Effects of Identification and Pronunciation Training Methods on L2 Speech Perception and Production: Training Adult Japanese Speakers to Perceive and Produce English /r/-/l/

Zeyu Feng[1]

ABSTRACT

Adult Japanese speakers often experience difficulty learning English /r/-/l/. Previous research has reported the effectiveness of using perception-based high-variability identification training with natural speech stimuli on improving Japanese speakers' perception and production of /r/-/l/. Recent studies have also shown the effectiveness of using production-based multipronged pronunciation training with explicit instruction on articulatory movements and visualized spectrograms showing F3 frequencies of English /r/-/l/. The proposed study will compare the effects of identification training and pronunciation training on Japanese speakers' English /r/-/l/ learning, as well as the generalization of learning gains to novel phonetic environments. Further, the proposed study will contribute to speech perception and production research by exploring the relationship between perceptual learning and production learning.

Keywords: pronunciation training, L2 speech perception, L2 speech production

INTRODUCTION

Adult Japanese speakers often experience difficulty distinguishing and learning English /r/-/l/ (MacKain et al., 1981; Miyawaki et al., 1975; Mochizuki, 1981; Strange & Dittmann, 1984; Takagi & Mann, 1995). Previous research has reported the effectiveness of using perception-based high-variability identification training with natural speech stimuli on improving Japanese speakers' perception and production of /r/-/l/ (Bradlow et al., 1997; Iverson et al., 2005; Lively et al., 1994; Logan et al., 1991; Shinohara & Iverson, 2018). Such training has been shown to facilitate the development of correct perception by helping trainees focus on the most important acoustic cue, the third formant (F3), which signals the /r/-/l/ contrast. In addition to perceptual training, recent studies have shown the effectiveness of using production-based multipronged pronunciation training with explicit instruction on articulatory movements and visualized spectrograms showing F3 frequencies of English /r/-/l/ productions (Akahane-Yamada, 1998; Hattori, 2009). The proposed study will replicate some robust elements in previous research

[1] Zeyu Feng received her MA in Applied Linguistics from Teachers College, Columbia University in May, 2020. She is interested in second language acquisition of phonetics and phonology, speech perception and production. Correspondence should be sent to E-mail: zf2197@tc.columbia.edu

(Bradlow et al., 1997; Hattori, 2009; Lively et al., 1994; Logan et al., 1991; Shinohara & Iverson, 2018) and compare the effects of identification training and pronunciation training on Japanese speakers' English /r/-/l/ learning. Further, the proposed study will contribute to speech perception and production research by exploring the relationship between perceptual learning and production learning. In particular, it will examine whether perceptual training of /r/-/l/ can promote modifications in articulation, as well as whether production training of /r/-/l/ can lead to better accuracy in perceptual identification and discrimination. Lastly, the proposed study will investigate whether perceptual and production learning gains achieved in one phonetic environment (word-initial) can be generalized to other environments (word-initial, word-final, word-initial consonant cluster).

REVIEW OF LITERATURE

English /r/-/l/ vs. Japanese /r/

In English, the liquids /r/ and /l/ in syllable-initial position are produced by raising the tongue toward the alveolar ridge (Raphael et al., 2011). These two sounds are made distinctive by the configuration of the lips and the tongue and position of the tongue tip. For the lateral /l/, the lips are relaxed. The tongue is relatively flat, and the tongue tip slightly touches the alveolar ridge. For the retroflex /r/, the lips are usually rounded. The tongue is grooved, and the tongue tip does not touch the alveolar ridge. These articulatory gestures result in differences in acoustic characteristics of these two sounds, especially the third formant frequency – F3 frequency, which is mainly caused by lip rounding (Delattre & Freeman, 1968). Epsy-Wilson's (1992) measurement of the formant frequencies of /r/ and /l/ in American English averaged across genders showed that the F3 frequencies of the prevocalic /r/ (1779 Hz) and the intervocalic /r/ (1720 Hz) were significantly lower than those of the prevocalic /l/ (2553 Hz) and the intervocalic /l/ (2640 Hz), respectively. No significant difference was found in the first two formants (F1 and F2) between /r/ and /l/.

Yet, the English /r/-/l/ contrast is not a distinctive contrast in Japanese. In the Japanese consonant inventory, there is no phoneme categorized as /l/ (Kubozono, 2015). Although there is a phoneme categorized as /r/, it is phonetically different from English /r/. Depending on the phonetic environment, Japanese /r/ is realized acoustically as either a stop consonant or a flap /ɾ/ (Kubozono, 2015; Lotto et al., 2004; Price, 1981; Vance, 2008; Yamada & Tohkura, 1992). Moreover, from a perception perspective, Price's (1981) research showed that American English-speaking listeners generally labeled Japanese /r/ as similar examples of English flap /t/ or /d/.

L2 Speech Perception and Learning

Difficulties in L2 Speech Perception

Japanese-speaking second language (L2) learners of English have been found to experience difficulty distinguishing English /r/-/l/ since neither of these sounds exists in Japanese. (Best & Strange, 1992; Bradlow et al., 1997; Guion et al., 2000; Hattori & Iverson, 2009; Iverson et al., 2003; Iverson et al., 2005; MacKain et al., 1981; Mochizuki, 1981; Sheldon

& Strange, 1982; Shinohara & Iverson, 2018; Strange & Dittmann, 1984). Based on the Perceptual Assimilation Model (PAM) (Best, 1995), Shinohara (2014) proposed that the difficulty for Japanese listeners to perceive the English /r/-/l/ contrast might depend on how these two sounds are assimilated to the first language (L1) phoneme categories. The PAM posits that when encountering novel speech sounds in a new language, naïve listeners seek similarities and dissimilarities between the novel phonemes and their native phonemes (Best, 1995). The unfamiliar speech sounds are perceptually assimilated into the most similar L1 category(s), and such assimilation patterns predict the relative accuracy for non-native listeners to differentiate novel phoneme contrasts.

Thus, according to PAM, the degree of difficulty in perceiving non-native phoneme contrasts is determined by the perceived similarity between the novel phonemes and the native categories. For example, excellent discrimination accuracy is predicted if two non-native sounds are assimilated to different native phoneme categories (Two-Category). However, poor discrimination is predicted if two phonemes of a non-native contrast are perceived similarly as two tokens of one native phoneme (Single-Category), and such assimilation pattern prevents listeners from noticing the discrepancies in gestural and phonetic details between novel and native phonemes. Further, the PAM posits that even if two non-native sounds are assimilated to the same native category, they might differ in terms of the degree of goodness-of-fit to the ideal native phoneme (Category-Goodness Difference). Very good discrimination is predicted if one non-native sound is perceived as a perfect exemplar of the native phoneme, whereas the other non-native sound is perceived as deviant. Poor discrimination is predicted if two non-native sounds fall out of the familiar phonological space (Uncategorized-Uncategorized), which means these two sounds are not assimilated to any native category. However, very good discrimination is predicted if one non-native sound falls in unfamiliar phonological space but the other non-native sound is assimilated to a native category (Uncategorized-Categorized).

Within the PAM framework, Shinohara (2014) noticed that the assimilation patterns observed in previous studies might explain the challenge Japanese speakers encounter in differentiating English /r/-/l/. For example, previous studies (Best & Strange, 1992; Guion et al., 2000) reported Japanese listeners' poor performance on English /r/-/l/ discrimination, and a similar assimilation pattern was identified – the two non-native sounds (English /r/ and /l/) were categorized into one native category (Japanese /r/). In the study by Best and Strange (1992), the Single-Category assimilation pattern was found: Japanese listeners assimilated English /r/ and /l/ to Japanese /r/ or /w/ categories. Takagi (1993) observed that English /r/ and /l/ occurring in syllable initial position were perceived as Japanese /r/. Komaki et al. (1999) and Guion et al. (2000), however, found that the English /r/-/l/ contrast were perceived as the Uncategorized-Uncategorized type since both phonemes fell "in between specific L1 categories" (Guion et al., 2000, p. 2720): English /r/ and /l/ were perceived as examples between Japanese /ɯɾ/ and /ɾ/ (/ɯ/ - a high back unrounded vowel; /ɾ/ - an alveolar flap). Therefore, such "two-to-one" perceptual assimilation patterns might prevent Japanese listeners from capturing the acoustic and phonetic details of English /r/ and /l/.

In a recent study, however, Hattori (2009) questioned whether the identification accuracy of non-native phonemes is determined by category assimilation patterns between L1 and L2 phoneme categories. Hattori's (2009) study showed that although Japanese speakers showed a stronger tendency to assimilate English /l/ to Japanese /r/ than English /r/, such perceptual assimilation pattern did not predict accuracy in English /r/-/l/ identification. Instead, Hattori (2009) noticed that Japanese-speaking L2 learners' phonetic representations of the F3 values of

English /r/ and /l/ were predictive in determining one’s ability to distinguish these two sounds. Hattori (2009) pointed out that although the PAM posits that the perceived similarity between the non-native contrasts and the native categories determines how well L2 speakers differentiate non-native sounds, such perceptual assimilation processes might not directly lead to difficulties in L2 speech perception and production. Instead, Hattori (2009) proposed that perception in the F3 dimension might explain the difficulty for Japanese speakers to contrast English /r/ and /l/.

Difficulties in L2 Speech Learning

Although Japanese listeners have been found to perpetually assimilate English /r/ and /l/ to Japanese /r/, the degree to which non-native sound is perceptually similar to Japanese /r/ seems to vary. Previous research using goodness-of-fit ratings (listeners rated tokens containing English /r/ and /l/ according to goodness-of-fit to Japanese /r/; the higher the rating, the higher perceptual similarity of the sounds) has found that compared to English /r/, English /l/ is perceptually more similar to Japanese /r/. For example, using natural speech stimuli, Takagi (1993) and Komaki et al. (1999) found that Japanese speakers offered higher ratings to English /l/ than English /r/ in goodness-of-fit to Japanese /r/. Further, using synthesized stimuli, Iverson et al. (2003) also observed higher goodness-of-fit ratings for English /l/ than English /r/ in correspondence to Japanese /r/.

Over the years, the Speech Learning Model (SLM) (Flege, 2005) have been applied to explain the difficulty for Japanese speakers to distinguish English /r/ and /l/, and such difficulty might be attributed to the perceived phonetic (dis)similarity of L2 phonemes (here, English /r/-/l/) and the closest native category (here, Japanese /r/) (Aoyama et al., 2004; Shinohara, 2014). The purpose of the SLM is to account for the variation in L2 learning of phonetic segments – the non-native sounds that can or cannot be perceived and produced accurately (Flege, 2005). The SLM rests on the following premises: 1) L2 learners can accurately perceive the phonetic details of non-native speech sounds given sufficient input; 2) L2 speech production is governed by the perceptual representations of speech sounds developed overtime; 3) the mechanisms responsible for L1 speech acquisition (e.g., the ability to create new phoneme categories) “remain intact and accessible” (Flege, 2005, p. 93) throughout an individual’s lifespan; 4) the native and non-native phoneme categories are placed in a “common phonological space” (Flege, 2005, p. 93) and mutually interact with each other. The SML hypothesizes that a new phoneme category will be likely to be created if there is a great perceived discrepancy between a non-native sound and the closest native sound (Flege, 2005). If a non-native sound, however, is perceived as an exemplar of the closest L1 phoneme, “category assimilation” (Flege et al., 2003, p. 469) will occur (i.e., the L1 and L2 phonemes will be assimilated into one phoneme category) and might prevent the formation of a new L2 category. Therefore, according to the SLM, since English /l/ is perceptually similar to Japanese /r/, English /l/ will be more challenging for Japanese learners to acquire than English /r/, which is perceptually more distinct to Japanese /r/.

The SLM hypothesis has been supported by previous studies on Japanese speakers’ learning of English /r/ and /l/ – Japanese speakers tend to identify English /r/ more successfully than English /l/ (Aoyama et al., 2004; Bradlow et al., 1997; Flege et al., 1996; Mochizuki, 1981; Sheldon & Strange, 1982; Shinohara, 2014). For example, two studies by Mochizuki (1981) and Sheldon and Strange (1982) reported an asymmetry in the error pattern of Japanese speakers’ identification of word-initial English /r/-/l/ – English /r/ was misidentified less frequently than English /l/. Later studies have shown that with adequate language experience, English /r/ might be more acquirable than English /l/. In a study on the interaction between English-language experience and speech perception, Flege et al. (1996) investigated the identification of word-initial English /r/-/l/ by experienced Japanese speakers (who lived in the U.S. for 21 years) and inexperienced Japanese speakers (who lived in the U.S. for 2 years). Results showed that although both groups identified English /l/ less accurately than the native English speakers, the experienced Japanese group outperformed more than the inexperienced group on identifying English /r/. Based on the findings, Flege et al. (1996) proposed that Japanese adults could form a new phoneme category for English /r/ even though such phonetic representation might be different from that of native English speakers. Further, in a longitudinal study, Aoyama et al. (2004) investigated English /r/-/l/ perception and production by native Japanese adults and children to verify whether English /r/ is less challenging to acquire than English /l/ for Japanese-speaking L2 learners of English. The study included two tests with an interval of one year. Results showed that over the one-year period, the Japanese children's /r/-/l/ discrimination improved significantly. Moreover, the Japanese adults and children showed greater performance in producing English /r/ than English /l/, which agreed with the findings by Cochrane (1980), who also observed a more accurate production of English /r/ than English /l/ from Japanese adults and children who were studying English in the U.S.

In addition to rich language exposure, artificial phonetic training has been shown to be more effective in improving production of English /r/ than English /l/ (Bradlow et al., 1997; Hattori, 2009; Shinohara, 2014). In a production-based training study, Hattori (2009) observed a slight improvement in Japanese speakers' English /l/ productions but a substantial improvement in English /r/ productions – the production intelligibility (rated by native English speakers) of /r/ increased more significantly than that of /l/ (/r/: pretest 64.4%, posttest 95.7%; /l/: pretest 92%, posttest 96%). In a perceptual training study, Bradlow et al. (1997) reported an asymmetrical distribution of /r/-/l/ identification accuracy in the pretest: Japanese speakers identified English /r/ more accurately than English /l/. Moreover, Japanese speakers showed significant improvement from the pretest to the posttest on /l/ perceptual identification, indicating that the /l/ sound might be correctly perceived through artificial perceptual training, if not through English communication experience. However, the improvement in /l/ perception failed to translate into production, that is, production evaluation results (production judged by native English speakers) showed that the /l/ tokens produced by Japanese speakers were less accurate than the /r/ tokens. Similarly, Shinohara (2014) also found that the effects of perceptual training were more applicable to enhancing /r/ production than /l/ production. In this study, after the perceptual training, Japanese speakers lowered the F3 of English /r/ for both production tasks, including a word-reading task and a passage-reading task. The subjects, however, only raised the F3 of English /l/ for the word-reading task. Based on the SLM, Shinohara (2014) attributed such difference in /r/-/l/ production to the perceptual similarity of the non-native sounds and the native category. That is, English /r/ is relatively easier to be improved since it is more perceptually dissimilar to Japanese /r/ than English /l/. Therefore, previous findings seemed to support the SLM hypothesis – the learning of L2 phonetic segments is influenced by the degree of perceptual phonetic (dis)similarity between L1 and L2 speech sounds (Flege, 2005).

Phonetic Training on English /r/-/l/ for Japanese Speakers

Perceptual Training

Research has shown that intensive English communication experience might help Japanese speakers distinguish between English /r/ and /l/. For example, using identification and discrimination tasks, MacKain et al. (1981) investigated the categorical perception of synthetic /r/-/l/ stimuli by three groups: American-English speakers and Japanese bilinguals with two English language experience levels, experienced and in-experienced in English conversations with native speakers. The inexperienced Japanese group failed to perceive /r/ and /l/ categorically and showed “chance performance” on all tasks. In contrast, the experienced Japanese group showed similar identification performance as the American-English control group, but their performance on discrimination tasks was relatively less accurate than the American English listeners. Ingvalson et al. (2011) also found a positive relationship between Japanese speakers’ /r/-/l/ perception in natural speech and increased experience in English communication, as measured by the length of residence in America.

However, although /r/-/l/ identification accuracy might improve through gaining English conversational experience, the phonetic processing of F3 might be resistant to be altered. In Ingvalson et al.’s (2011) study, no difference was observed in Japanese speakers’ use of F3 cue in /r/-/l/ perception and production in terms of different length of residence in America, age of arrival in American, years of education in English-speaking contexts, or use of Japanese. Additionally, although longer length of residence and greater F3 reliance predicted more accurate /r/-/l/ perception, length of residence and F3 reliance were not correlated, which indicated that the long-term Japanese residents’ more native-like performance might be due to a change in assimilation patterns or communication strategies rather than modifying F3 cue-weighting or adding new phoneme categories to native categories (Ingvalson et al., 2011; Iverson & Evans, 2009). Therefore, it might be necessary to enhance Japanese listeners’ /r/-/l/ perception using artificial training, especially on the phonetic processing of F3.

To help Japanese listeners’ effectively contrast English /r/-/l/, previous studies have employed two types of perceptual training methods: the discrimination and the identification training methods (Bradlow et al., 1997; Lively et al., 1994; Logan et al., 1991; Shinohara & Iverson, 2018; Strange & Dittmann, 1984). Strange and Dittmann (1984) carried out extensive training (14-18 sessions) using a Same-Different discrimination task with immediate feedback to modify Japanese listeners’ within-category perception of /r/-/l/. The training materials included a synthetic *rock-lock* stimulus series (the values of F2 onset, F3 onset, and F1 steady-state duration were manipulated). For the “Different” trials, the first stimulus (the standard) in each trial was kept constant, and the second stimulus was varied. The standard stimulus was used twice for the “Same” trials. A pretest-posttest design with naturally produced real words contrasting /r/ and /l/ was used to measure the effectiveness of training. Results indicated that although every subject’s discrimination performance improved after training, the learning gains did not translate into the perception of natural speech stimuli. As Strange and Dittmann (1984) concluded: “We cannot conclude that this training experience generalized to perception of the phoneme contrast in real speech by a native AE speaker” (p. 141).

However, the identification training method using highly variable natural speech has been shown to be more effective than the discrimination training method using synthetic speech (Bradlow et al., 1997; Lively et al., 1994; Logan et al., 1991). Using a high-variability identification task but keeping the same test items for the pretest and posttest used by Strange and Dittmann (1984), Logan et al. (1991) observed significant improvement in their participants’ identification accuracy. Logan et al. (1991) identified two reasons why Strange and Dittmann’s (1984) training effect failed to generalize to natural speech stimuli: 1) the Same-Different discrimination task focused on low-level and sensory-based information, which might not help improve /r/-/l/ identification accuracy; 2) the training stimuli lacked variability because the target sounds only occurred in syllable-initial position.

To circumvent these issues in Strange and Dittmann’s (1984) study, Logan et al. (1991) revised the training procedure. First, a two-alternative forced-choice identification task was used to encourage listeners to form phonetic codes and classify stimuli into categories. Second, Logan et al. (1991) claimed that the impoverished acoustic cues in synthesized speech might provide listeners with deficient information signaling the target phoneme categories. Therefore, instead of using synthesized stimuli, real English words were included as the training stimuli. Third, considering the role of speech variability in training perception, the stimuli consisted of 207 /r/-/l/ minimal pairs produced by five native English speakers, providing listeners with various acoustic cues featuring these two phonemes across different phonetic contexts, as well as different speakers. Further, to measure the identification performance changes during the training phase, response time for correct responses was calculated, and results indicated that as response time decreased, identification accuracy increased. Finally, Logan et al. (1991) carried out a generalization test using novel stimuli produced by a new speaker and a speaker whose voice was heard during the training. Such modification of the training procedure helped Japanese listeners focus their attention on the important acoustic attributes of /r/ and /l/ and develop stable perceptual representations to cope with speech variability (Logan et al., 1991).

Lively et al. (1994) identified three limitations of Logan et al.’s (1991) study: 1) the sample size was relatively small since only six subjects were trained; 2) only three subjects participated in the generalization test, which limited the possibility of generalizing the training effect to other subjects or a larger population; 3) the training effect might be influenced by subjects’ English conversational experience since their length of residence in American varied from 6 months to 3 years. As a replication of Logan et al.’s (1991) study, Lively et al. (1994) adopted the identification training with highly variable natural speech with 19 Japanese speakers who had studied English grammar but had no experience receiving instructions on English conversations or living abroad. The effectiveness of such training method was verified: the Japanese listeners shifted their attention to the acoustic characteristics cuing the /r/-/l/ contrast and developed more accurate perception. Additionally, Lively et al. (1994) used delayed posttests to assess retention of newly-developed phoneme categories, and the results showed that the improvement made during training was retained six months after the 3-week training period.

Built on the studies by Logan et al. (1991) and Lively et al. (1994), Bradlow et al. (1997) reinvestigated the effect of using the high-variability identification training method with naturally produced speech stimuli on 11 Japanese listeners’ perceptual learning of English /r/ and /l/. None of the subjects had lived abroad or received any English conversational training. After the training, however, all subjects made significant perceptual learning gains. Moreover, to assess the effectiveness of perceptual learning on making changes in speech production, Bradlow et al. (1997) included both perception and production tasks in the pretest and posttest phases. Japanese speakers’ oral productions in the pretest and posttest were evaluated by native American English speakers. A paired-comparison method was used: each trial started with a target English word in orthography, and the raters listened to two versions of the word produced by one subject, one produced in the pretest, and one produced in the posttest. Results showed that the learning gains achieved during perceptual training could transfer to improvement in speech production: the /r/-/l/ productions during the posttest phase were perceived by native raters as clearer and more intelligible, compared to those during the pretest phase. However, there was a poor correlation between the degree of improvement in perception and production of English /r/ and /l/. Based on the findings, Bradlow et al.'s (1997) proposed that there is "a unified, common mental representation that underlies both speech perception and production" (p. 2308). Further, Bradlow et al. (1999) investigated the long-term retention of perceptual training effects on English /r/-/l/ perception and production, and results showed that Japanese speakers retained improvements in perceptual identification accuracy and overall intelligibility of /r/-/l/ productions three months after completing the perceptual training procedure.

The relationship between perceptual training and speech production was further explored by Shinohara and Iverson (2018). In this study, Shinohara and Iverson (2018) compared the effects of high-variability identification training method (with natural speech stimuli) and discrimination training method (with both natural and synthetic speech stimuli) using a pre-training – during-training – post-training test design. In total, 41 Japanese subjects were recruited, 22 trained in the UK and 19 trained in Japan. The subjects were divided into two groups (which were balanced according to age, sex, training locations, and length of residence/education in English-speaking countries), one received the identification-discrimination sequence and one received the discrimination-identification sequence. Results showed that all subjects showed significant improvement in correctly perceiving and producing English /r/-/l/, which was in line with the earlier finding by Bradlow et al. (1997). However, no significant difference was found between the discrimination and identification training methods in improving accuracy in /r/-/l/ perception and production, and little combined training effect of these two methods was observed. Shinohara and Iverson (2018) proposed that identification and discrimination training can be equally effective, provided that highly variable speech stimuli were adopted in the training procedure.

There are three robust elements of the experiment design by Shinohara and Iverson (2018). First, this study recruited a more diverse subject group, which might increase the generalization of the findings to a larger population or a different context. Second, Shinohara and Iverson (2018) varied the natural speech stimuli used in the training and testing phases. The training stimuli included words with /r/ and /l/ in word-initial position, while the test stimuli contained new words contrasting /r/ and /l/ in word-initial, word-medial, and consonant cluster positions. By doing so, the generalization of perceptual knowledge gained in one consonantal context to unfamiliar contexts could be assessed. Third, to track the changes in the manipulation of acoustic cues during training, Shinohara and Iverson (2018) used Praat software to analyze the acoustic characteristics of the Japanese subjects' /r/-/l/ productions.

Production Training

Given that previous studies using perceptual training (Bradlow et al., 1997; Shinohara & Iverson, 2018) have demonstrated that such training procedure leads to improved performance in perceiving and producing non-native phonetic segments, speech perception and production might share certain common mental representations in processing speech (Bradlow et al., 1997; Shinohara & Iverson, 2018). Based on the research by Bradlow et al. (1997), Hattori (2009) hypothesized that if such common mental representations exist, production-based training might also yield similar results as perception-based training in promoting increased accuracy in /r/-/l/ perception (e.g., improved /r/-/l/ identification accuracy and discrimination sensitivity at category boundaries) and production (e.g., lowered F3 for /r/, raised F3 for /l/). In a production-based training study by Akahane-Yamada et al. (1998), visual-acoustic spectrographic representations of the main acoustic cue to English /r/-/l/ were used to draw Japanese speakers' attention – trainees recorded their productions and compared the spectrograms to those of native English speakers' productions. Results showed that Japanese speakers made significant improvement in /r/-/l/ productions after training, and Akahane-Yamada et al. (1998) argued that such production training with visualized acoustic characteristics provided by spectrograms was effective for improving L2 speech production.

However, Akahane-Yamada et al. (1998) did not examine whether the effects of production training could transfer to perceptual learning. To further verify whether speech perception and production share common representations, Hattori (2009) adapted Akahane-Yamada et al.'s (1998) production training study to investigate whether one-to-one pronunciation training helps Japanese speakers' English /r/-/l/ perception and production. The training procedure involved explicit instructions (by a native Japanese-English bilingual who was phonetically trained to provide precise pronunciation feedback), immediate feedback, and real-time acoustic spectrograms (i.e., such spectrograms are used to monitor the F3 of the /l/-/l/ productions). Results showed that such pronunciation training failed to improve Japanese subjects' English /r/-/l/ perceptual accuracies. Hattori (2009) proposed five possible explanations to the finding: 1) Japanese speakers received limited amount of listening (subjects' original and improved recordings and the instructor's productions), which might not help with the modification of perceptual knowledge of English /r/-/l/; 2) Japanese speakers received low-variability speech stimuli (three minimal-pair words, and two talkers including the subject and the instructor); 3) Japanese speakers received only within-category pronunciation training, which might help improve identification accuracy but not discrimination sensitivity across /r/-/l/ boundaries; 4) if speech perception and production share common representations, there might be only one direction of transfer available - perceptual learning to production learning, but the reverse direction of transfer is not available; lastly, 5) if speech perception and production share independent mental representations, it takes time for L2 learners to establish mental connections between perceptual and production knowledge. Among all possible explanations, Hattori (2009) hypothesized that the fifth one might best explain why Japanese speakers' /r/-/l/ perception did not improve after pronunciation training, as well as the low correlation between perceptual learning and production learning observed in a previous study by Bradlow et al. (1997).

Although little improvement in perceptual behaviors was found, Hattori (2009) observed that pronunciation training led to successful production learning, especially for English /r/. In particular, Hattori (2009) proposed that explicit instructions and feedback on articulatory gestures (e.g., tongue position, tongue shape, lip shape) effectively helped Japanese speakers to lower F3 for English /r/, leading to improved productions. In addition to articulatory movements, with explicit instruction and feedback, Japanese speakers were also capable of noticing and making changes in terms of temporal aspects of their productions: they increased the transition period for /r/ and extended closure period for both /r/ and /l/. Further, in this study, all Japanese speakers demonstrated improvement through the training, especially those who had poor performance in /r/-/l/ perception and production before training, indicating that the Japanese adult learners maintained their ability to learn non-native phonetic segments (Hattori, 2009). Hattori (2009) suggested that such finding might support the SLM premise - the mechanisms responsible for L1 speech acquisition remain intact across one's lifespan.

THE PROPOSED STUDY

Previous research employing perceptual training has shown the effectiveness of high-variability identification training with natural speech stimuli in helping Japanese listeners differentiate English /r/ and /l/ (Bradlow et al., 1997; Iverson et al., 2005; Lively et al.,1994; Logan et al., 1991; Shinohara & Iverson, 2018). In addition, research based on production training has shown the effectiveness of pronunciation training with explicit instruction and visualized acoustic information (Akahane-Yamada et al., 1998; Hattori, 2009). However, the individual and combined effects of perceptual training and production training have not been fully explored since most studies have focused on one training approach. The proposed study attempts to compare the effectiveness of perceptual training and production training on Japanese listeners' perception and production of English /r/ and /l/ by replicating certain robust elements of the experiment design of previous studies. The experimental treatment will include: 1) perceptual training adapted from Shinohara and Iverson's (2018) study; and 2) production training adapted from Hattori's (2009) study. To monitor performance changes, the response time for correct responses will be calculated (Logan et al., 1991). To assess training effects, the proposed study will include three experiment groups: 1) Perceptual-only group receiving only perceptual training; 2) Production-only group receiving only production training; and 3) Perceptual-Production group receiving both perceptual and production training (to counterbalance order effect, half of the Perceptual-Production group will receive perceptual training first, and the other half will receive production training first). The study will include a pre-training – during-training – post-training test method (Lively et al.,1994; Shinohara & Iverson, 2018). To examine the generalization of training effect, the training stimuli will involve minimal pairs contrasting /r/ and /l/ in word-initial position, whereas the test stimuli will involve new minimal pairs contrasting /r/ and /l/ in word-initial, word-medial, and word-initial consonant cluster positions (Shinohara & Iverson, 2018). To assess the retention of training effect, a delayed posttest will be conducted five weeks after the last training session (adapted from Lively et al., 1994). Further, to examine the relationship between perceptual learning and production learning, perception and production tasks will be included in every test (Shinohara & Iverson, 2018). In particular, for production data, the F3 values of /r/ and /l/ in target words will be analyzed acoustically (Praat analysis), as designed in Shinohara and Iverson's (2018) study.

Research Questions

Specifically, the proposed study aims to address three research questions: 1) whether perception-based high-variability identification training improves Japanese listeners' perception and production accuracy of English /r/ and /l/, and how? 2) whether production-based multipronged one-on-one pronunciation training improves Japanese listeners' perception and production accuracy of English /r/ and /l/, and how? 3) whether perceptual and production knowledge gained from exposure to stimuli involving one phonetic environment can generalize to novel phonetic environments? 4) whether English /r/ is easier to learn than English /l/ through phonetic training?

Based on the findings by previous studies (Bradlow et al., 1997; Hattori, 2009; Shinohara & Iverson, 2018), it is hypothesized that perceptual learning can transfer to production learning, but not the reverse direction. Thus, the Perceptual-only group might show improved /r/-/l/ perception and production after training. The Production-only group might show enhanced /r/-/l/

production after training, but their /r/-/l/ perception performance might not change. The Perceptual-Production group might show improvement in both perception and production of English /r/-/l/. Further, it is anticipated that trainees might be able to generalize the perceptual or production knowledge gained from phonetic training to new word contexts (words involving /r/-/l/ in various syllable positions). Lastly, according to the SLM (Flege, 2005), it is hypothesized that the training procedure might help Japanese speakers perceive and produce English /r/ more successfully than English /l/, which is perceptually more similar to Japanese /l/ and difficult for Japanese speakers to establish a new phoneme category.

METHOD

Participants

The subjects will be 48 native Japanese speakers (24 males and 24 females), ranging in age from 18 to 45 years. Thesubjects will be recruited from the Community English Program at Teachers College (TC), Columbia University in the City of New York. To include a diverse subject group, research flayers will be distributed to four proficient levels: beginner, elementary, intermediate, and advanced levels. The subjects will also vary in their length of residence in English-speaking countries, years of education in English-medium institutions, and exposure to English in instructional and social settings. An online questionnaire will be used to collect information about participants' language background (see Appendix A). Before training, all subjects will sign a consent form (see Appendix B) and complete a pretest. Subjects who scored above 75% on the pretest will be eliminated from the experiment, following the inclusion criterion used by Shinohara and Iverson (2018). In total, about 48 subjects will be randomly divided into three groups: 1) the Perceptual-only group (16 subjects); 2) the Production-only group (16 subjects); and 3) the Perceptual-Production group (8 subjects receiving perceptual training first, 8 subjects receiving production training first).

Instruments

Stimuli for Perceptual Training & Perceptual Tests

Real English words will be used as stimuli (used in Shinohara and Iverson 2018). The training stimuli will include 50 minimal pairs contrasting /r/-/l/ in word-initial position (*right-light*) (see Appendix C). The test stimuli will include a new series of 60 minimal pairs contrasting /r/-/l/ in three consonantal contexts: 20 word-initial (*wrist-list*), 20 word-medial (*pirate-pilot*), and 20 word-initial consonant cluster (*graze-glaze*) (see Appendix D). Eight native American English speakers (4 males and 4 females) will be recruited from TC to record the stimuli, six will produce the training stimuli, and two for the test stimuli. Each talker will read every target word (presented in random order) in a semantically neutral carrier sentence five times in clear and natural speech: "*I said _____ this time*" (Zhang et al., 2008). The talkers will be recorded individually using a digital audio recorder (TASCAM, DR44-WL) with a built-in microphone. During recording, the device will be positioned approximately 20 cm from the talkers' lips. Each recording will be conducted with a sampling frequency of 44100 Hz and saved as a WAV file format with a quantization of 24 bits. The target words will be extracted using

Praat (Boersma and Weenink, 2011). For each word produced by each talker, the three most clear tokens will be selected as stimuli. In total, there will be 1,800 tokens for the training stimuli (100 words x 3 tokens x 6 speakers) and 720 tokens for the test stimuli (120 words x 3 tokens x 2 speakers). To ensure that the stimuli are intelligible, ten native English speakers will be recruited to participate in a preliminary listening test. Following the design by Lively et al. (1994), listeners will hear each stimulus and type the corresponding word on laptops/computers. The stimuli with more than 15% error rate across all listeners will be eliminated.

Stimuli for Production Training

For training, six word-initial /r/-/l/ monosyllables (i.e., /rɑ/-/lɑ/, /ri/-/li/, /ru/-/lu/) and six word-initial /r/-/l/ minimal-pair words (i.e., *room-loom*, *read-lead*, *rock-lock*) will be used as training words, which are adapted from Hattori's (2009) study. During each session, participants will practice pronouncing and recording these monosyllables and words. For testing, the production section of each test will include two tasks: 1) reading 40 word-initial /r/-/l/ words (used as the perception testing stimuli) in a carrier sentence: *I said* ____ *this time*; 2) reading part of *The Rainbow Passage* by Fairbanks (1941), as originally used in Hattori's (2009) study (see Appendix E).

Procedure

For each subject, the entire training procedure will be four weeks (see Figure 1 for the procedures for each group). The Perceptual-only group will complete 12 perceptual training sessions within four weeks. The Production-only group will complete 12 production sessions within four weeks. The Perceptual-Production group will complete: 1) six perceptual within two weeks; and 2) six production training sessions within two weeks.

FIGURE 1
Diagram of the Grouping, Training, and Assessment Design

Randomly assigned groups		**Pre-training test**		**Training**	**During-training test**	**Training**	**Post-training test**		**Retention test**
Perceptual-only Group (16 subjects)	Team A (4 subjects)	Test #1		6 Perceptual Training Sessions	Test #2	6 Perceptual Training Sessions	Test #3		Test #4
	Team B (4 subjects)	Test #2			Test #3		Test #4		Test #1
	Team C (4 subjects)	Test #3			Test #4		Test #1		Test #2
	Team D (4 subjects)	Test #4			Test #1		Test #2		Test #3
Production-only Group (16 subjects)	Team A (4 subjects)	Test #1		6 Production Training Sessions	Test #2	6 Production Training Sessions	Test #3		Test #4
	Team B (4 subjects)	Test #2			Test #3		Test #4		Test #1

	Team C (4 subjects)	Test #3			Test #4		Test #1		Test #2
	Team D (4 subjects)	Test #4			Test #1		Test #2		Test #3
Perception-Production Group (16 subjects)	Team A (4 subjects)	Test #1		6 Perceptual Training Sessions (for teams A & B); 6 Production Training Sessions (for teams C & D)	Test #2	6 Perceptual Training Sessions (for teams C & D); 6 Production Training Sessions (for teams A & B)	Test #3		Test #4
	Team B (4 subjects)	Test #2			Test #3		Test #4		Test #1
	Team C (4 subjects)	Test #3			Test #4		Test #1		Test #2
	Team D (4 subjects)	Test #4			Test #1		Test #2		Test #3
		1 week	1 week	2 weeks	4 weeks	2 weeks	1 week	5 weeks	1 week

Perceptual Training

Following the identification training design by Shinohara and Iverson (2018), a two-alternative forced-choice minimal-pair word identification task will be used in training. Each perceptual training session will last about 40 minutes. There will be a different talker for each session, and the talker sequence will be kept consistent for all trainees. Each training session will include 300 two-alternative forced-choice trials (randomly distributed), consisting of 50 /r/-/l/ minimal pairs with three repetitions for each pair. For each trial, the orthography of a minimal pair will be displayed on the screen (one on the left, one on the right), followed by a spoken stimulus including one member of the minimal pair. Subjects need to click the word they heard based on a one-time listening. For incorrect response, zero score is obtained, and a message of “Wrong” will show up on the screen. For correct response, one score is obtained, and a message of “Correct” will show up on the screen. After each trial, the correct answer will be displayed with the spoken stimulus repeated twice. The response time for correct responses will be used for further statistical analyses. Throughout the training, the score will be presented on the screen. After each training session, there will be an identification test of 20 additional trials without feedback. The score will be revealed after the test.

Production Training

Each production training session will last about 40 minutes. A student from the Speech-Language Pathology program will be recruited as the instructor for all subjects. The instructions and feedback used in the proposed study will be adapted from Hattori’s (2009) study. Considering comprehensibility and time constraint, instructions on manipulating temporal correlates (closure and transition durations for /r/-/l/), which were included in Hattori’s (2009) study, will not be included in the proposed study. The training will primarily emphasize on modifying the most important acoustic cue (F3) to English /r/-/l/ of by drawing trainees’ attention on articulatory movements. Specifically, the training will focus on the F3 of the closure duration of /r/-/l/ (see Figures 2 and 3). Based on Hattori’s (2009) study and Epsy-Wilson’s

(1992) measurement, the appropriate F3 range for /r/ during the closure period will be 1200-1800 Hz, and the range for /l/ will be 2500 Hz -3600 Hz. Praat acoustic spectrograms will be used to monitor F3 values of /r/-/l/productions. Given the spectrogram information, the instructor will provide immediate feedback on articulatory movements. For example, if the instructor finds that a high F3 (e.g., over 1800 Hz) for a subject's /r/ production, the instructor will guide the subject to modify the lip shape and tongue shape. If the F3 value appears to be in the appropriate range, the instructor will encourage the subjects to maintain the articulation gestures and keep pronouncing the consonants.

Each training session will include a 15-minute /r/ practice, a 15-minute /l/ practice, and a 10-minute review-test. For the /r/ practice, the subjects will watch a video clip (in natural speed and slow motion) of a native English speaker pronouncing *ray*. The instructor will ask the subjects to pay attention to how the talker slightly protrudes and rounds the lips. To provide articulatory instruction on the shape and position of the tongue, the instructor will show the subjects a side-face animation of /r/ (see Figure 4) and provide modeling – demonstrating the articulation of /r/ by extending the sound and holding articulators steady. Then, the subjects will practice pronouncing monosyllables (i.e., /r/ with a following vowel: /rɑ/, /ri/, /ru/) and target words (i.e., *rock*, *read*, *room*). The subjects will repeat after the instructor at least three times before practicing independently. The instructor will correct mispronounced vowels since the F3 of /r/ will be affected by incorrect vowel pronunciation. To overcome coarticulation difficulties, the subjects will be asked to slowly articulate the syllables with exaggerated pronunciation and observe their articulatory gestures using a hand mirror. The instructor will provide feedback on articulatory movements based on the F3 values showed on the spectrogram.

For the /l/ practice, the subjects will watch a video clip (in natural speed and slow motion) of a native English speaker pronouncing *lay*. The instructor will show the subjects how the talker keeps the tongue body flat, moves the tongue tip toward the alveolar ridge, and releases it after remaining steady for about 100 ms (English /l/ has a longer closure period than Japanese flap /r/). Similar to the /r/ instructional procedure, the instructor will show the subjects a side-face animation of /l/ (see Figure 5) and provide modeling articulation. Then, the subjects will practice pronouncing monosyllables (i.e., /l/ with a following vowel: /lɑ/, /li/, /lu/) and target words (i.e., *lock*, *lead*, *loom*) and receive explicit feedback on articulatory movements. Lastly, for the review test, the subjects will record themselves pronouncing the six monosyllables and the six target words for /r/ and /l/ (i.e., for /r/: /rɑ/, /ri/, /ru/, *rock*, *read*, *room*; for /l/: /lɑ/, /li/, /lu/, *lock*, *lead*, *loom*) five times in a carrier sentence "*I said _____ this time*" (60 tokens). The instructor will provide immediate feedback by checking the F3 production using Praat spectrogram.

FIGURE 2
A Spectrogram of *read* Produced by a Female American English Speaker (The green area indicates the closure period of /r/. The yellow area indicates the transition to the following vowel /ɪ/)

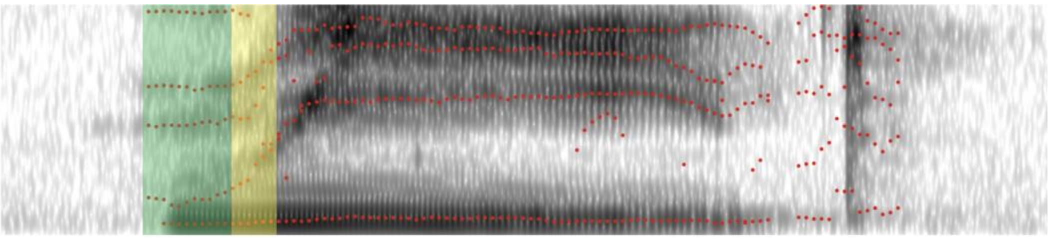


FIGURE 3
A Spectrogram of *lead* Produced by a Female American English Speaker (The green area indicates the closure period of /l/. The yellow area indicates the transition to the following vowel /ɪ/)

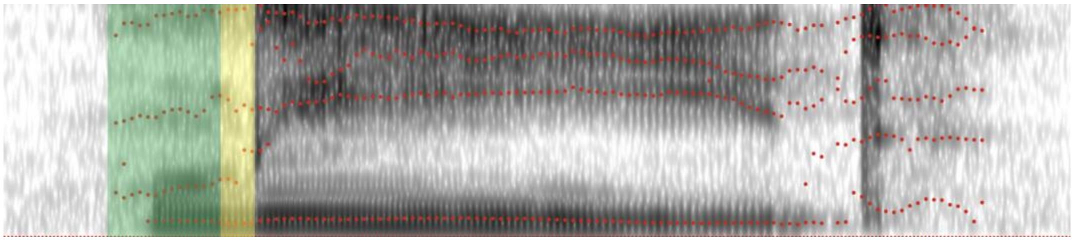


FIGURE 4
A Side-face Animation of /r/ Production (Phonetics: The Sounds of American English, retrieved from https://soundsofspeech.uiowa.edu/main/english)

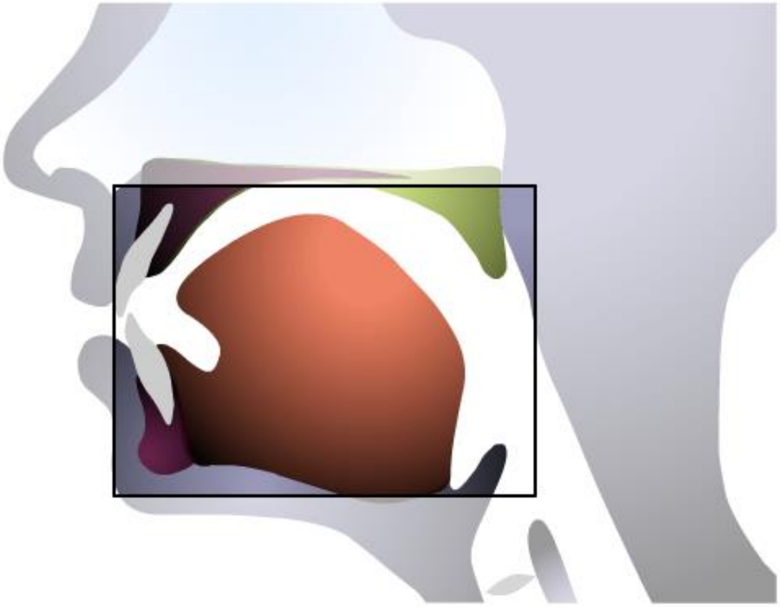
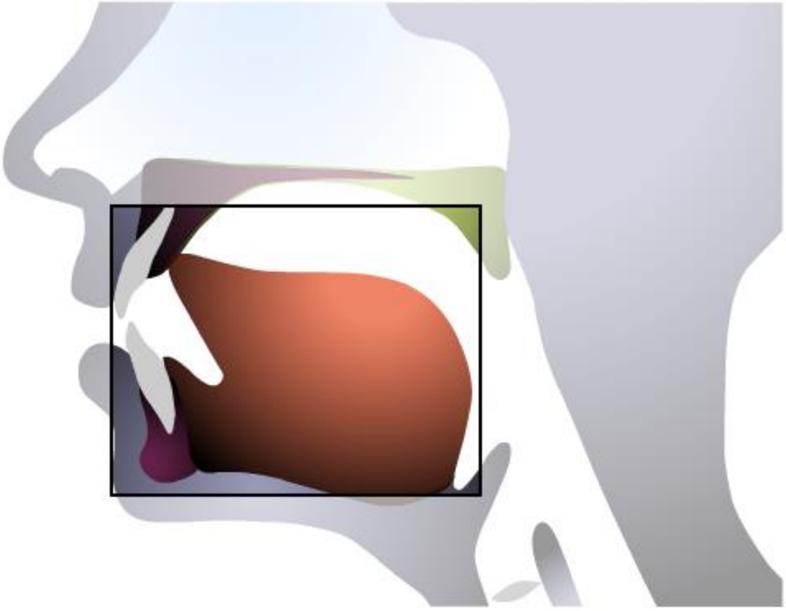


FIGURE 5
A Side-face Animation of /l/ Production (Phonetics: The Sounds of American English, retrieved from https://soundsofspeech.uiowa.edu/main/english)



Assessment

All subjects will complete a pretest, a during-training test, a posttest, and a retention test, which will be conducted one week before training, one week after the sixth training session, one week after the training, and five weeks after the training, respectively. Following the test design by Shinohara and Iverson's (2018), identical test materials will be used for each test, including a perception section and a production section. There will be four test versions, and each one differs only in the sequence of test items. To control for possible discrepancies between tests, all test versions will be administered during each test (see Figure 1). From the pretest, each experimental group will be randomly divided into four teams, and each team will receive one version of the tests. All subjects will have completed all four versions by the end of the retention test.

The perception section will include a two-alternative forced-choice identification task. Subjects will complete two blocks of trials: 1) 20 word-initial /r/-/l/ minimal pairs with two repetitions for each pair; 2) 20 word-medial /r/-/l/ and 20 word-initial consonant cluster /r/-/l/ minimal pairs with two repetitions for each pair. The trials for each block will be randomly distributed. In total, the perception test will include 120 trials. The procedure for the test trials will be the same as that for the training stimuli. No feedback will be provided after each trial, and the score will be revealed after the last trial.

The production section will include two tasks: 1) reading 40 word-initial /r/-/l/ words; 2) reading part of *The Rainbow Passage* by Fairbanks (1941). Subjects' productions in four tests will be analyzed using Praat. The target words will be extracted from the carrier sentence or the reading passage. Following Shinohara and Iverson's (2018) design, for the word-reading task, based on the predicted accuracy of the pronunciation of the following vowel of a word, 10 words which are relatively easy to pronounce will be analyzed. For the passage-reading task, 13 words involving the phonemes /r/ and /l/ will be analyzed. See Appendix F for a complete list of the target words for analysis. The average F3 of the closure period of /r/ and /l/ in the target words will be measured (see Figures 2 and 3).

Planned Data Analysis & Anticipated Results

Tests

To examine whether there is a significant difference between the Perceptual-only, Production-only group, and the Perceptual-Production group before training, an ANOVA will be conducted to compare three groups' performance on the pre-training test. If no significant difference is found, the researcher will have greater confidence to attribute the improvement achieved in the following tests to the training effect.

To examine improvement in perceptual learning, the scores on the perception section of each test will be analyzed. A two-factor Analysis of Variance (ANOVA) will be conducted, with test (pre-training, during-training, post-training, and retention tests) and training method (Perceptual-only, Production-only, and Perceptual-Production) as factors, as used by Bradlow et al. (1997). Post hoc pairwise comparison analysis (Fisher's PLSD) will also be conducted to examine whether three groups are statistically different on each test, as well as whether each group performed statistically differently on different tests. To further examine perceptual learning results, the three experimental groups' /r/-/l/ identification performance on the perception section of the four tests will be analyzed separately, as used by Bradlow et al. (1997). A multi-factor repeated measures ANOVA will be conducted, with test (pre-training, during-training, post-training, and retention tests) as the repeated measure and phoneme (/r/, /l/), phonetic environments (word-initial, word-medial, word-initial consonant cluster), training method (Perceptual-only, Production-only, and Perceptual-Production) as within-group factors. Such analysis will show whether there is a significant relationship between the four factors and the perception performance on each test. In a previous study, Bradlow et al. (1997) found that the main effect of test was significant (because of the improved identification accuracy on the posttest), as well as phoneme (/l/ was identified less accurately than /r/ because /l/ was perceptually more similar to Japanese /r/) and environment (accuracy decreased from initial to media to cluster). Therefore, it is hypothesized that these three factors might influence perception performance. Additionally, based on the findings in Hattori's (2009) study (production training failed to improve /r/-/l/ perception), it is anticipated that the training method might influence perception performance. Thus, compared to the Production-only group, the Perception-only and the Perception-Production groups might perform better on perception tasks after receiving training. However, if the Production-only group shows improved perception, it will indicate the transfer of production training to perception learning.

To examine improvement in production learning, the average F3 of the closure duration of /r/-/l/ in target words produced during the production section of four tests will be measured using Praat. For each phoneme, the F3 values will be analyzed using ANOVA to compare the manipulation of this acoustic cue across test (pre-training, during-training, post-training, and retention tests), training method (Perceptual-only, Production-only, and Perceptual-Production), and phonetic environment (word-initial, word-medial, word-initial consonant cluster). Previous studies have shown that F3 is an important acoustic characteristic that differentiates native English speakers' production of /r/ and /l/ - the F3 frequency for /r/ is lower than that of /l/ (Hattori & Iverson, 2009; Iverson et al., 2003). Such manipulation of F3 has been found to be acquired through perceptual training. Shinohara and Iverson (2018) found that after identification training, the Japanese subjects raised F3 for /l/ and lowered F3 for /r/. Therefore, if the subjects of the proposed study show similar patterns in their productions of F3 in the four tests, it will indicate that they have learned to use F3 to distinguish English /r/ and /l/ in speech production.

Further, it is hypothesized that if a significant difference is found between the Perceptual-only group, the Production-only group, and the Perceptual-Production group on the perception section or the production section of the during-training test, it will indicate that one type of training might be more effective in improving /r/-/l/ perception or production. Additionally, if the Perceptual-Production group shows significant improvement on the posttests, and there is no significant difference between the subjects receiving perceptual training first and those receiving production training first, it will indicate a combined effect of two training approaches (regardless of the sequence of two training methods). Moreover, since the tests include novel stimuli (not involved in the training materials) involving /r/ and /l/ in new consonantal contexts, the improved performance on the posttests will also indicate that the learning gains can be generalized to novel stimuli. Further, as discussed in the literature review section, compared to English /r/, English /l/ is perceived as more similar to Japanese /r/. If subjects show greater improvement on the perception and production of English /r/ than English /l/, it will support the SLM hypothesis that the greater the perceived similarity between the non-native sound and the closest native phonetic category, the more likely a new category of the non-native sound will be created.

Training

To measure performance changes during perceptual training, the scores on six training sessions will be analyzed (for the Perception-only group, the 1st, 3rd, 5th, 7th, 9th, 11th sessions; for the Perception-Production group, all perceptual training sessions). An ANOVA will be conducted to compare the scores across talker (male, female), week of training (weeks 1 to 4), phoneme (/r/, /l/), and training method (Perceptual-only, Perceptual-Production). The classification of male and female talker categories is based on the findings in Logan et al.'s (1991) study, which found "talker-specific influence" – the stimuli produced by make takers were identified less accurately than those produced by female talkers. Therefore, the proposed study attempts to reinvestigate the relationship between Japanese listeners' perceptual learning and training talkers. Additionally, based on the findings in previous studies (Lively et al., 1994; Logan et al.,1991), it is predicted that significant improvement will be observed during the second week of training. Finally, subjects' response times to correct responses during the training phase will be analyzed. Adapted from Logan et al. (1991) design, an ANOVA will be conducted to compare the response times across talker (male, female), week of training (weeks 1 to 4), and phoneme (/r/, /l/). Therefore, changes in response rate during training will be observed.

To measure performance changes during production training, the productions on six training sessions will be analyzed (for the Production-only group, the 1st, 3rd, 5th, 7th, 9th, 11th sessions; for the Perception-Production group, all production training sessions). At the end of each production training session, the subjects will record themselves reading monosyllables and target words containing /r/-/l/ in word-initial position and save as one recording. The recordings by every subject will be used for further acoustic analyses using Praat. Considering practicality, for each recording, two tokens of each target word (12 tokens) will be extracted, and the average F3 of the closure duration of /r/ and /l/ in these words will be measured. The F3 values for /r/ and /l/ will be analyzed separately using ANOVA to compare the manipulation of this cue across week of training (weeks 1 to 4), and phoneme (/r/, /l/), and training method (Production-only, Perceptual-Production).

CONCLUSION

In conclusion, the ultimate goal of the proposed study is to contribute to the development of speech training programs by exploring effective training techniques that can help L2 speakers modify their phonological system and improve their ability to accurately perceive and produce non-native phonemes. To achieve this goal, the proposed study focuses on L2 learning of English /r/-/l/ for adult Japanese speakers. Based on an understanding of the difficulties in L2 speech perception and learning predicted by the Speech Perception Model by Best (1995) and the Speech Learning Model by Flege (2005), the study intends to investigate the relative and combined effects of two training methods on /r/-/l/ learning: 1) perception-based high-variability identification training with natural speech stimuli; and 2) production-based pronunciation training with explicit instruction and acoustic spectrograms. By monitoring Japanese speakers' performance on English /r/-/l/ identification and pronunciation tasks before, during, and after training, the researcher can further examine the relationship between perceptual learning and production learning in L2 speech development. That is, the proposed study can investigate whether the modification of /r/-/l/ perception and perceptual knowledge obtained through perceptual training promotes improvement in production, and whether changes in /r/-/l/ production skills and production knowledge gained through production training leads to better perception. Further, according to the Speech Learning Model (Flege, 2005), the more similar a new L2 phoneme is perceived to an L1 category, the more challenging for L2 speakers to create a new L2 phoneme category. Therefore, it is hypothesized that Japanese speakers might be more likely to experience difficulties learning English /l/ than English /r/ because English /l/ is perceptually more similar to Japanese /r/. In addition to verifying whether such asymmetry in /r/-/l/ learning for Japanese speakers exists, the proposed study will investigate whether Japanese speakers' /l/ perception and production can be enhanced through phonetic training.

Certain elements of the proposed study could be improved. First, because of the nature of phonetic training, subject performance might decrease during training due to fatigue. One way to solve this problem is to control the training time within 40 minutes. For instance, for each perceptual trial, the subject will have a maximum of 10 seconds to click on the word they heard. To keep subjects attentive, it would be helpful to make the training more entertaining by including animated characters, as used in the studies by Iverson et al. (2005) and Shinohara and Iverson (2008). Additionally, it is anticipated that some subjects might not complete the retention test, as happened to previous studies (Lively et al., 1994; Logan et al., 1991; Shinohara & Iverson, 2018). One way to address this attrition issue is to schedule the retention test before the final week of the English language program. Second, there might be an accumulative testing effect on test performance due to the fact that identical materials will be used in all tests. One way to mitigate this problem is to make the test stimuli randomly distributed for each identification trial of the perception section. Third, for the production section, participants might have difficulty pronouncing certain words due to unfamiliarity, and the mispronunciation of the neighboring vowels would affect F3 frequencies of /r/ and /l/. To address this issue, Bradlow et al. (1997) provided the subjects with both visual (orthographic words) and auditory prompts (spoken stimuli). The rationale was to provide modeling of the pronunciation of the entire word. Since Japanese listeners have difficulty distinguishing between /r/ and /l/, it was assumed that the Japanese subjects would not merely rely on the auditory prompt without referring to the visual prompt to determine whether a stimulus contained /r/ or /l/. Therefore, the researcher might consider providing both visual and auditory prompts because the subject groups include various proficiency levels. Additionally, to increase talker-variability for production training, it would be ideal to recruit two to four trainers, and each trainer will conduct training sessions to a certain number of subjects from each experimental group.

For further research, to complement the acoustic measurement of oral productions using artificial software, it would be ideal to include human-rater evaluation by recruiting native English speakers to evaluate the intelligibility and goodness of Japanese speakers' /r/-/l/ productions, as employed in the studies by Bradlow et al. (1997) and Hattori (2009). Further, it would be interesting to investigate the effectiveness of different types of combined training, such as comparing the following two combinations: 1) having a series of perception/production training sessions first, following by a series of production/perception training sessions; 2) having one perception/production training session followed by one proception/perceptual training session and repeating the cycle for multiple times. Additionally, over recent years, acoustic spectrograms have become increasingly popular in the field of Speech-Language Pathology. Current studies have shown the effectiveness of using acoustic spectrograms in providing visual biofeedback intervention for English /r/ articulatory treatment (Byun, 2017; Preston et al., 2018). Therefore, further research could carry out cross-linguistic studies to examine the effect of pronunciation training with explicit instructions and acoustic spectrograms on the learning of other L2 speech sounds at segmental (e.g., consonants, vowels) and suprasegmental (e.g., tones, stress) levels. Lastly, to further improve Japanese speakers' /r/-/l/ production skills in spontaneous speech, it might be necessary to increase the variety of pronunciation training materials. For example, Hattori (2009) proposed that it might be useful to adopt English tongue twisters and conversation tasks involving /r/-/l/. Such tasks should be considered because they consist of challenging continuous speech, which might encourage trainees to achieve the correct configuration of articulators when articulating the target sounds (Hattori, 2009).

REFERENCES

Akahane-Yamada, R., Adachi, T., Kawahara, H., Pruitt, J. S., & McDermott, E. (1998). Toward the optimization of computer-based second language production training. *ETRW on Speech Technology in Language Learning*, pp. 111–114.

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

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

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

Boersma, P., and Weenink, D. (2011). *Praat: Doing Phonetics by Computer.* Retrieved from http://www.fon.hum.uva.nl.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/and /1/: Long-term retention of learning in perception and production. *Perception and Psychophysics, 61*(5), 977–985. doi:10.3758/BF03206911

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. *Journal of the Acoustical Society of America, 101*(4), 2299–2310. doi:10.1121/1.418276

Byun, T. M. (2017). Efficacy of visual-acoustic biofeedback intervention for residual rhotic errors: A single-subject randomization study. *Journal of Speech, Language, and Hearing Research, 60*(5), 1175–1193. doi:10.1044/2016_JSLHR-S-16-0038

Cochrane, R. M. (1980). The acquisition of /r/ and /l/ by Japanese children and adults learning English as a second language. *Journal of Multilingual and Multicultural Development, 1*(4), 331–360. doi: 10.1080/01434632.1980.9994031

Delattre, P., & Freeman, D. C. (1968). A dialect study of American R's by X-ray motion picture. *Linguistics: An Interdisciplinary Journal of the Language Sciences, 44*, 29–68.

Epsy-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 Pt 1), 736–757.

Fairbanks, G. (1941). Voice and articulation drillbook. *The Laryngoscope, 51*(12), 1141. doi:10.1288/00005537-194112000-00007

Feng, Z. (2017). *Acoustic Characteristics of English Word Stress Produced by English and Mandarin speakers: "OBject" or "obJECT"?* (Unpublished Bachelor's thesis).

Flege, J. E. (2005). Origins and development of the speech learning model [Lecture notes]. Retrieved from http://www.jimflege.com/files/SLMvancouver_updated.pdf

Flege, J. E., Schirru, C., & MacKay, I. R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication, 40*(4), 467–491. doi:10.1016/S0167-6393(02)00128-0

Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /l. *The Journal of the Acoustical Society of America, 99*(2), 1161–1173. doi:10.1121/1.414884

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America, 107*(5), 2711–2724. doi:10.1121/1.428657

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

Hattori, K., & Iverson, P. (2009). English /R/-/L/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *Journal of the Acoustical Society of America, 125*(1), 469–479. doi:10.1121/1.3021295

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. doi:10.1016/j.wocn.2011.03.003

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America, 126*(2), 866–877. doi:10.1121/1.3148196

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. *Journal of the Acoustical Society of America, 118*(5), 3267–3278.

doi: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. doi:10.1016/S0010-0277(02)00198-1

Komaki, R., Akahane-Yamada, R., & Choi, Y. (1999). Effects of native language on the perception of American English /r/ and /l/: Cross-language comparison between Korean and Japanese. *The Institute of Electronics, Information and Communication Engineers*, pp. 39–46.

Kubozono, H. (2015). *Handbook of Japanese phonetics and phonology*. Boston: De Gruyter, Inc.

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. *Journal of the Acoustical Society of America, 96*(4), 2076–2087. doi:10.1121/1.410149

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89*(2), 874–886. doi: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 /ɹ/ and /l/. In J. Slifka, S. Manuel, & M. Matthies (Eds.), *From sound to sense: 50+ years of discoveries in speech communication* (pp. C381–C386). Cambridge, MA: Research Laboratory of Electronics at MIT.

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.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). The effect of linguistic experience: The discrimination of (r) and (l) by native speakers of Japanese and English. *Perception & Psychophysics, 18*(5), 331–340. doi:10.3758/BF03211209

Mochizuki, M. (1981). The identification of /r/ and /l/ in natural and synthesized speech. *Journal of Phonetics, 9*(3), 283–303.

Preston, J. L., McAllister, T., Phillips, E., Boyce, S., Tiede, M., Kim, J. S., & Whalen, D. H. (2018). Treatment for residual rhotic errors with high-and low-frequency ultrasound visual feedback: A single-case experimental design. *Journal of Speech, Language, and Hearing Research, 61*(8), 1875–1892. doi:10.1044/2018_JSLHR-S-17-0441

Phonetics: The Sounds of American English. Retrieved from https://soundsofspeech.uiowa.edu/main/english

Price, P. J. (1981). *A cross-linguistic study of flaps in Japanese and in American English*. Unpublished doctoral dissertation. University of Pennsylvania.

Raphael, L. J., Borden, G. J., Harris, K. S. (2011). *Speech science primer: physiology, acoustics, and perception of speech*. New York: Williams & Wilkins.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics, 3*(3), 243–261.

Shinohara, Y. (2014). *Perceptual training of English /r/ and /l/ for Japanese adults, adolescents and children.* (Doctoral dissertation, University College London). Retrieved from https://discovery.ucl.ac.uk/id/eprint/1421176/3/final_submission_thesis_yasuaki_shinohara.pdf

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

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics, 36*(2), 131–145. doi:10.3758/BF03202673

Student Questionnaire on Language Background. (n.d.). Retrieved from https://dept.writing.wisc.edu/wac/student-questionnaire-on-language-background/

Takagi, N. (1993). *Perception of American English /r/ and /l/ by adult Japanese learners of English: A unified view*. (Doctoral dissertation, University of California).

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

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

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. doi:10.3758/BF03206698

Zhang, Y., Nissen, S. L., & Francis, A. L. (2008). Acoustic characteristics of English lexical stress produced by native Mandarin speakers. *The Journal of the Acoustical Society of America, 123*(6), 4498–4513.

APPENDIX A

Language Background Questionnaire
(Adapted from Student Questionnaire, University of Wisconsin Madison)

1. Gender: □ Male □ Female □ Other
2. Year of Birth: ____________
3. Difficulty with hearing or speech: □ Yes □ No
4. Have you studied in English-speaking countries before coming to Teachers College?
   □ No □ Yes
5. If your answer to question #5 is “yes”, where did you study? How long?
   ____
6. First Language(s): ____________
7. How long have you been in the United States? ____________
8. Have you lived in other English-speaking countries? If so, how long?

Country #1	Length	Country #2	Length	Country #3	Length

9. How long have you been learning English? ____________
10. How have you learned English? (check all that apply)

□ Through formal classroom instruction

□ Through interacting with people

□ Online from chatting, messaging, or emailing

□ From TV, music, or movies

□ Other, specify: ____________

11. Have you taken (or are you talking) the following courses offered at Teachers College?
□ Intermediate Conversation □Advanced Conversation
□ Intermediate Pronunciation □Advanced Pronunciation

12. Have you taken other courses in pronunciation and conversation?

13. On a scale of 1-10, please select your level of proficiency in speaking, pronunciation, reading and listening in English

English speaking	(a scale bar will be displayed)
English pronunciation	(a scale bar will be displayed)
English reading	(a scale bar will be displayed)
English listening	(a scale bar will be displayed)

14. In a typical day, which languages do you use and at what percent?
Language: ____________ □ <25% □ 25%-50% □ 50%-75% □ >75%
Language: ____________ □ <25% □ 25%-50% □ 50%-75% □ >75%
Language: ____________ □ <25% □ 25%-50% □ 50%-75% □ >75%

15. Which languages do you use in the following activities?
Listening to radio, watching TV or movies: ___________________________
Reading for work/school: ___________________________
Reading on the Internet: ___________________________
Writing emails to or chatting online: ___________________________
Writing papers or assignments: ___________________________

16. Is there anything else that you feel is interesting or important about your language background that you'd like me to know? _________________________________

APPENDIX B

Consent Form
(Adapted from Feng, 2017)

Consent to Participate in Research

Project Name: English Word Stress Production by English and Mandarin speakers
Investigator: Zeyu Feng **E-mail**: zf2197@tc.columbia.edu Telephone: (1) 917-513-6026

Introduction

You are invited to consider participating in this research study. We will investigate the effects of perceptual training and production training on Japanese speakers' perception and production of English /r/ and /l/. This form describes the purpose and nature of the study and your rights as a participant in the study. The decision to participate or not is yours. If you decide to participate, please sign and date the last line of this form.

Explanation of the study

Research has shown that native Japanese speakers often have difficulty distinguishing the English /r/-/l/ contrast because both sounds are not available in the Japanese sound system. As part of the study, all participants will be invited to six perceptual training sessions and six production training sessions. Each training session will last 30-40 minutes, and the entire training procedure will last four weeks. The researcher will examine whether and how these two types of training approaches help improve Japanese speakers' performance on perceiving and producing the English /r/-/l/ contrast. All participants will complete four tests including listening and reading tasks. The first test will be carried out one week before training. The second and third tests will be carried out during the training period. The fourth test will be carried out five weeks after the completion of the entire training.

Confidentiality
All of the information collected will be confidential and will only be used for research purposes. This means that your identity will be anonymous; in other words, no one besides the researcher will know your name. Whenever data from this study are published, your name will not be used. The data will be stored on a computer, and only the researcher will have access to it.

Your participation
Participating in this study is strictly voluntary. If at any point you change your mind and no longer want to participate, you can tell the investigator. You will not be paid for participating in this study. If you have any questions about the research, you can contact Zeyu Feng by telephone at (1) 917-513-6026, or by e-mail zf2197@tc.columbia.edu.

Investigator's statement
I have fully explained this study to the participants. I have discussed all the issues and answered all the questions that the participants asked.
Signature of investigator ________________________________ Date _____________

Participant's consent
I have read the information provided in this Informed Consent Form. All my questions were answered to my satisfaction. I voluntarily agree to participate in this study.
Your signature ________________________________ Date _____________

APPENDIX C

Training Materials – 50 /r/-/l/ Minimal Pairs (used by Shinohara, 2014; Shinohara & Iverson, 2018. Originally created by Iverson et al., 2005)

Word-initial					
/r/	/l/	/r/	/l/	/r/	/l/
rack	lack	rear	leer	roves	loaves
rad	lad	rent	lent	rob	lob
rag	lag	rice	lice	robe	lobe
raid	laid	rick	lick	rock	lock
rake	lake	rid	lid	wrong	long
ram	lamb	rise	lies	rook	look

rain	lane	rife	life	room	loom
rank	lank	rift	lift	ross	loss
rate	late	right	light	rot	lot
raft	laughed	rim	limb	rowed	loud
roars	laws	rhyme	lime	rout	lout
ray	lay	rhine	line	row	low
raise	laze	rind	lined	rose	lows
reach	leach	rink	link	rump	lump
reef	leaf	rip	lip	rush	lush
reek	leak	writ	lit	rust	lust
red	led	roan	loan		

APPENDIX D

Test Materials – 60 /r/-/l/ Minimal Pairs
(used by Shinohara, 2014; Shinohara & Iverson, 2018)

All 60 /r/-/l/ minimal pairs will be used in the perception section; 20 word-initial r/-/l/ minimal pairs used in the “word-reading task” in the production section.

Word-initial		Word-medial		Word-initial consonant cluster	
/r/	/l/	/r/	/l/	/r/	/l/
race	lace	arouse	allows	brand	bland
ramp	lamp	arrive	alive	broom	bloom
raps	lapse	bereave	believe	brunt	blunt
raw	law	berries	bellies	brush	blush
reap	leap	boring	bawling	cramp	clamp
rest	lest	coring	calling	crime	climb
ride	lied	correct	collect	crowd	cloud
road	load	erect	elect	frame	flame
roam	loam	fairy	fairly	fresh	flesh
roared	lord	farrow	fallow	froze	flows
root	loot	horror	holler	fruit	flute
rope	lope	marrow	mallow	grass	glass
rude	lewd	mirror	miller	graze	glaze
rug	lug	parrot	palate	grew	glue
rune	loon	pirate	pilot	grow	glow
rung	lung	poring	palling	praise	plays
ruse	lose	starring	starling	prank	plank
wrap	lap	tarry	tally	prod	plod
wrens	lens	terror	teller	spray	splay
wrist	list	whirring	whirling	sprint	splint

APPENDIX E

Test Materials – The Rainbow Passage (The passage was cited from Fairbanks, 1941. It was originally used by Hattori, 2009. It was also used by Shinohara, 2014 and Shinohara & Iverson, 2018.)

The Rainbow Passage

When the sunlight strikes raindrops in the air, they act as a prism and form a rainbow. The rainbow is a division of white light into many beautiful colors. These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon. There is, according to legend, a boiling pot of gold at one end. People look, but no one ever finds it. When a man looks for something beyond his reach, his friends say he is looking for the pot of gold at the end of the rainbow. Throughout the centuries, people have explained the rainbow in various ways. Some have accepted it as a miracle without physical explanation.

APPENDIX F

Target Words in Production Test Tasks for Acoustic Analysis

Task #1: Word-reading

/r/	/l/
race	lace
road	load
root	loot
rung	lung
wrist	list

Task #2: Passage-reading

/r/	/l/
raindrops	legend
reach	light
round	long
Rainbow * 4 (occurred 4 times)	look
	looks
	looking