DOES ULTRASOUND TRAINING LEAD TO IMPROVED PERCEPTION OF A NON-NATIVE SOUND CONTRAST?: EVIDENCE FROM JAPANESE LEARNERS OF ENGLISH[*]

Miwako Tateishi and Stephen Winters
University of Calgary

1. Introduction

Both the perception and production of speech sounds in a non-native language can be challenging to adult second language learners due to long-time experience with their native language. Numerous studies have investigated how the ability to perceive and produce speech sounds can be modified during adulthood. This study explores whether production training using ultrasound as visual feedback can lead to improved perception and production of a non-native speech contrast in the absence of perceptual training. To this end, Japanese learners of English who were beginning ESL students in Canada were trained to produce English /r/ and /l/.

It has been well documented that native Japanese speakers are likely to have difficulty in discriminating between English /r/ and /l/ (e.g., Goto 1971, Miyawaki et al. 1975) due to the perceived similarity between these phonemes and the Japanese liquid /r/ (e.g., Best and Strange 1992). This claim is often made despite the fact that the Japanese /r/ is phonetically an apico-alveolar tap that is distinct from the English liquids (Vance 2008). The primary acoustic cue that differentiates English /r/ and /l/ is the third formant (F3), which is lower for /r/ and higher for /l/ (e.g., O'Conner et al. 1957). Those phonemes also differ in the second formant (F2), which is slightly lower for /r/ and slightly higher for /l/; however, this does not appear to be a reliable cue in discriminating the phonemes for native English speakers (O'Conner et al. 1957). Compared to native English speakers, native Japanese speakers are less sensitive to the F3 difference (Miyawaki et al. 1975, Best and Strange 1992, Iverson et al. 2003) but are more sensitive to the F2 difference, which may be crucial in identifying the Japanese tap (Iverson et al. 2003).

English /r/ and /l/ can also be difficult for native Japanese speakers to produce (e.g., Goto 1971, Sheldon and Strange 1982). This could be due in part to their unfamiliarity with the accurate configurations of the articulatory gestures required for these sounds (Bradlow 2008). Lotto et al. (2004) acoustically analyzed Japanese speakers' productions of /r/ and /l/ by plotting multiple

[*] We are grateful to the Linguistics Department at the University of Calgary for their supports for this research. We are also grateful to Dr. Sonya Bird and the members of the Speech Research Laboratory at the University of Victoria for their help with the production training experiment. And many thanks to the students at the University of Calgary who helped with recordings, pilot testing, or data analyses, and to the research participants in Victoria.
Actes du congrès annuel de l'Association canadienne de linguistique, 2013.
Proceedings of the 2013 annual conference of the Canadian Linguistic Association.

tokens of these phonemes in terms of F2 and F3. The analysis revealed that the tokens were not clearly separated on the F3 continuum, while they were more distinct on the F2 continuum. Therefore, Japanese speakers' productions of these phonemes are likely to confuse native English listeners due to a lack of separation in the primary acoustic cue used to distinguish these phonemes.

Japanese speakers have been shown to improve their perceptual ability to discriminate between /r/ and /l/ after undergoing a short-term, intensive laboratory perceptual training protocol called High Variability Phonetic Training (HVPT) (e.g., Logan et al. 1991). In this protocol, Japanese learners listen to multiple instances of /r/ and /l/ in a variety of phonetic contexts in natural speech, as produced by multiple native English speakers. After training, the Japanese learners correctly identified the phonemes in speech significantly more often than before training. Subsequently, Bradlow et al. (1997) demonstrated that HVPT improved both the perception and production of /r/ and /l/, even though production is not targeted in training. If perceptual training alone can improve both perception and production of the same phonemes, it seems possible that production training alone can improve the production and perception of the same phonemes, even in the absence of perceptual training.

Hattori (2009) examined whether training Japanese learners of English to produce /r/ and /l/ with the help of acoustic spectrograms could lead to improved production and perception of these phonemes. In training, the instructor monitored real-time spectrograms of the learner's speech and provided feedback on their production. After the learners made recordings of training targets, they saw spectrograms of their recorded productions and received feedback in terms of visible features of the spectrograms. Analyses of the perception and production of /r/ and /l/ by the learners before and after training revealed that the learners' productions significantly improved after training; however, the training did not improve perception of the same phonemes.

In recent years, ultrasound technology has been used in production training. Ultrasound allows learners to see the appropriate configuration of articulatory gestures for a speech sound by providing direct, dynamic images of tongue movements in both front-to-back and side-to-side views of the vocal tract. Ultrasound technology has been incorporated in research that has investigated native Japanese speakers' ability to learn how to produce English /r/ and /l/. Gick et al. (2008) and Tsui (2012) demonstrated that Japanese learners of English with varying degrees of experience with English can improve their production of /r/ and /l/ with ultrasound. However, Gick et al. (2008) did not examine whether the improved production led to improved perception of the same phonemes in the Japanese learners. Additionally, Tsui (2012) conducted an exploratory investigation of the learners' perceptual ability, but only used inconsistent numbers of perceptual tasks throughout the experiment.

The goals of this study were thus to explore: 1) whether production training using ultrasound imaging as visual feedback leads to improved production of /r/ and /l/ by Japanese learners of English in terms of F2 and F3; 2) whether training improves the intelligibility of the Japanese learners' productions of the phonemes; and 3) whether training improves perception of

the same phonemes by the Japanese learners in the absence of perceptual training. It was predicted that ultrasound training would lead to improved production quality. That is, the training should lead to changes in the F2 and F3 frequencies of Japanese learners' productions of /r/ and /l/ that would make them more closely approximate the F2 and F3 in native English speakers' productions of the same phonemes (Tsui 2012). Second, it was predicted that the training would improve the intelligibility of the Japanese learners' productions of /r/ and /l/ for native English listeners (Gick et al. 2008, Tsui 2012). Finally, it was predicted that utilizing visualization of tongue shape and movements as feedback in production training would facilitate the perception of /r/ and /l/ (Adler-Bock et al. 2007, Tsui 2012).

2. General Experiment Design

The production training experiment comprised three stages: 1) a pre-training perception test and production recordings; 2) production training; and 3) a post-training perception test and production recordings. The training comprised five separate sessions. Each session lasted approximately 30 minutes, and only one session took place per day. The entire experiment took place over a three-week period. The pre-training perception test and recordings were conducted on the day before the first production training session, and the post-training perception test and recordings were completed immediately after the fifth training session. The experiment was conducted in the Speech Research Laboratory at the University of Victoria. Recordings of auditory stimuli for the pre-/ post-training perception tests and auditory prompts for the production recordings were made in the Phonetics Laboratory at the University of Calgary.

3. Production Training

3.1 Method

3.1.1 Participants

Participants were 10 native Japanese speakers (four male and six female) ranging from 18 to 30 years of age (mean age: 24.6 years). All of the participants were attending ESL programs offered at schools in Victoria. All the participants and their parents were born and raised in Japan. Participants had been living in Canada no more than four months (except one who had been living in Victoria for nine months), and none had lived in any other English-speaking countries before coming to Canada. None spoke a language other than Japanese and English fluently, and none reported speech or hearing impairments.

3.1.2 Apparatus

For the production training, a LOGIQe portable ultrasound machine (GE Healthcare) was used. Ultrasound images of soft tissue are obtained through the echo patterns of ultra high-frequency sound waves emitted by and reflected back to piezoelectric crystals contained under the upper surface of a transducer (Gick

2002). In order to image tongue shapes and movements, the transducer is placed against soft tissue under the chin; by rotating it 90 degrees, both front-to-back and side-to-side views of the tongue can be captured (Gick 2002). During the training, the transducer was hand-held by the learners themselves.

3.1.3 Training Targets

Isolated /r/ and /l/, six consonant-vowel (CV) syllables (/ri/, /li/, /ru/, /lu/, /ræ/ and /læ/), and six monosyllabic minimal-pair words contrasting /r/ and /l/ word-initially (*reek, leak, room, loom, rack,* and *lack*) were selected as a total of 14 targets for the production training. A native English (NE) speaker (male) recorded ultrasound images of his own production of the targets by using video recording and editing software (Sony Vegas Pro) at the Speech Research Laboratory at the University of Victoria. These ultrasound images were provided to learners during training as a model of tongue shapes and movements to use in the production of the target sounds and words. The NE speaker produced each target six times, and these six utterances were recorded as a single video clip. He recorded the tongue movements for the first three utterances in a front-to-back view and for the next three utterances in a side-to-side view. Audio signals of his production were simultaneously recorded with the video clip. During the training, the ultrasound machine and a lap-top computer displaying the recorded ultrasound images of the NE speaker's productions were placed side by side.

3.1.4 Procedure

Native Japanese (NJ) learners underwent the production training individually. The training progressed from 1) production of isolated /r/ and /l/ to 2) production of the CV syllables, and ultimately to 3) production of the monosyllabic words.

The first training session began with instructions on correct articulatory movements for /r/ and /l/, and each of the subsequent training sessions began with a review of what the learners had learned in the previous session. In each training session, ultrasound images of learners' productions and the corresponding audio signals were selectively recorded in order for the experimenter (the first author), who is a phonetically trained, English-Japanese bilingual, to evaluate progress and identify difficulties for each learner to work on in successive training sessions. Following the approach used by Gick et al. (2008), the recorded images were also shown to the learners themselves for discussions with the experimenter in order to promote intellectual involvement in the training process and self-awareness of their own articulations. During these discussions, the learner was asked to describe similarities and differences between his or her productions and the NE speaker's productions by referring to general tongue shapes, shapes of specific parts of the tongue, and movements of various tongue parts.

At the beginning of the first training session, the experimenter described the articulatory gestures used in the production of /r/ and /l/. After these initial instructions, learners sat in front of the ultrasound machine and were instructed on how to hold and place the transducer for front-to-back and side-to-side views of the tongue. The learners were then presented with ultrasound images of a model production of the target. Finally, they practiced producing the target while looking at real-time images of their own production displayed on the ultrasound machine. They were allowed to look at the images of the model production again if they so desired.

3.1.5 Production Recordings

The NJ learners made production recordings individually in the sound attenuated booth in the Speech Research Laboratory at the University of Victoria. They were asked to articulate the prompt words after they were presented with the visual and auditory prompts. In each recording, they saw an orthographic representation of the word to be produced on the computer screen and heard the word through speakers while looking at the screen. They were allowed to listen to the word twice if necessary. Prompts and procedures were identical for the pre-test and the post-test.

3.1.6 Prompts

Twenty minimal-pair monosyllabic words contrasting /r/ and /l/ word-initially (e.g., *right* and *light*) and 20 non-minimal-pair monosyllabic words containing /r/ and /l/ word-initially were selected as a total of 40 prompts to be presented visually and auditorily for the production recordings. The minimal-pair words included the six monosyllabic words used as targets in the production training. A different male NE speaker recorded the auditory prompts in a sound-attenuated booth in the Phonetics Laboratory at the University of Calgary.

3.2 Acoustic Analysis

In order to assess changes in the Japanese learners' productions, acoustic measurements were made for the initial segments in each recorded production. Of a total of 800 utterances from the pre- and post-training recordings (40 prompts × 10 learners × 2 recording conditions), 46 utterances (23 utterances from the pre-training recordings and 23 utterances from the post-training recordings) were excluded from the analysis because the onset consonants were either missing or pronounced as stop consonants, in which formant frequencies were absent. For each of the remaining 754 utterances, the formant frequency values of the initial segments were measured by taking the average F2 and F3 values for the steady state of the segment using Praat (Boersma and Weenink 2009). For normative data, speech samples were collected from five NE speakers (two male and three female), who were students at the University of Calgary. They produced each of the 40 prompt words used for the pre- and post-

training recordings. The recordings were made in the Phonetics Laboratory at the University of Calgary. The F2 and F3 values of the initial segments for each of these 200 utterances (40 words × 5 speakers) were measured using Praat (Boersma and Weenink 2009).

In order to make equitable comparisons across formant frequencies from different speakers, all formant frequency measurements were normalized for each speaker, using the z-score transformation method proposed by Lobanov (1971).

3.3 Results

Figure 1 displays the average F2 measurements produced by both NJ learners and NE speakers. For the NJ learners' productions, there was a large decline in the mean F2 for /r/ from -0.24 (SD = 0.49) at pre-test to -0.52 (SD = 0.32) at post-test. Similarly, the mean F2 for /l/ largely declined from 0.59 (SD = 0.59) at pre-test to 0.23 (SD = 0.42) at post-test. However, for /r/, the mean F2 frequency for the NE speakers' productions (M = -0.05, SD = 0.28) was higher than the mean F2 frequencies for the NJ learners' productions at both pre-test and post-test. On the other hand, for /l/, the mean F2 frequency for the NE speakers' productions (M = 0.05, SD = 0.28) was lower than the mean F2 frequencies for the NJ learners' productions at both pre-test and post-test.

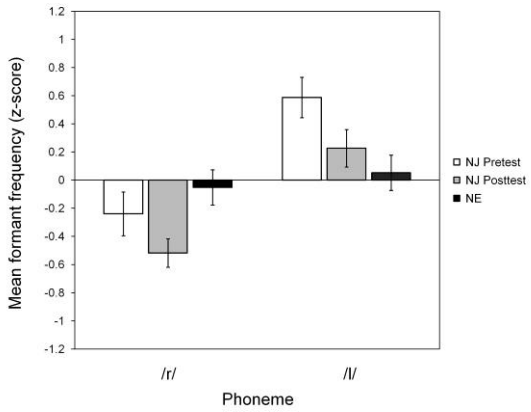

Figure 1. F2 frequencies for the NJ learners' productions at pre-test and at post-test and the NE speakers' productions. Error bars represent standard errors.

A two-way repeated ANOVA with phoneme (/r/, /l/) and testing session (pre-test, post-test) as within-subject factors was performed in order to examine the observed differences in F2 between the testing sessions for the NJ learners' productions. The main effect of phoneme was significant, $F(1, 9) = 21.79$, $p = .001$. However, there was no significant main effect of testing session, $F(1, 9) = 3.17$, $p = .109$, nor any interaction between phoneme and testing session, $F(1, 9) = 0.14$, $p = .715$.

In order to examine whether the F2 frequencies for the NJ learners' productions of /r/ and /l/ at pre-test and post-test significantly differed from the F2 frequencies for the NE speakers' productions of the same phonemes, Mann-Whitney tests were performed across language groups (pre-test NJ vs. NE, post-test NJ vs. NE) for each phoneme. For /r/, the F2 for the NJ groups' productions at pre-test was not significantly lower than the F2 for the NE group's productions, U = 22.00, z = -0.37, p = .768. However, the F2 for the NJ group's productions for /r/ at post-test was marginally lower than the F2 for the NE group's productions, U = 9.00, z = -0.961, p = .052. For /l/, the F2 for the NJ group's productions at pre-test was significantly higher than the F2 for the NE group's productions, U = 8.00, z = -2.08, p = .04. However, the F2 for the NJ group's productions at post-test was not significantly higher than the F2 for the NE group's productions, U = 20.00, z = -0.61, p = .594. Thus, the analysis suggests that F2 became lower overall after training. F2 for the NJ groups' productions became lower than F2 for the NE group's productions after the training for /r/, whereas F2 for the NJ group's productions became similar to F2 for the NE group's productions after the training for /l/.

Group F3 measurements from both pre-test and post-test are displayed in Figure 2. For the NJ group's productions, the mean F3 for /r/ declined from -0.65 (SD = 0.50) at pre-test to -0.74 (SD = 0.46) at post-test, whereas the mean F3 for /l/ showed negligible decline from 0.73 (SD = 0.22) at pre-test to 0.72 (SD = 0.34) at post-test. The mean F3 for the NE group's productions (M = -0.96, SD = 0.02) was lower than the mean F3 for the NJ group's productions at both pre-test and post-test for /r/, whereas the mean F3 for the NE group's productions (M = 0.96, SD = 0.03) was higher than the mean F3 for the NJ group's productions at pre-test and post-test for /l/.

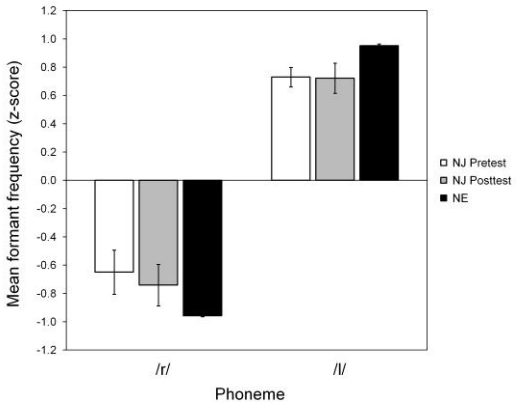

Figure 2. F3 frequencies for the NJ learners' productions at pre-test and at post-test and the NE speakers' productions. Error bars represent standard errors.

A two-way repeated ANOVA with phoneme (/r/, /l/) and testing session (pre-test, post-test) as within-subject factors was performed in order to examine

how the observed differences in F3 for the two segments were affected by the production training. The analysis revealed that the main effect of phoneme was significant, $F(1, 9) = 70.32$, $p < .001$. On the other hand, the main effect of testing session and the interaction of phoneme and testing session were not significant, $F(1, 9) = 0.12$, $p = .734$ for testing session, $F(1, 9) = 0.13$, $p = .723$ for phoneme and testing session.

Mann-Whitney tests were performed across language groups (pre-test NJ vs. NE, post-test NJ vs. NE) for each phoneme. For /r/, the difference in F3 between the NJ group's productions at pre-test and the NE group's productions was marginally significant, $U = 9.50$, $z = -1.90$, $p = .06$, whereas the F3 for the NJ group's productions at post-test was not significantly higher than the F3 for the NE group's productions, $U = 15.00$, $z = -1.23$, $p = .254$. For /l/, the F3 for the NJ group's productions at pre-test was not significantly lower than the F3 for the NE group's productions, $U = 11.00$, $z = -1.72$, $p = .099$. Likewise, the F3 for the NJ group's productions at post-test was not significantly lower than the F3 for the NE group's productions, $U = 13.00$, $z = -1.47$, $p = .165$. Therefore, the analysis suggests that the NJ group's F3 for /r/ became similar to the NE group's F3 after the training. Moreover, the NJ group's F3 was similar to the NE group's F3 for /l/ before and after the training.

4. Production Intelligibility Judgments by English Listeners

4.1 Method

4.1.1 NE Listeners

Three phonetically trained native English speakers, who were students at the University of Calgary, performed a phoneme identification task to perceptually evaluate intelligibility of the NJ learners' productions of /r/ and /l/. These students performed the identification task individually as volunteers in a testing room in the Phonetics Laboratory at the University of Calgary.

4.1.2 Stimuli

The stimuli for this task were a total of 800 utterances from the pre- and post-training recordings of the NJ learners (40 prompts × 10 learners × 2 testing sessions). Utterances from both testing sessions were randomly mixed by learner, and the presentation order was uniquely randomized for each listener.

4.1.3 Procedure

In the phoneme identification task, NE listeners heard the recorded utterances from the NJ learners, one at a time, and were asked to identify the sound that formed the initial segment of each utterance. During each trial, they saw the spelling of a word without the initial segment on a computer screen while listening to an utterance of the word from an NJ learner. They were asked to select one out of a set of sound categories (/r/, /l/, /d/, /b/, /t/, /w/, /ɾ/, and *other*) displayed on the screen for the missing segment. If the listeners selected *other*,

they were asked to describe the sound by typing in a description in a dialog box displayed on the screen.

4.2 Results

Figure 3 displays mean percent intelligibility scores for the NJ learners' productions of English /r/ and /l/ judged by the NE listeners. The mean intelligibility score for /r/ increased slightly from 74.17 (SD = 21.24) at pre-test to 75.33 (SD = 27.53) at post-test. There was a greater increase in the mean intelligibility score for /l/ from 75.83 (SD = 19.01) at pre-test to 89.00 (SD = 18.99) at post-test.

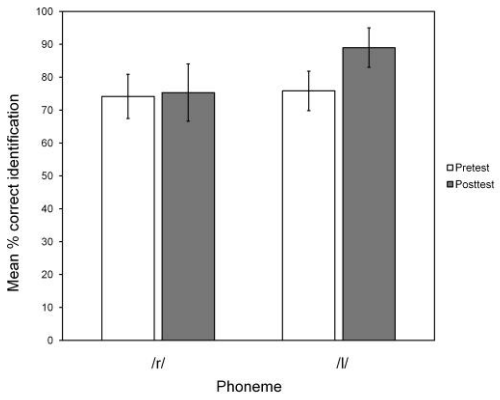

Figure 3. Percentages of intelligibility scores for the NJ learners' productions in the intelligibility judgment task as a function of phoneme and testing session. Error bars represent standard errors.

A two-way repeated ANOVA was performed with intelligibility scores as a dependent measure and phoneme (/r/, /l/) and testing session (pre-test, post-test) as within-subject factors. There were no significant main effects of phoneme, $F(1, 9) = 1.05$, $p = .333$, or testing session, $F(1, 9) = 2.74$, $p = .133$. Moreover, the interaction effect of phoneme and testing session was not significant, $F(1, 9) = 0.88$, $p = .374$. Although this analysis suggests that the training did not improve the intelligibility of the NJ learners' /r/ and /l/ productions, the lack of significance may be due to the large variability in the NE listeners' responses, as the standard deviation values indicate.

5. Perception of English /r/ and /l/ by Japanese Learners

5.1 Method

5.1.1 Stimuli

Sixty sets of minimal-pair monosyllabic English words contrasting /r/ and /l/ word-initially were selected as auditory stimuli for the perceptual tests (120

words in total). None of the words were used for the production recordings or the production training. The stimuli were recorded in a sound-attenuated booth in the Phonetics Laboratory at the University of Calgary by a female NE speaker and the male NE speaker who recorded the auditory prompts for the production recordings. A total of 240 stimuli (120 words × 2 speakers) were created for the perception tests. The stimuli were divided into two sets, with each set containing 120 stimuli, including 60 words produced by the male speaker and 60 words produced by the female speaker. That is, Set 1 included Pairs 1 to 30 produced by the male speaker and Pairs 31 to 60 produced by the female speaker. Set 2 included Pairs 1 to 30 produced by the female speaker and Pairs 31 to 60 produced by the male speaker. Each NJ learner was randomly assigned to either of the stimulus sets.

5.1.2 Procedure

NJ learners underwent perception tests individually in the sound-attenuated booth in the Speech Research Laboratory at the University of Victoria. At the beginning of each trial, orthographic representations of two words from a minimal pair were displayed on the computer screen. The word from the pair starting with /r/ was positioned at the bottom right, and the word from the pair starting with /l/ was positioned at the bottom left. While seeing the pair of words on the screen, learners heard one of the words over headphones and were asked to select the word that they thought they had heard by pressing a key corresponding to the word. Before the test, the learners completed a practice block of two trials in order to gain familiarity with the task. No feedback on the learners' responses was provided in the test trials and practice trials. The test comprised two blocks, and each block comprised 60 trials (2 blocks × 60 trials = 120 trials). Each stimulus was presented only once. Presentation order was randomized within block and across learners. Stimuli and procedures were identical for the pre-test and the post-test. The learners heard the same sets of words in both pre-test and post-test.

5.2 Analysis

Changes in the learners' perceptual sensitivity to the contrast between the phonemes were also assessed using d' (d-prime), a measure of sensitivity used in Signal Detection Theory (Green and Swets 1966, Macmillan and Creelman 2005). d' values were calculated for each learner for each testing session and were subsequently averaged across learners for each testing session. Additionally, changes in the learners' response bias were assessed using c (criterion location) (Green and Swets 1966, Macmillan and Creelman, 2005). c values were calculated for each learner for each testing session and were subsequently averaged across learners for each testing session.

5.3 Results

As Figure 4 shows, the mean percent correct identification of /r/ for the NJ learners declined from 66.50 (SD = 11.29) at pre-test to 61.50 (SD = 11.34) at post-test. On the other hand, the mean percent correct identification of /l/ increased from 55.83 (SD = 9.85) at pre-test to 60.67 (SD = 12.20) at post-test.

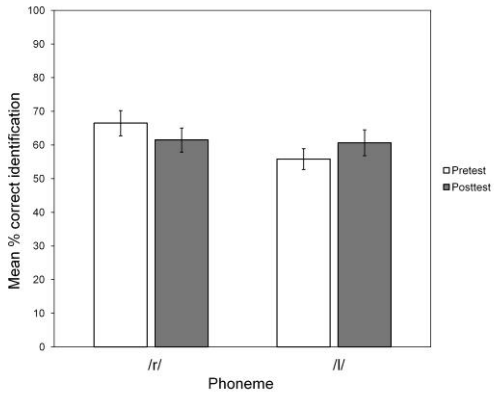

Figure 4. Percentages of correct identification scores for the NJ learners in the perceptual tests as a function of phoneme and testing session. Error bars represent standard errors.

In order to examine the observed changes, a 2 × 2 × 2 mixed ANOVA was conducted with phoneme (/r/, /l/) and testing session (pre-test, post-test) as within-subject factors, as well as stimulus set (Set 1, Set 2) as a between-subject factor. The stimulus set was included as a factor in the analysis in order to examine whether particular combinations of the NE talkers and stimulus words influenced the learners' perception. The analysis revealed no significant main effects of phoneme, $F(1, 8) = 1.70$, $p = .229$, testing session, $F(1, 8) = 0.001$, $p = .98$, or stimulus set, $F(1, 8) = 0.15$, $p = .71$. There were no significant interaction effects of 1) phoneme and stimulus set, $F(1, 8) = 0.49$, $p = .51$, 2) testing session and stimulus set, $F = 0.36$, $p = .564$, or 3) phoneme, testing session and stimulus set, $F(1, 8) = 0.04$, $p = .843$. However, an interaction of phoneme and testing session was significant, $F(1, 8) = 5.87$, $p = .042$. A simple effect analysis revealed that there was a marginally significant difference in identification accuracy between the phonemes at pre-test, $F(1, 8) = 5.07$, $p = .054$. However, the difference in identification accuracy between the phonemes at post-test was not significant, $F(1, 8) = 0.03$, $p = .871$. This indicates that although the NJ learners were more likely to accurately identify /r/ than /l/ before the training, this tendency disappeared after the training.

The mean d' value showed negligible decline from 0.60 (SD = 0.42) at pre-test to 0.59 (SD = 0.51) at post-test. In order to examine the observed change, a paired samples t-test was conducted with testing session as the within-subject factor. The difference in perceptual sensitivity (d') between the testing

sessions was not significant, $t(9) = 0.01$, $p = .996$, indicating that the perceptual sensitivity to the phoneme contrast did not improve significantly after the training. On the other hand, the mean c value increased from -0.15 (SD = 0.19) at pre-test to -0.01 (SD = 0.21) at post-test. Note that negative values indicate response bias toward /r/. A paired samples t-test was conducted with c as the dependent measure and testing session as the within-subject factor. The difference in response bias (c) was significant, $t(9) = -2.50$, $p = .034$. Therefore, the result suggests that the learners' bias to select /r/ became significantly reduced after the training.

6. Relationship between Production and Perception

6.1 Analysis

In order to analyze whether the learners' gain in production intelligibility aligned with the degree of change in their perceptual accuracy, Pearson's correlation coefficients and their levels of statistical significance were calculated for each phoneme. The degree of change in production intelligibility was calculated by subtracting the average percentage of each phoneme correctly identified for the pre-training productions from the average percentage of the corresponding phoneme correctly identified by the English listeners for the post-training productions for each Japanese learner. Likewise, the degree of change in perceptual accuracy was calculated by subtracting the percentage of times each phoneme was correctly identified in the pre-training perception test from the percentage of times the same phoneme was correctly identified in the post-training perception test by each Japanese learner.

6.2 Results

The correlation between the degree of change in production intelligibility and the degree of change in perceptual accuracy for /r/ was not significant, $r = -.59$, $p = .074$. Likewise, the correlation between the degree of change in production intelligibility and the degree of change in perceptual accuracy for /l/ was not significant, $r = -.08$, $p = .837$. These appear to suggest that the amount of improvement in production intelligibility was not related to the amount of improvement in perceptual accuracy for either of the phonemes.

7. Discussion

The results indicate that the ultrasound production training improved the Japanese learners' productions of /l/. The acoustic analyses suggest that the quality of the productions became more native-like. The learners' productions of /r/ and /l/ became more similar to the English speakers' productions of the same phonemes in terms of F3 after the training. On the other hand, the learners' productions of /r/ became less similar to the English speakers' production of /r/, whereas the learners' productions of /l/ became more similar to the English speaker's productions of /l/ in terms of F2. Thus, /r/ and /l/, as produced by the

learners, became more distinct from each other, but /l/ became more native-like than /r/.

The intelligibility of the Japanese learners' productions of /l/ did not improve significantly. However, the higher percentage of /l/ identified correctly by the English listeners for the post-test productions suggests that it is possible that the learners' productions of /l/ became more intelligible to native English speakers after the training.

This improvement in production, however, did not lead to improved perceptual accuracy for the same phoneme. This lack of improvement in perception confirms the previous finding in which improvement in the production of /r/ and /l/ did not transfer to the perception of the same phonemes (Hattori 2009). At the same time, the training in the present study helped the learners to reduce their bias to provide more /r/ responses without changing their perceptual sensitivity to the /r/-/l/ contrast.

Close inspection of individual Japanese learners' performance revealed considerable variation in degree and modality (i.e., perception and production) of improvement across learners. Further, the lack of significant correlation in the degree of change between production and perception is in line with previous production training studies (Hattori 2009, Baese-Bark 2010). These appear to suggest that production does not align with perception, which contradicts claims of Speech Learning Model (Flege, 1995) that production learning requires accurate auditory representations as targets, and the accuracy level of the auditory representation confines the accuracy level of production in L2 speech learning.

Some researchers have claimed that speech perception and production are tightly linked (e.g., Liberman and Mattingly 1985, Fowler 1986, Best 1995). If perception and production are associated through a tight linkage, improvement in production could transfer to perception. However, evidence for transfer of learning from production to perception was not observed in the present study, as well as in earlier studies (Hattori 2009, Baese-Bark 2010). On the other hand, transfer of perception learning to production with perception training has been observed (Bradlow et al. 1997, Wang et al. 2003, Baese-Bark 2010). Such transfer of learning across modalities could not occur if production and perception were not associated in any way. It is not clear, however, why there is such asymmetry in transfer between the two modalities.

The present study as well as earlier studies (Hattori 2009, Baese-Bark 2010) showed no correlations between perception and production. Such lack of correlations has also been observed with perceptual training (Bradlow et al. 1997, Iverson et al. 2012). An emerging account for such lack of correlation in improvement between production and perception is that the two modalities may make use of distinct developmental processes underlain by different representations (Iverson et al., 2012). This proposal appears to be in line with the considerable variability observed in the individual learners' data from the present study.

8. Implications

Although the present study showed improvement in production of /l/, it is possible that more robust production learning may have emerged for both phonemes with more training sessions and more learners. Additionally, perceptual learning might occur if learners with a longer length of residence in an English-speaking country and more proficiency with the sounds in question were included in a study such as this one. One implication of these results for L2 education is that improvement in production does not necessarily indicate that the learner has become able to perceive the phoneme to the same degree, which is also suggested by Sheldon and Strange (1982). In other words, the learner's perceptual ability needs to be assessed through a perception test, not based on his or her production ability, and perceptual training is necessary in order for perception to improve.

References

Adler-Bock, Marcy, Barbara May Bernhardt, Bryan Gick, and Penelope Bacsfalvi. 2007. The use of ultrasound in remediation of North American English /r/ in 2 adolescents. *American Journal of Speech-Language Pathology* 16: 128-139.

Boersma, Paul and David Weenink. 2009. Praat: Doing phonetics by computer (Version 5.1.17) [Computer program]. Amsterdam. http://www. praat.org/

Bradlow, Ann R. 2008. Training non-native language sound patterns: Lessons from training Japanese adults on the English /ɹ/-/l/ contrast. In *Phonology and second language acquisition*, Vol. 36, ed. Jette G. Hansen Edwards and Mary .L. Zampini, 287-308. Amsterdam: John Benjamins Publishing Company.

Baese-Berk, Melissa Michaud. 2010. An examination of the relationship between speech perception and production. Doctoral dissertation, Northwestern University.

Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In *Speech perception and linguistic experience: Issues in cross-language research*, ed. Winifred Strange, 171-204. Timonium: York Press.

Best, Catherine T and Winifred Strange. 1992. Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20:305-330.

Bradlow, Ann R., David B. Pisoni, Reiko Akahane-Yamada, and Yoh'ichi Tohkura. 1997. Training Japanese listeners to identify English /r/ and /l/ IV: Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America* 101: 2299-2310.

Flege, James Emil. 1995. Second language speech learning: Theory, findings, and problems. In *Speech perception and linguistic experience: Issues in cross-language research*, ed. Winifred Strange, 233-277. Timonium: York Press.

Fowler, Carol A. 1986. An event approach to the study of speech perception from a direct realist perspective. *Journal of Phonetics* 14: 3-28.

Gick, Bryan. 2002. The use of ultrasound for linguistic phonetic fieldwork. *Journal of the International Phonetic Association* 32(2): 113-121.

Gick, Bryan, Barbara May Bernhardt, Penelope Bacsfalvi, and Ian Wilson. 2008. Ultrasound imaging applications in second language acquisition. In *Phonology and second language acquisition*, Vol. 36, ed. Jette G. Hansen Edwards and Mary L. Zampini, 315-328. Amsterdam: John Benjamins Publishing Company.

Goto, Hiromu. 1971. Auditory perception by normal Japanese adults of the sounds “L” and “R”. *Neuropsychologia* 9: 317-323.

Green, David M. and John A. Swets. 1966. *Signal detection theory and psychophysics*. NewYork: Wiley.

Hattori, Kota. 2009. Perception and production of English /r/-/l/ by adult Japanese speakers. Doctoral dissertation, University College London.

Iverson, Paul, Patricia K. Kuhl, Reiko Akahane-Yamada, Eugen Diesch, Yoh’ichi Tohkura, Andres Kettermann, and Claudia Siebert. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87:B47-B57.

Iverson, Paul, Melanie Pinet, and Bronwen G. Evans. 2012. Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics* 33: 145-160.

Liberman, Alvin M. and Ignatius G. Mattingly. 1985. The motor theory of speech perception revised. *Cognition* 21: 1-36.

John S. Logan, Scott E. Lively, and David B. Pisoni. 1991. Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of Acoustical Society of America 89*: 874-886.

Lotto, Andrew J., Momoko Sato, and Randy L. Diehl. 2004. Mapping the task for the second language learner: The case of Japanese acquisition of /r/ and /l/. In *From sound to sense: 50+ years of discoveries in speech communication*, ed. J. Slifka, S. Manueal, and M. Matthies, C-181-C-186. Cambridge: MIT Research Laboratory in Electronics.

Lobanov, B. M. 1971. Classification of Russian vowels spoken by different speakers. *Journal of Acoustical Society of America* 49(2): 606-608.

Macmillan, Neil A. and C. Douglas Creelman. 2005. *Detection theory: A user’s guide*. 2nd ed. Mahwah: Lawrence Erlbaum Associates.

Miyawaki, Kuniko, Winifred Strange, Robert Verbrugge, Alvin M. Liberman, James J. Jenkins, and Osamu Fujimura. 1975. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psycholinguistics* 18: 331-340.

O’Conner, J. D., L. J. Gerstman, A. M. Liberman, P. C. Delattre, and F. S. Cooper. 1957. Acoustic cues for the perception of initial /w, j, r, l/ in English. *Word* 13: 25-43.

Sheldon, Amy and Winifred Strange. 1982. The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics* 3: 243-261.

Tsui, Haley May-Lai. 2012. Ultrasound speech training for Japanese adults learning English as a second language. Master’s thesis, University of British Columbia.

Vance, Timothy J. 2008. *The sounds of Japanese.* New York: Cambridge University Press.

Wang, Yue, Allard Jongman, and Joan A. Sereno. 2003. Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America* 113(2): 1033-1043.