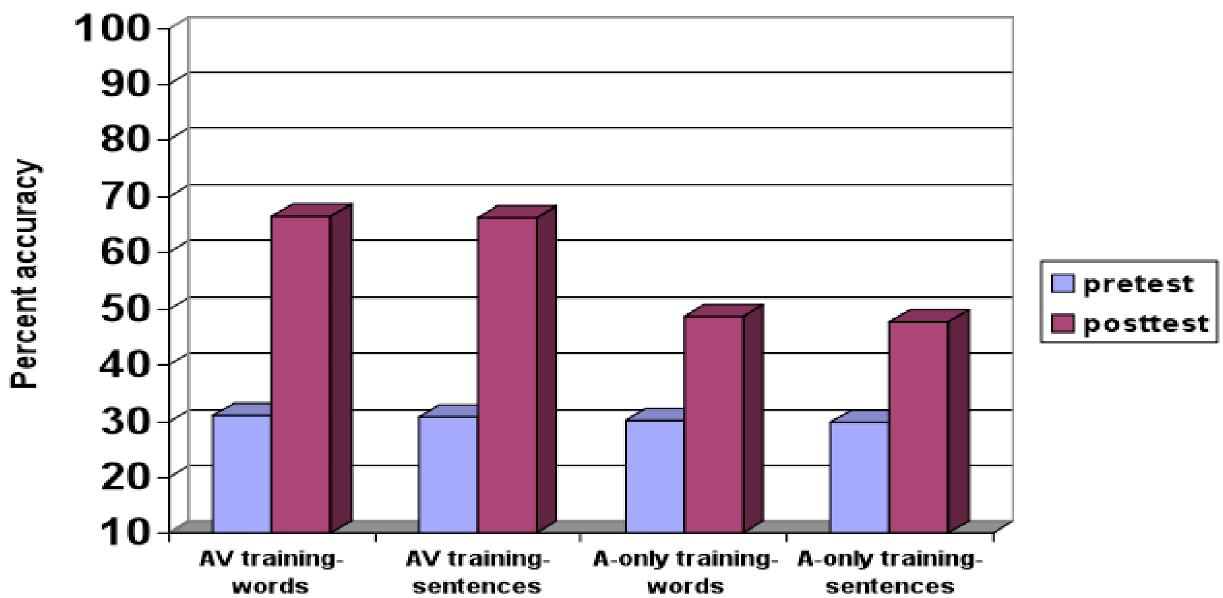


Figure 5. Effect of perception training on production: Percent accuracy of geminate production in the pretest and posttest for AV and A-only training groups in both conditions (isolated words, carrier sentences)

A-only Training Group

A three-factor within-subjects ANOVA was also carried out on the geminate production scores for the A-only training group. Variables were time (pretest, posttest), geminate type (/t, /k, /s/), and vowel type (/a/, /u/). There was a significant main effect of time, $F(1,14) = 54.743$, $p < .001$ and of geminate type, $F(2,28) = 21.094$, $p < .001$. As shown in Figure 5, posttest scores (49% for isolated words and 48% for carrier sentences) were higher than those in the pretest (30% in both conditions). Pairwise comparisons (with Sidak adjustment) of geminate productions revealed the highest overall mean was for geminate /k/, followed by /t/, and then /s/. Significant differences were found between geminate /t/ and /s/, and between /k/ and /s/, as was the case for the AV group. There was a significant Time x Geminate x Vowel interaction, $F(2,28) = 12.324$, $p < .001$. All geminate-vowel sequences improved from pretest to posttest, but to varying degrees. The largest improvement was for geminate /k/ when followed by /u/, and the smallest improvement was for geminate /s/ followed by /u/.

Singleton production accuracy for the A-only group in the pretest was 87% in the isolated word condition, which rose to 91% in the posttest, and 85% in the carrier sentence condition, which increased to 89% in the posttest. Again, there was very little variation in scores across contexts; therefore, no further analyses were conducted.

Analysis of Production Errors

Similar production errors were noted for the AV and A-only groups. Chi-square analysis of the pretest data revealed a significantly greater frequency of production inaccuracies involving the substitution of a singleton (vs. a long vowel) for a geminate (e.g., $\chi^2 = 1{,}009.0$, $df = 1$, $p < .001$ for the AV group). This is in contrast to the pattern of perception errors, which favored misperceptions of geminates as long vowels. In the production posttest, the number of inaccuracies was much lower; however, the pattern of producing more singletons than long vowels was retained, (e.g., $\chi^2 = 378.6$, $df = 1$, $p < .001$ for the AV group).

Perception-Production Relationship

In both perception and production of geminates, each participant showed some degree of improvement from pretest to posttest; however, individual variation was evident, which is consistent with the findings of previous training studies (e.g., Bradlow et al., 1997; Hardison, 2003). The relative perception and production improvement rates were compared for the learners in the current study. This effectiveness was expressed as the relative perceptual improvement as a result of training. Following Bradlow et al., *relative improvement* was defined as the measure of improvement as a proportion of the room for improvement; that is, it takes into account that a higher pretest score leaves less room for improvement than a lower one. In this approach, relative improvement for each learner is the result of posttest accuracy minus pretest accuracy divided by 100 minus the pretest accuracy; in other words, posttest-pretest/100-pretest. Spearman rank-order correlation revealed no significant relationship between relative perception and production improvement of geminates for learners in either training group, Spearman rho (r_s) = .136, $p > .05$ (AV), and $r_s = -.071$, $p > .05$ (A-only). Although a link was observed between perception and production based on the finding that perceptual learning transferred to improved production in the absence of explicit production instruction, there was no significant relationship between degrees of learning in these two domains due to individual variation.

We also investigated whether the pretest perception or production scores were indicators of how effective the training would be for these participants. For this analysis, data were collapsed across geminate-vowel sequences and conditions. For both the AV and A-only groups, there was a significant negative rank-order correlation between the learners' pretest perceptual accuracy and their relative perceptual improvement as a result of training, $r_s = -.837$, $p = .001$ (AV group); and $r_s = -.612$, $p < .05$ (A-only group). This negative correlation indicates that participants with the lowest pretest scores showed the highest rates of improvement. This is the opposite of the pattern found by Bradlow et al. (1997) whose results indicated that the higher pretest scores predicted greater effectiveness of their training. However, the two studies are difficult to compare. Bradlow et al. conducted an auditory training study of American English /r/ and /l/ with L1 Japanese speakers. Perceptual accuracy of sounds involving spectral differences such as /r/ and /l/ is usually more challenging for L2 learners than perception of contrasts involving temporal differences, which is the case in the current study, and has been found for other similar phenomena such as VOT differences (e.g., Bohn, 1995). In terms of the relationship between the production pretest results and production improvement after training, the Spearman rank-order correlation coefficient for both groups was negative and did not reach significance, $r_s = -.170$, $p > .05$ (AV group), and $r_s = -.038$, $p > .05$ (A-only group).

Post-Study Interview Comments

Each participant was interviewed by one of the researchers at the conclusion of the study. The primary objective was to determine how effective they perceived this type of training to be, and the ease of use of the Web-based format. Following are the predominant comments from the participants in both the AV and A-only training groups:

- Geminate /s/ was the most difficult of the geminates to identify.
- Training increased their awareness of the singleton-geminate difference. For those who received AV training, waveform displays showed them where to focus their attention.
- Participants liked the flexibility of the Web-based format.
- Most of them indicated they took the training using their computers in their dorm rooms in the evening after dinner.
- Several participants recommended the development of additional training programs using a similar Web-based format to address other areas of language learning.

GENERAL DISCUSSION

This study investigated the effectiveness of auditory-visual training using waveform displays for the improvement of L2 geminate perception by beginning learners of Japanese. Consistent with previous studies, results of the current investigation demonstrate that perception training is effective, generalizes to novel stimuli, and transfers to improvement in production. AV training provided significantly better results compared to A-only training. Findings emphasize that visual cues can be a valuable source of input in L2 learning. With minimal instruction, learners were able to benefit from waveform displays to focus their attention on critical duration cues.

The AV and A-only training group results showed several similarities. The most difficult geminate-vowel sequence for both groups in perception and production was geminate /s/ followed by /u/, which supports the findings of previous research (Hardison & Motohashi-Saigo, in press). Geminate /s/ was also the consonant that showed the greatest amount of improvement in perceptual accuracy by both groups. Neither group had difficulty perceiving or producing singletons. Scores in the pretest showed little variation across singleton-vowel sequences or conditions. In addition, contrary to one of our hypotheses, their identification accuracy of geminates did not differ significantly according to presentation condition (isolated words vs. carrier sentences). The sentences may not have been long enough to be sufficiently challenging to the speech segmentation process.

Although perception training resulted in improved production based on native-speaker rater judgments, the dominant error patterns differed between these skills. The majority of misperceptions of geminates for both groups favored the response option with the long vowel (vs. singleton consonant). These findings suggest that even beginning-level learners are able to perceive moraic weight but may have difficulty attributing it to the correct segment. Learners may have detected the longer duration of the vowel preceding the geminate and misidentified it as two morae instead of one. They may have treated this point as the identification point of the word, and failed to attend to the duration of the consonant. This consonant would then have been syllabified as a singleton to the second syllable of the word. For L1 English learners, vowel length is a strong cue that is compatible with a segmentation strategy that syllabifies the consonant as the singleton onset of the second syllable, lengthening the preceding vowel in the resulting open syllable. On the other hand, in terms of production, when learners in both groups failed to produce a geminate, the majority of their errors involved a singleton, suggesting they were not yet consistently producing the necessary duration difference.

The above findings support hypotheses put forth in earlier studies that learning in perception and production are linked (e.g., Bradlow et al., 1997; Hardison, 2003; Lively et al., 1993), and add to the growing body of literature on the performance variability between these domains across individual learners. Similar to the findings of Bradlow et al. (1997), the relative perception and production rates of improvement in the current study showed no significant relationship. This suggests that degree of learning in one domain is, to some extent, separate from that in the other. In contrast to the Bradlow et al. (1997) study, we found a negative rank-order correlation between learners' pretest perceptual accuracy and their relative perceptual improvement as a result of training. Those learners with the lowest pretest scores showed the greatest amount of improvement.

It may be somewhat difficult to generalize performance by the participants in the current study to other language learners. The learners in this study chose to participate in a study abroad program in Japan, and were not language majors. Although English was the language of the majority of their classes and of the interactions in the dormitory, they were motivated to improve their perception and production of the language. These factors may place them towards the high end of a motivation scale.

The data suggest that easily accessible Web-based tools that utilize auditory-visual input provide a boost for learners. Interview results confirmed that these students enjoyed being able to access this type of information at their convenience and to receive immediate feedback on their judgments. Although the researcher forfeits some control over this process, the user-friendly approach encourages participation and satisfaction. These types of tools are recommended as additional sources of learning for students to supplement classroom instruction and interaction opportunities with native speakers of the target language.

APPENDIXES

Appendix A: Stimuli (Tokyo dialect): Pretest, Posttest, Training

Singletons	Translation	Geminates	Translation
sasu	stab	*sassu*	guess
asu	(nonword)	*assu*	dominate
saku	tear (verb)	*sakku*	sack
aku	evil	*akku*	bad mouth (saying something one shouldn't)
satu	volume	*sattu*	(nonword)
atu	pressure	*attu*	(nonword)
sasa	(nonword)	*sassa*	quickly
asa	morning	*assa*	(nonword)
saka	refreshments	*sakka*	last summer
aka	red	*akka*	money of poor quality
sata	trouble	*satta*	went
ata	(nonword)	*atta*	existed

Note. All stimuli have a H(igh)-L(ow) accent pattern (HLL in the case of geminates). Filler items with long vowels (nonwords): kaaze, tachuu, kaate, seechi, kiike, kaaki.

Appendix B: Stimuli (Tokyo dialect): Test of Generalization

Singletons	Translation	Geminates	Translation
*sati/sachi**	happiness; products from nature (e.g., fish)	*satti/sacchi**	to know by guessing
sate	by the way; and then; a word that is said when one moves on to the next activity	*satte*	conjugated form (*te* form) of the verb *saru* 'to leave'
saki	the part of writing that is added later to show more detailed explanations	*sakki*	Meaning varies according to kanji used. Options: a little while ago; to record one's opinion/reaction from reading books; unhappiness
sake	salmon	*sake*	(nonword)
*sasi/sashi**	a bite; truncation of *sashimi* 'to eat raw fish'	*sassi/sasshi**	a thin book; blush; paper for printing
sase	imperative form of *sasu* 'to bite or pierce'	*sasse*	(nonword)
sapi	(nonword)	*sappi*	(nonword)
sape	(nonword)	*sappe*	(nonword)

Note. All stimuli have a H(igh)-L(ow) accent pattern (HLL in the case of geminates). Filler items with long vowels (nonwords): *saatsu, kazee, keeta, taatsu, kuute, taachu*.

* = spelling depends on the Romanization alphabet that is used; the second form represents the palatalized pronunciation of /t/ and /s/ in these contexts (Tsujimura, 1996).

NOTES

1. Portions of this study were presented at the *American Association for Applied Linguistics* Conference, Costa Mesa, California in April, 2007.

2. There are other types of special morae such as a moraic nasal, and the second half of a diphthongal vowel sequence. See Kubozono (1999) and Tsujimura (1996) for detailed discussions of the mora and syllable in Japanese and the mora's role in phonological processes. See also B. Hayes (1989) for a discussion of the status of the geminate in moraic phonology.

3. 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 [ɯ] ('a turned m') in Japanese.

4. There are different sonority scales, which show slight variation in some values (e.g., Blevins, 1995; Parker, 2002); however, there is consensus that vowels (voiced) are the most sonorous and stops are the least.

5. Previous research demonstrated that word status (i.e., real word vs. nonword) did not significantly influence segmental perception by L2 learners in a forced-choice identification task (e.g., Lively, Logan, & Pisoni, 1993). In such a task, the word serves as a carrier of phonetic context.

6. Created by Boersma (2001) and Weenink. Available free of cost at http://www.fon.hum.uva.nl/praat.

7. Usually in tests of generalization, researchers manipulate one of these variables at a time (i.e., a test involving novel stimuli, and a separate one involving familiar stimuli with a new talker). Due to time constraints, we decided to challenge the learners' perception abilities with a test of generalization involving both novel stimuli and a new voice.

8. Available at http://www.animo.co.jp/EN/products/analyze/index.jsp.

9. A forced-choice task (vs. an open-response task) was preferred in order to avoid the issue of participants' spelling not accurately representing their perception.

10. This tool was from a Web site in Japanese http://www.fureai.or.jp/~irie/webquiz. There are several Web quiz creating tools available in English.

ABOUT THE AUTHORS

Miki Motohashi-Saigo, Japanese Language Department, Kansai Gaidai University, Japan; Debra M. Hardison, Department of Linguistics and Languages, Michigan State University. Correspondence concerning this article should be addressed to Debra M. Hardison, Michigan State University, A-714 Wells Hall, East Lansing, MI 48824.

E-mail: hardiso2@msu.edu

REFERENCES

Blevins, J. (1995). The syllable in phonological theory. In J. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 206-224). Cambridge, MA: Blackwell.

Boersma, P. (2001). PRAAT: A system for doing phonetics by computer. *Glot International 5*, 341-345.

Bohn, O.-S. (1995). Cross-language perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 370-410). Timonium, MD: York Press.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R. A., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101,* 2299-2310.

Chun, D. M. (2002). *Discourse intonation in L2: From theory and research to practice.* Amsterdam: Benjamins.

Chun, D. M., Hardison, D.M., & Pennington, M. C. (2008). Technologies for prosody in context: Past and future of L2 research and practice. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 323-346). Amsterdam: Benjamins.

Cutler, A. (1997). The comparative perspective on spoken-language processing. *Speech Communication, 21,* 3-15.

Cutler, A., & Otake, T. (2002). Rhythmic categories in spoken-word recognition. *Journal of Memory and Language, 46,* 296-322.

Fukui, S. (1978). Perception for the Japanese stop consonants with reduced and extended durations. *The Bulletin of the Phonetic Society of Japan, 159,* 9-12.

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

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

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context and talker variability. *Applied Psycholinguistics, 24,* 495-522.

Hardison, D. M. (2004). Generalization of computer assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology, 8,* 34-52. Retrieved January 13, 2009 from http://llt.msu.edu/vol8num1/hardison.

Hardison, D. M. (2005). Contextualized computer-based L2 prosody training: Evaluating the effects of discourse context and video input. *CALICO Journal, 22,* 175-190.

Hardison, D. M., & Motohashi-Saigo, M. (in press). Development of perception of L2 Japanese geminates: Role of duration, sonority, and segmentation strategy. *Applied Psycholinguistics.*

Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry, 20,* 253-306.

Hayes, R. L. (2002). The perception of novel phoneme contrasts in a second language: A developmental study of native speakers of English learning Japanese singleton and geminate consonant contrasts. *Coyote Papers, 12,* 28-41. Retrieved January 13, 2009, from http://coyotepapers.sbs.arizona.edu/CPXII.htm.

Hirata, Y. (1990). Perception of geminated stops in Japanese word and sentence levels. *Journal of the Phonetic Society of Japan, 194,* 23-28.

Johnson, K. (1997). *Acoustic & auditory phonetics.* Cambridge, MA: Blackwell.

Kubozono, H. (1999). Mora and syllable. In N. Tsujimura (Ed.), *The handbook of Japanese linguistics* (pp. 31-61). Malden, MA: Blackwell.

Ladefoged, P. (2003). *Phonetic data analysis: An introduction to fieldwork and instrumental techniques.* Malden, MA: Blackwell.

Levis, J., M., & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System, 32,* 505-524.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, *94,* 1242-1255.

Minagawa, Y., & Kiritani, S. (1996). Discrimination of the single and geminate stop consonant in Japanese by five different language groups. *Annual Bulletin, Research Institute of Logopedics and Phoniatrics (RIPL), 30,* 23-28. Retrieved January 13, 2009 from http://www.umin.ac.jp/memorial/rilp-tokyo/R30/R30_023.pdf.

Ofuka, E. (2003). Perception of a Japanese geminate stop /tt/: The effect of pitch type and acoustic characteristics of preceding/following vowels. *Journal of the Phonetic Society of Japan, 7,* 70-76.

Parker, S. G. (2002). *Quantifying the sonority hierarchy*. Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Port, R. R., Dalby, J., & O'Dell, M. (1987). Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America, 81,* 1574-1585.

Selkirk, E. (1984). On the major class features and syllable theory. In M. Aronoff & R. T. Oehrle (Eds.), *Language sound structure* (pp. 107-136). Cambridge, MA: MIT Press.

Tsujimura, N. (1996). An introduction to Japanese linguistics. Cambridge, MA: Blackwell.

Wright, R. (2004). A review of perceptual cues and cue robustness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 34-57). Cambridge: Cambridge University Press.

Yamagata, A., & Preston, D. (1999). English learners' acquisition of the katakana spelling of English loanwords in Japanese. In M. Wysocka (Ed.), *On language theory and practice. In honor of Janusz Arabski on the occasion of his 60th birthday* (pp. 276-292). Katowice, Poland: University of Silesia.

Yoshida, Y. A. (2006). Accents in Tokyo and Kyoto Japanese vowel quality in terms of duration and licensing potency. *SOAS (School of Oriental and African Studies) Working Papers in Linguistics, 14,* 249-264.