Journal of Phonetics 89 (2021) 101108

Contents lists available at ScienceDirect

Journal of Phonetics

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

Research Article

The effect of age on English /r/-/l/ perceptual training outcomes for Japanese speakers ☆

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

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

ARTICLE INFO

Article history:
Received 11 August 2020
Received in revised form 7 September 2021
Accepted 27 September 2021

Keywords:
Age effects
High variability phonetic training
Speech perception and production
Second-language learning

ABSTRACT

Younger learners are better at acquiring second-language (L2) phoneme contrasts than are older learners, but this general correlation between age and learning ability is often confounded with factors such as how late learners use their L2 in daily life. The present study trained Japanese speakers across a wide age range (young children through adults) on English /r/-/l/, using a computer-based high variability phonetic training program, in order to control the /r/-/l/ inputs across age during the training period. The results demonstrated that training improved Japanese speakers' perception and production of the English /r/-/l/ contrast, and age affected their improvement in perception. Over the 10 training sessions, younger learners (children and adolescents) improved their perception more than adult learners, suggesting that L2 phoneme learning may indeed decline with age. Children did not improve their identification, perceptual sensitivity to the primary acoustic cue (F3), and category discrimination as much as adolescents, possibly due to their immature cognition and phonemic awareness.

1. Introduction

It is well known that second-language (L2) learning is affected by age. For example, individuals who move to a country with a different dominant language (e.g., Italians moving to English-speaking Canada) typically learn to speak their L2 with a more native-like accent if they move when they are younger (Flege et al., 1995; Flege, Yeni-Komshian, et al., 1999; Piske & MacKay, 1999). However, the cause of this phenomenon is not clear because age is confounded by many factors, such as biological neural maturation (Lenneberg, 1967; Oyama, 1976; Werker & Tees, 2005; Zhang & Wang, 2007), cognitive development (Heeren & Schouten, 2010; Wang & Kuhl, 2003; Zhang & Wang, 2007), quality and quantity of L2 input in L2 speaking countries (Flege, MacKay, et al., 1999; Flege & MacKay, 2011), amount of L1 use (Flege & MacKay, 2011; Flege, MacKay, et al., 1999; Piske & MacKay, 1999), development of L1 categories (Flege, 1995, 1999, 2002; Flege, Yeni-Komshian, et al., 1999; Flege & MacKay, 2011; Kuhl, 2011; Kuhl et al., 2006, 2008), and sociophonetic attitudes about speaking with an L2 accent (Crowther et al., 2015; Davies, 2017; Derwing, 2003; Janicka et al., 2005; Jenkins, 2007; Lippi-Green, 2012; Moyer, 2007).

Aoyama et al. (2004) examined the effect of age on Japanese speakers' perception and production of the English /r/-/l/ contrast by investigating how these abilities changed after spending one year in the US. They found that Japanese children improved more than did Japanese adults over this period. However, Aoyama et al. (2004) concluded that there is no single explanation for this age effect, because real language learning situations like this are affected by many uncontrolled factors. For example, all the Japanese adults had learned English before they arrived in the US, but most of the children had not learned English in Japan. Moreover, the language input and social setting for child and adult L2 learners differ; Japanese children typically interacted with native English-speaking friends and teachers at local schools, whereas adults typically did not have much opportunity to speak English. Thus, it is not clear to what extent the greater phonetic learning ability of children was due to maturational factors (e.g., general perceptual/cognitive abilities or less interference from the L1)

☆ A portion of this work was presented in "Computer-based English /r/-/l/ perceptual training for Japanese children" at the 21st International Congress on Acoustics, Montréal, Canada, 2–7 June 2013, and in "Effects of English /r/-/l/ perceptual training on Japanese children's production" at the 18th International Congress of Phonetic Sciences, Glasgow, UK, 10–14 August 2015.

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

https://doi.org/10.1016/j.wocn.2021.101108
0095-4470/

rather than how the L2 was learned and used. In the present study, we used a computer-based phonetic training program to control the quality and quantity of L2 input for Japanese speakers learning the English /r/-/l/ contrast. English usage is limited in Japan, such that we could reasonably assume that little additional English exposure occurred during the training period.

The English /r/-/l/ contrast is one of the most difficult L2 contrasts for Japanese speakers to identify and discriminate. Neither the voiced post-alveolar approximant /r/ nor the voiced alveolar lateral approximant /l/ exists in the Japanese language. The voiced apico-alveolar tap /ɾ/ is the only similar phoneme in Japanese (Vance, 2008), and it is more similar to English /l/ than /r/ (Aoyama et al., 2004; Aoyama & Flege, 2011; Hattori & Iverson, 2009). Compared to English /l/, Japanese /ɾ/ tends to be shorter (i.e., more rapid articulation), has a higher second-formant (F2) frequency, and has a lower third-formant (F3) frequency, although there is individual variation (Hattori & Iverson, 2009; Vance, 2008). Being exposed to a Japanese language environment during the first year of life seems to reduce sensitivity to the critical F3 contrast for English /r/ and /l/ (i.e., lower F3 for /r/ vs. higher F3 for /l/), which has knock-on effects for higher-level category learning (Hattori & Iverson, 2009; Iverson et al., 2003, 2005; Kuhl et al., 2006, 2008). Whereas being exposed to an English language environment during the first year of life increases sensitivity to the /r/-/l/ F3 contrast (Kuhl et al., 2006). Even after living in English-speaking countries for decades, Japanese adults cannot easily adapt their auditory phonetic processing of F3 (Ingvalson et al., 2011). Individual variation in their F3 processing difficulties are correlated with lower identification accuracy for natural recordings of English /r/ and /l/ (Hattori, 2009; Hattori & Iverson, 2009).

Phonetic training has long been used for Japanese speakers learning English /r/-/l/, mostly as an existence proof demonstrating that plasticity is retained by adults; learning is possible when adults are exposed to natural phonetic variation (i.e., stimuli produced by multiple talkers in multiple syllable contexts) in an identification task with feedback (Lively et al., 1993, 1994; Logan et al., 1991). That being said, performance improvements are typically limited (about 15%) and it is not clear exactly what level of processing is being changed as a result of phonetic training. Researchers initially hypothesized that high-variability training might work because people can adjust the cues they use for categorization, to rely most on cues that are robust across this variability (Lively et al., 1993, 1994; Logan et al., 1991). However, it appears that listeners do not make fine-grained category adjustments that match what they hear during training or create new phonological categories. Instead, they may learn to more consistently apply their L1 category knowledge, identifying English /l/ if the stimulus sounds similar to Japanese /ɾ/ and identifying English /r/ if it does not (Iverson et al., 2005; Shinohara & Iverson, 2018). We have attempted to improve training outcomes by using tasks that would separately target early auditory-phonetic processing (auditory and category discrimination) and linguistic categorization (identification tasks), but found that both tasks improve perception in similar ways (Shinohara & Iverson, 2018). Likewise, a perceptual training task can have broad effects, improving both perception and production (Bradlow et al., 1997, 1999; Shinohara & Iverson, 2018). The present study used a diverse approach to training, combining both discrimination and identification tasks in an attempt to produce improvements across age groups from the same controlled experience, rather than targeting a particular level of processing.

One could hypothesize that children will benefit from training more than adults, because younger learners experience less interference from their existing L1 categories. Newborns have language-universal abilities to discriminate speech sounds, but through development, infants improve in their sensitivity to native contrasts and their non-native phonetic perception declines (Kuhl et al., 2006, 2008; Tsao et al., 2006). Kuhl's Native Language Magnet model suggests that the increasing neural commitment of infants to their L1 makes L2 learning more difficult (Kuhl, 2011; Kuhl et al., 2008). Similarly, Flege's Speech Learning Model (SLM; Flege, 1995, 1999, 2002) claims that it becomes more difficult to form new L2 categories as the L1 categories become better established. Iverson et al.'s (2003) Perceptual Interference account suggests that L2 learning difficulties can occur because language experience affects auditory phonetic processing, which alters how easily higher-level phonetic categories can be learned.

Recently, Flege and his colleagues have revised the Speech Learning Model (i.e., SLM-r; Flege, 2021; Flege et al., 2021; Flege & Bohn, 2021). SLM-r has placed more emphasis on the precision with which L1 categories have been learned and more of a link to experience than to chronological age. When examining individual differences among Japanese adults learning English /r/ and /l/, SLM-r thus suggests that examining individual differences in L1 representations might be more relevant than age for understanding individual learning differences. That being said, a detailed assessment of L1 category precision is not simple with young children who likely have limited phonemic awareness, and it is likewise not clear how one could control for all the possible confounds of chronological age. Hattori and Iverson (2009) assessed individual differences in L1 and L2 representations in Japanese adults, and concluded that the main driver of English /r/-/l/ category learning difficulty was sensitivity to F3, rather than individual details in their representations of their Japanese tap /ɾ/ category. Further research along these lines would be interesting (e.g., see Aoyama & Flege, 2011). However, the present investigation focuses on age-related individual differences, even though age may act more as a proxy for the various experience-related changes in L1 phonetic processing that might affect L2 learning.

Despite these theories and findings, previous work on phonetic training has often not established clear advantages for young learners. Wang and Kuhl (2003) delivered a two-week Mandarin tone training program to four age groups of American English speakers (6, 10, 14 years old, and adults) and found that all age groups improved their discrimination accuracy but there was no difference in the degree of improvement seen between age groups. Heeren and Schouten (2010) examined training effects on the perception of the Finnish /t/-/tː/ length contrast by child and adult Dutch speakers and similarly found that the degree of improvement was not significantly different between groups. The relationship between the target L2 phonetic feature and L1 phonology may have effects on these results, and children's greater plasticity for L2 sounds may be counteracted by broader perceptual and cognitive factors (e.g., an immaturity of cognitive, auditory, and language processing) that make them less responsive to computer-based training. The development of phonemic awareness is also a potential factor affecting L2 phonetic learning (Shinohara, 2014; Shinohara & Iverson, 2013; Snowling & Hulme, 1994).

To our knowledge, there are few studies examining the effect of age on production improvement after perceptual training. For adults, perception and production abilities tend to be correlated before training, likely because both depend on common underlying factors such as the length of time spent in an L2-speaking country (Flege et al., 1997; Flege, MacKay, et al., 1999; Ingvalson et al., 2011; Ingvalson & Holt, 2011; Llisterri, 1995; Yamada, 1995; Yamada et al., 1994) and the amount of L2 use in an L2-speaking country (Aoyama & Flege, 2011; Flege & MacKay, 2004, 2011; Jia et al., 2006; Tsukada et al., 2005). However, degrees of perception and production improvement after training are often uncorrelated in adults, suggesting that learning may progress at a different rate within each individual (Bradlow et al., 1997, 1999; Shinohara & Iverson, 2018) or that there are different mechanisms for perception and production (Hattori, 2009). For children, it seems likely that any age-related advantages in perceptual learning would hold for production too (Baker et al., 2008; Kartushina & Frauenfelder, 2014). However, if the training transfer between perception and production is indirect in adults (e.g., improvement in general awareness or knowledge of this contrast, rather than improvement of a common underlying mechanism involved in speech production and perception), it is unclear whether these indirect mechanisms would operate similarly at earlier developmental stages.

The present study thus tested the effect of age on the improvement of Japanese speakers' perception and production of the English /r/-/l/ contrast. We took a combined training approach using both identification and discrimination tasks, and training was conducted within a game environment designed to be more interesting to participants, particularly children. There were several pre/post perceptual tests: identification, to test recognition accuracy for /r/ and /l/; auditory discrimination, to test sensitivity to primary, secondary, and irrelevant acoustic cues (i.e., F3 at the /r/-/l/ boundary, F2 at the /r/-/l/ boundary, F3 within the /r/ category); and category discrimination, to examine the primary acoustic cue use in perception. In addition, pre/post production was recorded and assessed in terms of identifiability by native speakers and F3 frequencies. The tests were designed to assess overall improvements in identification and production accuracy, and to provide finer-grained assessments of whether their use of acoustic cues is becoming more native-like through training (e.g., increased sensitivity to F3 contrasts between /r/ and /l/, and larger F3 differences for /r/ and /l/ in production).

2. Methods

2.1. Subjects

A total of 99 Japanese speakers were tested, all in the Kanto area, Japan. Of these 99 subjects, 14 subjects were not used in the analysis (5 dropped out during the training stages, and 9 completed too many training sessions per day). When subjects participated in this study, they were assigned to either the control group (i.e., they took part in only pre and post tests, without completing any training) or the training group (i.e., they took part in the pre test, 10 training sessions, and the post test). In order to balance subjects' age between the control and training groups, two subjects aged 59.6 and 48.3 from the control group were excluded from the analyses, leaving 83 subjects. Forty-seven subjects (24 females, 23 males) aged from 7.3 to 40.4 (median = 16.0 years) were in the training group, and 36 subjects (25 females, 11 males) aged from 6.9 to 43.3 (median = 16.1 years) were in the control group. Subjects who were assigned to the control group had 10–14 days between the pre and post tests without training.

The subjects were recruited such that all had pre-test /r/-/l/ identification scores of less than 75%, no experience living in English-speaking countries prior to the age of 18 years old, and no history of language/hearing impairments. Subjects that did not meet these criteria (e.g., six subjects with 75% or higher identification score at the pre test) were not included in the training or control group. Apart from four adults who had stayed in English-speaking countries for several years after turning 18, no subjects had lived outside Japan. All subjects had started learning English at school at 12 years of age except for child subjects under the age of 12, who had learned only some basic English at school (e.g. greetings, color names).

2.2. Apparatus and stimuli

Natural recordings were used for both identification and category discrimination tasks in training and testing, and they were the same recordings used in Iverson et al. (2005) and Shinohara and Iverson (2018). A total of 220 English /r/-/l/ minimal-pair words from each of nine adult Standard Southern British English (SSBE) speakers were digitally recorded with 44,100 16-bit samples per second. Those recordings were downsampled to 22,050 samples per second to reduce the computer memory demands of the training program. Of these 220 words, 100 word-initial /r/-/l/ minimal-pair words produced by each of five speakers were used as training stimuli. The remaining 120 minimal-pair words, produced by each of the other four speakers (two females and two males), were used as test stimuli. They comprised 40 word-initial /r/-/l/ minimal-pair words (e.g., *road* and *load*), 40 word-medial /r/-/l/ minimal-pair words (e.g., *correct* and *collect)*, and 40 word-initial /r/-/l/ consonant cluster minimal-pair words (e.g., *brand* and *bland*). Word-final /r/-/l/ minimal pairs were not included as either testing or training stimuli, since /r/ is fully deleted and /l/ is often vocalized in word-final position by speakers of southern British English (Cruttenden, 2014; Lindsey, 2019; McMahon, 2000; Wells, 1982). The speakers of the test stimuli were different from those of the training stimuli, and the test stimuli included minimal pairs of not only the trained word-initial positions but also the word-medial and consonant cluster positions to test the generalization effects for the speakers and phonetic environments (e.g., Shinohara & Iverson, 2018).

In addition, synthetic /rɑ/ and /lɑ/ stimuli were used for testing perceptual sensitivities in the auditory discrimination test.

They were the same stimuli as used in Hattori (2009) and Shinohara and Iverson (2018). Hattori (2009) examined the best exemplar locations and phoneme boundary location of the English /r/ and /l/ within a larger set of synthesized stimuli and generated identifiable /r/-/l/ stimuli based on these results. As displayed in Table 1, three stimulus pairs were manipulated with five acoustic cues (steady state duration, transition duration, F1, F2, and F3). The three stimulus pairs contrasted F3 at the /r/-/l/ phoneme boundary, F2 at the /r/-/l/ phoneme boundary, and F3 within the /r/ category. For the F3 contrast at the /r/-/l/ boundary, steady state duration, transition duration, F1, and F2 were set at 64 ms, 48 ms, 327 Hz, and 1196 Hz, respectively. The target F3 was set at 2639 Hz for the /r/ stimulus and 3328 Hz for the /l/ stimulus, because the /r/-/l/ phoneme boundary was at 2965 Hz on this female-voiced F3 stimulus continuum (Hattori, 2009). For the F2 contrast at the boundary, steady state duration, transition duration, F1, and F3 were set at 64 ms, 48 ms, 327 Hz, and 2965 Hz, respectively; F2 was set to 1051 and 1358 Hz. Finally, for F3 contrast within the /r/ category, steady state duration, transition duration, F1, and F2 were set at 31 ms, 81 ms, 327 Hz, and 1196 Hz, respectively; F3 was set to 1739 and 2212 Hz.

Resynthesized stimuli were used for auditory discrimination training, and had been used in a previous study (Shinohara & Iverson, 2018). All acoustic cues except F3 were equated for /r/-/l/ minimal-pair words. Specifically, the formant transitions except for F3 were averaged across each /r/-/l/ minimal pair, and these average transitions were imposed on the stimuli using LPC analysis and resynthesis in Praat. Four minimal pairs whose /r/ and /l/ stimuli were identified consistently as /r/ or /l/ by native speakers on 85% of trials were selected from each of five speakers, so that the stimuli bracketed the /r/-/l/ phoneme boundary.

All stimuli used for pre/post tests were presented over a pair of headphones (Sennheiser HD 205 or Sennheiser HD 201), at a user-controlled comfortable level.

2.3. Procedure

2.3.1. Training

Ten sessions of a computer-based perceptual training program, that we developed, were given to 47 Japanese speakers (i.e., training group). Children and adolescents aged 18 and under were trained in supervised sessions (i.e., in a room with the tester present) but adults had the option of completing the training on their own laptops. Each session included three tasks: identification, auditory discrimination, and category discrimination. In the identification task (90 trials), subjects heard a stimulus, a minimal pair was displayed on screen, and subjects judged which word they thought they had heard by clicking on R or L displayed next to minimal pair words (e.g., *rock* vs. *lock*). As Japanese children are not always able to read English letters, there were two mascot characters—R-man, wearing a blue hat, and L-man, wearing a yellow hat—next to the R and L minimal-pair words. For the first 10 trials, both words of an /r/-/l/ minimal pair were played before the trial to remind subjects of the phonemes. When each of the minimal-pair words was played, R-man and L-man jumped slightly to help subjects remember each of the /r/ and /l/ phonemes. For the remaining 80 trials, there were no model stimuli. When subjects clicked on the correct answer, a message せいかい ('Correct answer') was displayed on the screen and a cash register sound was played. The correct stimulus was replayed once with the correct mascot character slightly jumping. When subjects clicked on a wrong answer, the message ざんねん ('Bad luck') was displayed on the screen, two beep sounds with descending pitches were played, and the correct stimulus was replayed twice with the correct mascot character jumping. Each session had one speaker, such that each of the five English speakers was presented in two sessions over the 10-session training program. Each session had ninety trials (words) that were selected from the 100 /r/-/l/ minimal-pair words recorded from native English speakers, such that each word was used nine times during the 10 training sessions.

In the auditory discrimination task (48 trials), subjects heard three stimuli that were produced by the same talker, were resynthesized so that they contrasted only F3, and formed a minimal pair. They judged which of the three stimuli was different (e.g., *rain, rain, lane*), and the "same" stimuli were acoustically identical. Three objects numbered 1, 2, and 3 were displayed with a black cursor highlighting the number and object as each stimulus was being played. If subjects clicked on a correct answer, they heard a cash register sound with a message せいかい ('Correct answer') displayed on the screen. If subjects clicked on a wrong answer, they heard two descending beeps with a message ざんねん ('Bad luck') displayed on the screen and heard all three of the stimuli repeated once with the correct answer highlighted. Each session had one speaker and four minimal pairs, with 12 trials for each minimal pair, resulting in 48 trials in each session.

In the category discrimination task (60 trials), subjects heard three natural, non-resynthesized, recordings of a minimal pair produced by three different speakers, and they selected the one whose first phoneme was categorically different from the other two (e.g., *reek, reek, leak*). The "same" stimuli were not acoustically identical (i.e., produced by different talkers), so listeners needed to attend to phonological or phonetic differences in this task. There were three red devils displayed on

Table 1
Parameters of synthesized stimuli.

	Steady state duration (ms)	Transition duration (ms)	F1 (Hz)	F2 (Hz)	F3 (Hz)
F3 at the /r/-/l/ phoneme boundary	64	48	327	1196	2639 3328
F2 at the /r/-/l/ phoneme boundary	64	48	327	1051 1358	2965
F3 within the /r/ category	31	81	327	1196	1739 2212

the screen with a black cursor highlighting each devil as each stimulus was played. When subjects clicked on a correct answer, they were able to receive a treasure box from the devil, with a cash register sound played and a message せいかい ('Correct answer') displayed. When subjects clicked on a wrong answer, the red devil that produced the categorically different stimulus became larger than the other two devils, subjects heard two descending beeps, and saw a message ざんねん ('Bad luck'). All three stimuli were replayed once. On each trial, /r/-/l/ minimal-pair words produced by five SSBE speakers were randomly selected.

After completing all the three tasks, subjects performed a short identification test with 20 trials produced by the same speaker used in that particular session. During the short test, subjects did not receive any feedback but saw their percentage of correct answers after finishing the test to track their improvement through the 10 training sessions.

All records of the training sessions (date and time, stimuli, subjects' responses, etc.) were saved in password-locked log files on computers. Subjects were instructed not to complete two or more sessions in one day. The time of day at which training was conducted was not controlled between subjects, but they were not allowed to conduct multiple training sessions without an overnight sleep for consolidation (Earle & Myers, 2015; Qin & Zhang, 2019). The data of those who completed more than one session in a day were excluded from the analysis.

2.3.2. Pre/post test: perception

Subjects completed three perception tests before and after training (i.e., identification, auditory discrimination, and category discrimination), in a quiet room under the supervision by the first author. As with training, in the identification test, subjects heard a stimulus, saw a minimal pair displayed on screen accompanied by the characters of R-man and L-man, and selected either R or L based on what they thought they had heard. Since most child subjects had not learned English before training, some of them did not know how written English /r/ and /l/ look or how the English /r/ and /l/ sound. An English /r/-/l/ minimal pair (*rock* vs. *lock*) was played as an example several times for all subjects at both pre and post tests. To make sure all subjects understood the identification task, they completed two practice trials (*rock–lock, rack–lack*) without feedback. After this practice session, all subjects completed an identification test with 120 trials in total (i.e., 40 word-initial /r/-/l/ minimal-pair trials, 40 word-medial /r/-/l/ minimal-pair trials, and 40 word-initial consonant cluster minimal-pair trials). Two SSBE speakers (a female and a male), whose recordings were not included in the training corpus, each produced sixty different tokens (20 tokens for each /r/-/l/ position).

In the auditory discrimination test, subjects heard three synthesized /rɑ/-/lɑ/ stimuli produced by three identical white cats, and selected the cat that they thought produced a different stimulus from the other two. There were three stimulus pairs examining sensitivity to the primary acoustic cue for the /r/-/l/ contrast (i.e., high vs. low F3 at the /r/-/l/ phoneme boundary), a secondary/irrelevant acoustic cue which Japanese speakers often rely on when perceiving the contrast (i.e., high F2 vs. low F2 at the boundary), and another irrelevant acoustic cue (i.e., high F3 vs. low F3 within the /r/ category). Subjects completed 72 trials in total (3 stimulus pairs × 2 possible odd /rɑ/ and /lɑ/ stimuli × 3 odd stimulus positions × 4 repetitions for each).

In the category discrimination test, subjects heard three stimuli produced by different speakers and selected the one whose first phoneme they thought was different from the other two. There were three identical dogs displayed on the screen, each of them produced either of the /r/-/l/ minimal-pair words (e.g., *race, lace, race*), and subjects selected the categorically different stimulus (e.g., *lace*). Before starting the test session, all subjects practiced with 12 trials of the English /s/-/ɡ/ minimal pairs to make sure all subjects understood the task. The test stimuli included eight /r/-/l/ minimal pairs (*race–lace, reap–leap, ride–lied, road–load, roared–lord, rug–lug, rune–loon, wraps–lapse*) produced by four SSBE speakers not used in the training corpora. During the test, there was no feedback, and subjects were not allowed to listen to the same stimulus again. There were 48 trials in total (8 minimal pairs × 2 /r/-/l/ odd stimuli × 3 different odd stimulus positions).

2.3.3. Pre/post test: production

After the perception tests, all subjects' production of 40 word-initial English /r/-/l/ minimal-pair words were recorded in a quiet room under the supervision of the first author. For each item, the word was displayed on the screen, and subjects heard the word produced by a female native speaker of British English over a pair of headphones. This was followed by a recording of a male British English speaker saying "read the word after the tone" and a beep sound; subjects produced the word after they heard the beep. The interval between the word presentation and the beep was 2500 ms, with the intention that listeners would rely on their mental representations of /r/ and /l/ and rely less on auditory short-term memory (Flege, 2003; Gerrits & Schouten, 2004; Matthews & Stewart, 2009; Mora, 2007). Due to a technical error, one subject's post-test recordings were missing and excluded from the analysis. In addition, 39 tokens were omitted because the subjects did not speak on these trials (i.e., 40 words × 82 subjects × 2 tests – 39 tokens).

To measure the identifiability of English /r/-/l/ productions, 10 SSBE speakers identified the stimuli using an /r/-/l/ forced-choice identification task. They were asked to click on the phonemic representation that they thought they had heard. Other alternatives than /r/ and /l/ were not included to make it consistent with the perceptual identification test for Japanese speakers. Owing to the large number of recordings, each SSBE speaker identified all Japanese speakers' pre- and post-test productions, but only for two minimal pairs; different words were given to different English subjects.

The primary acoustic cue for the /r/-/l/ contrast (F3) was measured from the steady state part of the English /r/-/l/ to investigate acoustic changes from pre- to post-test production (Shinohara & Iverson, 2018). The steady state part was defined as the period at the onset where the spectrum remained relatively static. Fig. 1 displays example spectrograms of the *reap–leap* minimal pair produced by a Japanese subject. The steady state part is highlighted in yellow. This period is typical of the English /l/ lateral articulation (i.e., closure at the tongue tip), which is held at its target before a rapid transition into the vowel; this static period is often briefer for the English /r/. Ten minimal-pair words, whose orthographies were the

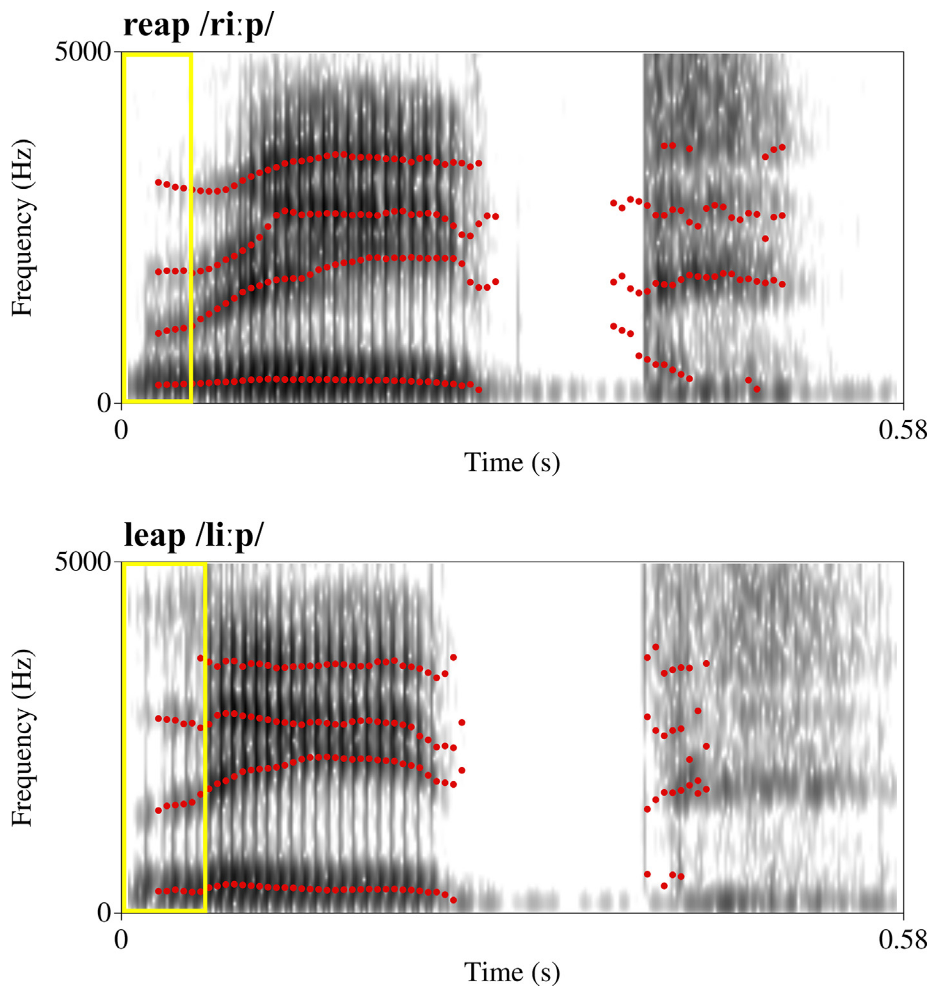

Fig. 1. Spectrograms of the *reap–leap* minimal pair produced by a Japanese subject at post test. F3 was measured from the steady state part in yellow.

same except the first letter, *r* or *l*, were selected and analyzed. The formant-tracking setting in Praat (Boersma & Weenink, 2012) was adjusted for each subject. The F2 frequencies of the following vowels were measured for each production and a scale factor was calculated by the log grand mean of F2 divided by each subject's F2 log mean. Their F3 frequency was then normalized, using the scale factor for each subject, and rescaled into Hertz.

3. Results

3.1. Perception

Fig. 2 displays identification accuracy for pre and post tests. The best-fitting logistic mixed effects model was selected with a top-down approach, using the R (version 4.0.4) package *lme4 (1.1-23)* (Bates, Mächler, et al., 2015). Fixed and random factors that did not influence the model fit were excluded from a model with all potential factors, based on Akaike's information criterion (AIC), at a significance level of 0.05 (Barr et al., 2013; Bates, Kliegl, et al., 2015; Bates, Mächler, et al., 2015; Kuznetsova et al., 2017; Zuur et al., 2009). The best-fitting model for identification accuracy included the fixed factors of group (training, control), testing block (pre, post), /r/-/l/ position (initial, medial, consonant cluster), and polynomial functions of age in a logarithmic scale (linear and quadratic regressions). Six 2-way interactions were included as well as three 3-way interactions (see Appendix A: Table 3). All the contrasts for the categorical variables were orthogonal. The random factors were crossed intercepts of subject and word. By-subject and by-word random slopes for testing block were not included, because the model failed to converge.

The best-fitting logistic mixed effects model for identification accuracy demonstrated that there was a significant 2-way interaction of group and testing block, β = 0.13, SE = 0.01, z = 8.96, $p < .001$. This means that Japanese speakers in the training group improved their English /r/-/l/ identification significantly more than did those in the control group (see the top-left boxplots in Fig. 2). In addition, the 3-way interaction of group, testing block, and quadratic function of age was significant, β = −4.90, SE = 2.06, z = −2.38, p = .018, as well as the single effect of linear function of age, β = 20.88, SE = 3.85, z = 5.43, $p < .001$. These results suggest that the training effects increase with an inverted-U shape. That is, older learners were better at identifying English /r/ and /l/ than younger learners at pre test, but adolescents increased their identification accuracy more than adults and children after training (see the scatterplots in Fig. 2). The significant 3-way interaction of group, testing block, and /r/-/l/ position (word-initial vs. the other two positions), β = −0.06, SE = 0.01, z = −5.24,

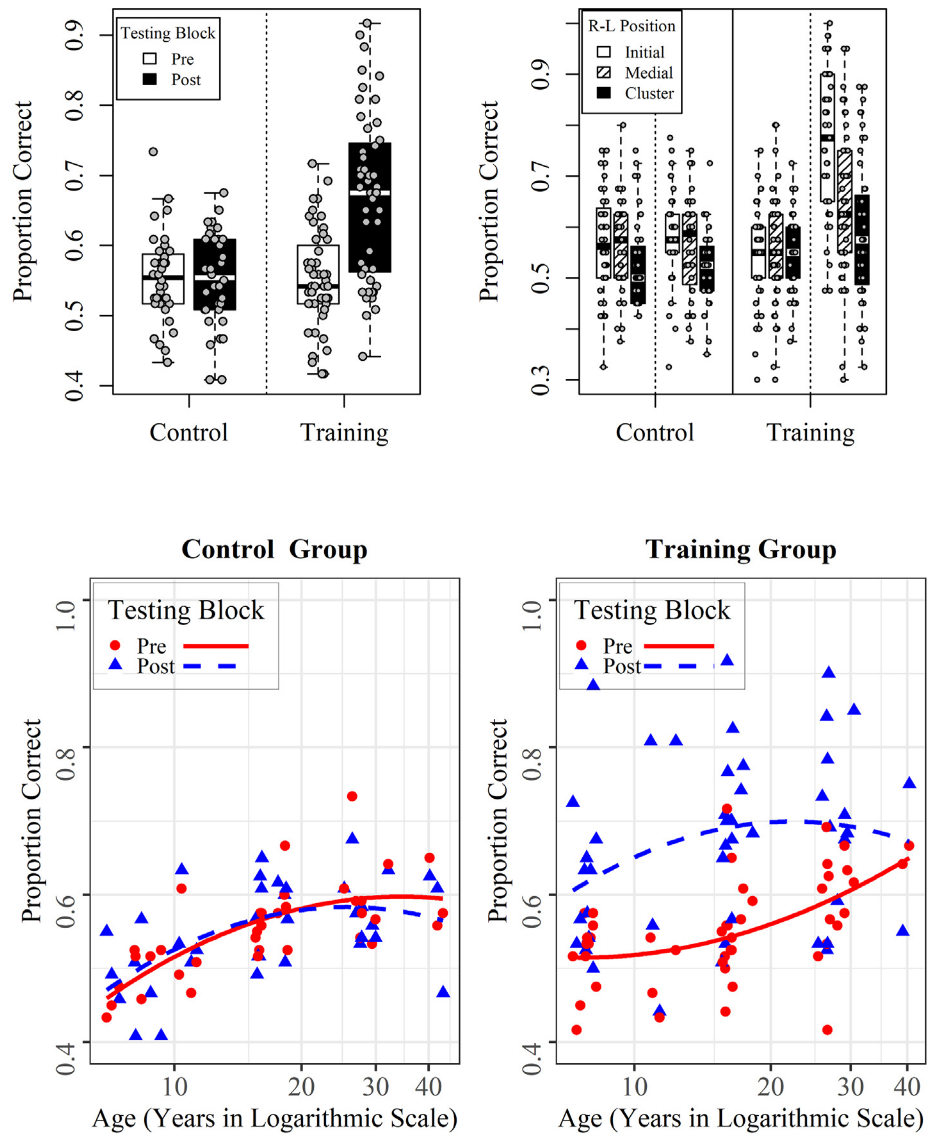

Fig. 2. Boxplots (top) of identification accuracy for the pre test (white boxes) and post test (black boxes) completed by Japanese speakers assigned in the control and training groups (top-left) and by each /r/-/l/ position (top-right): word-initial (white boxes), word-medial (cross-hatched boxes), and consonant cluster (black boxes); and scatterplots (bottom) for identification accuracy along with age in logarithmic scale for pre test (circle dots) and post test (triangle dots). The lines in the scatterplots represent the loess regression curves for pre (solid lines) and post (dashed lines) tests.

p < .001, demonstrated that subjects showed the greatest improvement in the trained word-initial position among the three different positions (see the top-right boxplots in Fig. 2). See Appendix A: Table 3 for detailed statistical results.

Fig. 3 displays the auditory discrimination accuracy for F3 at the English /r/-/l/ boundary, F2 at the boundary, and F3 within the English /r/ category. The logistic mixed effects model for each of these three stimulus pairs included the same fixed factors of group, testing block, polynomial (linear, quadratic) functions of log age, and all the interactions (see Appendix A: Table 4.1, 4.2, and 4.3). All the contrasts for the categorical variables were orthogonal. The random factors were crossed intercepts of subject and stimulus. Their random slopes for the testing block were included in the model testing F3 sensitivity at the phoneme boundary but excluded from the ones testing the F2 sensitivity at the phoneme boundary and F3 sensitivity within the /r/ category, due to the convergence failure and the singular fit.

The logistic mixed effects model for F3 sensitivity at the /r/-/l/ boundary demonstrated that there was a significant interaction between group and testing block, β = 0.10, SE = 0.04, z = 2.32, p = .020. This suggests that Japanese speakers in the training group increased their F3 sensitivity at the phoneme boundary (i.e., the primary acoustic cue) significantly more than those in control group (see the boxplots in Fig. 3). The 3-way interaction of group, testing block, and the quadratic function of age was also significant, β = −5.24, SE = 2.61, z = −2.00, p = .045, suggesting that adolescents improved their perceptual sensitivity to the primary acoustic cue for the /r/-/l/ contrast more than did adults and children after training (see the left scatterplot in Fig. 3).

The logistic mixed effects model for F2 sensitivity at the phoneme boundary (i.e., a secondary/irrelevant acoustic cue for the /r/-/l/ contrast for English speakers) demonstrated that although the 2-way interaction between group and block was not significant, β = −0.01, SE = 0.03, z = −0.37, p > .05, the

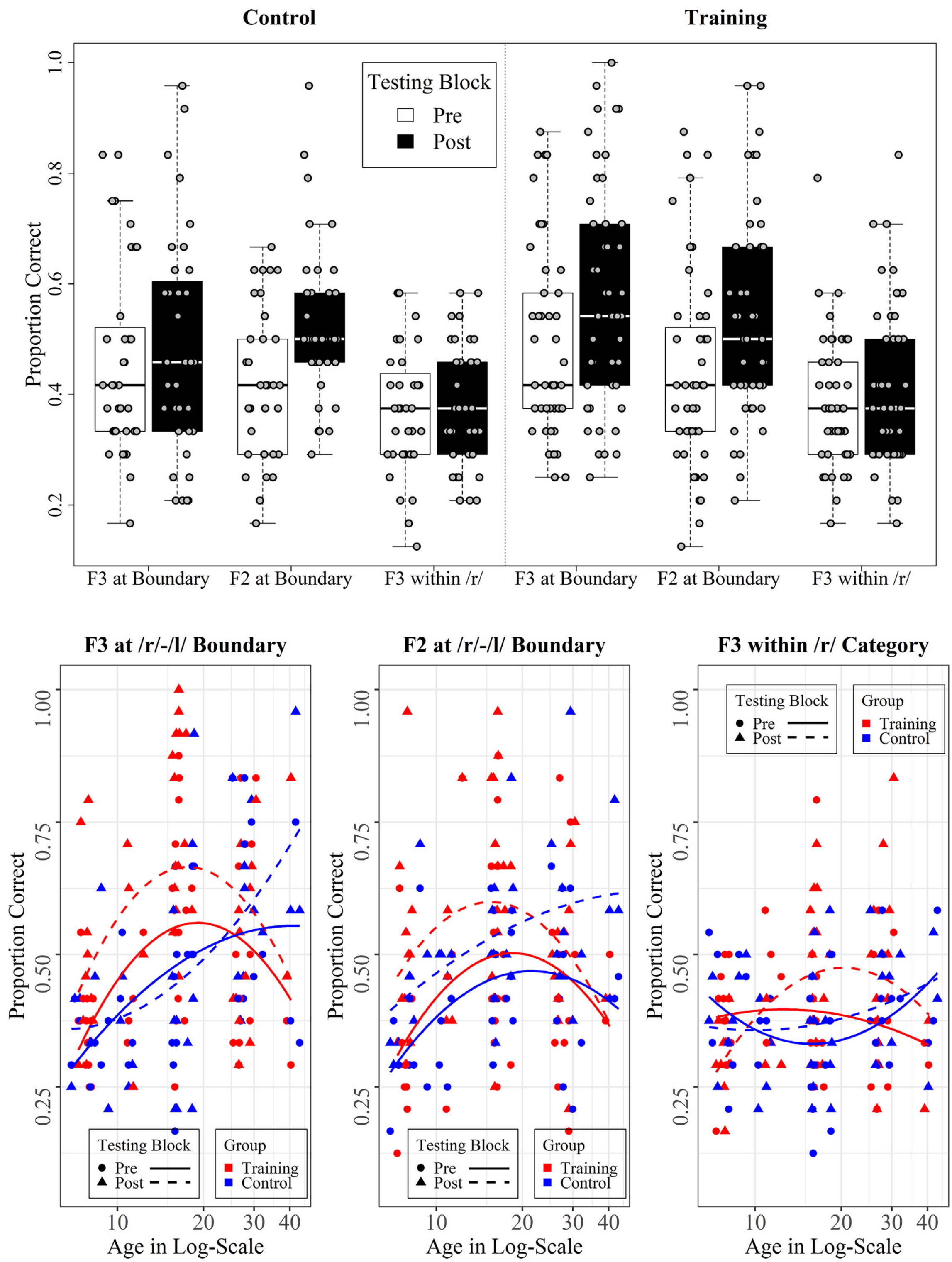

Fig. 3. Boxplots (top) depicting subjects' auditory discrimination accuracies with regards to F3 at the /r/-/l/ phoneme boundary, F2 at the /r/-/l/ phoneme boundary, and F3 within the /r/ category for the pre test (white boxes) and post test (black boxes) completed by Japanese speakers assigned in the control (left) and training groups (right), and scatterplots (bottom) of the auditory discrimination accuracies along with age in logarithmic scale for pre test (circle dots) and post test (triangle dots). The lines in the scatterplots represent the loess regression curves for pre test (solid lines) and post test (dashed lines).

3-way interaction of group, block, and the linear function of age was significant, β = −4.28, SE = 2.14, z = −2.00, p = .046. These results suggest that both the training group and the control group increased their F2 sensitivity from pre to post test (see the boxplots in Fig. 3) and that the increase in F2 sensitivity by training was higher for younger learners than for older learners.

The logistic mixed effects model for F3 sensitivity within the /r/ category (i.e., an irrelevant acoustic cue for the /r/-/l/ contrast for both English and Japanese speakers) demonstrated that

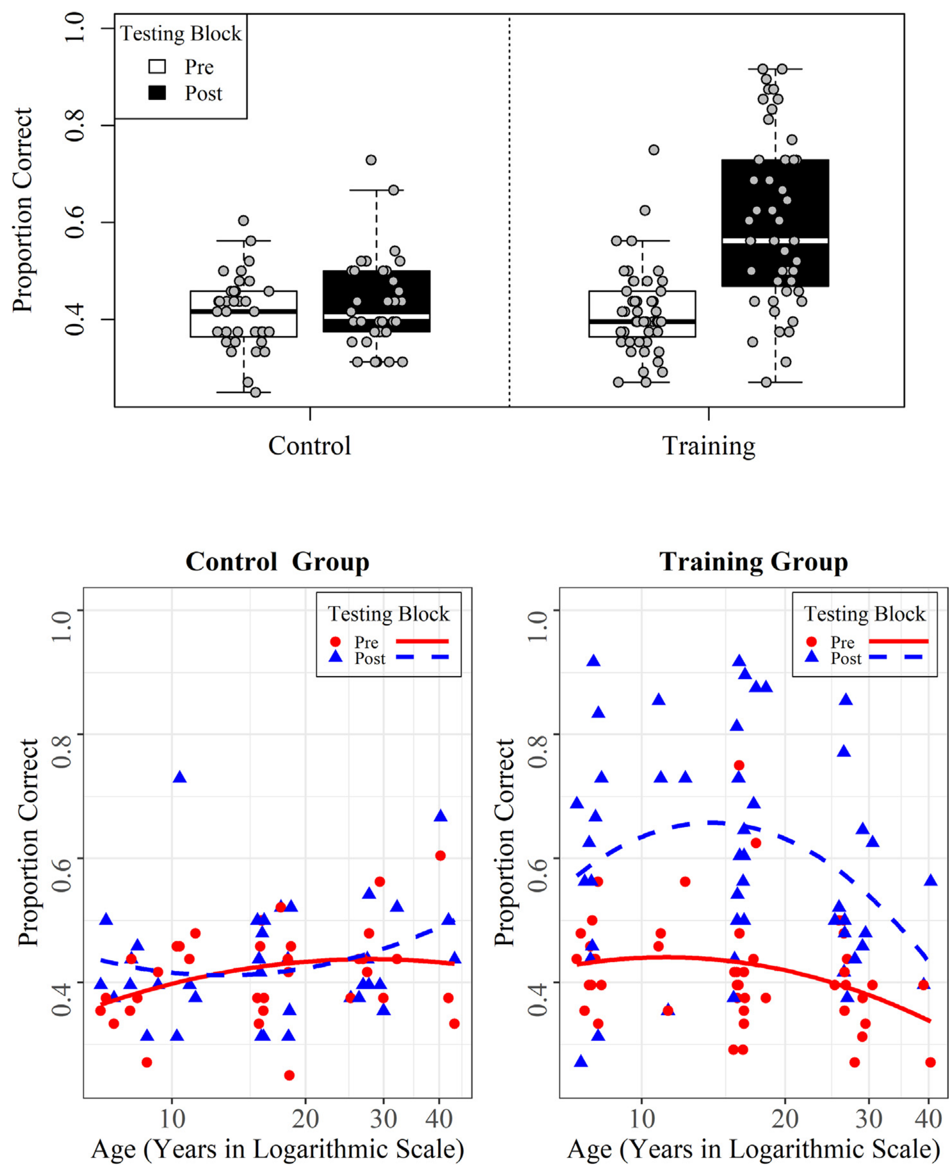

Fig. 4. Boxplots (top) of category discrimination accuracies for the pre test (white boxes) and post test (black boxes) completed by Japanese speakers assigned to the control (left) and training (right) groups, and scatterplots (bottom) of the category discrimination accuracies along with age in logarithmic scale for pre test (circle dots) and post test (triangle dots). The lines in the scatterplots represent the loess regression curves for pre test (solid lines) and post test (dashed lines).

there was no significant 2-way interaction between group and testing block, $\beta = 0.02$, SE = 0.03, $z = 0.59$, $p > .05$. The 3-way interaction of group, testing block, and the linear function of age was also not significant, $\beta = 2.77$, SE = 2.13, $z = 1.30$, $p > .05$, nor was the 3-way interaction of group, testing block, and the quadratic function of age, $\beta = -0.92$, SE = 2.13, $z = -0.43$, $p > .05$. These results suggest that training did not increase subjects' sensitivity to F3 within the /r/ category (see the boxplots in Fig. 3). Detailed results of the statistical analyses for the auditory discrimination accuracy are reported in Appendix A (Tables 4.1, 4.2, and 4.3).

Fig. 4 displays the category discrimination accuracy for Japanese speakers at pre and post tests. The logistic mixed effects model for category discrimination included the fixed factors of group, testing block, the polynomial (linear, quadratic) functions of log age, and all possible interactions of these factors (see Appendix A: Table 5). All the contrasts for the categorical variables were orthogonal. The random factors were crossed intercepts of subject and stimulus, and their random slopes for testing block were included. The logistic mixed effects model demonstrated that there was a significant interaction between group and testing block, $\beta = 0.19$, SE = 0.04, $z = 5.10$, $p < .001$, suggesting that training had a significant effect on category discrimination accuracy (see the boxplots in Fig. 4). In addition, the 3-way interaction of group, testing block, and quadratic function of age was also significant, $\beta = -6.75$, SE = 3.41, $z = -1.98$, $p = .048$, suggesting that adolescents improved their category discrimination accuracy more than adults and children did after completing training sessions (see Appendix A: Table 5 for detailed statistical results).

Fig. 5 displays how age affected Japanese speakers' improvement in identification accuracy over the course of 10 training sessions. At the end of each training session, a short identification test with only 20 trials was conducted. The

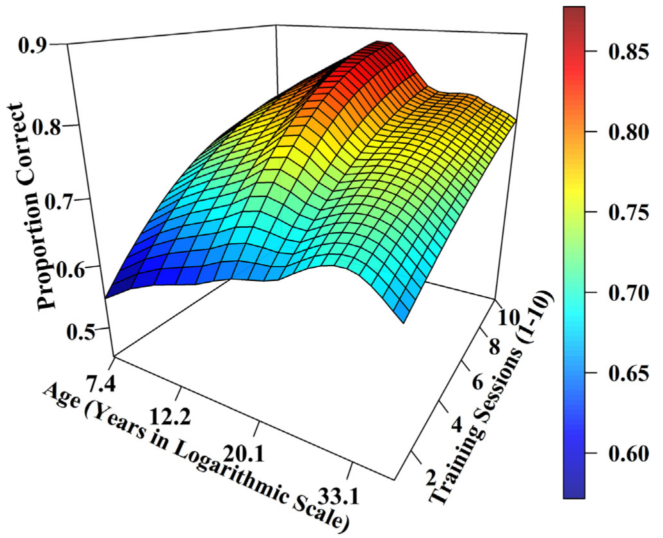

Fig. 5. Three-dimension loess regression curve for the identification accuracy improvement through the 10 training sessions along with the log-scaled age.

three-dimensional figure illustrates the relationship of identification accuracy, age, and number of training sessions completed. This prediction curve was calculated based on the subjects' identification scores at each training session.

The best-fitting logistic mixed effects model for this analysis included the fixed factors of training session (1 to 10, coded linearly), polynomial functions of log age (linear, quadratic), and their interaction. The random factors were the intercept of subject and the intercept of word nested under speaker. Their random slopes for the training session were not included due to the singular fit. The logistic mixed effects model demonstrated that there was a significant effect of training session, $\beta = 0.14$, SE = 0.02, $z = 8.21$, $p < .001$, suggesting that Japanese speakers improved their identification accuracy through the course of 10 training sessions. The interaction between training session and the linear function of age was significant, $\beta = -3.30$, SE = 0.95, $z = -3.47$, $p < .001$, as well as the one between training session and the quadratic function of age, $\beta = 2.80$, SE = 0.98, $z = 2.86$, $p < .01$. These results suggest that older learners were better at identifying the /r/-/l/ contrast than younger learners at the beginning, but as Japanese speakers received more training sessions, younger learners improved their identification accuracy more than older learners, and adolescents improved their identification accuracy more than children (see Appendix A: Table 6 for detailed statistical results).

3.2. Production

Fig. 6 displays the identifiability of Japanese speakers' English /r/-/l/ productions for pre and post tests judged by 10 SSBE speakers. For the statistical analysis, a logistic mixed effects model based on correct/incorrect binomial responses was used for production identifiability. The best-fitting model included the fixed factors of group, testing block, polynomial functions of log age (linear, quadratic), and two 2-way interactions between group and testing block, and between testing block and age (see Appendix A: Table 7). All of the contrasts for the categorical variables were orthogonal. The random factors were crossed intercepts of subjects and words, and their random slopes for the testing block were also included.

The logistic mixed effects model demonstrated that there was a significant 2-way interaction of group and testing block, $\beta = 0.17$, SE = 0.05, $z = 3.72$, $p < .001$, suggesting that Japanese speakers in the training group increased their production identifiability significantly more than did those in the control group (see the boxplots). Although the 2-way interaction between testing block and linear function of age was marginally significant, $\beta = -6.17$, SE = 3.63, $z = -1.70$, $p = .089$, the 3-way interaction of group, testing block and age was not included in the best-fitting model, suggesting that the effect of age on improvement was not significant (see Appendix A: Table 7 for detailed statistical results).

Fig. 7 displays the F3 frequencies of Japanese speakers' English /r/-/l/ productions for pre and post tests. A linear mixed effects model was used for the primary acoustic cue (F3). It included the fixed factors of group, consonant (i.e., English /r/, English /l/), testing block, and polynomial (linear, quadratic) functions of log age. Six 2-way interactions and three 3-way interactions were also included (see Appendix A: Table 8). All the contrasts for the categorical variables were orthogonal. The random factors were crossed intercepts of subject and word, and their random slopes for testing block were not included.

The results demonstrated that the 3-way interaction of group, consonant, and testing block was significant, $\beta = -20.97$, SE = 9.84, $t = -2.13$, $p = .033$. This suggests that training increased Japanese speakers' F3 contrast for English /r/-/l/ productions (see the boxplots). Although the 3-way interaction of consonant, testing block, and linear function of age was marginally significant, $\beta = 687.20$, SE = 391.75, $t = 1.75$, $p = .080$, neither the 3-way interaction of group, consonant, and age, nor the 4-way interaction of group, consonant, testing block, and age was included in the best-fitting model. These results suggest that age did not affect the degree of the F3 training effect (see Appendix A: Table 8 for detailed statistical results).

For brevity, only the F3 changes are reported here, but please see the Appendix B for F2, F3–F2 difference, and duration cues. Japanese speakers improved their /r/-/l/ production distinction in F2, steady state duration, and transition duration. There were age effects on the /r/-/l/ distinction in F2, steady state duration, and transition duration. Adolescents increased the distinction in F2 and transition duration more than adults and children, but children increased it in the steady state duration more than did older learners.

3.3. Link between perception and production

Table 2 shows the correlation matrix of changes in each variable from pre to post test by Japanese speakers in the training group. There were seven variables: perceptual identification accuracy (Perc id), production identifiability (Prod id), F3 sensitivity at the English /r/-/l/ boundary (Perc F3 bnd), F2 sensitivity at the English /r/-/l/ boundary (Perc F2 bnd), F3 sensitivity within the English /r/ category (Perc F3R), category discrimination (Perc cd), and F3 difference for the /r/-/l/ contrast in production (Prod F3). Pearson's r was used for the correlation tests; the r values are displayed in Table 2. As 11 correlations

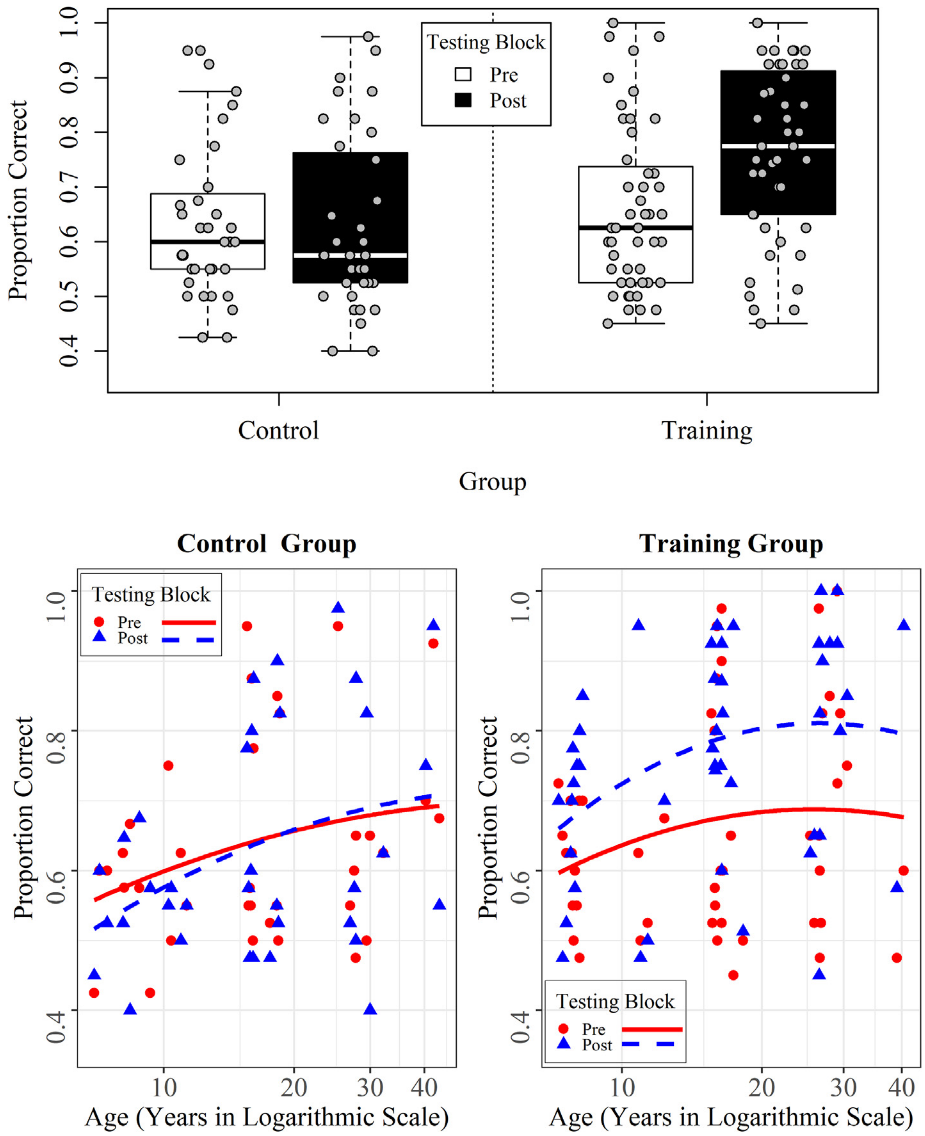

Fig. 6. Boxplots (top) of the identifiability of English /r/ and /l/ produced by Japanese speakers in the control (left) and training (right) groups for pre test (white boxes) and post test (black boxes), judged by 10 Standard Southern British English (SSBE) speakers, and scatterplots (bottom) of the identifiability, along with age for pre test (circle dots) and post test (triangle dots). The lines in the scatterplots represent the loess regression curves for pre test (solid lines) and post test (dashed lines).

were tested, *p* values were adjusted using the False Discovery Rate (FDR) method (Benjamini & Hochberg, 1995).

As shown in Table 2, there was a significant positive correlation between the perceptual identification accuracy improvement and the production identifiability improvement, $r = 0.39$, df = 45, $p = .017$. This result suggests that Japanese speakers who improved their perceptual identification accuracy also tended to improve their production identifiability. There was also another positive significant correlation between the improvements in subjects' production identifiability and F3 sensitivity at the /r/-/l/ boundary, $r = 0.41$, df = 45, $p = .016$, suggesting that Japanese speakers who became more sensitive to the primary acoustic cue often produced a more intelligible /r/-/l/ contrast. In addition, the significant positive correlation between production identifiability and F3 distinction for subjects' /r/-/l/ productions, $r = 0.50$, df = 43, $p < .01$, suggests that Japanese speakers who increased their F3 contrast (e.g., lowering for /r/ and raising for /l/) also tended to produce consonants that could better be labeled as /r/ and /l/ by English listeners. Finally, there was a significant positive correlation between perceptual identification and perceptual category discrimination, $r = 0.56$, df = 45, $p < .001$.

To explore age effects for these correlations, three subgroups were created (16 children aged under 13 years, 16 adolescents aged 15–18 years, 15 adults aged 25 years and over). Fig. 8 displays the correlation in the improvement between production identifiability and perceptual sensitivity to F3 at the /r/-/l/ boundary in each age group. The correlation was stronger for adolescents than for children, but the difference between the correlations was only marginally significant, $z = 1.81$, $p = .070$. There were no other correlation coefficient differences between age groups, suggesting that age did not affect the link between perception and production improvement. However, the lack of significant differences could be a result of a power issue, given the numbers within each subgroup.

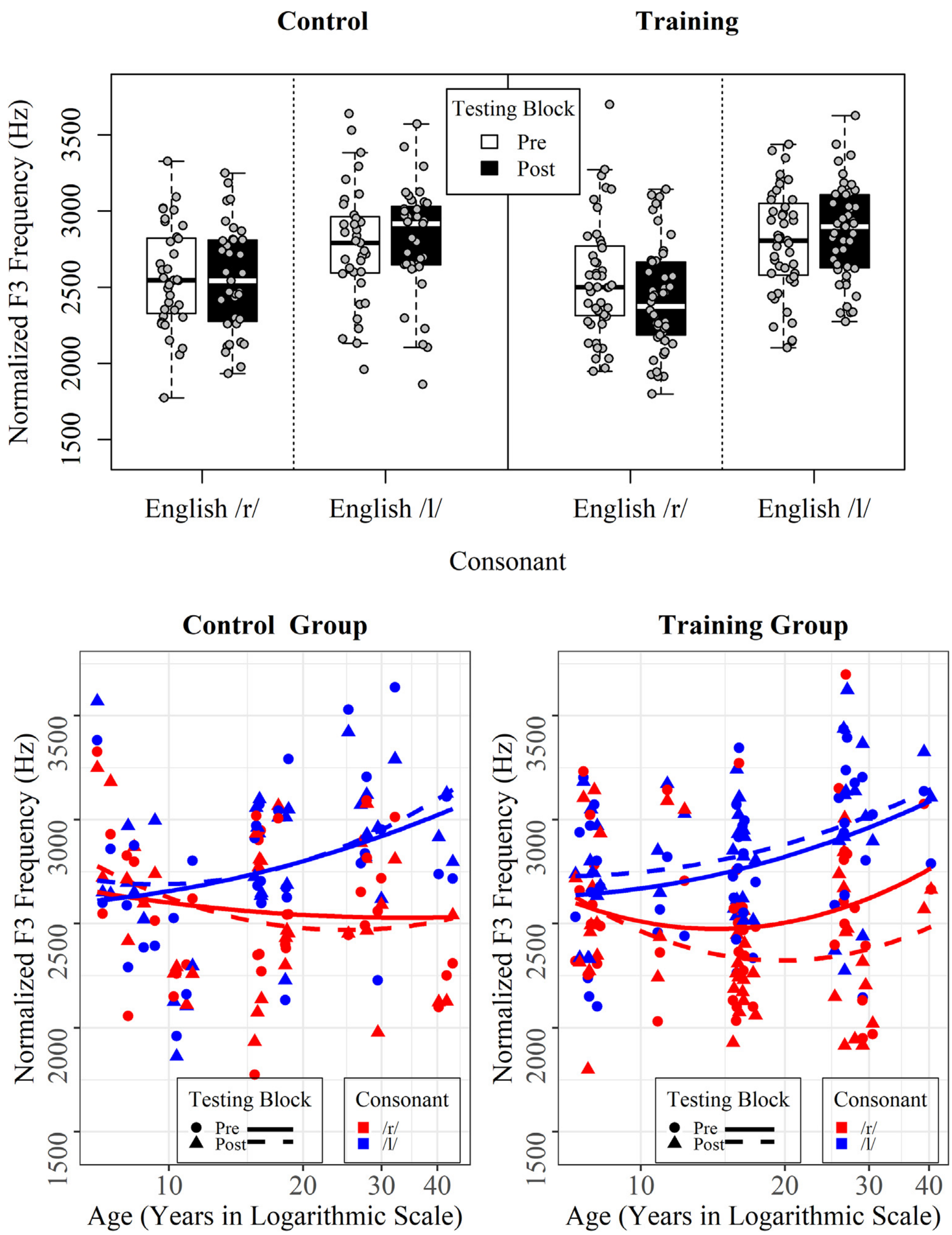

Fig. 7. Boxplots (top) of the normalized F3 frequency of English /r/ and /l/ produced by Japanese speakers assigned in the control (left) and training (right) groups, for pre test (white boxes) and post test (black boxes), and scatterplots (bottom) of the normalized F3 frequency of the English /r/ (red) and /l/ (blue) along with age for pre test (circle dots) and post test (triangle dots). The lines in the scatterplots represent the loess regression curves for pre test (solid lines) and post test (dashed lines).

4. Discussion

This study examined how age affects Japanese speakers' improvement in their perception and production of English /r/ and /l/, using computer-based high variability phonetic training. We expected that training would be more effective for children, given that they likely have less L1 interference due to less established L1 categories and processing, even though previous child training results have been equivocal (e.g., Heeren & Schouten, 2010; Wang & Kuhl, 2003). Our results demonstrated that training improved Japanese speakers' perceptual identification, auditory discrimination, and category discrimination, and that age affected the degree of improvement. Adolescents improved more than adults and children in terms of identification and category discrimination and had greater improvement in their sensitivity to F3 at the /r/-/l/ phoneme boundary.

These results are not consistent with previous results examining the effects of age on training the discrimination of L2 consonantal duration (Heeren & Schouten, 2010) and L2 lexical tone perception (Wang & Kuhl, 2003). These previous studies demonstrated that improvement was comparable between adults and children, but the present study clearly showed advantages for children and adolescents compared to adults. It is worth noting, however, that all these studies involved very different acoustic cues, which potentially has an impact on how

Table 2
A correlation matrix (Pearson's r) of changes from pre-training to post-training tests in perceptual identification (Perc id), production identifiability (Prod id), perceptual F3 sensitivity at the English /r/-/l/ boundary (Perc F3 bnd), perceptual F2 sensitivity at the English /r/-/l/ boundary (Perc F2 bnd), perceptual F3 sensitivity within the English /r/ category (Perc F3R), perceptual category discrimination (Perc CD), and F3 difference between /r/ and /l/ in production (Prod F3). All data included here are from the training group only.

CORRELATIONS (Pearson's *r*)		Perc F3 bnd	Perc F2 bnd	Perc F3R	Perc CD	Prod id	Prod F3
All (N = 47)	Perc id	0.09	−0.06	0.05	0.56***	0.39*	−0.02
	Prod id	0.41*	−0.03	0.08	0.28		0.50**

*** $p < .001$, ** $p < .01$, * $p < .05$

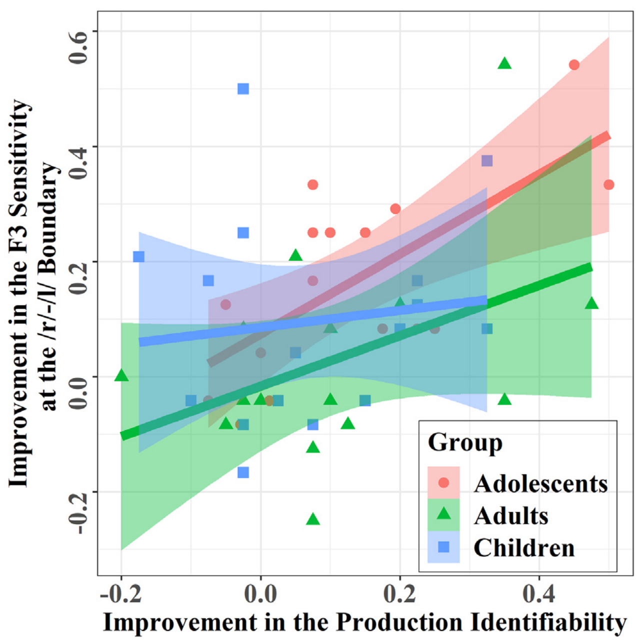

Fig. 8. Scatterplots of the improvements in the primary acoustic cue sensitivity (F3 at the /r/-/l/ phoneme boundary) and the production identifiability by each age group (16 Children aged under 13, 16 Adolescents aged 15–18, and 15 Adults aged 25 and over).

these cues are learned. Iverson et al. (2003; Iverson et al., 2016) suggested that F3 sensitivity for English /r/ and /l/ arises from early changes in perceptual sensitivity that make learning higher-level phonemic categories difficult. Lexical tone perception has been demonstrated to have language-specific aspects even earlier (i.e., at the brainstem; Krishnan et al., 2005; Krishnan, Gandour, et al., 2009; Krishnan, Swaminathan, et al., 2009), but on the other hand, it is clear that all non-native groups can perceive linguistic pitch given that it is important to intonation. It is thus possible that non-native listeners can discern enough F0 information to learn, at least to some extent, phonological categories for tone. A similar conjecture could be made for duration, although the improvement observed by Heeren and Schouten (2010) was small for both adults and children. Given that the learning difficulties for these phonemes might arise from different levels of processing, it is less surprising that the effects of age on training are not uniform, despite the general prediction that learning ought to be easier for individuals who have yet to fully establish their L1 categories (e.g., Flege, 1995, 1999, 2002; Kuhl, 2000; Zhang et al., 2005).

There is some evidence in our data that learning at earlier ages did indeed involve positive changes in sensitivity to the underlying perceptual dimensions, more than for adults. There was a decrease in sensitivity among children to the irrelevant acoustic cue (i.e., F3 within the /r/ category) and an increase in their sensitivity to the primary acoustic cue (i.e., F3 at the /r/-/l/ boundary), suggesting that they may have begun to warp their perception for these phonemes in a native-like manner (e.g., Iverson et al., 2003). That being said, children also increased their F2 sensitivity more than did adults, suggesting that children and adolescents are not solely tuning into the critical F3 distinction but have greater overall auditory plasticity.

Interestingly, our results did not completely support *the younger the better* hypothesis. In this study, younger children did not improve their perception and production more than adolescents, which is contrary to the sensitive/optimal period theory (Oyama, 1976; Werker & Tees, 2005) and SLM (Flege, 1995, 1999, 2002). This is possibly due to the immaturity of phonemic awareness and the auditory sensory system of younger children. According to Ogawa and Tanemura (2001), basic Japanese moraic awareness tends to be acquired by the age of four, but it takes more time to develop children's phonemic awareness (Endo, 1991; Matsumoto, 2004, 2006). For example, Japanese children aged 6–7 years old, who are in the first grade in elementary school, cannot often delete a word-initial consonant; those who are 9–10 years old in the fourth grade of elementary school are better at doing this task (Mann, 1986). This suggests that the younger children in this study may have had difficulty understanding the concept of the word-initial /r/-/l/ phonemes in English minimal-pair words used in the testing and training programs. Individual variation in Japanese children's phonetic development (see e.g., Arai & Mugitani, 2016) may have also affected their /r/-/l/ learning (Flege et al., 2021).

In addition, it should be noted that training never provides the same language environment that native monolinguals have when they learn their L1. For example, high-variability training techniques do not fully provide the range of phonetic input that L1 monolinguals receive through their language development. Moreover, training requires learners to focus on contrasts (e.g., /r/ vs. /l/), which is not an explicit part of L1 language learning (Flege, 2021).

That being said, there was a learning advantage in the present training study. Previous /r/-/l/ training work had found a ceiling of about 15% in identification improvement regardless of the number of training sessions (Bradlow et al., 1997, 1999; Iverson et al., 2005; Lively et al., 1993, 1994; Logan et al., 1991; MacKain et al., 1981). In our adult work, we have unsuccessfully attempted to exceed this 15% ceiling by manipulating acoustic cues (Iverson et al., 2005) and by combining identification and discrimination training methods (Shinohara & Iverson, 2018). The present study trained younger learners, and we were able to exceed the 15% ceiling: Japanese speakers from 7 to 18 years old (i.e., both children and adolescents) improved word-initial /r/-/l/ identification accuracy by 24.7% on average (SD = 17.2). Although we continue to hope that better training outcomes will be found for adults, one way forward might be to increase the use of computer-based training while new English learners are still in school.

Japanese speakers also improved their production identifiability and the degree of F3 distinction in their production. We hypothesized that Japanese speakers would improve both identifiability and F3 distinction for the contrast after training, and that the younger learners who improved their perception more than adults would also improve their production more than adults (Baker et al., 2008; Kartushina & Frauenfelder, 2014). Although the former hypothesis was supported, no effects of age were observed in the improvements of production identifiability or F3 production distinction. Given this null result, it could be concluded that production and perception improvements are not related. However, the improvement rate of production identifiability was positively correlated with that of perceptual identification and that of F3 sensitivity at the English /r/-/l/ boundary. While only two significant correlations may not strongly suggest that there is a common underlying factor governing all perception and production abilities, it may be unlikely that these two abilities are completely independent. Bradlow et al. (1997) trained adult Japanese speakers using a high variability identification training method but found there was no correlation between their improvements in perception and production. Shinohara and Iverson (2018) had a similar finding for identification and discrimination training for Japanese speakers learning the English /r/-/l/ contrast. Those previous studies concluded that the learning may progress at different rates within each subject. The cause of the inconsistent result between the present and previous studies is not clear.

A question remains open on how L2 perception and production are related. SLM proposes that perception learning precedes production and that L2 perception accuracy places an upper limit on production accuracy (Flege & Bohn, 2021). The results of the /r/-/l/ training studies for Japanese speakers could be seen to support this hypothesis: /r/-/l/ perceptual training improves production (Bradlow et al., 1997; Shinohara & Iverson, 2018), and the improvement in perception is correlated with the improvement in production in the present study. However, finding correlations in improvement between perception and production does not mean that perceptual improvement always precedes production improvement. Indeed, Hattori (2009) demonstrated that Japanese speakers can improve their /r/-/l/ production by production training without perception improvement. It is difficult to test correlations with small sample sizes and to examine whether there is an underlying common factor governing both L2 perception and production learning. We did not find an age effect on correlation between perception and production improvements, but the important fact is that people have the ability to integrate their perception and vocal tract gestures for their L2 learning (Flege & Bohn, 2021).

In summary, we found that younger Japanese speakers, particularly adolescents, had an advantage in terms of improving their perception of the English /r/-/l/ contrast, but there remain uncertainties about what actually improved. Previous work has identified F3 sensitivity as the main driver of individual differences in English /r/-/l/ identification (Hattori & Iverson, 2009), but, in the present study, changes in F3 sensitivity were more strongly linked to the improvement of production than that of perception. Iverson et al. (2005) concluded that phonetic training for Japanese learners worked by applying the existing L1 category representations for the /r/-/l/ identification, but that does not explain why younger learners, who likely have less developed L1 representations, learned better. It may be that better understanding the initial state of the learner (e.g., quality and quantity of the input, and details of existing category representations) will provide a way forward. It may also be true that the extensive literature on phonetic training with adults is not completely relevant for young learners, and that more general auditory and attention factors (e.g., Holt et al., 2018) might be more influential for young learners than are factors related to L1 category interference.

CRediT authorship contribution statement

Yasuaki Shinohara: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Paul Iverson:** Methodology, Resources, Writing - review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The first author of this research (Yasuaki Shinohara) was funded by the Japan Student Services Organization, when this study was conducted (JASSO, ID: L10326890002). We thank our subjects and colleagues for their support.

Appendix

A. Statistical analyses reports

Table 3
Statistical analysis for the identification test.

Fixed factor (contrast)	Estimate	SE	z	p	
(Intercept)	0.36	0.04	8.49	< .001	***
Group (training vs. control)	−0.14	0.03	−4.52	< .001	***
Block (pre vs. post)	−0.13	0.01	−8.88	< .001	***
/r/-/l/ Position (initial vs. others)	−0.08	0.02	−3.49	< .001	***
/r/-/l/ Position (medial vs. cluster)	−0.08	0.04	−2.04	.041	*
Log age (linear)	20.88	3.85	5.43	< .001	***
Log age (quadratic)	−5.94	4.16	−1.43	> .05	
Group (training vs. control): Block (pre vs. post)	0.13	0.01	8.96	< .001	***
Group (training vs. control): /r/-/l/ Position (initial vs. others)	0.04	0.01	3.78	< .001	***
Group (training vs. control): /r/-/l/ Position (medial vs. cluster)	−0.01	0.02	−0.56	> .05	
Block (pre vs. post): /r/-/l/ Position (initial vs. others)	0.07	0.01	6.31	< .001	***
Block (pre vs. post): /r/-/l/ Position (medial vs. cluster)	0.03	0.02	1.78	.075	.
Group (training vs. control): Log age (linear)	1.07	3.36	0.32	> .05	
Group (training vs. control): Log age (quadratic)	−2.87	4.23	−0.68	> .05	
Block (pre vs. post): Log age (linear)	3.14	2.13	1.47	> .05	
Block (pre vs. post): Log age (quadratic)	5.88	2.09	2.82	< .01	**
/r/-/l/ Position (initial vs. others): Log age (linear)	1.05	1.48	0.71	> .05	
/r/-/l/ Position (medial vs. cluster): Log age (linear)	−3.99	2.50	−1.60	> .05	
/r/-/l/ Position (initial vs. others): Log age (quadratic)	1.59	1.49	1.06	> .05	
/r/-/l/ Position (medial vs. cluster): Log age (quadratic)	−2.36	2.50	−0.95	> .05	
Group (training vs. control): Block (pre vs. post): /r/-/l/ Position (initial vs. others)	−0.06	0.01	−5.24	< .001	***
Group (training vs. control): Block (pre vs. post): /r/-/l/ Position (medial vs. cluster)	−0.01	0.02	−0.75	> .05	
Group (training vs. control): Block (pre vs. post): Log age (linear)	0.004	2.09	0.002	> .05	
Group (training vs. control): Block (pre vs. post): Log age (quadratic)	−4.90	2.06	−2.38	.018	*
Group (training vs. control): /r/-/l/ Position (initial vs. others): Log age (linear)	−5.13	1.47	−3.50	< .001	***
Group (training vs. control): /r/-/l/ Position (medial vs. cluster): Log age (linear)	5.28	2.54	2.08	.037	*
Group (training vs. control): /r/-/l/ Position (initial vs. others): Log age (quadratic)	−3.38	1.48	−2.29	.022	*
Group (training vs. control): /r/-/l/ Position (medial vs. cluster): Log age (quadratic)	−0.11	2.48	−0.04	> .05	

Table 4.1
Statistical analysis for the auditory discrimination test for the primary acoustic cue (F3 sensitivity at the English /r/-/l/ phoneme boundary).

Fixed factor (contrast)	Estimate	SE	z	p	
(Intercept)	0.01	0.13	0.08	> .05	
Group (training vs. control)	−0.15	0.07	−2.03	.042	*
Block (pre vs. post)	−0.14	0.04	−3.28	< .01	**
Log age (linear)	18.15	4.58	3.97	< .001	***
Log age (quadratic)	−10.00	4.60	−2.18	.030	*
Group (training vs. control): Block (pre vs. post)	0.10	0.04	2.32	.020	*
Group (training vs. control): Log age (linear)	7.59	4.56	1.66	.096	.
Group (training vs. control): Log age (quadratic)	11.44	4.60	2.49	.013	*
Block (pre vs. post): Log age (linear)	−0.25	2.57	−0.10	> .05	
Block (pre vs. post): Log age (quadratic)	−2.18	2.59	−0.84	> .05	
Group (training vs. control): Block (pre vs. post): Log age (linear)	−2.66	2.57	−1.03	> .05	
Group (training vs. control): Block (pre vs. post): Log age (quadratic)	−5.24	2.61	−2.00	.045	*

Table 4.2
Statistical analysis for the auditory discrimination test for the secondary acoustic cue (F2 sensitivity at the English /r/-/l/ phoneme boundary).

Fixed factor (contrast)	Estimate	SE	z	p	
(Intercept)	−0.10	0.12	−0.83	> .05	
Group (training vs. control)	−0.03	0.06	−0.50	> .05	
Block (pre vs. post)	−0.22	0.03	−6.53	< .001	***
Log age (linear)	9.39	4.06	2.32	.021	*
Log age (quadratic)	−12.07	4.08	−2.96	< .01	**
Group (training vs. control): Block (pre vs. post)	−0.01	0.03	−0.37	> .05	
Group (training vs. control): Log age (linear)	6.02	4.05	1.49	> .05	
Group (training vs. control): Log age (quadratic)	5.11	4.08	1.25	> .05	
Block (pre vs. post): Log age (linear)	1.72	2.15	0.80	> .05	
Block (pre vs. post): Log age (quadratic)	−1.42	2.12	−0.67	> .05	
Group (training vs. control): Block (pre vs. post): Log age (linear)	−4.28	2.14	−2.00	.046	*
Group (training vs. control): Block (pre vs. post): Log age (quadratic)	−2.32	2.13	−1.09	> .05	

Table 4.3
Statistical analysis for the auditory discrimination test for the irrelevant acoustic cue (F3 sensitivity within the English /r/ category).

Fixed factor (contrast)	Estimate	SE	z	p	
(Intercept)	−0.49	0.13	−3.83	< .001	***
Group (training vs. control)	−0.05	0.04	−1.14	> .05	
Block (pre vs. post)	−0.05	0.03	−1.46	> .05	
Log age (linear)	4.06	2.77	1.46	> .05	
Log age (quadratic)	−1.11	2.76	−0.40	> .05	
Group (training vs. control): Block (pre vs. post)	0.02	0.03	0.59	> .05	
Group (training vs. control): Log age (linear)	−0.68	2.78	−0.25	> .05	
Group (training vs. control): Log age (quadratic)	7.40	2.76	2.68	< .01	**
Block (pre vs. post): Log age (linear)	−5.54	2.14	−2.59	< .01	**
Block (pre vs. post): Log age (quadratic)	4.00	2.14	1.87	.061	.
Group (training vs. control): Block (pre vs. post): Log age (linear)	2.77	2.13	1.30	> .05	
Group (training vs. control): Block (pre vs. post): Log age (quadratic)	−0.92	2.13	−0.43	> .05	

Table 5
Statistical analysis for the category discrimination test.

Fixed factor (contrast)	Estimate	SE	z	p	
(Intercept)	−0.15	0.10	−1.53	> .05	
Group (training vs. control)	−0.21	0.05	−4.31	< .001	***
Block (pre vs. post)	−0.24	0.05	−5.30	< .001	***
Log age (linear)	−2.61	4.35	−0.60	> .05	
Log age (quadratic)	−6.83	4.45	−1.54	> .05	
Group (training vs. control): Block (pre vs. post)	0.19	0.04	5.10	< .001	***
Group (training vs. control): Log age (linear)	8.50	4.30	1.98	.048	*
Group (training vs. control): Log age (quadratic)	8.17	4.32	1.89	.059	.
Block (pre vs. post): Log age (linear)	1.00	3.45	0.29	> .05	
Block (pre vs. post): Log age (quadratic)	1.62	3.42	0.48	> .05	
Group (training vs. control): Block (pre vs. post): Log age (linear)	−0.53	3.43	−0.15	> .05	
Group (training vs. control): Block (pre vs. post): Log age (quadratic)	−6.75	3.41	−1.98	.048	*

Table 6
Statistical analysis for the improvement through 10 training sessions.

Fixed factor (contrast)	Estimate	SE	z	p	
(Intercept)	1.11	0.19	5.74	< .001	***
Session (linear)	0.14	0.02	8.21	< .001	***
Log age (linear)	3.94	8.67	0.45	> .05	
Log age (quadratic)	−13.14	9.46	−1.39	> .05	
Session (linear): Log age (linear)	−3.30	0.95	−3.47	< .001	***
Session (linear): Log age (quadratic)	2.80	0.98	2.86	< .01	**

Table 7
Statistical analysis for the English /r/-/l/ production identifiability.

Fixed factor (contrast)	Estimate	SE	z	p	
(Intercept)	0.92	0.14	6.60	< .001	***
Group (training vs. control)	−0.24	0.09	−2.73	< .01	**
Block (pre vs. post)	−0.17	0.05	−3.57	< .001	***
Log age (linear)	23.04	7.12	3.24	< .01	**
Log age (quadratic)	−6.39	7.26	−0.88	> .05	
Group (training vs. control): Block (pre vs. post)	0.17	0.05	3.72	< .001	***
Block (pre vs. post): Log age (linear)	−6.17	3.63	−1.70	.089	.
Block (pre vs. post): Log age (quadratic)	1.23	3.65	0.34	> .05	

Table 8
Statistical analysis for the F3 distinction for the English /r/-/l/ production.

Fixed factor (contrast)	Estimate	SE	t	p	
(Intercept)	2677.95	54.19	49.42	< .001	***
Group (training vs. control)	−5.20	35.11	−0.15	> .05	
Consonant (R vs. L)	140.26	42.43	3.31	.011	*
Block (pre vs. post)	−6.28	9.84	−0.64	> .05	
Log age (linear)	1688.22	1420.56	1.19	> .05	
Log age (quadratic)	1152.93	1418.59	0.81	> .05	
Group (training vs. control): Consonant (R vs. L)	−29.90	9.84	−3.04	< .01	**
Group (training vs. control): Block (pre vs. post)	10.34	9.84	1.05	> .05	
Consonant (R vs. L): Block (pre vs. post)	30.06	9.82	3.06	< .01	**
Group (training vs. control): Log age (linear)	−624.09	1420.56	−0.44	> .05	

Table 8 (*continued*)

Fixed factor (contrast)	Estimate	SE	*t*	*p*	
Group (training vs. control): Log age (quadratic)	184.91	1418.58	0.13	> .05	
Consonant (R vs. L): Log age (linear)	3259.88	391.74	8.32	< .001	***
Consonant (R vs. L): Log age (quadratic)	−304.12	392.63	−0.78	> .05	
Block (pre vs. post): Log age (linear)	−849.15	398.71	−2.13	.033	*
Block (pre vs. post): Log age (quadratic)	420.62	397.48	1.06	> .05	
Group (training vs. control): Consonant (R vs. L): Block (pre vs. post)	−20.97	9.84	−2.13	.033	*
Group (training vs. control): Block (pre vs. post): Log age (linear)	262.95	398.72	0.66	> .05	
Group (training vs. control): Block (pre vs. post): Log age (quadratic)	253.11	397.47	0.64	> .05	
Consonant (R vs. L): Block (pre vs. post): Log age (linear)	687.20	391.75	1.75	.080	.
Consonant (R vs. L): Block (pre vs. post): Log age (quadratic)	−142.79	392.52	−0.36	> .05	

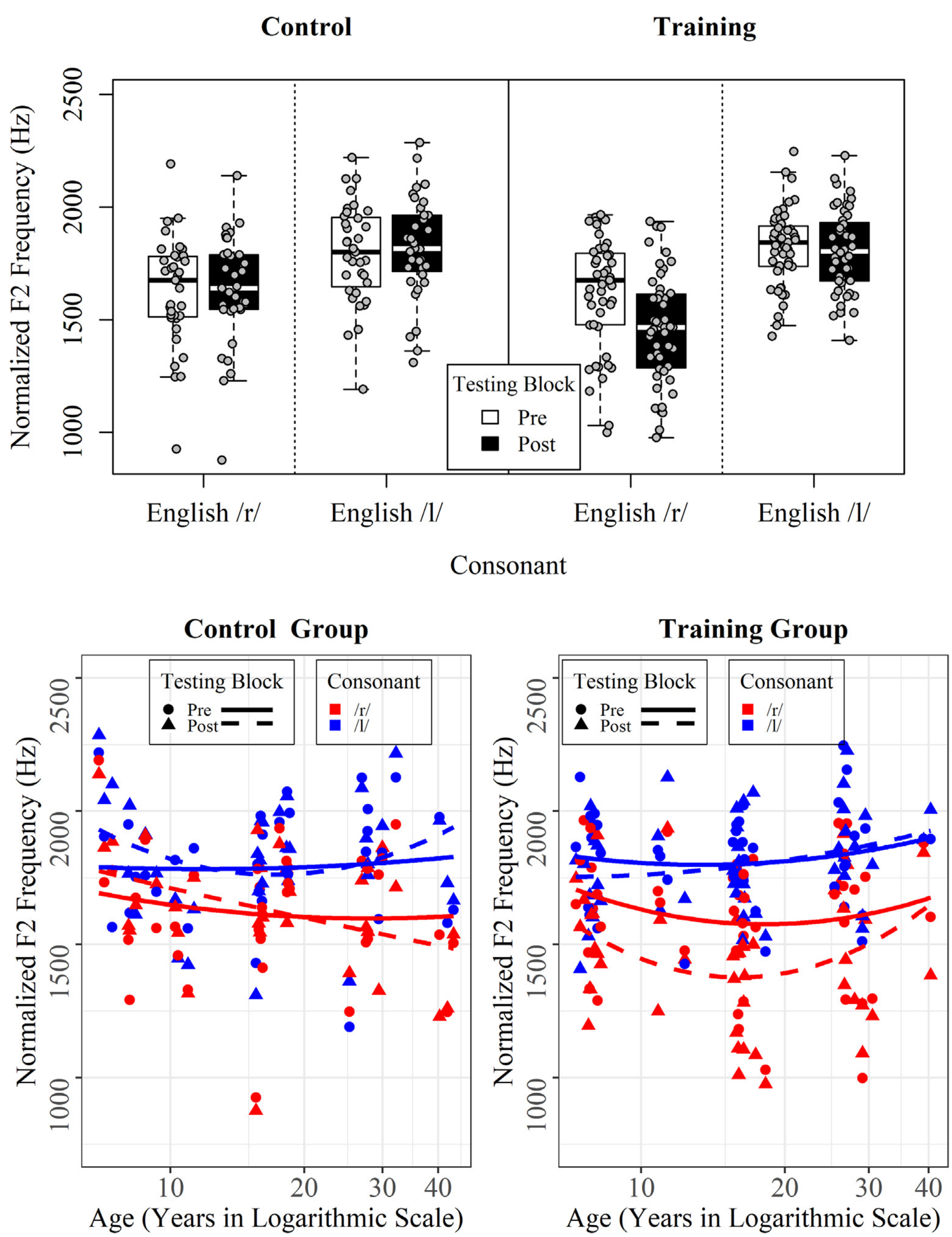

Fig. 9. Boxplots (top) of the normalized F2 frequency of English /r/ and /l/ produced by Japanese speakers assigned to control (left) and training (right) groups, for pre test (white boxes) and post test (black boxes), and scatterplots (bottom) of the normalized F2 frequency of English /r/ (red) and /l/ (blue) along with age for pre test (circle dots) and post test (triangle dots). The lines in the scatterplots represent the loess regression curves for pre test (solid lines) and post test (dashed lines).

Table 9
Statistical analysis for the F2 distinction for the English /r/-/l/ production.

Fixed factor (contrast)	Estimate	SE	*t*	*p*	
(Intercept)	1699.43	42.40	40.08	< .001	***
Consonant (R vs. L)	112.30	37.50	3.00	.017	*
Group (training vs. control)	19.47	20.77	0.94	> .05	
Block (pre vs. post)	−16.24	6.34	−2.56	.010	*
Log age (linear)	−385.52	840.75	−0.46	> .05	
Age (quadratic)	1199.74	839.09	1.43	> .05	
Consonant (R vs. L): Group (training vs. control)	−24.56	6.33	−3.88	< .001	***
Consonant (R vs. L): Block (pre vs. post)	18.97	6.33	3.00	< .01	**
Group (training vs. control): Block (pre vs. post)	26.40	6.34	4.17	< .001	***
Consonant (R vs. L): Log age (linear)	932.64	257.63	3.62	< .001	***
Consonant (R vs. L): Log age (quadratic)	−192.65	255.81	−0.75	> .05	
Group (training vs. control): Log age (linear)	−818.75	840.74	−0.97	> .05	
Group (training vs. control): Log age (quadratic)	−405.80	839.09	−0.48	> .05	
Block (pre vs. post): Log age (linear)	−132.16	257.74	−0.51	> .05	
Block (pre vs. post): Log age (quadratic)	362.72	255.83	1.42	> .05	
Consonant (R vs. L): Group (training vs. control): Block (pre vs. post)	−15.83	6.33	−2.50	.013	*
Consonant (R vs. L): Group (training vs. control): Log age (linear)	87.99	257.61	0.34	> .05	
Consonant (R vs. L): Group (training vs. control): Log age (quadratic)	684.60	255.80	2.68	< .01	**
Consonant (R vs. L): Block (pre vs. post): Log age (linear)	242.58	257.60	0.94	> .05	
Consonant (R vs. L): Block (pre vs. post): Log age (quadratic)	22.85	255.80	0.09	> .05	
Group (training vs. control): Block (pre vs. post): Log age (linear)	−672.09	257.74	−2.61	< .01	**
Group (training vs. control): Block (pre vs. post): Log age (quadratic)	17.60	255.83	0.07	> .05	
Consonant (R vs. L): Group (training vs. control): Block (pre vs. post): Log age (linear)	107.00	257.60	0.42	> .05	
Consonant (R vs. L): Group (training vs. control): Block (pre vs. post): Log age (quadratic)	529.12	255.80	2.07	.039	*

B. Other acoustic analyses of the English /r/-/l/ production

Fig. 9 displays the F2 frequencies of English /r/ and /l/ sounds produced for pre and post tests by Japanese speakers assigned to control and training groups. It was hypothesized that Japanese adults would learn to produce a lower F2 for /r/ and a higher F2 for /l/ after training, as do native English speakers (Hattori, 2009; Hattori & Iverson, 2009; Lotto et al., 2004; Polka & Strange, 2005; Yamada & Tohkura, 1992). That being said, English-speaking children aged from 5 to 8 years tend to raise F2 for /r/ and to lower F2 for /l/, as well as producing a lower F3 for /r/ and a higher F3 for /l/ (Dalston, 1975; Idemaru & Holt, 2013, 2011; Ingvalson et al., 2011; Lotto et al., 2004; McGowan et al., 2004); Japanese children could improve similarly.

The best-fitting linear mixed effects model for the F2 analysis demonstrated that there was a significant 3-way interaction of consonant, group, and testing block, $\beta = -15.83$, SE = 6.33, $t = -2.50$, $p = .013$, suggesting that Japanese speakers increased their F2 distinction for the English /r/-/l/ contrast after training (higher F2 for /l/ and lower F2 for /r/). In addition, the 4-way interaction of consonant, group, testing block, and the quadratic function of age was also significant, $\beta = 529.12$, SE = 255.80, $t = 2.07$, $p = .039$. This suggests that adolescents in the training group increased F2 distinction more than did adults and children, compared to the control group (see the scatterplots). See Table 9 for detailed statistical results.

Fig. 10 displays the difference between the F3 and F2 frequencies of Japanese speakers' /r/-/l/ productions for pre and post tests. It was hypothesized that both Japanese adults and children would learn to produce the smaller F3–F2 difference for /r/ and the larger difference for /l/ after training, as do native English speakers (Aoyama et al., 2019; Espy-Wilson, 1992). The 3-way interaction of group, consonant, and testing block was not significant, $\beta = -5.84$, SE = 9.26, $t = -0.63$, $p > .05$, and the 4-way interaction of group, consonant, testing block, and the quadratic function of age was marginally significant, $\beta = -646.18$, SE = 372.98, $t = -1.73$, $p = .083$. These results suggest that there were no training effects on the F3–F2 difference (see the boxplots), and that age did not significantly affect the training effects (see the scatterplots). See Table 10 for detailed statistical results.

Fig. 11 displays the steady state duration of Japanese speakers' English /r/-/l/ productions for pre and post tests. Steady state duration was defined as the period from the onset of periodicity to the beginning of transition (Ingvalson & Holt, 2011). Previous studies show that Japanese speakers tend to be more sensitive to duration cues than to the primary acoustic cue (i.e., F3) (Hattori, 2009; Iverson et al., 2005). Therefore, it was hypothesized that Japanese speakers would learn to produce a shorter steady state duration for /r/ and a longer steady state duration for /l/, as do native English speakers (Hattori, 2009; Hattori & Iverson, 2009; Kent & Read, 2002).

The linear mixed effects model for the steady state duration analysis demonstrated that there was a significant 3-way interaction of group, consonant, and testing block, $\beta = -1.77$, SE = 0.82, $t = -2.16$, $p = .031$. This suggests that Japanese speakers in the training group improved their /r/-/l/ production distinction (i.e., longer duration for /l/ and shorter duration for /r/) but that those in the control group did not change their distinction as much (see the boxplots). Although the 4-way interaction of group, consonant, testing block, and quadratic function of age was marginally significant, $\beta = 54.48$, SE = 33.03, $t = 1.65$, $p = .099$, the 4-way interaction of group, consonant, testing block, and the linear function of age was significant, $\beta = 107.95$, SE = 33.10, $t = 3.26$, $p < .01$. These results suggest that younger learners improved their /r/-/l/ production distinction more than did the older learners in the training group but that this age effect was not observed in the control group. See Table 11 for detailed statistical results.

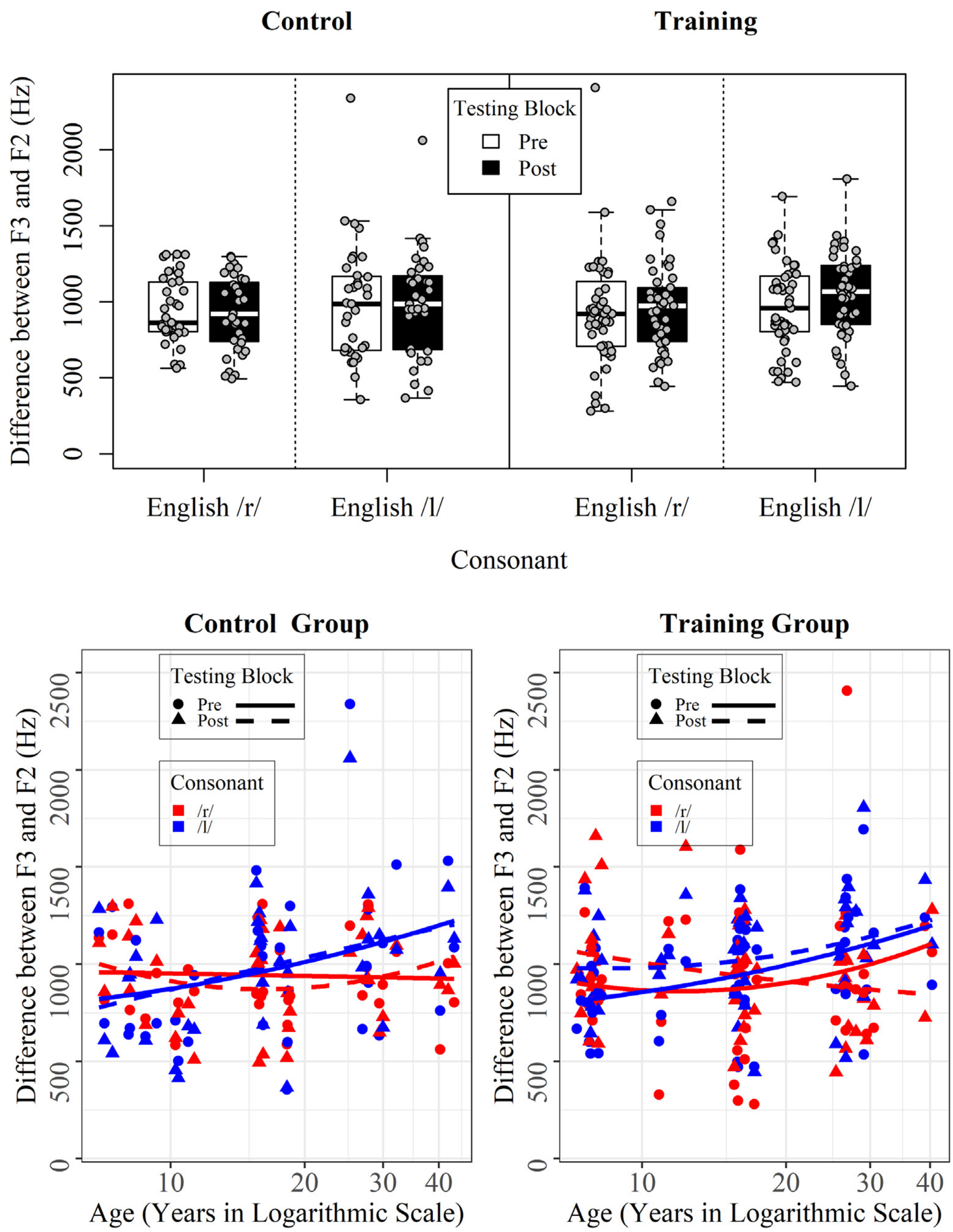

Fig. 10. Boxplots (top) of the F3–F2 difference of English /r/ and /l/ produced by Japanese speakers assigned to control (left) and training (right) groups, for pre test (white boxes) and post test (black boxes), and scatterplots (bottom) of the F3–F2 difference of the English /r/ (red) and /l/ (blue) along with age for pre test (circle dots) and post test (triangle dots). The lines in the scatterplots represent the loess regression curves for pre test (solid lines) and post test (dashed lines).

Table 10
Statistical analysis for the F3–F2 difference for the English /r/-/l/ production.

Fixed factor (contrast)	Estimate	SE	*t*	*p*	
(Intercept)	957.18	27.26	35.12	< .001	***
Group (training vs. control)	−17.78	25.58	−0.70	> .05	
Consonant (R vs. L)	31.46	13.21	2.38	.044	*
Block (pre vs. post)	11.88	9.26	1.28	> .05	
Log age (linear)	1904.17	1033.05	1.84	.069	.
Log age (quadratic)	544.85	1033.65	0.53	> .05	
Group (training vs. control): Consonant (R vs. L)	−9.63	9.26	−1.04	> .05	
Group (training vs. control): Block (pre vs. post)	−16.28	9.26	−1.76	.079	.
Consonant (R vs. L): Block (pre vs. post)	12.11	9.26	1.31	> .05	
Group (training vs. control): Log age (linear)	169.25	1033.05	0.16	> .05	

(*continued on next page*)

Table 10 (*continued*)

Fixed factor (contrast)	Estimate	SE	*t*	*p*	
Group (training vs. control): Log age (quadratic)	−65.29	1033.65	−0.06	> .05	
Consonant (R vs. L): Log age (linear)	2296.43	374.16	6.14	< .001	***
Consonant (R vs. L): Log age (quadratic)	−264.80	373.09	−0.71	> .05	
Block (pre vs. post): Log age (linear)	−828.26	374.21	−2.21	.027	*
Block (pre vs. post): Log age (quadratic)	59.23	373.02	0.16	> .05	
Group (training vs. control): Consonant (R vs. L): Block (pre vs. post)	−5.84	9.26	−0.63	> .05	
Group (training vs. control): Consonant (R vs. L): Log age (linear)	391.17	374.15	1.05	> .05	
Group (training vs. control): Consonant (R vs. L): Log age (quadratic)	−167.76	373.08	−0.45	> .05	
Group (training vs. control): Block (pre vs. post): Log age (linear)	855.05	374.21	2.29	.023	*
Group (training vs. control): Block (pre vs. post): Log age (quadratic)	364.91	373.02	0.98	> .05	
Consonant (R vs. L): Block (pre vs. post): Log age (linear)	477.15	374.01	1.28	> .05	
Consonant (R vs. L): Block (pre vs. post): Log age (quadratic)	−116.07	372.97	−0.31	> .05	
Group (training vs. control): Consonant (R vs. L): Block (pre vs. post): Log age (linear)	−199.86	374.02	−0.53	> .05	
Group (training vs. control): Consonant (R vs. L): Block (pre vs. post): Log age (quadratic)	−646.18	372.98	−1.73	.083	.

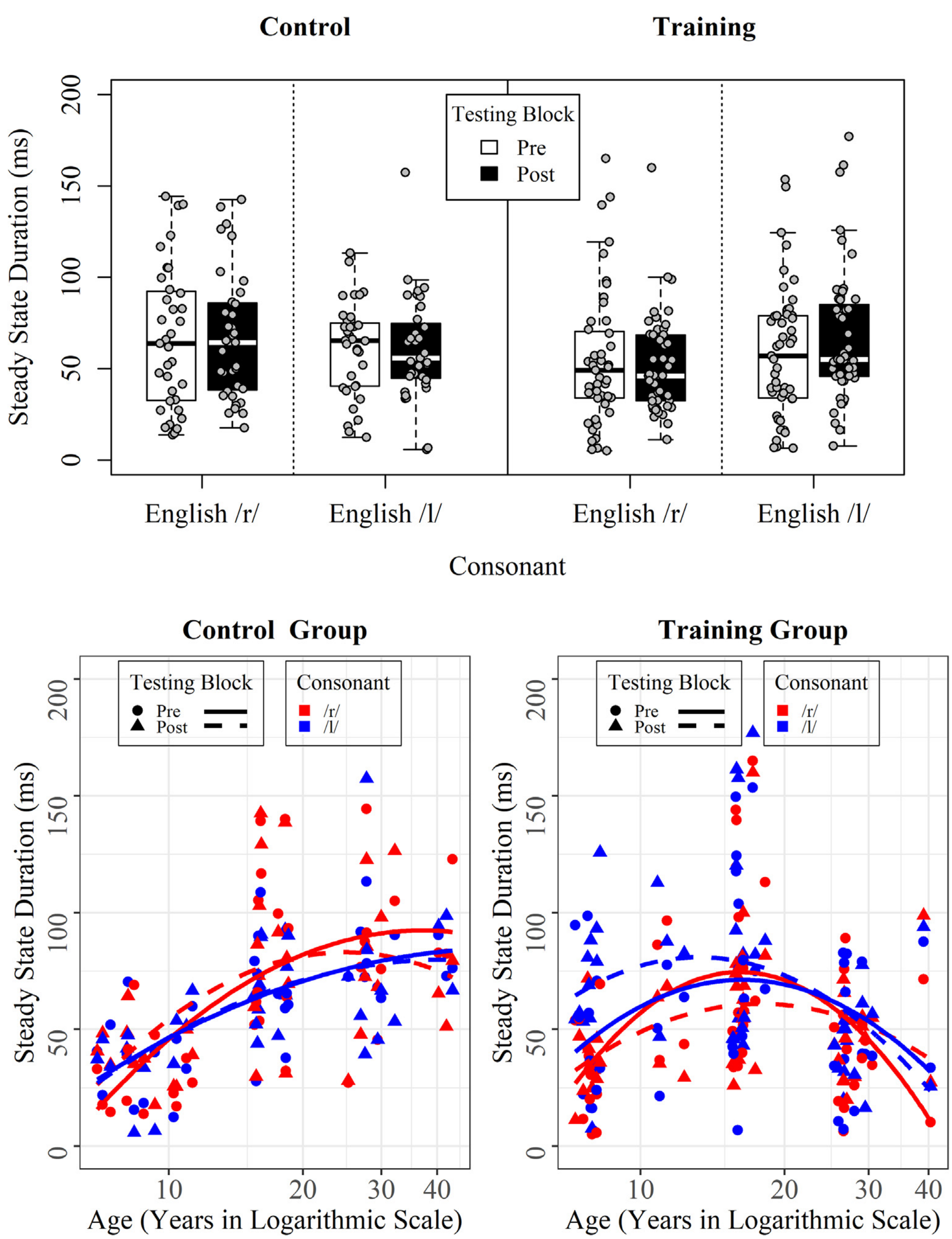

Fig. 11. Boxplots (top) of the steady state duration of English /r/ and /l/ produced by Japanese speakers assigned to control (left) and training (right) groups, for pre test (white boxes) and post test (black boxes), and scatterplots (bottom) of the steady state duration of the English /r/ (red) and /l/ (blue) along with age for pre test (circle dots) and post test (triangle dots). The lines in the scatterplots represent the loess regression curves for pre test (solid lines) and post test (dashed lines).

Table 11
Statistical analysis for the steady state duration of the English /r/-/l/ production.

Fixed factor (contrast)	Estimate	SE	*t*	*p*	
(Intercept)	60.34	3.63	16.62	< .001	***
Group (training vs. control)	3.18	2.73	1.16	> .05	
Consonant (R vs. L)	1.28	2.53	0.51	> .05	
Block (pre vs. post)	0.76	0.82	0.93	> .05	
Log age (linear)	351.55	110.51	3.18	< .01	**
Log age (quadratic)	−383.23	110.37	−3.47	< .001	***
Group (training vs. control): Consonant (R vs. L)	−3.80	0.82	−4.64	< .001	***
Group (training vs. control): Block (pre vs. post)	−0.42	0.82	−0.52	> .05	
Consonant (R vs. L): Block (pre vs. post)	1.08	0.82	1.32	> .05	
Group (training vs. control): Log age (linear)	390.51	110.51	3.53	< .001	***
Group (training vs. control): Log age (quadratic)	142.74	110.37	1.29	> .05	
Consonant (R vs. L): Log age (linear)	−102.13	33.10	−3.09	< .01	**
Consonant (R vs. L): Log age (quadratic)	62.59	33.04	1.89	.058	.
Block (pre vs. post): Log age (linear)	−79.11	33.11	−2.39	.017	*
Block (pre vs. post): Log age (quadratic)	45.28	33.03	1.37	> .05	
Group (training vs. control): Consonant (R vs. L): Block (pre vs. post)	−1.77	0.82	−2.16	.031	*
Group (training vs. control): Consonant (R vs. L): Log age (linear)	24.82	33.10	0.75	> .05	
Group (training vs. control): Consonant (R vs. L): Log age (quadratic)	22.50	33.04	0.68	> .05	
Group (training vs. control): Block (pre vs. post): Log age (linear)	−15.75	33.11	−0.48	> .05	
Group (training vs. control): Block (pre vs. post): Log age (quadratic)	−68.48	33.03	−2.07	.038	*
Consonant (R vs. L): Block (pre vs. post): Log age (linear)	−25.28	33.10	−0.76	> .05	
Consonant (R vs. L): Block (pre vs. post): Log age (quadratic)	−47.35	33.03	−1.43	> .05	
Group (training vs. control): Consonant (R vs. L): Block (pre vs. post): Log age (linear)	107.95	33.10	3.26	< .01	**
Group (training vs. control): Consonant (R vs. L): Block (pre vs. post): Log age (quadratic)	54.48	33.03	1.65	.099	.

Fig. 12 displays the transition duration of Japanese speakers' English /r/-/l/ productions for pre and post tests. The transition duration was defined as the period from the offset of the steady state of the target consonant to the onset of the steady state of the following vowel (Ingvalson et al., 2011). It was hypothesized that Japanese speakers would learn to produce a longer transition for /r/ and a shorter transition for /l/ after training, as do native English speakers (Hattori, 2009; Hattori & Iverson, 2009; Kent & Read, 2002).

The linear mixed effects model for the transition analysis demonstrated that there was a significant 3-way interaction of group, consonant, and testing block, $\beta = 1.57$, SE = 0.31, $t = 5.12$, $p < .001$. This result suggests that Japanese speakers in the training group improved their production distinction (longer transition for /r/ and shorter transition for /l/) after training, compared to those in the control group (see the boxplots). The significant 4-way interaction of group, consonant, testing block, and the quadratic function of age, $\beta = -48.58$, SE = 12.33, $t = -3.94$, $p < .001$, and the marginally significant interaction of group, consonant, testing block, and the linear function of age, $\beta = -22.10$, SE = 12.36, $t = -1.79$, $p = .074$, suggest that age affected the training effect. Younger learners improved their distinction more than older learners, but children did not improve as much as adolescents (see the scatterplots). See Table 12 for detailed statistical results.

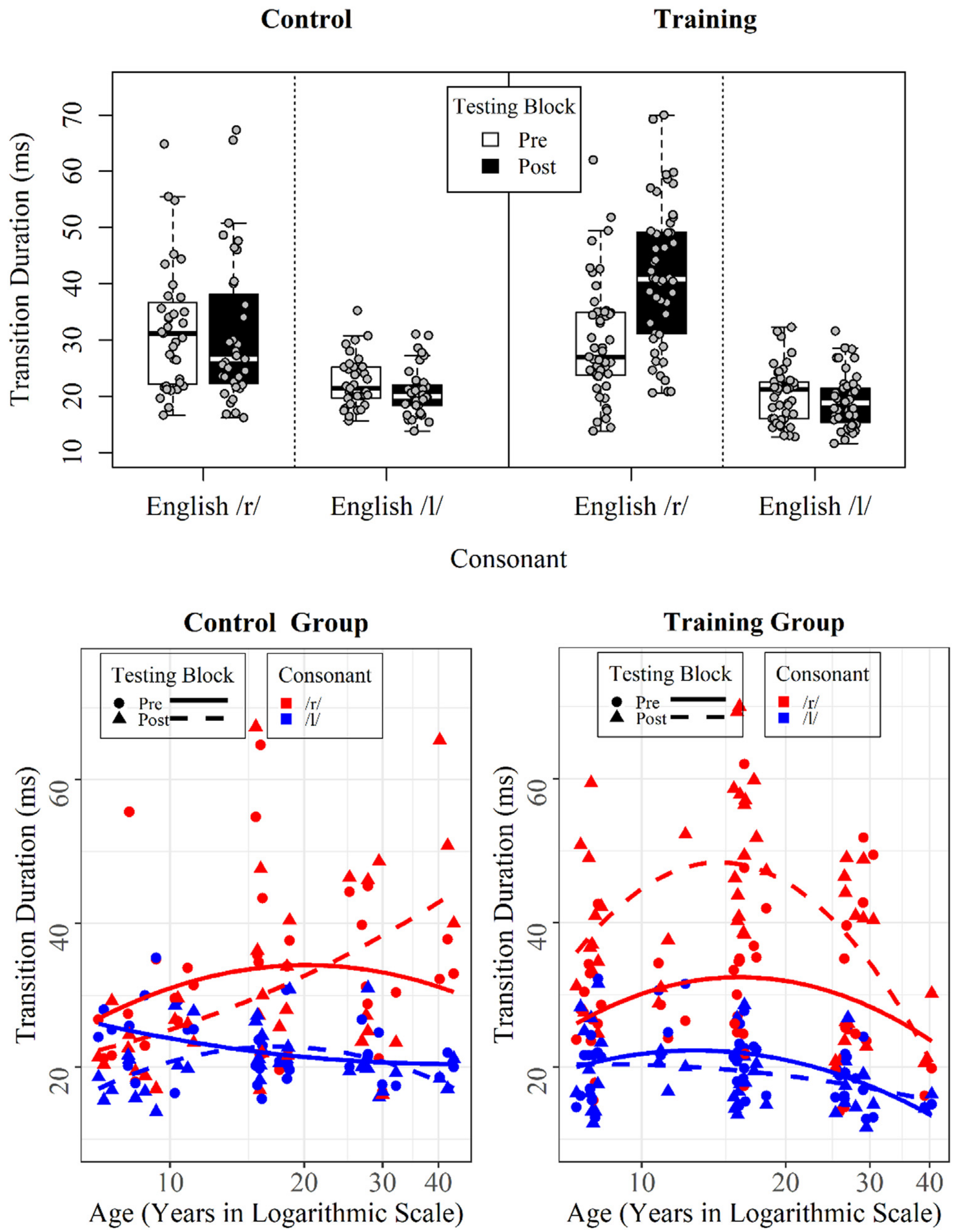

Fig. 12. Boxplots (top) of the transition duration of English /r/ and /l/ produced by Japanese speakers assigned to control (left) and training (right) groups, for pre test (white boxes) and post test (black boxes), and scatterplots (bottom) of the transition duration of the English /r/ (red) and /l/ (blue) along with age for pre test (circle dots) and post test (triangle dots). The lines in the scatterplots represent the loess regression curves at pre test (solid lines) and post test (dashed lines).

Table 12
Statistical analysis for the transition duration of the English /r/-/l/ production.

Fixed factor (contrast)	Estimate	SE	*t*	*p*	
(Intercept)	26.95	0.99	27.25	< .001	***
Group (training vs. control)	−0.40	0.62	−0.65	> .05	
Consonant (R vs. L)	−6.36	0.83	−7.65	< .001	***
Block (pre vs. post)	0.83	0.41	2.02	.068	.
Log age (linear)	4.73	24.94	0.19	> .05	
Log age (quadratic)	−67.54	24.89	−2.71	< .01	**
Group (training vs. control): Consonant (R vs. L)	1.53	0.31	5.02	< .001	***
Group (training vs. control): Block (pre vs. post)	−1.71	0.35	−4.85	< .001	***
Consonant (R vs. L): Block (pre vs. post)	−1.51	0.37	−4.04	< .01	**
Group (training vs. control): Log age (linear)	61.39	24.94	2.46	.016	*

Table 12 (*continued*)

Fixed factor (contrast)	Estimate	SE	*t*	*p*	
Group (training vs. control): Log age (quadratic)	52.17	24.89	2.10	.039	*
Consonant (R vs. L): Log age (linear)	−45.94	12.36	−3.72	< .001	***
Consonant (R vs. L): Log age (quadratic)	31.31	12.33	2.54	.011	*
Block (pre vs. post): Log age (linear)	18.32	14.22	1.29	> .05	
Block (pre vs. post): Log age (quadratic)	−15.35	14.18	−1.08	> .05	
Group (training vs. control): Consonant (R vs. L): Block (pre vs. post)	1.57	0.31	5.12	< .001	***
Group (training vs. control): Consonant (R vs. L): Log age (linear)	−45.11	12.36	−3.65	< .001	***
Group (training vs. control): Consonant (R vs. L): Log age (quadratic)	−41.22	12.33	−3.34	< .001	***
Group (training vs. control): Block (pre vs. post): Log age (linear)	44.87	14.21	3.16	< .01	**
Group (training vs. control): Block (pre vs. post): Log age (quadratic)	13.34	14.18	0.94	> .05	
Consonant (R vs. L): Block (pre vs. post): Log age (linear)	7.22	12.36	0.58	> .05	
Consonant (R vs. L): Block (pre vs. post): Log age (quadratic)	5.08	12.33	0.41	> .05	
Group (training vs. control): Consonant (R vs. L): Block (pre vs. post): Log age (linear)	−22.10	12.36	−1.79	.074	.
Group (training vs. control): Consonant (R vs. L): Block (pre vs. post): Log age (quadratic)	−48.58	12.33	−3.94	< .001	***

References

Aoyama, K., & Flege, J. E. (2011). Effects of L2 experience on perception of English /r/ and /l/ by native Japanese speakers. *Journal of the Phonetic Society of Japan, 15* (3), 5–13. https://doi.org/10.24467/ONSEIKENKYU.15.3_5.

Aoyama, K., Flege, J. E., Akahane-Yamada, R., & Yamada, T. (2019). An acoustic analysis of American English liquids by adults and children: Native English speakers and native Japanese speakers of English. The Journal of the Acoustical Society of America, 146(4), 2671–2681. 10.1121/1.5130574.

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. https://doi.org/10.1016/S0095-4470(03)00036-6.

Arai, T., & Mugitani, R. (2016). The acoustic environment and spoken language development of children. *Journal of Acoustical Society of Japan, 72*(3), 129–136.

Baker, W., Trofimovich, P., Flege, J. E., Mack, M., & Halter, R. (2008). Child-adult differences in second-language phonological learning: The role of cross-language similarity. *Language and Speech, 51*(4), 317–342. https://doi.org/10.1177/0023830908099068.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious Mixed Models. *Journal of Statistical Software, 67*(1) http://arxiv.org/abs/1506.04967.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models Using lme4. *Journal of Statistical Software, 67*(1), 1–51. https://doi.org/10.18637/jss.v067.i01.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B, 57*(1), 289–300.

Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer [Computer program]. Version 5.3.13. http://www.praat.org/.

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. R., Pisoni, D. B., Akahane-Yamada, R., & 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.

Crowther, D., Trofimovich, P., Isaacs, T., & Saito, K. (2015). Does a speaking task affect second language comprehensibility? *The Modern Language Journal, 99*(1), 80–95. https://doi.org/10.1111/modl.12185.

Cruttenden, A. (2014). Gimson's pronunciation of English. *Routledge*. https://doi.org/10.4324/9780203784969.

Dalston, R. M. (1975). Acoustic characteristics of English /w, r, l/ spoken correctly by young children and adults. *The Journal of the Acoustical Society of America, 57*(2), 462–469. https://doi.org/10.1121/1.380469.

Davies, A. (2017). Commentary on the native speaker status in pronunciation research. In T. Isaacs & P. Trofimovich (Eds.), Second language pronunciation assessment: Interdisciplinary perspectives (pp. 185–192). Multilingual Matters. 10.21832/ISAACS6848.

Derwing, T. (2003). What do ESL students say about their accents? *Canadian Modern Language Review, 59*(4), 547–567. https://doi.org/10.3138/cmlr.59.4.547.

Earle, F. S., & Myers, E. B. (2015). Overnight consolidation promotes generalization across talkers in the identification of nonnative speech sounds. *The Journal of the Acoustical Society of America, 137*(1), EL91–EL97. https://doi.org/10.1121/1.4903918.

Endo, M. (1991). Phonological awareness of Japanese young children and learning to read and write yoo-on (small-sized kana character). *Japanese Journal of Educational Psychology, 39*, 448–454.

Espy-Wilson, C. Y. (1992). Acoustic measures for linguistic features distinguishing the semivowels /w j r l/ in American English. *The Journal of the Acoustical Society of America, 92*(2), 736–757. https://doi.org/10.1121/1.403998.

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). York Press.

Flege, J. E. (1999). Age of learning and second language speech. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 101–131). Lawrence Erlbaum Associates.

Flege, J. E. (2002). Interactions between native and second-language phonetic systems. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development: Papers in honor of Henning Wode* (pp. 217–244). Wissenschaftlicher Verlag Trier.

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

Flege, J. E. (2021). New methods for second language (L2) speech research. In R. Wayland (Ed.), *Second language speech learning* (pp. 119–156). Cambridge University Press. https://doi.org/10.1017/9781108886901.004.

Flege, J. E., Aoyama, K., & Bohn, O.-S. (2021). The revised Speech Learning Model (SLM-r) applied. In R. Wayland (Ed.), *Second language speech learning* (pp. 84–118). Cambridge University Press. https://doi.org/10.1017/9781108886901.003.

Flege, J. E., & Bohn, O.-S. (2021). The revised speech learning model (SLM-r). In R. Wayland (Ed.), *Second language speech learning* (pp. 3–83). Cambridge University Press. https://doi.org/10.1017/9781108886901.002.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25*(4), 437–470. https://doi.org/10.1006/jpho.1997.0052.

Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26*(1), 1–34. https://doi.org/10.1017/S0272263104261010.

Flege, J. E., & MacKay, I. R. A. (2011). What accounts for "age" effects on overall degree of foreign accent? In M. Wrembel, M. Kul, & K. Dziubalska-Ko□aczyk (Eds.), *Achievements and perspectives in SLA of speech: New Sounds 2010* (pp. 65–82). Peter Lang.

Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America, 106*(5), 2973–2987. https://doi.org/10.1121/1.428116.

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Effects of age of second-language learning on the production of English consonants. *Speech Communication, 16*(1), 1–26. https://doi.org/10.1016/0167-6393(94)00044-B.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language, 41*(1), 78–104. https://doi.org/10.1006/jmla.1999.2638.

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics, 66*(3), 363–376. https://doi.org/10.3758/BF03194885.

Hattori, K. (2009). Perception and production of English /r/–/l/ by adult Japanese speakers [Doctroral thesis, University College London]. https://discovery.ucl.ac.uk/id/eprint/19204/.

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. 10.1121/1.3021295.

Heeren, W. F. L., & Schouten, M. E. H. (2010). Perceptual development of the Finnish /t–t:/ distinction in Dutch 12-year-old children: A training study. *Journal of Phonetics, 38* (4), 594–603. https://doi.org/10.1016/j.wocn.2010.08.005.

Holt, L. L., Tierney, A. T., Guerra, G., Laffere, A., & Dick, F. (2018). Dimension-selective attention as a possible driver of dynamic, context-dependent re-weighting in speech processing. *Hearing Research, 366*, 50–64. https://doi.org/10.1016/j.heares.2018.06.014.

Idemaru, K., & Holt, L. L. (2013). The developmental trajectory of children's perception and production of English /r/-/l/. *The Journal of the Acoustical Society of America, 133*(6), 4232–4246. https://doi.org/10.1121/1.4802905.

Idemaru, K., & Holt, L. L. (2011). Production and perception of English /l/ and /r/ by 4-, 5-, and 8-year old children. In W. S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 926–929). ICPhS Permanent Council.

Ingvalson, E. M., & Holt, L. L. (2011). Perception-production relationship for /r-l/ by native Japanese speakers. In W. S. Lee & E. Zee (Eds.), The proceedings of the 17th international congress of phonetic sciences (pp. 938–941). ICPhS Permanent Council.

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. https://doi.org/10.1016/j.wocn.2011.03.003.

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. https://doi.org/10.1121/1.2062307.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*(1), B47–B57. https://doi.org/10.1016/S0010-0277(02)00198-1.

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. https://doi.org/10.1121/1.4944755.

Janicka, K., Kul, M., & Weckwerth, J. (2005). Polish students' attitudes to native English accents as models for EFL pronunciation. In K. Dziubalska-Ko□aczyk & J. Przedlacka (Eds.), English pronunciation models: A changing scene (2nd ed., pp. 251–292). Peter Lang.

Jenkins, J. (2007). *English as a Lingua Franca: Attitudes and identity*. Oxford University Press.

Jia, G., Strange, W., Wu, Y., Collado, J., & Guan, Q. (2006). Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *The Journal of the Acoustical Society of America, 119*(2), 1118–1130. https://doi.org/10.1121/1.2151806.

Kartushina, N., & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology, 5*(NOV), 1–17. https://doi.org/10.3389/fpsyg.2014.01246.

Kent, R. D., & Read, C. (2002). *The acoustic analysis of speech* (2nd ed.). Thomson Learning.

Krishnan, A., Gandour, J. T., Bidelman, G. M., & Swaminathan, J. (2009). Experience-dependent neural representation of dynamic pitch in the brainstem. *NeuroReport, 20*(4), 408–413. https://doi.org/10.1097/WNR.0b013e3283263000.

Krishnan, A., Swaminathan, J., & Gandour, J. T. (2009). Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *Journal of Cognitive Neuroscience, 21*(6), 1092–1105. https://doi.org/10.1162/jocn.2009.21077.

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research, 25*(1), 161–168. https://doi.org/10.1016/j.cogbrainres.2005.05.004.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences, 97*(22), 11850–11857. https://doi.org/10.1073/pnas.97.22.11850.

Kuhl, P. K. (2011). Brain mechanisms underlying the critical period for language: Linking theory and practice. In A. M. Battro, S. Dehaene, & W. J. Singer (Eds.), Human neuroplasticity and education (Vol. 117, pp. 33–59). The Pontifical Academy of Sciences.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1493), 979–1000. https://doi.org/10.1098/rstb.2007.2154.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9*(2), F13–F21. https://doi.org/10.1111/j.1467-7687.2006.00468.x.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. Journal of Statistical Software, 82(13). 10.18637/jss.v082.i13.

Lenneberg, E. H. (1967). *Biological foundations of language*. Wiley.

Lindsey, G. (2019). English after RP: Standard British pronunciation today. 10.1007/978-3-030-04357-5.

Lippi-Green, R. (2012). *English with an accent: Language, ideology, and discrimination in the United States*. Routledge.

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. https://doi.org/10.1121/1.408177.

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. https://doi.org/10.1121/1.410149.

Llisterri, J. (1995). Relationships between speech production and speech perception in a second language. In *Proceedings of the 13th International Congress of Phonetic Sciences* (pp. 92–99).

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. https://doi.org/10.1121/1.1894649.

Lotto, A. J., Sato, M., & Diehl, R. L. (2004). Mapping the task for the second language learner: The case of Japanese acquisition of /r/ and /l/. In J. Slifka, S. Manuel, & M. Matthies (Eds.), *From sound to sense: 50 +years of discoveries in speech communication* (pp. C181–C186). MIT Research Laboratory in Electronics.

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. https://doi.org/10.1017/S0142716400009796.

Mann, V. A. (1986). Phonological awareness: The role of reading experience. *Cognition, 24*(1-2), 65–92. https://doi.org/10.1016/0010-0277(86)90005-3.

Matsumoto, H. (2004). Development of phonemic awareness and the beginning process of learning to read in Japanese speaking children. *Research and Clinical Center for Child Development Annual Report, 26*, 67–77.

Matsumoto, H. (2006). Relationship between phonemic awareness and syllabic awareness during the initial stage of learning to read in Japanese children. *Research and Clinical Center for Child Development Annual Report, 28*, 77–87.

Matthews, W. J., & Stewart, N. (2009). The effect of interstimulus interval on sequential effects in absolute identification. *Quarterly Journal of Experimental Psychology, 62*(10), 2014–2029. https://doi.org/10.1080/17470210802649285.

McGowan, R. S., Nittrouer, S., & Manning, C. J. (2004). Development of [ɹ] in young, Midwestern, American children. *The Journal of the Acoustical Society of America, 115*(2), 871–884. https://doi.org/10.1121/1.1642624.

McMahon, A. (2000). Lexical Phonology and the History of English. 10.1017/CBO9780511486432.

Mora, J. C. (2007). Methodological issues in assessing L2 perceptual phonological competence. *Proceedings of Phonetics Teaching & Learning Conference (PTLC) 2007*, 1–5 https://www.phon.ucl.ac.uk/ptlc/proceedings/ptlcpaper_34e.pdf.

Moyer, A. (2007). Do language attitudes determine accent? A study of bilinguals in the USA. *Journal of Multilingual and Multicultural Development, 28*(6), 502–518. https://doi.org/10.2167/jmmd514.0.

Ogawa, A., & Tanemura, J. (2001). Contributions of phoneme operating ability in acquisition of kana reading. *The Japan Journal of Logopedics and Phoniatrics, 42*(3), 220–226.

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research, 5*(3), 261–283. https://doi.org/10.1007/BF01067377.

Piske, T., & MacKay, I. R. A. (1999). Age and L1 use effects on degree of foreign accent in English. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 1433–1436). ICPhS Permanent Council.

Polka, L., & Strange, W. (2005). Perceptual equivalence of acoustic cues that differentiate /r/ and /l/. *The Journal of the Acoustical Society of America, 78*(4), 1187–1197. https://doi.org/10.1121/1.392887.

Qin, Z., & Zhang, C. (2019). The effect of overnight consolidation in the perceptual learning of non-native tonal contrasts. PLOS ONE, 14(12), e0221498. 10.1371/journal.pone.0221498.

Shinohara, Y. (2014). Perceptual training of English /r/ and /l/ for Japanese adults, adolescents and children [Doctoral thesis, University College London]. https://discovery.ucl.ac.uk/id/eprint/1421176/.

Shinohara, Y., & Iverson, P. (2018). High variability identification and discrimination training for Japanese speakers learning English /r/–/l/. *Journal of Phonetics, 66*, 242–251. https://doi.org/10.1016/j.wocn.2017.11.002.

Shinohara, Y., & Iverson, P. (2013). Computer-based English /r/-/l/ perceptual training for Japanese children. Proceedings of meetings on acoustics, 19, article(060049). 10.1121/1.4800136.

Snowling, M., & Hulme, C. (1994). The development of phonological skills. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 346*(1315), 21–27. https://doi.org/10.1098/rstb.1994.0124.

Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2006). Perception of native and non-native affricate-fricative contrasts: Cross-language tests on adults and infants. *The Journal of the Acoustical Society of America, 120*(4), 2285–2294. https://doi.org/10.1121/1.2338290.

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. E. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics, 33*(3), 263–290. https://doi.org/10.1016/j.wocn.2004.10.002.

Vance, T. J. (2008). *The sounds of Japanese*. Cambridge University Press.

Wang, Y., & Kuhl, P. K. (2003). Evaluating the "Critical Period" hypothesis: Perceptual learning of Mandarin tones in American adults and American children at 6, 10 and 14 years of age. In M. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international congress on phonetic sciences* (pp. 1537–1540). ICPhS Permanent Council.

Wells, J. C. (1982). *Accents of English 1: An introduction*. Cambridge University Press.

Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychology, 46*(3), 233–251. https://doi.org/10.1002/dev.20060.

Yamada, R. A. (1995). Age and acquisition of second language speech sounds perception of American English /ɹ/ and /l/ by native speakers of Japanese. In W. Strange (Ed.), *Speech perception and language experience: Issues in cross-language research* (pp. 305–320). York Press.

Yamada, R. A., Strange, W., Magnuson, J. S., Pruitt, J. S., & Clarke, W. D. (1994). The intelligibility of Japanese speakers' production of American English /r/, /i/, and /w/, as evaluated by native speakers of American English. Proceedings of the 3rd international conference of spoken language processing, 2023–2026. https://www.isca-speech.org/archive/icslp_1994/i94_2023.html.

Yamada, R. A., & Tohkura, Y. (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Perception & Psychophysics, 52*(4), 376–392. https://doi.org/10.3758/BF03206698.

Zhang, Y., Kuhl, P. K., Imada, T., Kotani, M., & Tohkura, Y. (2005). Effects of language experience: Neural commitment to language-specific auditory patterns. *NeuroImage, 26*(3), 703–720. https://doi.org/10.1016/j.neuroimage.2005.02.040.

Zhang, Y., & Wang, Y. (2007). Neural plasticity in speech acquisition and learning. *Bilingualism: Language and Cognition, 10*(2), 147–160. https://doi.org/10.1017/S1366728907002908.

Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). Mixed effects models and extensions in ecology with R. In Smart society: A sociological perspective on smart living. Springer. 10.1007/978-0-387-87458-6.