

Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production

ANN R. BRADLOW

Northwestern University, Evanston, Illinois

REIKO AKAHANE-YAMADA

ATR Human Information Processing Research Laboratories, Kyoto, Japan

DAVID B. PISONI

Indiana University, Bloomington, Indiana

and

YOH'ICHI TOHKURA

ATR Human Information Processing Research Laboratories, Kyoto, Japan

Previous work from our laboratories has shown that monolingual Japanese adults who were given intensive high-variability perceptual training improved in both perception and production of English /r/-/l/ minimal pairs. In this study, we extended those findings by investigating the long-term retention of learning in both perception and production of this difficult non-native contrast. Results showed that 3 months after completion of the perceptual training procedure, the Japanese trainees maintained their improved levels of performance on the perceptual identification task. Furthermore, perceptual evaluations by native American English listeners of the Japanese trainees' pretest, posttest, and 3-month follow-up speech productions showed that the trainees retained their long-term improvements in the general quality, identifiability, and overall intelligibility of their English /r/-/l/ word productions. Taken together, the results provide further support for the efficacy of high-variability laboratory speech sound training procedures, and suggest an optimistic outlook for the application of such procedures for a wide range of "special populations."

Over the past decade, several important advances have been made toward establishing effective laboratory training procedures for modifying the identification of difficult non-native phonetic categories (for recent reviews, see Akahane-Yamada, 1996; Jamieson, 1995; Logan & Pruitt, 1995; Pisoni & Lively, 1995; Pisoni, Lively, & Logan, 1994). In addition to the benefits that such speech sound training procedures present for second-language learners, this general research agenda also provides important new information regarding the extent to which the adult phonetic system is plastic and thus capable of undergoing linguistically meaningful modifications. In our laboratories, we have focused our efforts on the acquisition of the English /r/-/l/ contrast by monolingual Japanese speakers. This

contrast was selected as a test case for assessing novel approaches to non-native speech contrast training because of its extreme difficulty for Japanese speakers (Goto, 1971; MacKain, Best, & Strange, 1981; Miyawaki et al., 1975; Mochizuki, 1981; Sheldon & Strange, 1982; Yamada & Tohkura, 1992), and because it had been shown in previous studies to be resistant to modification after discrimination training with synthetic consonant-vowel stimuli (see Strange & Dittmann, 1984).

Accordingly, a laboratory training procedure that included several novel features was developed in our laboratories (Logan, Lively, & Pisoni, 1991). In this training procedure, a minimal-pair identification task was used in order to encourage classification into broad phonetic categories, rather than emphasize discrimination of fine-grained within-category acoustic differences. Furthermore, the trainees were presented with naturally produced tokens of English /r/ and /l/ words with the target segment in a variety of phonetic environments. Finally, all training stimuli were uttered by multiple talkers of General American English. In this manner, the trainees were exposed to the full range of category variability that they could expect to encounter in real-world English /r/ and /l/ exemplars. More importantly, the training task closely matched the

We are grateful to Luis Hernandez and Takahiro Adachi for technical support, and to Rieko Kubo and Melissa Kluck for subject running. This work was supported by NIDCD Training Grant DC-00012 and by NIDCD Research Grant DC-00111 to Indiana University. Correspondence concerning this article should be addressed to A. R. Bradlow, Department of Linguistics, Northwestern University, 2016 Sheridan Rd., Evanston, IL 60208-4090 (e-mail: abradlow@nwu.edu).

—Accepted by previous editor, Myron L. Braunstein

demands of the identification task used to assess changes in spoken word recognition performance before and after training. The results of several initial training studies using this "high-variability" perceptual training procedure demonstrated that Japanese trainees could acquire robust /r/ and /l/ phonetic categories, which they could apply more generally in understanding novel talkers and novel tokens (Lively, Logan, & Pisoni, 1993; Logan et al., 1991; Yamada, 1993). Moreover, these changes were retained for several months after the completion of training (see Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994).

More recently, we also showed that the improved English /r/-/l/ identification that resulted from these perceptual training procedures transferred to improved production of English /r/ and /l/ words spoken immediately after perception training was completed (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997). Specifically, using a "playback" design in which the Japanese trainees' productions were presented to native American English listeners, the Japanese trainees' posttest productions of English /r/ and /l/ words were judged to be "better pronounced" than the corresponding pretest productions. The posttest productions were also more accurately identified in a forced-choice minimal-pair identification task than the pretest productions. Thus, the perceptual changes that resulted from the high-variability training procedure extended beyond the perceptual domain and also produced changes in speech production and motor control used in the articulation of these non-native phonetic categories. This finding has provided important new information about the relationship between speech perception and production, and it suggests that the perceptually oriented training program resulted in modifications of an underlying perceptuomotor, phonetic representation that is common to or shared by both speech perception and speech production mechanisms.

The goal of the present study was to extend this latest finding by investigating whether the observed changes in speech production would be retained several months after the perceptual identification training was completed. Lively et al. (1994) showed that improvement in perceptual identification was retained for 3 months after training. We were therefore interested in replicating and extending this result by examining the long-term retention of the improvement in both production and perception that we observed immediately following training (Bradlow et al., 1997). A second goal of this study was to investigate whether the production improvement was also present in an open-set transcription task in which the American English listeners were given no clues regarding the identity of the intended word. This perceptual assessment of the production improvement after perceptual identification training would allow us to assess the extent to which the Japanese trainees showed an improvement in overall word intelligibility, in addition to the improvement in general quality and minimal-pair identifiability that we have reported previously (Bradlow et al., 1997). The results of these investigations would

allow us to develop a more detailed understanding of the long-term phonetic changes that resulted from the high-variability perceptual identification training procedure.

METHOD

Subjects

A group of 11 native Japanese speakers (5 female, 6 male) served as the trained subjects. Of these 11 trained subjects, 9 returned for the 3-month follow-up test. They ranged in age from 19 to 22 years and were recruited from Doshisha University in Kyoto, Japan. None had ever lived abroad or had any special English language training. A comparable group of 12 subjects (6 female, 6 male) served as untrained controls. Of these 12 Japanese controls, 7 returned to the laboratory for the 3-month follow-up test. Finally, native American English listeners were recruited from the Indiana University student population to serve as judges of the Japanese subjects' pre- and posttest utterances. For each Japanese subject, separate panels of 10 American English listeners served as judges in each of the three production evaluation tests described below. Thus, there were 90 American English listeners for each evaluation test of the Japanese trainees' productions (10 listeners \times 9 trainees), plus 70 American English listeners for each evaluation test of the Japanese controls' productions (10 listeners \times 7 controls).

Perception Training

The stimuli and procedure used to train these subjects have been described in detail in our earlier papers (see Bradlow et al., 1997; Lively et al., 1993; Lively et al., 1994; Logan et al., 1991; Yamada, 1993). In the present report, therefore, we will provide only a brief description of our training methodology, and we refer the reader to the previous papers for additional details. The speech stimuli were selected from a large digital database of naturally produced /r/-/l/ minimal pairs that was originally recorded in the Speech Research Laboratory at Indiana University (see Logan et al., 1991). The pretest stimuli consisted of 16 English minimal pairs that contrasted /r/ and /l/ in four phonetic environments, plus four additional minimal pairs that contrasted other English phonemes. The original word list was taken from Strange and Dittmann (1984), but new recordings of the items were made at Indiana University, using a male speaker of General American English. The stimuli for the training phase consisted of 68 minimal pairs that contrasted /r/ and /l/ in five phonetic environments. These utterances were spoken by five speakers of General American English (three males and two females). At the posttest phase, the subjects were presented with three sets of stimuli: the original pretest stimuli, plus two sets of generalization stimuli. The stimuli for the first test of generalization (TG-1) consisted of an additional 96 words that placed /r/ or /l/ in five different phonetic environments spoken by a new talker (i.e., not one of the talkers who had produced the training stimuli). The stimuli for the second test of generalization (TG-2) consisted of an additional 99 words (five phonetic environments) spoken by an old talker (i.e., one of the talkers who had produced the training stimuli). In order to assess retention of improved perceptual identification abilities, a 3-month follow-up test was administered in which the subjects were tested with the original pretest stimuli as well as the stimuli for the two tests of generalization.

All perception training and testing was done at ATR Human Information Processing Research Laboratories in Kyoto, using individual subject cubicles that were equipped with NeXT workstations and headphones (STAX-SR-Lambda Signature). On each trial, the two members of an English /r/-/l/ minimal pair appeared on the screen in standard English orthography. The spoken test word was then presented over headphones, and the subjects had 10 seconds to

identify the stimulus by pressing "1" for the word on the left of the screen or "2" for the word on the right of the screen. During training, feedback was provided in the form of a buzzer (incorrect response) or a chime (correct response). As an additional motivation to perform well on the training task, the trainees received a one yen bonus for each correct response (at the time of testing, one yen = approximately one U.S. cent). There was no feedback for the pretest, posttest, tests of generalization, or 3-month follow-up perceptual tests. The training phase took place over a period of 3–4 weeks, during which time the trainees returned to the laboratory 15 times for training sessions. On each of the 15 training days, the trainees were given 3 training sessions for a total of 45 training sessions. In each training session, the trainees performed the two-alternative identification task described above with the full set of training stimuli from a single talker. Thus, on each of the 15 training days, the subjects were exposed to stimuli produced by three talkers. Each session took approximately 20–30 min, for a total of 15–22.5 h of training (45 sessions of 20–30 min each).

Speech Production Recordings

At the time of the pretest, posttest, and 3-month follow-up phases, the Japanese trainees were also asked to produce a set of 55 English /r/-/l/ minimal pairs. This set of stimuli included words that placed the target /r/ and /l/ segments in a variety of phonetic environments (see Bradlow et al., 1997, for additional details). The audio recordings were made in an anechoic chamber at ATR Human Information Processing Research Laboratories. An imitation task was used to elicit these utterances. The subjects were presented with both visual and auditory prompts. For each word, the visual prompt was simply the target English word displayed in standard English orthography on a cardboard panel, and the auditory prompt was a recording of a male speaker of General American English producing the target word. This auditory prompt was provided to ensure consistent pronunciation of the rest of the words (aside from the /r/ or /l/) across subjects. Once collected and stored digitally at ATR, these digital speech files were transferred to the Speech Research Laboratory at Indiana University, where they were presented to native speakers of General American English for perceptual evaluation.

Perceptual Evaluations of Trainee Productions

Three independent perceptual evaluation tests were carried out in which native speakers of General American English were asked to judge the Japanese trainees' pretest, posttest, and 3-month follow-up utterances. These playback tests included a preference rating task, a minimal-pair identification task, and an open-set transcription task. For each test, independent groups of 10 listeners evaluated the productions of each Japanese subject. Thus, each American English listener evaluated the utterances of only 1 Japanese subject. Furthermore, no listener participated in more than one evaluation test. All of the perceptual evaluation tests were carried out in the Speech Research Laboratory at Indiana University in Bloomington.

Preference rating task. In the preference rating task, the American English listeners were asked to directly compare the relative phonetic qualities of two versions of a single Japanese subject's productions (e.g., pretest vs. posttest, pretest vs. 3-month). In this task, the listener heard two tokens of an English /r/ or /l/ word (e.g., pretest and posttest), and then indicated on a 7-point scale which version sounded "better." The target word appeared in standard English orthography on a CRT monitor so that the listener was aware of the Japanese subject's intended pronunciation. A response of "1" indicated that the first version sounded *much better than the second*, a response of "7" indicated that the second version sounded *much better than the first*, and a response of "4" indicated that there was *no noticeable difference*. The order of presentation of the two versions was counterbalanced across trials. We expected that this test would pro-

vide a very sensitive measure of any improvement in the general quality of English /r/ and /l/ words produced by the Japanese subjects.

Minimal-pair identification task. In the minimal pair-identification task, the American English listeners were asked to identify and categorize the Japanese subjects' productions using a two-alternative forced-choice presentation format. On each trial, the listeners saw the two members of the minimal pair in standard English orthography on a CRT monitor. They then heard one of the members of the pair spoken by the Japanese subject and responded by identifying the stimulus with one of the two written words. The order of the response alternatives was counterbalanced across trials, so that on half the trials the /l/ word was on the left of the CRT monitor, and on the other half, it was on the right. This perceptual test provided a quantitative measure of segment-specific improvement in /r/ and /l/ articulation.

Open-set transcription task. The open-set transcription task was a dictation task in which the listener was given no context for the identity of the Japanese subject's utterance. In this test, the listeners heard a word spoken by the Japanese subject and then responded by typing what they heard on the keyboard. A word was scored as correctly transcribed if, and only if, the transcription exactly matched the intended word (aside from any obvious typographical or spelling errors). This test provided a strict test of overall word intelligibility without context.

Taken together, the three perceptual assessment tests provided us with a converging set of behavioral measures of the changes in speech production that resulted from the perceptual identification training procedure. These measures allowed us to assess the extent of the changes in general quality (preference rating task), in segment-specific articulation (minimal pair-identification task), and overall speech intelligibility (open-set transcription task).

RESULTS

Perceptual Learning

Figure 1 shows the perceptual identification accuracy scores at pretest, at posttest, and on the 3-month follow-up test for the 9 trained (left panel) and 7 control (right panel) subjects who participated in these three phases of the study. As this figure shows, the trained subjects improved substantially above pretest scores in their ability to identify English /r/ and /l/ words in the posttest and 3-month follow-up phases, whereas the control subjects showed no change in perceptual identification accuracy across these conditions. The trained group did not differ in performance on the posttest and 3-month follow-up. Furthermore, the trained and control groups' pretest accuracy did not differ, indicating that the two groups were indeed comparable at the time of pretest. A two-factor repeated-measures analysis of variance (ANOVA), with test (pre-, post-, 3-month) as the repeated measure and group (trained, control) as the between-groups factor showed main effects of test [$F(2,28) = 13.851, p < .001$] and of group [$F(1,14) = 5.65, p = .032$]. The group \times test interaction was also significant [$F(2,28) = 15.17, p < .001$] because of the difference in accuracy scores across tests for the trained group, but not for the control group. Paired t tests showed a significant improvement for the trained group from pretest to posttest [$t(8) = -7.392, p < .005$], and from pretest to 3-month follow-up [$t(8) = -3.905, p < .005$].

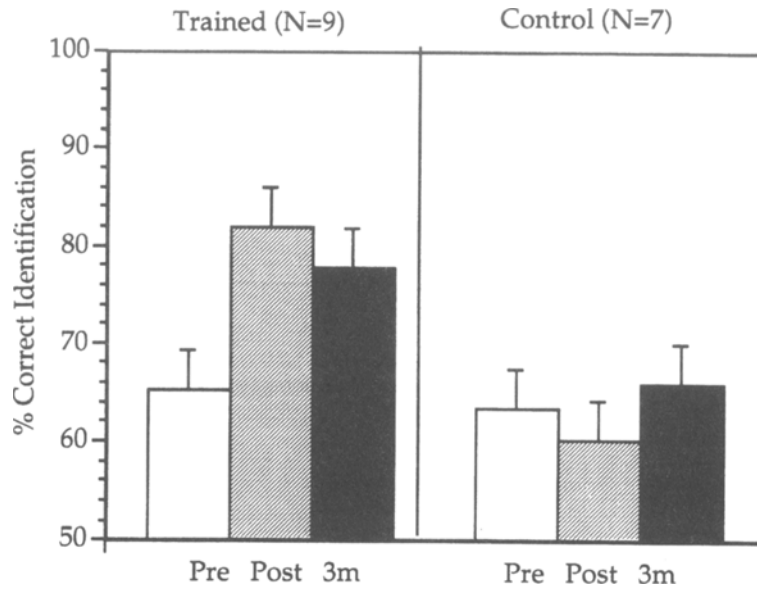


Figure 1. Percent correct perceptual identification performance for trained (left) and control (right) subjects at pretest, posttest, and 3-month follow-up for the subjects who participated in all three test phases. The error bars represent one standard error of the mean.

Table 1 provides the complete set of identification scores for all of the individual trained and control subjects on the pretest, posttest, and 3-month follow-up test, as well as on the two tests of generalization that were administered only

in the posttest and 3-month follow-up phases (see also Bradlow et al., 1997). An examination of the individual subject data shows that, of the 9 trained subjects who returned to the laboratory for the 3-month follow-up test, 7

Table 1
Pretest, Posttest, and 3-Month Follow-Up Perceptual Identification Accuracy Scores for All Japanese Trained and Control Subjects; Also Shown Are Scores for the Two Tests of Generalization at Posttest and 3-Month Follow-Up

Subjects	Pretest	Posttest	3-Months	Post-TG1	3mo.-TG1	Post-TG2	3mo.-TG2
Trained							
1	67.19	81.25	82.10	89.90	80.81	85.42	85.42
2	85.94	95.31	95.31	96.97	95.96	97.92	96.88
3	56.25	78.12	57.81	59.60	48.49	50.00	51.04
4	82.81	96.88	90.63	96.97	95.96	96.88	96.88
5	65.63	76.56	67.19	78.79	68.67	86.46	70.83
6	56.25	76.56	75.00	81.82	79.80	72.92	68.75
7	51.56	59.38	59.38	66.67	59.60	61.46	56.25
8	68.75	92.19	85.94	95.96	89.90	89.58	84.38
9	56.25	62.50	—	62.63	—	61.46	—
10	57.81	84.38	89.06	88.89	82.82	89.58	78.13
11	67.19	92.19	—	93.94	—	87.50	—
<i>M</i>	65.06	81.39	78.05	82.92	78.00	79.92	76.51
Control							
1	57.81	59.38	—	54.55	—	57.29	—
2	64.06	60.94	65.63	57.58	61.62	59.38	55.21
3	62.50	51.57	—	58.59	—	58.33	—
4	67.19	62.50	67.19	68.69	67.68	65.63	72.92
5	62.50	53.13	—	59.60	—	53.13	—
6	73.44	62.50	—	66.67	—	64.58	—
7	71.88	71.88	68.75	69.70	62.63	76.04	61.46
8	54.69	48.44	54.69	48.48	55.56	52.08	57.29
9	57.81	53.13	65.63	64.65	61.62	54.17	62.50
10	54.69	56.25	—	49.50	—	56.25	—
11	67.19	62.50	70.31	57.58	56.57	66.67	63.54
12	73.44	68.75	70.31	61.62	61.62	68.75	70.83
<i>M</i>	63.93	59.25	66.07	64.65	61.04	57.29	63.39

Note—TG, generalization test; TG1, new words, new talker; TG2, new words, old talker.

maintained a level of performance that was at least eight percentage points higher than their pretest level of performance. Only 2 subjects showed a decrease in identification accuracy back to their pretest level of performance (Trainees 3 and 5). Furthermore, at both the posttest and 3-month follow-up phases, the gains made in perceptual identification accuracy scores generalized to tests with novel items and novel talkers. Thus, the information that these subjects learned in the training generalized well beyond the specific items used in the perceptual training task.

These data replicate the earlier findings of Lively et al. (1993), who showed that a group of trained subjects maintained the improved level of identification ability even 3 months after perceptual identification training had been completed. In contrast, the group of control subjects showed no change in perceptual identification accuracy from pretest to posttest or to the 3-month follow-up test. Having established that the perceptual training procedure produced long-term changes in these Japanese trainees' ability to identify English /r/ and /l/ words, we now turn to an examination of the long-term changes in production of English /r/ and /l/ words that resulted from the perceptual identification training.

Production Improvement

Figure 2 shows the results of the preference rating task in which American English listeners directly compared the Japanese subjects' pretest utterances with their posttest utterances, and their pretest utterances with their 3-month follow-up utterances. The data shown here are for the 9 trained subjects and the 7 controls who participated in all three phases of the study. Recall that in this preference rating test the American English listeners heard two tokens (e.g., one pretest token and one posttest token) of a given word spoken by 1 Japanese subject, and they responded by indicating on a 7-point scale which version was "better articulated." The figure shows the percentage of trials in which the American English listeners judged the pretest token to be better than the posttest (or 3-month) token, versus the percentage of trials in which they judged the posttest (or 3-month) token to be better than the pretest token. (The trials that received a rating of 4, indicating no preference, are not shown in the figure.) As the figure shows, the trained subjects' posttest and 3-month tokens were preferred over the pretest tokens in 44.4% and 43.9% of the trials, respectively, whereas their pretest tokens were preferred over the posttest tokens in only 33.1% of

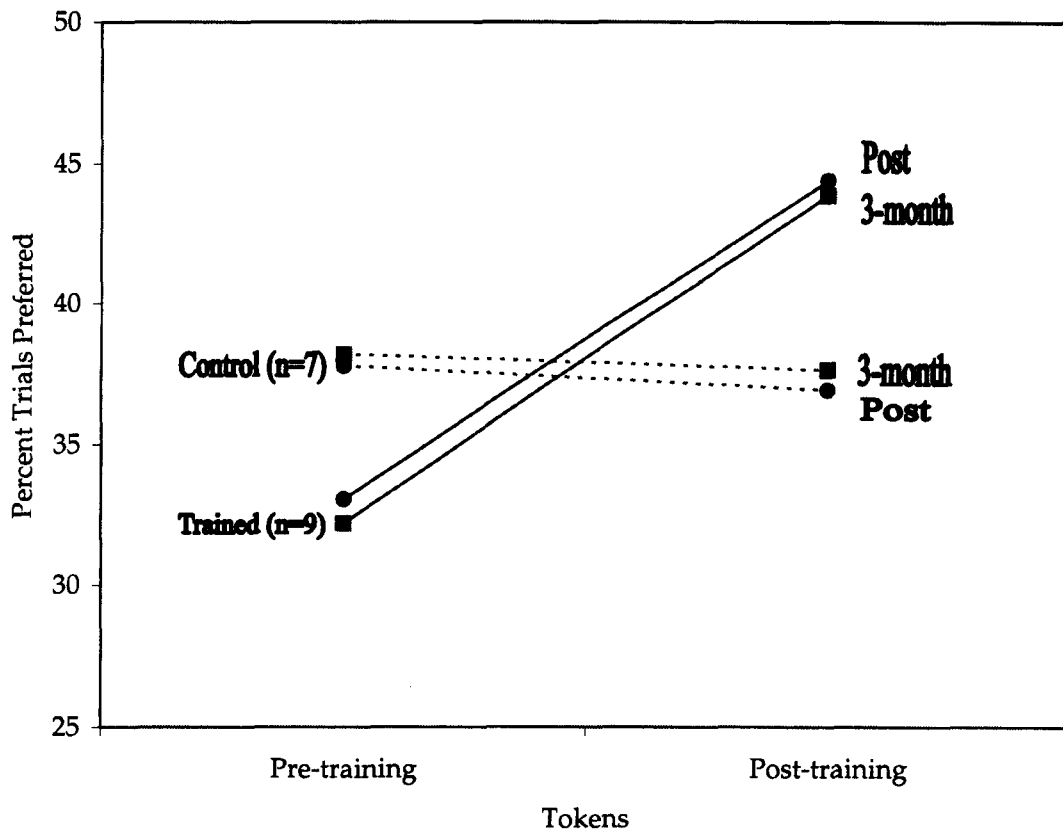


Figure 2. Proportion of trials in which the pretest tokens were preferred versus the proportion of trials in which the posttest tokens (circles) or 3-month follow-up tokens (squares) were preferred for both trained and control subjects.

the trials and over the 3-month tokens in only 32.2% of the trials. In contrast, the control subjects' data showed a much more even distribution: in both tests (pre vs. post, and pre vs. 3-month), the pretraining tokens and the post-training tokens were equally likely to be judged better than the other (around 37% of the trials).

To quantify the pattern observed in Figure 2, the preference ratings were recoded so that a response of "1," "2," or "3" always indicated a preference for the pretest token and a response of "5," "6," or "7" always indicated a preference for the posttest (or 3-month follow-up) token. This recoding simply took into account the counterbalanced order of presentation of the two tokens and allowed us to compare the mean and median rating values in each test. The skewing of the judgments of the trained subjects' productions in favor of the posttest and 3-month productions over the pretest productions was supported by strongly negative Pearson coefficients of skewness (pretest vs. posttest = -2.378 , pretest vs. 3-month = -2.272), indicating significantly greater median than mean values. In contrast, the Pearson coefficients of skewness for the control subjects' data were very close to zero, indicating similar mean and median values (pretest vs. posttest = 0.0069 , pretest vs. 3-month = 0.0215). Taken together, these data provide evidence that even 3 months after training had been completed, the perceptual learning was retained and the improved general quality of the Japanese trainees' /r/ and /l/ words was still intact.

The second production evaluation test, the minimal pair identification test, allowed us to investigate changes in production that were specific to /r/ and /l/ articulation. In this test, American English listeners identified and categorized each word produced by a Japanese subject as either the /r/ or the /l/ member of an /r/-/l/ minimal pair. For each Japanese trainee, two separate tests were carried out with separate groups of American English listeners. In the first, the pretest and posttest productions were presented; in the second, the pretest and 3-month follow-up productions were presented. Thus, the posttest and 3-month follow-

up productions were each identified in data collection sessions that also included the pretest productions. A comparison of the identification accuracies for the two presentations of the pretest productions showed no significant difference between the session with the posttest productions and the session with the 3-month follow-up productions, so in the final data analysis the two sets of pretest scores were averaged.

Table 2 shows the identification accuracy scores from the American English listeners' judgments of the pretest, posttest, and 3-month follow-up recordings from the 9 Japanese trainees who returned 3 months after training had been completed. A one-factor repeated measures ANOVA with test (pretest, posttest, 3-month) as the repeated measure showed a main effect of test [$F(2,16) = 6.381$, $p < .01$]. Paired t tests showed significant increases in identification accuracy between pretest and posttest [$t(8) = -2.516$, $p < .05$] and between pretest and 3-month follow-up test [$t(8) = -2.601$, $p < .05$], but no difference between posttest and 3-month follow-up. These data demonstrate that, even 3 months after the perceptual identification training had been completed, the segment-specific improvement in /r/ and /l/ articulation was retained by these subjects.

The control subjects' 3-month follow-up recordings were not submitted to the minimal-pair identification task (or the open-set transcription production evaluation task) since the data from the preference rating task indicated no discriminable change in the overall quality of the control subjects' productions from pretest to posttest or from pretest to 3-month follow-up test. Furthermore, in an earlier paper (Bradlow et al., 1997), we reported the results of the minimal-pair identification task for the control subjects' pretest and posttest productions ($n = 12$) which showed no difference in the American English listeners' identification accuracies for the pretest and posttest productions. Thus, there was strong a priori evidence that the control subjects' productions did not change at all from pretest to posttest to 3-month follow-up test. We therefore eliminated their productions from any further production evaluation tests under the assumption that if there were no reliable differences at posttest there would also be no differences on the 3-month follow-up test.

The third, and final, production evaluation test allowed us to examine overall word intelligibility of the Japanese trainees' pretest, posttest, and 3-month follow-up productions in the absence of any context or response constraints. In this open-set transcription task, the American English listeners heard a word spoken by a Japanese trainee and then typed on the keyboard what they heard. Table 2 shows the percent correct transcription scores for the trainees' pretest, posttest, and 3-month follow-up productions. This production evaluation test provides a very stringent measure of overall word intelligibility using an open-set response format. Thus, the overall percent correct transcription scores were considerably lower than the percent correct identification scores that were obtained in

Table 2
Individual Trainee Minimal-Pair Identification and Open-Set Transcription Scores as Judged by American English Listeners at Pretest, Posttest, and 3-Month Follow-Up

Trainee	Minimal-Pair Identification			Open-Set Transcription		
	Pretest	Posttest	3-Month	Pretest	Posttest	3-Month
1	55.95	73.00	80.05	26.01	35.97	39.18
2	95.75	95.18	97.23	53.87	55.28	57.00
3	60.43	65.41	73.86	27.03	36.63	38.27
4	98.50	98.95	98.32	74.09	71.18	83.73
5	59.86	60.91	59.23	34.39	36.10	36.18
6	62.50	72.14	73.55	36.99	42.27	45.27
7	60.75	60.18	58.05	30.93	27.94	29.27
8	75.73	81.32	85.82	47.10	54.10	59.00
9	60.00	76.09	—	34.64	38.29	—
10	56.64	62.55	68.27	29.26	34.38	35.18
11	56.50	60.55	—	30.79	28.20	—
<i>M</i>	67.51	73.30	77.15	38.65	41.85	47.01

the minimal pair(s) identification test, which had a chance level of 50%. Nevertheless, we observed a significant improvement in the overall intelligibility of the Japanese trainees' productions from pretest to posttest, and this improved level of performance was maintained even 3 months after perceptual identification training had been completed. A one-factor repeated measures ANOVA with test (pretest, posttest, 3 month) as the repeated measure, showed a main effect of test [$F(2,16) = 10.576, p < .005$]. Paired t tests showed a significant increase in identification accuracy between pretest and posttest [$t(8) = -2.356, p < .05$] and between pretest and 3-month follow-up test [$t(8) = -4.155, p < .05$].

In summary, the three perceptual evaluation tests provided independent and converging support for the claim that the "high-variability" perceptual identification training procedure produced long-term changes in the Japanese trainees' control over production of English /r/ and /l/ words. The first perceptual test, the preference rating task, was a highly sensitive, relative measure of the general quality of the Japanese subjects' productions. The second perceptual test, the minimal-pair identification task, was a segment-specific probe that provided direct evidence for improvement in the Japanese trainees' /r/ and /l/ articulations. The final test, the open-set transcription test, provided a measure of improvement in overall word intelligibility in the absence of any contextual cues for the listener regarding the identity of the target word. Thus, each of these production evaluation tests provided us with a different assessment of the changes in speech production that the perceptual identification training procedure produced in the Japanese trainees. The improvements were both general and segment specific, and they resulted in higher overall word intelligibility retained even 3 months after training had been completed.

DISCUSSION

The primary goal of this study was to investigate and assess the nature of changes in perceptual identification and production of English /r/-/l/ minimal pairs following intensive perceptual identification training and to measure the retention of this knowledge over time. The findings showed that the "high-variability" perceptual training procedure did indeed produce long-term modifications in both perception and production of a difficult non-native phonetic contrast. These findings demonstrate the effectiveness of stimulus variability for the acquisition and retention of fine phonetic details of these segmental contrasts. Furthermore, the transfer and retention of knowledge across receptive and expressive domains implies a close link between speech perception and production during perceptual learning of novel phonetic contrasts.

At this point, we can identify three major generalizations regarding speech sound learning that have emerged from our efforts to train Japanese speakers to acquire the English /r/-/l/ contrast in a laboratory setting. First, the

consistent success of the "high-variability" training procedure demonstrates that the adult phonetic system displays sufficient neuroplasticity to undergo substantial modification through laboratory listening training alone. However, it is also important to note that the Japanese trainees who participated in our studies have consistently failed to reach native-like abilities to identify English /r/ and /l/. Thus, although the adult phonetic system apparently maintains the ability to change in response to novel stimuli, it also appears to be subjected to certain limitations imposed by the native language phonetic system.

Second, the robust nature of the perceptual learning exhibited by the trainees in our studies has established that the high-variability training approach is an effective means of producing generalized long-term changes in the underlying phonetic system. Specifically, the improvements in /r/-/l/ identification generalized to novel items and novel talkers, and this knowledge was retained for at least 3 months after training. The key elements of this training approach that are apparently responsible for the robust learning are the stimuli (i.e., a wide range of /r/ and /l/ exemplars produced by multiple talkers) and the task (i.e., a minimal-pair identification task that encourages classification into broad phonetic categories rather than a discrimination task that encourages perception of fine-grained within-category differences). While the unique impact of the high-variability approach to perceptual learning has been established experimentally (Lively et al., 1993; Magnuson et al., 1995), it is still an open question whether the high-variability approach is more effective in promoting long-term improvement in production than other, "low-variability" approaches. The major strength of the high-variability training approach for perceptual learning is that it produces highly generalized learning. Because speech production also requires knowledge of how phonetic categories vary across segmental and prosodic environments, the high-variability approach should also be particularly effective in improving speech production.

The third generalization regarding speech sound learning is that the learning produced via the perceptual modality produces long-term modifications to both perception and production of the trained contrast. We believe that the underlying mechanism that facilitates the transfer and retention of learning in the perceptual domain to the production domain is due to training-induced modifications to a common mental representation that underlies both speech perception and speech production. Although our data do not address the nature (acoustic vs. articulatory) of the underlying phonetic representation, the fact that we observed immediate and long-term changes in production after perceptually oriented training suggests that the production improvement is the result of modified underlying representations rather than modified "output monitors" that are activated only during speech production. For example, speech produced under a noise load undergoes various modifications relative to speech produced in quiet (Draeger, 1951; Hanley & Steer, 1949; Lane & Tranel,

1971). However, in that case, the speech production modifications are due to the talker's on-line monitoring of the output rather than to modification of the underlying phonetic representations, as we claim to be the case in the present study.

This claim is consistent with the motor theory of speech perception (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985, 1989), which claims that listeners perceive speech in terms of the articulatory gestures of their own that would produce the perceived sound. It is also consistent with the direct realist approach to speech perception (e.g., Best, 1995; Fowler, 1986), which claims that listeners directly perceive the speaker's articulatory gestures via the structure that they impart to the acoustic medium. Both theories posit integrated perception and production systems and thus provide principled mechanisms by which perceptual training alone produces generalized changes that affect diverse speech processing operations, all of which are retained for several months after training has been completed.

The overall pattern of results suggests a very encouraging scenario for the design and application of laboratory speech sound training procedures for second-language learners, as well as for other "special populations" who exhibit difficulties with speech sound perception and production. Available data from a variety of populations that exhibit phonological problems (including second-language learners, pediatric cochlear implant users, and language-impaired children) suggest that performance on speech perception tasks tends to correlate positively with performance on corresponding speech production tasks. For example, in a recent investigation of the factors that correlate with superior performance in aural-oral language acquisition by prelingually deafened children with cochlear implants, Pisoni, Svirsky, Kirk, and Miyamoto (1997) reported a strong positive correlation between performance on a word recognition test and performance on a test of speech intelligibility. Similarly, Stark and Heinz (1996) found that impaired stop consonant perception by language-impaired children relative to normal children was associated with the presence of corresponding speech articulation errors. Furthermore, Yamada, Strange, Magnuson, Pruitt, and Clarke (1994) found a positive correlation between perception and production of the English /r/-/l/ contrast in a large group of Japanese speakers. The present investigation extends this general finding by establishing a perception-production link such that successful perceptual learning leads directly to corresponding improvement in speech production—specifically, in speech motor control and articulation. Taken together, these results support the claim that phonological acquisition via auditory-perceptual input involves concurrent development in both speech perception and speech production. Moreover, our findings suggest that the high-variability training procedure holds great promise as a general approach to the development of laboratory training procedures for the ac-

quisition of difficult phonological categories in a wide range of "phonologically disabled" populations.

REFERENCES

- AKAHANE-YAMADA, R. (1996). Learning non-native speech contrasts: What laboratory training studies tell us [Abstract]. *Journal of the Acoustical Society of America*, **100** (4, Pt. 2), 2728.
- BEST, C. T. (1995). A direct-realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language speech research* (pp. 171-206). Timonium, MD: York Press.
- 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**, 2299-2310.
- DRAEGERT, G. L. (1951). Relationships between voice variables and speech intelligibility in high level noise. *Speech Monographs*, **18**, 272-278.
- FOWLER, C. A. (1986). An even approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, **14**, 3-28.
- GOTO, H. (1971). Auditory perception by normal Japanese adults of the sounds "j" and "r." *Neuropsychologia*, **9**, 317-323.
- HANLEY, T. D., & STEER, M. D. (1949). Effect of level of distracting noise upon speaking rate, duration and intensity. *Journal of Speech & Hearing Disorders*, **14**, 363-368.
- JAMIESON, D. G. (1995). Techniques for training difficult non-native speech contrasts. *Proceedings of the International Congress of Phonetic Sciences*, **4**, 100-107.
- LANE, H. L., & TRANEL, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech & Hearing Research*, **14**, 677-709.
- LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. (1967). Perception of the speech code. *Psychological Review*, **97**, 431-461.
- LIBERMAN, A. M., & MATTINGLY, I. G. (1985). The motor theory of speech perception revised. *Cognition*, **21**, 1-36.
- LIBERMAN, A. M., & MATTINGLY, I. G. (1989). A specialization for speech perception. *Science*, **245**, 489-494.
- LIVELY, S. E., LOGAN, J. S., & PISONI, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, **94**, 1242-1255.
- 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**, 2076-2087.
- 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**, 874-886.
- LOGAN, J. S., & PRUITT, J. S. (1995). Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language speech research* (pp. 351-378). Timonium, MD: York Press.
- MACKAIN, K. S., BEST, C. T., & STRANGE, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, **2**, 369-390.
- MAGNUSON, J. S., YAMADA, R. A., TOHKURA, Y., PISONI, D. B., LIVELY, S. E., & BRADLOW, A. R. (1995). The role of talker variability in non-native phoneme training. *Proceedings of the 1995 Spring Meeting of the Acoustical Society of Japan* (pp. 393-394). Tokyo: Acoustical Society of Japan.
- MIYAWAKI, K., STRANGE, W., VERBRUGGE, R., LIBERMAN, A. M., JENKINS, J. J., & FUJIMURA, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, **18**, 331-340.
- MOCHIZUKI, M. (1981). The identification of /r/ and /l/ in natural and synthesized speech. *Journal of Phonetics*, **9**, 283-303.

- PISONI, D. B., & LIVELY, S. E. (1995). Variability and invariance in speech perception: A new look at some old problems in perceptual learning. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language speech research* (pp. 433-462). Timonium, MD: York Press.
- PISONI, D. B., LIVELY, S. E., & LOGAN, J. S. (1994). Perceptual learning of non-native speech contrasts: Implications for theories of speech perception. In J. Goodman & H. C. Nusbaum (Eds.), *Development of speech perception: The transition from speech sounds to spoken words* (pp. 121-166). Cambridge, MA: MIT Press.
- PISONI, D. B., SVIRSKY, M. A., KIRK, K. I., & MIYAMOTO, R. T. (1997, May). Looking at the "stars": A first report on the intercorrelations among measures of speech perception, intelligibility, and language in pediatric cochlear implant users. Paper presented at the Vth International Cochlear Implant Conference, New York.
- 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*, 243-261.
- STARK, R. E., & HEINZ, J. M. (1996). Perception of stop consonants in children with expressive and receptive-expressive language impairments. *Journal of Speech & Hearing Research*, *39*, 676-686.
- STRANGE, W., & DITTMANN, S. (1984). Effects of discrimination training on the perception of /r/-/l/ by Japanese adults learning English. *Perception & Psychophysics*, *36*, 131-145.
- YAMADA, R. A. (1993). Effects of extended training on /r/ and /l/ identification by native speakers of Japanese [Abstract]. *Journal of the Acoustical Society of America*, *93* (4, Pt. 2), 2391.
- YAMADA, R. A., STRANGE, W., MAGNUSON, J. S., PRUITT, J. S., & CLARKE, W. D., III (1994). The intelligibility of Japanese speakers' productions of American English /r/, /l/, and /w/, as evaluated by native speakers of American English. In *Proceedings of the International Conference on Spoken Language Processing* (pp. 2023-2026). Yokohama: Acoustical Society of Japan.
- 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*, 376-392.

(Manuscript received July 21, 1997;
revision accepted for publication May 3, 1998.)