Journal of Phonetics 66 (2018) 242–251

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

High variability identification and discrimination training for Japanese speakers learning English /r/–/l/ ☆

Yasuaki Shinohara [a,*], Paul Iverson [b]

[a] Waseda University, 1-6-1, Nishi-waseda, Shinjuku-ku, Tokyo 169-8050, Japan
[b] University College London, Department of Speech, Hearing and Phonetic Sciences, 2 Wakefield Street, London WC1N 1PF, United Kingdom

ARTICLE INFO

Article history:
Received 15 March 2017
Received in revised form 15 November 2017
Accepted 16 November 2017
Available online 7 December 2017

Keywords:
Speech perception
Speech production
Second language
Phonetic training

ABSTRACT

Second-language (L2) learners can benefit from exposure to phonetically variable speech during computer-based training. Moreover, this training can be effective even for L2 learners who have extensive exposure to their L2 in daily life, suggesting that there is something specific about the training task that aids learning. The present study compared traditional identification training with discrimination training to evaluate whether discrimination training could be effective, and whether different types of focused attention (i.e., on categorization vs. perceptual differences) could combine to provide a greater increase in learning. Adult Japanese speakers were given 10 sessions of identification and discrimination training, with pre/mid/post tests of identification, auditory discrimination, category discrimination, and /r/–/l/ production. The results demonstrated that both identification and discrimination training increased accuracy of Japanese speakers' perception and production of English /r/–/l/ in similar ways, but that there was little added benefit to using the two training methods in combination. It thus appears that identification and discrimination training have similar effects in second-language learners, as long as both training methods incorporate high variability.

1. Introduction

It is well established that learning second-language (L2) phonetic contrasts in adulthood can be difficult, particularly when the phonetic contrasts mismatch the processing and representations that have been developed for one's first language (L1; e.g., Best, 1995; Flege, 1995; Iverson & Kuhl, 1995). That being said, learning is possible for difficult contrasts after many years of real-world experience (Flege, Takagi, & Mann, 1996; Ingvalson, McClelland, & Holt, 2011; MacKain, Best, & Strange, 1981; Takagi & Mann, 1995) or intensive exposure to naturalistic variability in computer-based phonetic training programs (e.g., Lively, Logan, & Pisoni, 1993; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Logan, Lively, & Pisoni, 1991). One could assume that high-variability phonetic training works because it simulates the kind of exposure to phonetically variable speech that individuals receive during real-life experience. However, even the earliest studies hypothesized that there were specific aspects of the high-variability training methods, such as the opportunity for focused attention, that likely affected learning (e.g., Logan et al., 1991). Subsequent work has demonstrated that phonetic training improves perception even for experienced L2 speakers who use their L2 daily (e.g., Iverson, Pinet, & Evans, 2012), which would only happen if phonetic training made a unique contribution to the learning process beyond exposure to phonetically variable speech.

The exact contribution of phonetic training to L2 learning is still unclear, but it tends to have more focused effects than real-world learning. For example, individual differences in perception and production of L2 contrasts tend to be highly correlated among language learners (i.e., better perceivers tend to be better speakers), but phonetic training often improves production and perception more idiosyncratically, with low individual-differences correlations in improvement (e.g., Bradlow, Pisoni, Yamada, & Tohkura, 1997; Bradlow, Akahane-Yamada, Pisoni, and Tohkura, 1999; Huensch & Tremblay, 2015; Iverson et al., 2012); phonetic training thus seems to produce less global improvements in learning. Sadakata and McQueen (2013) have suggested that high-variability phonetic

☆ A portion of this work was presented in "Perceptual training effects on production of English /r/–/l/ by Japanese speakers" at the Phonetics Teaching and Learning Conference, London, UK, 8–10 August 2013.

* Corresponding author.
E-mail address: y.shinohara@waseda.jp (Y. Shinohara).

https://doi.org/10.1016/j.wocn.2017.11.002
0095-4470/© 2017 The Authors. Published by Elsevier Ltd.

training specifically improves abstract category representations rather than tuning auditory processing for speech. Iverson, Hazan, and Bannister (2005) have likewise argued that high-variability phonetic training does not change lower-level processing, but have further suggested that training improves how automatically listeners can apply their existing category knowledge and processing to real speech, rather than fundamentally changing category representations (e.g., cue weightings).

The aim of the present study was to produce larger improvements in L2 phonetic processing by combining different training methods (i.e., identification and discrimination) that may affect different underlying processes involved in phonetic perception and production. For the most part, improvements after standard high-variability phonetic identification training, particularly for Japanese adults learning to distinguish English /r/–/l/, have been limited to about 15 percentage points, suggesting that improvements due to this method alone can reach a ceiling and that other processes which underlie L2 phonetic perception and production need to be improved (Bradlow et al., 1999, 1997; Iverson et al., 2005; Lively et al., 1993, 1994; Logan et al., 1991; MacKain et al., 1981). Beyond searching for practical improvements in learning, we aimed to compare discrimination and identification methods to examine to what extent the different types of focused attention required during these tasks are able to improve different aspects of L2 phonetic processing.

Previous attempts at discrimination training have had limited success. For example, Strange and Dittmann (1984) gave Japanese adults same-different discrimination training along a synthesized *rock-lock* continuum. The results demonstrated that discrimination training improved identification and discrimination at the /r/–/l/ category boundary for these synthetic stimuli, but this improvement did not generalize to novel stimuli or natural recordings. In contrast, subsequent work by Logan et al. (1991) found that generalization could be obtained from identification training that used natural recordings of minimal-pair words spoken by multiple talkers. Logan et al. hypothesized that their technique was more effective because it allowed listeners to form robust categories that extend to varied phonetic contexts, whereas discrimination training primarily increases sensory resolution overall, including to within-category variation that likely interferes with categorization. However, there were many differences between the techniques of Strange and Dittmann (1984) and Logan et al. (1991), particularly related to the variability of the stimuli, making it unclear to what extent the differences in the findings were due to discrimination vs. identification training. If discrimination training was able to alter sensitivity to acoustic variation in a targeted way (e.g., improving the primary acoustic cue sensitivity at category boundaries without raising within-category sensitivity), then this might prove to be an effective supplement to identification training.

The English /r/–/l/ categories are primarily distinguished by L1 speakers in terms of F3 frequency, with a lower F3 frequency for /r/ and a higher frequency for /l/, along with secondary cues such as closure duration (shorter for /r/ and longer for /l/) and transition duration (longer for /r/ and shorter for /l/), and variation in F2 frequency related to dark/light articulation that is mostly orthogonal to the /r/–/l/ contrast (e.g., Hattori & Iverson, 2009; Underbakke, Pola, Gottfried, & Strange, 1988; Yuan & Liberman, 2011). It has been claimed that Japanese adults have difficulty learning this distinction because both English /r/ and /l/ are assimilated into their L1 tap category (Best & Strange, 1992), or because their L1 tap is similar enough to English /l/ to block the formation of a new English /l/ category (Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004). However, Hattori and Iverson (2009) found that individual differences in L1–L2 category assimilation, or the distance between these categories, is poorly correlated with identification accuracy. Moreover, they found that Japanese adults are able to learn secondary cues for /r/ and /l/, but have a highly specific difficulty with learning F3. Ingvalson et al. (2011) similarly demonstrated a positive correlation between F3 cue weighting and /r/–/l/ identification, but also showed that F3 reliance is not changed as a function of length of residence in an English-speaking country, age of arrival in an English-speaking country, the amount of Japanese use, or length of student status in an English environment (English education). It has been argued that this difficulty with F3 primarily stems from a pre-categorical level of phonetic processing (Iverson et al., 2003; Iverson, Wagner, & Rosen, 2016). That is, Japanese adults tend to be less sensitive to F3 variation near the English /r/–/l/ category boundary and more sensitive to irrelevant variation, and it is this perceptual warping of the cue variance at an early level that affects how the cues are represented at a more abstract categorical level. It would thus be desirable if this earlier level of processing could be altered through training, rather than only using identification training to improve more abstract representations.

In the present study, Japanese speakers were trained on both identification (ID) and discrimination (DIS). The training for each method lasted for five sessions, and the order of the training programs was balanced across subjects so that half of the subjects performed identification training first (ID-DIS) and the other half performed discrimination training first (DIS-ID). Identification training, which was intended to improve category representations (Gerrits & Schouten, 2004; Logan et al., 1991; Sadakata & McQueen, 2013; Sjerps, McQueen, & Mitterer, 2013), used a standard high-variability technique from a previous study (Iverson et al., 2005). The discrimination training, which was intended to improve pre-categorical processing (Gerrits & Schouten, 2004; Logan et al., 1991; Sadakata & McQueen, 2013; Sjerps et al., 2013; Strange & Dittmann, 1984), used 20 stimulus continua based on signal-processed natural stimuli (four contrasts each from five talkers), as well as fully natural stimuli. Listeners performed three types of three-interval oddity tasks: auditory discrimination with natural stimuli (i.e., two natural stimuli that were identical and one that was different, where the acoustic differences were uncontrolled), auditory discrimination with signal-processed stimuli (where the different stimulus was specifically different in terms of F3), and category discrimination (three natural stimuli that were all acoustically different, but one started with a different phoneme). It was intended that the auditory discrimination tasks would improve F3 perception at the boundary in a way that would generalize to other stimuli, and that category discrimination would additionally decrease sensitivity to irrelevant acoustic variation (e.g., Flege, 2003; Gerrits & Schouten, 2004; Højen & Flege, 2006; Iverson et al., 2003, 2012; Logan et al., 1991; Sadakata & McQueen, 2013; Sjerps et al., 2013).

The training effects were evaluated using three perception tasks: (1) identification of English /r/–/l/ minimal-pair words; (2) auditory discrimination of F3 at the English /r/–/l/ boundary, F2 at the English /r/–/l/ boundary and F3 within the English /r/ category; and (3) category discrimination. Production was tested by having listeners read isolated words as well as longer stories, as part of the battery of tests that might expose differences between identification and discrimination training.

2. Method

2.1. Subjects

Fifty-five native Japanese speakers completed the pre-training test. Of these subjects, 12 were not trained because their identification accuracy for /r/ and /l/ was more than 75% correct, and two were excluded because they were not able to finish all sessions. This left 41 subjects in total for the data analysis.

All subjects were native Japanese speakers with no self-reported hearing impairments. As shown in Table 1, the two trainer order groups (ID-DIS, DIS-ID) were balanced in terms of age, sex, tested location (UK, Japan), English experience (i.e., length of living in English speaking countries), and English word-initial /r/–/l/ identification accuracy (M = 63.50%, SD = 8.97 for ID-DIS; M = 61.31%, SD = 11.11 for DIS-ID). Twenty-two subjects were recruited in London, UK, and 19 subjects were collected in the Kanto area of Japan.

2.2. Stimuli

The natural stimuli were the same as those recorded by Iverson et al. (2005). One hundred word-initial /r/–/l/ minimal-pair words (e.g., *lay* and *ray*) produced by each of 10 standard southern British English (SSBE) speakers (5 females, 5 males) were used as training stimuli. An additional 120 stimuli produced by two SSBE speakers (1 female, 1 male) were used for testing (i.e., not part of the training set), consisting of 40 word-initial /r/–/l/ minimal-pair words (e.g., *rake* and *lake*, 20 words × 2 speakers), 40 word-medial /r/–/l/ minimal-pair words (e.g., *berries* and *bellies*, 20 words × 2 speakers), and 40 consonant cluster /r/–/l/ minimal-pair words (e.g., *fresh* and *flesh*, 20 words × 2 speakers).

Signal-processed versions of the natural recordings were created for auditory discrimination training using LPC analysis and resynthesis in Praat. In short, LPC filters were calculated from natural recordings (i.e., estimating how the formants changed over time for each utterance), then were manipulated (e.g., altering F3), and used to filter a neutral LPC residual. Specifically, the LPC filters for each stimulus had formant information from F1 to F4, and all the formant contours except F3 were averaged between /r/–/l/ minimal-pair words, creating stimuli with only F3 differences. F3 was interpolated in six steps for each minimal pair. After filtering, each stimulus was given pitch and amplitude envelopes that were averaged between the /r/ and /l/ minimal-pair words. The stimuli with only F3 differences were piloted with a group of 13 native British English speakers, to establish how well they could be identified. Four minimal pairs whose identification function curves were steep enough to have a clear phoneme boundary were selected for each speaker. After the four minimal pairs were selected, one hundred stimuli were interpolated between /r/ and /l/ for each minimal-pair, and two stimulus pairs for each series (i.e., close and distant pairs) were selected that bracketed the identification boundary for use in training. Training stimuli have to be close enough to the phoneme boundary to increase the F3 sensitivity at the phoneme boundary, and the close pairs were selected to be identified correctly as /r/ or /l/ by native speakers on 70% of trials. The distant pairs were farther from the boundary to be identified correctly as /r/ or /l/ by native speakers on 85% of trials. These 70% and 85% were selected based on the results of the pilot study.

Synthetic English /rɑ/ and /lɑ/ stimuli were created for an auditory discrimination test (Klatt & Klatt, 1990), which were adapted from stimuli used in a previous study (Hattori & Iverson, 2009). Stimulus pairs were created that contrasted F3 at the English /r/–/l/ boundary, F2 at the English /r/–/l/ boundary, and F3 within the English /r/ category. As described in Table 2, the three stimulus pairs were manipulated with five acoustic cues: closure duration, transition duration, F1, F2, and F3. For the stimulus pair contrasting F3 at the boundary, the acoustic cues of closure duration, transition duration, F1, and F2 were set at 64 ms, 48 ms, 327 Hz, and 1196 Hz. The target F3 was set at 2639 Hz for the /r/ stimulus and 3328 Hz for the /l/ stimulus. English speakers are better at discriminating F3 at the boundary than are Japanese speakers (e.g., Iverson et al., 2003), and F3 sensitivity at the boundary thus could selectively improve with training. For the stimulus pair contrasting F2 at the boundary, the acoustic cues of closure duration, transition duration, F1, and F3 were set at 64 ms, 48 ms, 327 Hz, and 2965 Hz. Low F2 was set at 1051 Hz and high F2 was set at 1358 Hz. Japanese speakers are better at discriminating F2 than English speakers (Iverson et al., 2003), and F2 sensitivity thus could decline with training. Finally, for the stimulus pair contrasting F3 within the English /r/ category, the acoustic cues of closure duration, transition duration, F1, and F2 were set at 31 ms, 81 ms, 327 Hz, and 1196 Hz. The low F3 was set at 1739 Hz and the high F3 was set at 2212 Hz; both stimuli were on the /r/ side of the boundary. Similarly to F2, Japanese speakers are relatively more sensitive to F3 within the /r/ category compared to English speakers, so that it was expected for them to decrease its perceptual sensitivity.

Table 1
Subject information.

Trainer order	Number of subjects (female, male)	Age range (median)	Tested location (number of subjects)	Length of living in English-speaking countries (median)
Identification-Discrimination (ID-DIS)	20 (14, 6)	20–61 years (25.5 years)	UK (11) Japan (9)	0–257 months (2 months)
Discrimination-Identification (DIS-ID)	21 (14, 7)	21–57 years (25 years)	UK (11) Japan (10)	0–245 months (2.4 months)

Table 2
Stimulus information for the auditory discrimination test.

Stimulus pair	Acoustic contrast	Closure duration (ms)	Transition duration (ms)	F1 (Hz)	F2 (Hz)	F3 (Hz)
F3 at the boundary	High F3 (English /l/)	64	48	327	1196	3328
	Low F3 (English /r/)					2639
F2 at the boundary	High F2	64	48	327	1358	2965
	Low F2				1051	
F3 within English /r/ category	High F3 within /r/	31	81	327	1196	2212
	Low F3 within /r/					1739

2.3. Procedure

2.3.1. Training

All subjects took part in both identification and discrimination training, assigned to either the identification-discrimination (ID-DIS) or the discrimination-identification (DIS-ID) order. Subjects in the identification-discrimination (ID-DIS) trainer order group were given five sessions of identification training first followed by five sessions of discrimination training; subjects in the discrimination-identification (DIS-ID) trainer order group had the reverse order. Each training session took approximately 30 min, and subjects completed all 10 training sessions within 10 to 28 days. The speaker of the training stimuli was different each day through the 10 sessions, but the speaker order within each training program was the same between subjects.

The identification training program was the same as that used in Iverson et al. (2005). The subjects completed 300 two-alternative forced choice trials per session that comprised 100 word-initial /r/–/l/ minimal-pair words (50 pairs) repeated 3 times each. The computer screen displayed a minimal pair (e.g., *rock* and *lock*) with a single auditory token (e.g., *rock*; no opportunity to repeat stimuli), the listeners clicked the word that they thought they heard, and they received feedback. When they clicked on a correct answer, they saw a message *Correct* on screen and heard a cash register sound. The answer was highlighted, and the stimulus was replayed once. When they clicked on a wrong answer, they saw a message *Wrong* on the screen and heard two descending beep sounds. The correct answer was highlighted, and the stimulus was replayed twice.

In the discrimination training, there were three tasks: auditory discrimination with natural recordings (40 trials, 20% of all trials), auditory discrimination with signal-processed stimuli (80 trials, 40% of all trials), and category discrimination (80 trials, 40% of all trials). All tasks used a three-alternative forced choice discrimination judgment with feedback. Three numbers were displayed on the screen, and listeners clicked on the one whose word-initial phoneme was categorically different from the other two. The first auditory discrimination task used natural recordings of minimal-pair trials spoken by the same talkers, so that subjects could use any acoustic differences to discriminate the odd stimulus (e.g., *red*, *red, led*). The minimal-pair stimuli were randomly chosen from the 50 minimal pairs that were used for identification training. The second auditory discrimination task used signal-processed stimuli, and F3 was the only difference between minimal-pair trials; subjects needed to discern the F3 difference between the three stimuli to answer correctly. The stimuli were the close and distant pairs described in the stimulus section. The final task, category discrimination, had three different words spoken by the same talker; subjects needed to focus on the critical phonetic differences between stimuli and ignore irrelevant acoustic variation. For example, if the three stimuli were *red*, *light*, and *link*, subjects were supposed to choose *red*, because the initial phoneme of *red* was different from the other two stimuli. Subjects did not have to identify the phoneme but needed to focus on phonological differences (Strange & Shafer, 2008). For all the discrimination tasks, subjects were not able to repeatedly listen to stimuli and feedback was given for each trial, as in identification training, except that the three-stimulus sequences were replayed once for wrong answers and not replayed for correct answers. In total, the discrimination training program had 200 trials, such that listeners spent approximately the same amount of time, not the same number of trials, as with the identification training program.

After completing each training session, subjects performed a short identification test of 20 trials. The stimuli were spoken by the same talker that was used for training in that particular session, and the percentage correct in the short test was shown to subjects at the end to track their improvements through the course of the 10 training sessions. Due to a technical problem, the short identification test results at the 10th session from four subjects in the ID-DIS order and two subjects in the DIS-ID order were missing and not included in the analysis.

2.3.2. Pre/Mid/Post test

All subjects were tested three times: at the start of the experiment (pre test), after five sessions of the first training program (mid test), and after another five sessions of the second training program (post test). There were three perceptual tasks in each test: identification, auditory discrimination, and category discrimination. For identification, subjects heard English /r/–/l/ minimal-pair words and chose /r/ or /l/ based on what they thought they heard. To investigate generalization, the stimuli were untrained English minimal-pair words contrasting /r/–/l/ at initial, medial, and consonant cluster positions. Forty minimal-pair words (20 pairs) produced by two SSBE speakers (20 words from a female and the remaining 20 words from a male) were used for each position, so that subjects completed 120 trials at each of the pre, mid, and post tests. The two SSBE speakers were not included in the training corpora and each of the speakers produced different words. For auditory discrimination, subjects heard three synthetic stimuli and chose the one that sounded different from the other two. They were tested with three stimulus pairs to examine F3 sensitivities at the boundary, F2 sensitivities at the boundary, and F3 sensitivities within the /r/ category. Each subject was given 72 trials (i.e., 3 stimulus pairs $\times$ 2 possible odd stimuli $\times$ 3 possible odd stimulus positions $\times$ 4 repetitions for each). For category discrimination, subjects heard three different English words and chose the one beginning with a different phoneme from the other two stimuli. The same four sets of stimuli were used at pre, mid, and post tests. They completed 48 trials with two possible odd stimuli (/r/ or /l/), three positions for the odd stimulus, four sets of stimuli, and two talkers (a female and a male). During the tests, subjects did not receive any feedback, and they were not allowed to repeatedly listen to trials. Each of identification, discrimination and category discrimination tests took approximately 5–10 min.

There were two production tasks of word and passage reading. In the word-reading task, Japanese speakers read 40 /r/–/l/ minimal-pair words randomly displayed on the computer screen one by one, and 10 of these minimal-pair words (i.e., *race-lace*, *road-load*, *root-loot*, *rung-lung*, and *wrist-list*) were acoustically analyzed. In the passage-reading task, subjects read an excerpt of "The Rainbow Passage" (Fairbanks, 1960); 13 word-initial /r/–/l/ words in the text were analyzed, as in Hattori (2009). Those were seven /r/-words (i.e., *raindrops*, *reach*, *round*, and *rainbow* × 4) and six /l/-words (i.e., *legend*, *light*, *long, look*, *looking*, and *looks*). In total, F3 frequencies of 2829 tokens were measured from the closure part of English /r/–/l/ (i.e., 10 for word reading and 13 for passage reading produced by 41 subjects at pre, mid, and post tests) in Praat, but 146 were excluded from the analyses due to difficulties with measuring F3, leaving 2683 tokens.

3. Results

Fig. 1 displays the identification accuracy of Japanese speakers in the two trainer orders at the pre, mid, and post tests. Although identification training improved identification accuracy more than did discrimination training, both identification and discrimination training improved identification accuracy.

Fig. 1. Boxplots of identification accuracy for the English /r/–/l/ contrast in two trainer orders, ID-DIS (left) and DIS-ID (right), at pre test (white boxes), mid test after five sessions of a first training method (cross-hatched boxes), and post test after five sessions of a second training method (black boxes), with individual subjects' data points (dots).

A logistic mixed effects model based on correct/incorrect binomial responses was used for the perception analyses. The best-fitting model was selected with a top-down approach (i.e., excluding ineffective random and fixed factors from a model with all potential factors) based on Akaike Information Criteria (AIC) with the alpha level of 0.05. The potential fixed factors were testing block (pre, mid, post), trainer order (ID-DIS, DIS-ID), /r/–/l/ position (word-initial, word-medial, consonant cluster) and all the possible interactions of these three factors. The best-fitting model included the fixed factors of testing block, trainer order, /r/–/l/ position, and two 2-way interactions between testing block and trainer order and between trainer order and /r/–/l/ position. The random factors were crossed intercepts for subject and word. The word factor was nested into speaker. The random factors also included both by-subject and by-word random slopes for testing block.

The logistic mixed effects model demonstrated that there was a significant main effect of testing block, $\chi^2(2) = 37.86$, $p < .001$, suggesting that high-variability training significantly improved Japanese speakers' English /r/–/l/ identification ability. Although there was no significant main effect of trainer order, $\chi^2(1) = 1.20$, $p > .05$, the interaction between testing block and trainer order was significant, $\chi^2(2) = 11.38$, $p < .01$. This suggests that the degree of the improvement in identification accuracy was different by trainer order.

Post-hoc analyses were conducted by repeating the mixed-model analysis for each pair-wise comparison. For the first five training sessions (pre vs. mid test), both identification and discrimination training methods significantly improved identification accuracy: identification, $\chi^2(1) = 38.47$, $p < .001$, *M*pre = 60.6%, *M*mid = 71.9%, discrimination, $\chi^2(1) = 7.29$, $p < .01$, *M*pre = 59.9%, *M*mid = 65.1%. However, the significant interaction between testing block (pre vs. mid) and trainer order (identification vs. discrimination) demonstrated that identification training improved the identification accuracy significantly more than discrimination training, $\chi^2(1) = 7.09$, $p < .01$. For the last five sessions (mid vs. post test), the identification accuracy significantly improved overall, $\chi^2(1) = 9.73$, $p < .01$, and the interaction between testing block and trainer order was marginally significant, $\chi^2(1) = 3.68$, $p = .055$. Identification training after having five sessions of discrimination training significantly improved the identification accuracy, $\chi^2(1) = 11.95$, $p < .001$, *M*mid = 65.1%, *M*post = 70.8%, whereas discrimination training after the five sessions of identification training did not significantly improve the identification accuracy, $\chi^2(1) = 0.39$, $p > .05$, *M*mid = 71.9%, *M*post = 72.3%. That is, both identification and discrimination training were useful for the improvement of identification accuracy, but identification training was more effective. Despite the combination of two different training programs, the improvement in word-initial identification accuracy was still limited to approximately 15% after the 10 training sessions, i.e., 62.38% at pre test to 77.32% at post test.

Although both the main effect of /r/–/l/ position, $\chi^2(2) = 9.84$, $p < .01$, and the interaction between trainer order and /r/–/l/ position, $\chi^2(2) = 12.09$, $p < .01$, significantly affected identification accuracy, neither the 2-way interaction between testing block and /r/–/l/ position nor the 3-way interaction of testing block, /r/–/l/ position and trainer order was included in the best-fitting model. This suggests that there was no significant difference in the improvement between the three /r/–/l/ positions. In short, training effects transferred to untrained /r/–/l/ positions.

Fig. 2 displays the increase of the identification accuracy through the 10 training sessions by Japanese speakers in two trainer orders. At the end of each training session, subjects had 20 trials of the identification test, and the stimuli used in the short tests were a part of training corpora produced by a speaker used in the particular session. To investigate how Japanese speakers improve their identification accuracy through both identification and discrimination training, each of the short test results were analyzed with a logistic mixed effects model. The best-fitting model included the fixed factors of training method (identification, discrimination), training session (1–5), training block (first five sessions, last five sessions), and two 2-way interactions between training session and training block and between training method and training block. The random factors were crossed intercepts of subject and word which was nested into speaker. The mixed-effects model demonstrated that although the main effect of training method was not significant, $\chi^2(1) = 2.35$, $p > .05$, both the main effects of training session, $\chi^2(1) = 3.89$, $p = .048$, and training block, $\chi^2(1) = 32.63$, $p < .001$, were significant. These results suggest that there was a significant improvement in the identification accuracy through the course of training sessions. Interestingly, the interaction between training method and training session did not fit in the model. This suggests that the identification accuracy improvement was not significantly different between the identification and discrimination training methods. However, the interaction between training session and training block was significant, $\chi^2(1) = 5.07$, $p = .024$, suggesting that Japanese speakers improved their identification accuracy through the first five training sessions more than the last five sessions (i.e., the improvement slope was steeper for the first five sessions than the last five sessions). The interaction between training method and training block was also significant, $\chi^2(1) = 9.57$, $p < .01$, but the 3-way interaction of training method, training session and training block did not fit in the model. In short, although the results of pre, mid, and post tests demonstrated that the identification training increased identification accuracy more than did the discrimination training, such difference was not observed in the identification data collected after each training session.

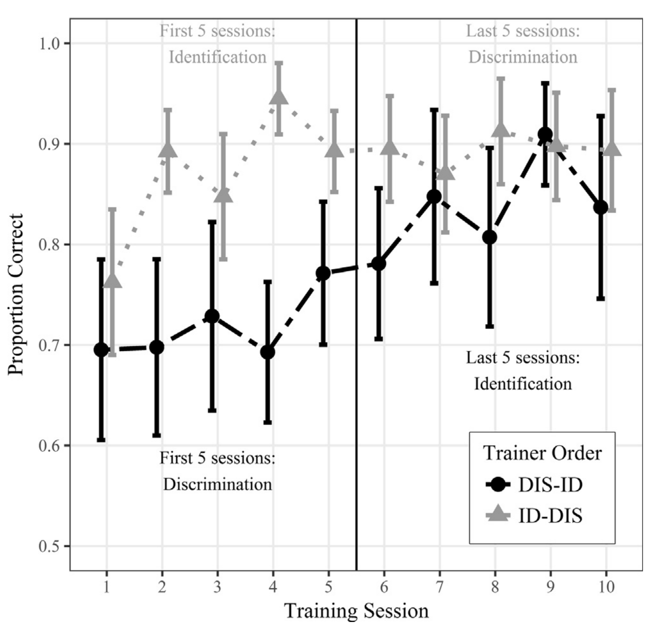

Fig. 2. Confidence intervals of the English /r/–/l/ identification accuracy at the end of each training session in two trainer orders, ID-DIS (gray) and DIS-ID (black).

Fig. 3 displays the auditory discrimination accuracy of each stimulus pair by Japanese speakers in two trainer orders at pre, mid, and post tests. As shown in the figure, Japanese speakers increased their perceptual sensitivities to F3 at the English /r/–/l/ boundary and F2 at the boundary. The best-fitting logistic mixed effects model for auditory discrimination included the fixed factors of testing block, trainer order, stimulus pair (F3 sensitivity at boundary, F2 sensitivity at boundary, F3 sensitivity within /r/ category), and two 2-way interactions between testing block and stimulus pair and between trainer order and stimulus pair. The random factors were crossed intercepts for subject and stimulus, and both by-subject and by-stimulus random slopes for testing block were included. The best-fitting model demonstrated that Japanese adults changed their discrimination with training, $\chi^2(2) = 12.52$, $p < .01$, but the interaction between testing block and trainer order did not fit in the best model, suggesting that there was no significant difference in the sensitivity change between the two trainer orders or between the two training methods. That is, both identification and discrimination training increased auditory discrimination accuracy to similar extents.

There was also a main effect of stimulus pair, $\chi^2(2) = 30.36$, $p < .001$, as well as a significant interaction between testing block and stimulus pair, $\chi^2(4) = 10.31$, $p = .036$. This suggests that there was a difference in the effect of training among the three stimulus pairs. Post-hoc analyses comparing the training effects between stimulus pairs demonstrated that the increase in the F2 sensitivity at the boundary was significantly higher than the increase in the F3 sensitivity within the /r/ category, $\chi^2(2) = 7.73$, $p = .021$, but there was no significant difference in the increase between F3 sensitivity at the boundary and F2 sensitivity at the boundary, $\chi^2(2) = 1.40$, $p > .05$, or between F3 sensitivity at the boundary and F3 sensitivity within the /r/ category, $\chi^2(2) = 3.84$, $p > .05$. Although both the F3 sensitivity at the boundary, $\chi^2(2) = 6.77$, $p = .034$, and F2 sensitivity at the boundary, $\chi^2(2) = 16.02$, $p < .001$, were significantly increased, the F3 sensitivity within the English /r/ category was not significantly changed, $\chi^2(2) = 2.47$, $p > .05$. These results suggest that Japanese speakers do not selectively improve only F3 sensitivity at the phoneme boundary, the primary acoustic cue for the English /r/–/l/ contrast.

Fig. 4 displays the category discrimination accuracy by Japanese speakers at pre, mid, and post tests. Similarly to the identification test, both identification and discrimination training methods improved Japanese speakers' category discrimination. The best-fitting logistic mixed effects model for the category discrimination included only testing block as the fixed factor, and the random factors were crossed intercepts for subject and stimulus, as well as by-subject and by-stimulus random slopes for testing block. The best-fitting model demonstrated that Japanese speakers significantly improved their category discrimination accuracy, $\chi^2(2) = 27.08$, $p < .001$. The interaction between testing block and trai-

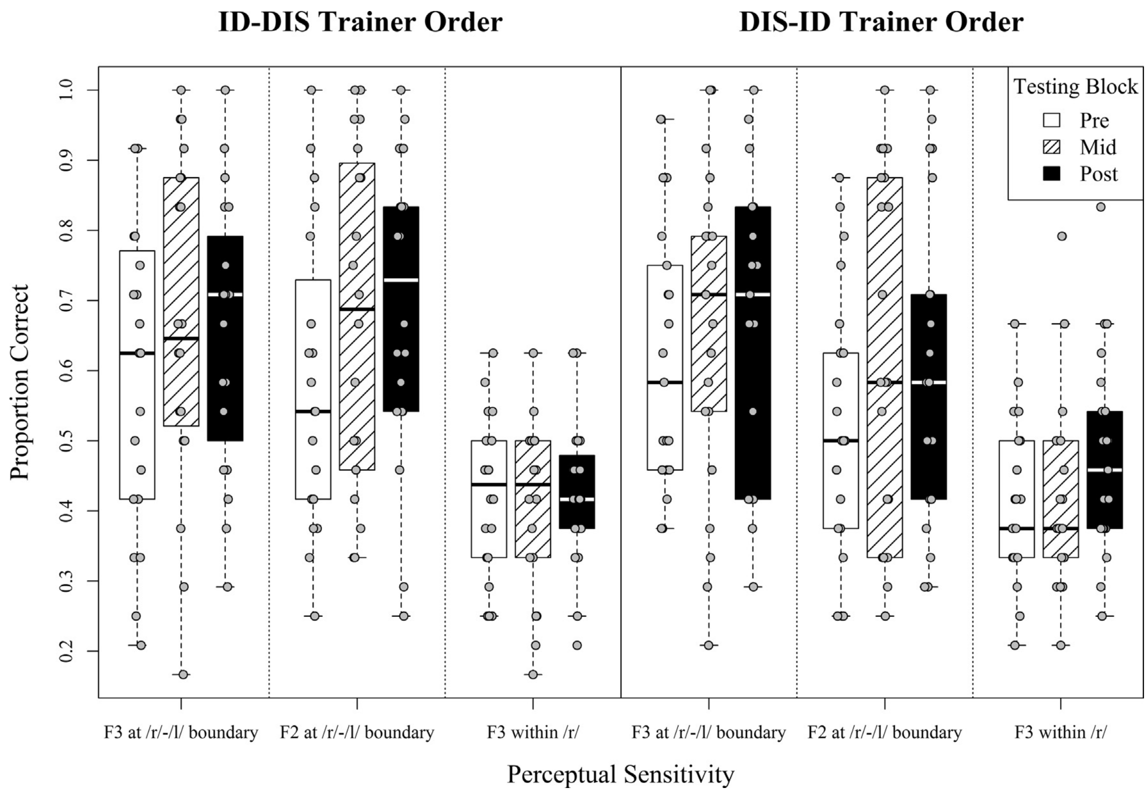

Fig. 3. Boxplots of auditory discrimination accuracy in three stimulus pairs testing F3 sensitivity at the English /r/–/l/ phoneme boundary, F2 sensitivity at the phoneme boundary, and F3 sensitivity within the English /r/ category, at pre (white boxes), mid (cross-hatched boxes), and post (black boxes) tests by Japanese speakers in two trainer orders, ID-DIS (left) and DIS-ID (right), with their individual data points (dots).

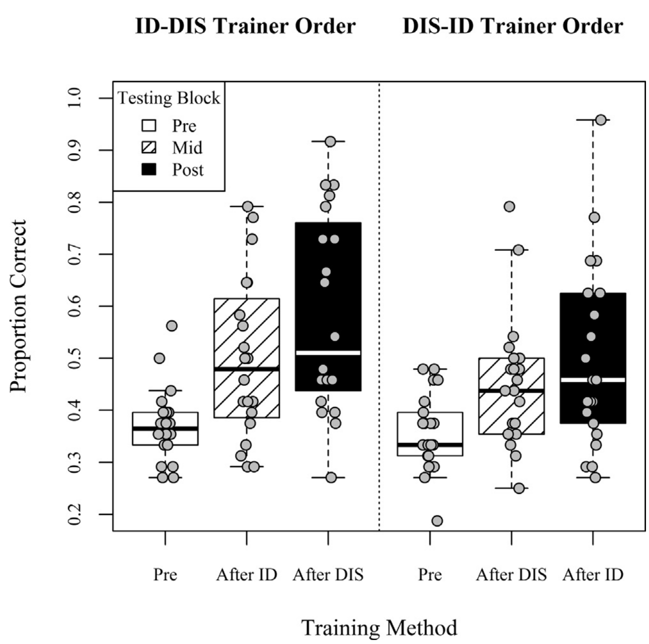

Fig. 4. Boxplots of category discrimination accuracy at pre (white boxes), mid (cross-hatched boxes), and post (black boxes) tests in two trainer orders, ID-DIS (left) and DIS-ID (right), with individual subjects' data points (dots).

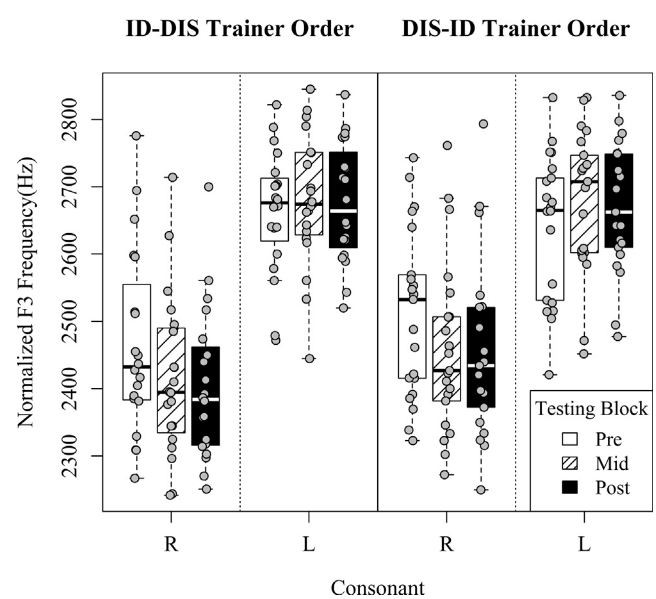

Fig. 5. Boxplots of normalized F3 frequencies of English /r/ and /l/ produced by Japanese speakers in two trainer orders, ID-DIS (left) and DIS-ID (right), at pre (white boxes), mid (cross-hatched boxes), and post (black boxes) tests, with individual subjects' data points (dots).

ner order did not fit in the best model, suggesting that there was no significant difference in the improvement between the two trainer orders or between the two training methods; identification and discrimination training improved category discrimination accuracy to similar extents.

Fig. 5 displays the acoustic measurements (F3 frequency) of the subjects' productions of English /r/ and /l/ at pre, mid, and post tests. The F3 values of all /r/–/l/ tokens were normalized to the median F3 in the passage for each subject. The median F3 frequency was measured from the passage recording for each subject, and it was subtracted from the F3 frequency of each token. The normalized F3 frequencies were rescaled for the analysis. Japanese speakers improved their production of the English /r/–/l/ contrast, lowering F3 for English /r/ and raising F3 for English /l/. A linear mixed effects model based on the normalized F3 measurements for English /r/–/l/ production was used for the analysis, and the fixed factors included in the best model were consonant (/r/, /l/), testing block, testing material (word, passage), trainer order, and 4 two-way interactions between consonant and testing block, between consonant and testing material, between testing block and testing material, and between consonant and trainer order. The random factors were intercepts of subject and word nested into subject, to take into account the articulatory difference between subjects.

The best-fitting model demonstrated that Japanese speakers significantly changed their F3 productions of these consonants with training (i.e., interaction between consonant and testing block), $\chi^2(2) = 50.70$, $p < .001$. There were likewise significant main effects of consonant, $\chi^2(1) = 516.24$, $p < .001$, and testing block, $\chi^2(2) = 10.49$, $p < .01$. Post-hoc analysis for each consonant demonstrated that Japanese speakers significantly lowered F3 for English /r/, $\chi^2(2) = 61.83$, $p < .001$, $M_{\text{pre to mid}} = -51$ Hz, $M_{\text{mid to post}} = -8$ Hz, and significantly raised F3 for English /l/, $\chi^2(2) = 8.06$, $p = .018$, $M_{\text{pre to mid}} = 2$ 7 Hz, $M_{\text{mid to post}} = -8$ Hz, after perceptual training. These results suggest that they improved the F3 distinction for English /r/ and /l/ in production.

However, there was no significant effect of the type of training. Neither the main effect of testing material, $\chi^2(1) = 2.01$, $p > .05$, nor trainer order, $\chi^2(1) = 0.63$, $p > .05$, was significant. All of the other three 2-way interactions were significant: consonant and testing material, $\chi^2(1) = 84.07$, $p < .001$, testing block and testing material, $\chi^2(2) = 6.85$, $p = .033$, and consonant and trainer order, $\chi^2(1) = 9.86$, $p < .01$, but the two 3-way interactions of consonant, testing block and trainer order and of consonant, testing block and testing material were excluded from the best-fitting model. This suggests that there was no significant improvement difference in the F3 distinction for the English /r/–/l/ contrast between the two trainer orders or between the two testing materials. In other words, both identification and discrimination training lowered F3 for English /r/ and raised F3 for English /l/ to similar extents for both word and passage reading tasks.

Finally, Pearson's correlations were conducted to compare individual differences in the degree of improvement with training (post test minus pre test) on each of the seven perception and production measures (i.e., identification, F3 sensitivity at the /r/–/l/ boundary, F2 sensitivity at the boundary, F3 sensitivity within /r/ category, category discrimination, F3 for the English /r/ production, F3 for the English /l/ production). Since all the possible 21 correlation tests were performed with the seven variables, p values were adjusted with Bonferroni correction. The results demonstrated that the improvement in identification was significantly correlated with the improvement in category discrimination, $r = 0.56$, df = 39, $p < .01$, which is consistent with the idea that these two tasks depended on common underlying abilities. The increase of F3 sensitivity at the English /r/–/l/ boundary was significantly correlated with the increase of F2 sensitivity at the boundary, $r = 0.66$, df = 39, $p < .001$. This suggests that Japanese subjects did not acquire native-like phonetic perception (i.e., higher sensitivity to the primary acoustic cue but lower sensitivity to irrelevant acoustic cues); they may have just increased their overall sensitivity to acoustic differences. There were no other significantly correlated variables.

4. Discussion

One main finding of this study is that high-variability discrimination training method can improve Japanese speakers' English /r/–/l/ identification for novel natural stimuli spoken by untrained talkers. Previous studies demonstrated that discrimination training did not generalize to untrained stimuli (Strange & Dittmann, 1984), whereas later studies showed that identification training did generalize (Lively et al., 1993, 1994; Logan et al., 1991). The early attempts at discrimination training used only synthetic stimulus continua (Strange & Dittmann, 1984), whereas identification training used English /r/–/l/ stimuli spoken by multiple talkers in a variety of phonetic environments (Lively et al., 1993, 1994; Logan et al., 1991). However, the present study combined discrimination training with a high-variability approach, and found that it was effective.

We had hypothesized that the high-variability discrimination training method could focus attention to a more perceptual level than identification training. That is, the Japanese speakers' mistuned auditory-phonetic perceptual processing may contribute to their /r/–/l/ learning difficulties (Bradlow, 2008; Hattori & Iverson, 2009; Ingvalson et al., 2011; Iverson et al., 2003, 2005, 2012), and discrimination training could target this auditory-phonetic level. Identification training, on the other hand, could improve English /r/–/l/ representations at phonological levels, since it forces subjects to categorize the phonemes (Gerrits & Schouten, 2004; Logan et al., 1991; Sadakata & McQueen, 2013; Sjerps et al., 2013). However, our perception and production results indicated that the two training methods improved performance in similar ways. Both identification and discrimination training increased Japanese speakers' identification, auditory discrimination, and category discrimination accuracy as well as improved their production, and the training effects differed between the two training methods only for the identification test. This one difference could have had a relatively superficial task-specific cause rather than indicating substantial differences in learning, given that the pre/post identification task was the same as that used in identification training. There was also no training method difference for the short identification test at the end of each training session, although this test was performed on the same talker and stimuli that they had just heard (i.e., no generalization). Moreover, the identification improvement was still limited by approximately 15%, similar to previous studies that used an identification training method alone (Bradlow et al., 1999, 1997; Iverson et al., 2005; Lively et al., 1993, 1994; Logan et al., 1991). The two training methods thus had similar effects on perception, contrary to our predictions (c.f., Gerrits & Schouten, 2004; Sadakata & McQueen, 2013; Sjerps et al., 2013).

It thus seems likely that both identification and discrimination training affected similar underlying processes, but it is less clear what processes improved. One possibility is that both methods improved auditory-phonetic sensitivity, particularly considering that both training techniques improved F3 sensitivity at the English /r/–/l/ phoneme boundary as well as category discrimination accuracy. However, this improvement was not very selective; an irrelevant acoustic cue (F2 sensitivity at the phoneme boundary) also improved. Moreover, it is possible that the category discrimination task was affected by changes in phonological labeling (i.e., listeners may covertly label the phonemes when performing this task; e.g., Flege, 2003; Højen & Flege, 2006; Iverson et al., 2012; Logan et al., 1991). This possibility is particularly plausible given that individual differences in category-discrimination improvement were correlated with improvements in phoneme identification.

Another possibility is that both training methods taught listeners how to cope with stimulus variability without necessarily making fundamental changes to their underlying categories. Iverson et al. (2005) suggested that identification training with high stimulus variability does not produce changes in the category representations for English /r/ and /l/ for Japanese speakers. They found that Japanese adults improve in their identification of English /r/–/l/ after training and that their use of acoustic cues changed, but they didn't change in a way that was systematically related to how these cues varied in the stimuli used for training (e.g., they shifted toward identifying a short-closure stimulus as /l/ even when the closure was lengthened in the training set). Instead, listeners became more systematic at identifying a stimulus as English /l/ when its acoustics approximated the Japanese L1 apico-alveolar tap [ɾ], regardless of the acoustic details of the training stimuli. That is, training on highly variable speech made them more consistent and automatic at applying the categories that they already had before training, rather than causing them to remap the cues that they used for these categories to match the training set. It is possible that training in the present study had a similar effect; individuals may have been trained by both methods to more consistently judge highly variable /r/–/l/ stimuli, but the exact judgments they made and the locus of their focused attention may have been less important.

This interpretation is consistent with other studies. For example, Ingvalson et al. (2011) demonstrated that F3 reliance and the length of residence in English speaking countries can predict Japanese speakers' English /r/–/l/ identification accuracy, but that F3 reliance is not correlated with their length of residence in English speaking countries. These results suggest that Japanese speakers can improve through experience on the use of secondary acoustic cues, or adopt some other strategy, even though their use of F3 does not become remapped based on the distribution of English /r/ and /l/ phonemes that they hear during daily speech communication. Without correcting Japanese speakers' mistuned phonetic processing, improvement due to experience or high-variability training may reach a ceiling (e.g., about 15 percentage points).

However, there remains the question of why Japanese speakers improve their production of the English /r/–/l/ contrast through perceptual training, because coping with speaker variability likely has less importance to production (i.e., speakers know what they intend to say). Given that production and perception both improved, it could be that Japanese speakers improved their underlying phonetic/phonological representations to some extent with training. Callan et al. (2003) examined changes in Japanese speakers' brain activity through a month of English /r/–/l/ perceptual training and suggested that the improvement may be partially attributed to the acquisition of auditory-articulatory (perceptual-motor) mappings. Although the present study showed that the individual differences in perceptual improvement were not correlated with production improvement, these kinds of non-significant perception-production relationships could be a result of perception and production learning proceeding at different rates within each subject. That is, the lack of a significant individual-differences correlation does not fully prove that there is no perception-production link at all (Bradlow et al., 1997, 1999; Draxler, Jügler, Möbius, & Zimmerer, 2015; Huensch & Tremblay, 2015; Hwang & Lee, 2015). It is possible that an improvement in coping with stimulus variability perceptually somehow improves the automaticity of phonetic processes in a way that makes it easier for speakers to produce the correct acoustic contrasts when producing speech.

The improvement in production is in line with the predictions of SLM (Flege, 1995). That is, Aoyama et al. (2004) suggested that it is easier to improve the production of English /r/ than English /l/, since the Japanese tap is further from English /r/ than English /l/. The current results demonstrated that although Japanese speakers significantly changed F3 for both phonemes (i.e., lowered F3 for /r/ and raised F3 for /l/), the F3 change for English /r/ was larger than for English /l/. It may be that, if a Japanese speaker produces English /l/ using many of the features of their Japanese tap, then this is sufficient to make it identifiable for English listeners. Raising F3 for English /l/ may not be so important or necessary, but their English /r/ production requires greater lowering in F3 frequency (Hattori & Iverson, 2009). Moreover, according to SLM, this lowering in F3 frequency may be relatively easy because this change does not cause it to interact with existing Japanese categories (i.e., no L1 liquid has a lowered F3).

Although discrimination and identification training appeared to have similar effects in the present study, one positive aspect of this result is that there may be applications in which discrimination training is easier to implement. For example, Japanese children who do not have categories already for /r/ and /l/ may have difficulty with a training program that requires them to respond with /r/ and /l/ labels; a discrimination task might be easier for them to perform yet also produce improvements in perception and production (Bradlow, 2008). Likewise, it is possible to imagine phonetic-training computer games (Lim & Holt, 2011) in which a discrimination task would be easier to implement than identification. Having two methods of achieving similar outcomes might thus add to the flexibility of training designs.

Acknowledgement

The first author of this research was funded by Japan Student Services Organization (JASSO).

References

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics, 32*(2), 233–250.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20*(3), 305–330.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Timonium, MD: York Press.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics, 61*(5), 977–985.

Bradlow, A. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/–/l/ contrast. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 287–308). Philadelphia, PA: John Benjamins.

Bradlow, A. R., Pisoni, D. B., Yamada, R. A., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/. IV: Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America, 101*(4), 2299–2310.

Callan, D. E., Tajima, K., Callan, A. M., Kubo, R., Masaki, S., & Akahane-Yamada, R. (2003). Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *NeuroImage, 19*(1), 113–124.

Draxler, C., Jügler, J., Möbius, B., & Zimmerer, F. (2015). The effect of high-variability training on the perception and production of French stops by German native speakers. *Proceedings of the Interspeech, 2015*(2), 806–810.

Fairbanks, G. (1960). *Voice and articulation drillbook* (2nd ed. p. 127). New York: Harper & Row.

Flege, J. E. (2003). A method for assessing the perception of vowels in a second language. In E. Fava & A. Mioni (Eds.), *Issues in clinical linguistics* (pp. 19–44). Padova: UniPress.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and language experience: Issues in cross-language research* (pp. 233–277). Baltimore: York Press.

Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /l/. *The Journal of the Acoustical Society of America, 99*(2), 1161–1173.

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics, 66*(3), 363–376.

Hattori, K. (2009). *Perception and production of English /r/–/l/ by adult Japanese speakers* (Doctoral thesis, University College London, United Kingdom). University College London. Retrieved from http://discovery.ucl.ac.uk/19204/1/19204.pdf.

Hattori, K., & Iverson, P. (2009). English /r/–/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *The Journal of the Acoustical Society of America, 125*(1), 469–479.

Højen, A., & Flege, J. E. (2006). Early learners' discrimination of second-language vowels. *The Journal of the Acoustical Society of America, 119*(5), 3072–3084.

Huensch, A., & Tremblay, A. (2015). Effects of perceptual phonetic training on the perception and production of second language syllable structure. *Journal of Phonetics, 52*, 105–120.

Hwang, H., & Lee, H. Y. (2015). The effect of high variability phonetic training on the production of English vowels and consonants. *Proceedings of the 18th international congress of phonetic sciences*, 1041–1045.

Ingvalson, E. M., McClelland, J. L., & Holt, L. L. (2011). Predicting native English-like performance by native Japanese speakers. *Journal of Phonetics, 39*(4), 571–584.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/–/l/ to Japanese adults. *The Journal of the Acoustical Society of America, 118*(5), 3267–3278.

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *The Journal of the Acoustical Society of America, 97*(1), 553–562.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*(1), B47–B57.

Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics, 33*(1), 145–160.

Iverson, P., Wagner, A., & Rosen, S. (2016). Effects of language experience on pre-categorical perception: Distinguishing general from specialized processes in speech perception. *The Journal of the Acoustical Society of America, 139*(4), 1799–1809.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America, 87*(2), 820–857.

Lim, S. J., & Holt, L. L. (2011). Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. *Cognitive Science, 35*(7), 1390–1405.

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. *The Journal of the Acoustical Society of America, 94*(3), 1242–1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America, 96*(4), 2076–2087.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America, 89*(2), 874–886.

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics, 2*(4), 369–390.

Sadakata, M., & McQueen, J. M. (2013). High stimulus variability in nonnative speech learning supports formation of abstract categories: Evidence from Japanese geminates. *The Journal of the Acoustical Society of America, 134*(2), 1324–1335.

Sjerps, M. J., McQueen, J. M., & Mitterer, H. (2013). Evidence for precategorical extrinsic vowel normalization. *Attention, Perception, & Psychophysics, 75*(3), 576–587.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r–l/ by Japanese adults learning English. *Perception & Psychophysics, 36*(2), 131–145.

Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 159–198). Philadelphia, PA: John Benjamins.

Takagi, N., & Mann, V. A. (1995). The limits of extended naturalistic exposure on the perceptual mastery of English /r/ and /l/ by adult Japanese learners of English. *Applied Psycholinguistics, 16*(4), 379–405.

Underbakke, M., Pola, L., Gottfried, J., & Strange, W. (1988). Trading relations in the perception of /r/–/l/ by Japanese learners of English. *The Journal of the Acoustical Society of America, 84*(1), 90–100.

Yuan, J., & Liberman, M. (2011). /l/ variation in American English: Corpus approach. *The Journal of Speech Sciences, 1*(2), 35–46.