RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 21 (1996-1997)
Indiana University

Training Japanese Listeners to Identify English /r/and /l/: Long-term Retention of Learning in Perception and Production[1]

Ann R. Bradlow,[2] Reiko Akahane-Yamada,[3] David B. Pisoni and Yoh'ichi Tohkura[3]

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by NIH-NIDCD Training Grant DC-00012 and NIH-NIDCD Research Grant DC-00111 to Indiana University. We are grateful to Luis Hernandez and Takahiro Adachi for technical support, and to Rieko Kubo and Melissa Kluck for subject running.
[2] Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL
[3] ATR Human Information Processing Research Laboratories, Kyoto, Japan

Training Japanese Listeners to Identify English /r/and /l/: Long-term Retention of Learning in Perception and Production

Abstract. Previous work from our laboratories has shown that monolingual Japanese adults who were subjected to intensive high-variability perceptual training improved in both perception and production of English /r/-/l/ minimal-pairs. This study 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 three 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, post-test, and three-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."

Introduction

Over the past decade, several important advances have been made towards 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 and Pruitt, 1995; Pisoni and Lively, 1995; Pisoni, Lively and Logan, 1995). 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; Miyawaki et al. 1975; Mochizuki, 1981; MacKain et al., 1981, Sheldon and Strange, 1982; Yamada and Tohkura, 1992), and because it had been shown in previous studies to be resistant to modification after discrimination training with synthetic CV stimuli (see Strange and Dittmann, 1984).

Accordingly, a laboratory training procedure that included several novel features was developed in our laboratories (Logan, Lively and 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 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 that generalized to novel talkers and novel tokens (Logan et al., 1991; Lively et al., 1993; Yamada, 1993). Moreover, these changes were retained for several months after the completion of training (see Lively et al, 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 (Bradlow et al., 1997). Specifically, using a "playback" design, the Japanese trainees' post-test productions of English /r/ and /l/ words were judged by American English listeners to be "better pronounced" than the corresponding pretest productions. The post-test 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 regarding the relationship between speech perception and production, and suggested that the perceptually-oriented training program resulted in modifications of an underlying perceptuomotor, phonetic representation that is common or shared by both speech perception and production mechanisms.

The goal of the present study was to extend this latest finding by investigating whether the observed changes in speech production were retained for several months after the perceptual identification training was completed. Lively et al. (1994) showed that improvement in perceptual identification was retained for three months after training. We were therefore interested in comparing the retention of production improvement with that of perception improvement. A second goal of this study was to investigate whether the production improvement was also present in an open-set transcription task where 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 observed in our earlier study (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.

Methods

Perception Training

The stimuli and procedure used to train these subjects have been described in detail in our earlier papers (see Logan et al, 1991; Lively et al, 1993; 1994; Yamada, 1993; Bradlow et al., 1997). Therefore, in the present report we provide only a brief description of our training methodology, and 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 contrast /r/ and /l/ in four phonetic environments, plus four additional minimal-pairs that contrast other English phonemes (Strange and Dittmann, 1984). These words were all spoken by a male speaker of General American English. The stimuli for the training phase consisted of 68 minimal-pairs that contrast /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 post-test phase, 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 that 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 that produced the training stimuli). In order to assess retention of improved perceptual identification abilities, a three-month follow-up test was administered in which the subjects were tested using 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, Japan 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 ten 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. There was no feedback for the pretest, post-test, tests of generalization, or the three-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.

Speech Production Recordings

At the time of the pretest, post-test, and three-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). Additionally, a set of 25 non-words were recorded. These non-words placed the target /r/ or /l/ adjacent to five vowels (/i, e, a, o, u/) and in a variety of syllable contexts (CV, CCV, VCV, VC, VCC). These non-words were collected primarily for future acoustic analysis, and were therefore omitted from the perceptual evaluation tests by American English listeners. The audio recordings were made in an anechoic chamber at ATR Human Information Processing Research Laboratories. The procedure used to elicit these utterances was an imitation task that presented the subjects 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 in order 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, post-test and three-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 ten listeners evaluated the productions of each Japanese subject. Thus, each American English listener evaluated the utterances of only one 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 versus post-test, pretest versus three- month). In this task, the listener heard two tokens of an English /r/ or /l/ word (e.g. pretest and post-test), and then indicated on a seven-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 counter-balanced across trials. This test provided 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 counter-balanced 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 of the monitor. 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 regarding 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 into the keyboard. The responses were scored such that a word was counted 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).

Subjects

A group of eleven native Japanese speakers (five females and six males) served as the trained subjects. Of these eleven trained subjects, nine returned for the three-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 twelve subjects (six females and six males) served as untrained controls. Of these twelve Japanese controls, seven returned to the laboratory for the three-month follow-up test. Finally, all of the native American English listeners who served as judges of the Japanese subjects' pre- and post-test utterances were recruited from the Indiana University student population. For each Japanese subject, separate panels of ten American English listeners served as judges in each of the production evaluation tests.

Results

Perceptual Learning

Figure 1 shows the perceptual identification accuracy scores at pretest, post-test, and at the three-month follow-up test for the nine trained (left panel) and seven control (right panel) subjects who participated in all three phases of the study (pretest, post-test, and three-month follow-up). Table I provides the complete set of identification scores for all of the individual trained and control subjects on the pretest, post-test, three-month follow-up test, as well as on the two tests of generalization that were administered only at the post-test and three-month follow-up phases (see also Bradlow et al., 1997). As shown in this figure, the trained subjects improved substantially above pretest scores in their ability to identify English /r/ and /l/ words at the post-test and three-month follow-up phases, whereas the control subjects showed no change in perceptual identification accuracy across these conditions. A two-factor repeated-measures ANOVA, with test (pre, post, 3-month) as the repeated measure and group (trained, control) as the between-groups factor showed a main effect of test ($F(2,28)=13.851$, $p<.001$), and a main effect of group ($F(1,14)=5.65$, $p=.032$). The group x test interaction was also significant ($F(2,28)=15.17$, $p<.001$) due to 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 post-test ($t(8)=-7.392$, $p<.005$), and from pretest to three-month follow-up ($t(8)=-3.905$, $p<.005$). There was no difference in performance for the trained group between the post-test and three-month follow-up scores. Furthermore, there was no difference between the trained and control groups' pretest accuracy scores, indicating that the two groups were indeed comparable at the time of pretest.

Insert Figure 1 about here.

An examination of the individual subject data (see Table I) shows that, of the nine trained subjects who returned to the laboratory for the three-month follow-up test, seven maintained a level of performance that was at least eight percentage points higher than their pretest level of performance. Only two subjects showed a decrease in identification accuracy back to their pretest level of performance. Furthermore, at both the post-test and three-month follow-up phases, the gains made in perceptual identification accuracy scores generalized to novel items and novel talkers (see Table I). Thus, the information 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 the group of trained subjects maintained the improved level of identification ability even three months after perceptual identification training was completed. In contrast, the group of control subjects showed no change in perceptual identification accuracy from pretest to post-test or to the three-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.

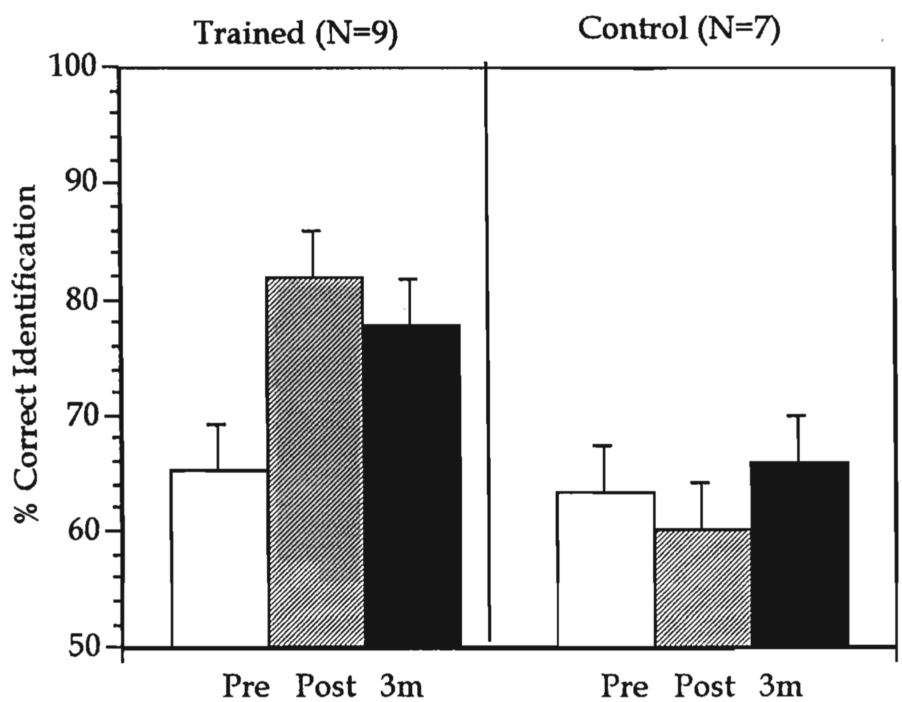

Figure 1. Percent correct perceptual identification performance for trained (left panel) and control (right panel) subjects at pretest, post-test, and three month follow-up for the subjects who participated in all three test phases. The error bars represent one standard error from the mean.

Table I.

Pretest, post-test, and three-month follow-up perceptual identification accuracy scores for all Japanese trained and control subjects. Also shown are the scores for the two tests of generalization (TG1=new words, new talker; TG2=new words, old talker) that were given at the post-test and three-month follow-up phases.

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

Production Improvement

Figure 2 shows the results of the preference rating task in which American English listeners directly compared the Japanese subjects' pretest versus post-test utterances (panels (a) and (c)) and the pretest versus the three-month follow-up utterances (panels (b) and (d)). The data shown here are for the nine trained subjects (upper panels) and the seven controls (lower panels) that 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 post-test token) of a given word spoken by one Japanese subject, and they responded by indicating on a seven-point scale which version was "better articulated." In order to take into account the counter-balanced presentation order of the two versions, the data were all 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 post-test (or three-month follow-up) token. The figure shows the distribution of rating responses across the seven response categories represented as percentages of the total number of responses from all listeners.

Insert Figure 2 about here.

As shown in Figure 2, the distribution of the ratings for the Japanese trained subjects' utterances was skewed in favor of consistently higher ratings for the test that compared the pretest versus the post-test utterances (panel (a)), as well as for the test that compared the pretest versus the three-month follow-up utterances (panel (b)). This tendency towards higher ratings indicated that the American English listeners reliably preferred the Japanese trainees' post-test and three-month follow-up utterances over their pretest utterances. In contrast, for both tests with the control subjects utterances (panels (c) and (d)), the ratings were normally distributed across the seven response categories, indicating no general preference for either the post-test or the three-month follow-up utterances over the pretest utterances. From this pattern of ratings, we conclude that the tokens from the control subjects were indiscriminable to native speakers of English.

Chi-square statistics were performed on the distribution of responses across all seven response categories using the distribution of responses to the control subjects' productions as the expected distributions, and the distribution of responses to the trained subjects productions as the observed distributions. These analyses yielded highly significant chi-squares for the pretest versus post-test comparison (chi-square=3033.46, $p(6)<.001$), and for the pretest versus three-month follow-up comparison (chi-square=1938.87, $p(6)<.001$). Taken together, these data provide evidence that even three months after training was 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 using separate groups of American English listeners. In the first, the pretest and post-test productions were presented; in the second, the pretest and three-month follow-up productions were presented. Thus, the post-test and three-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 show no significant difference between the session with the post-test productions and the session with the three-month follow-up productions, therefore in the final data analysis the two set of pretest scores were averaged.

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

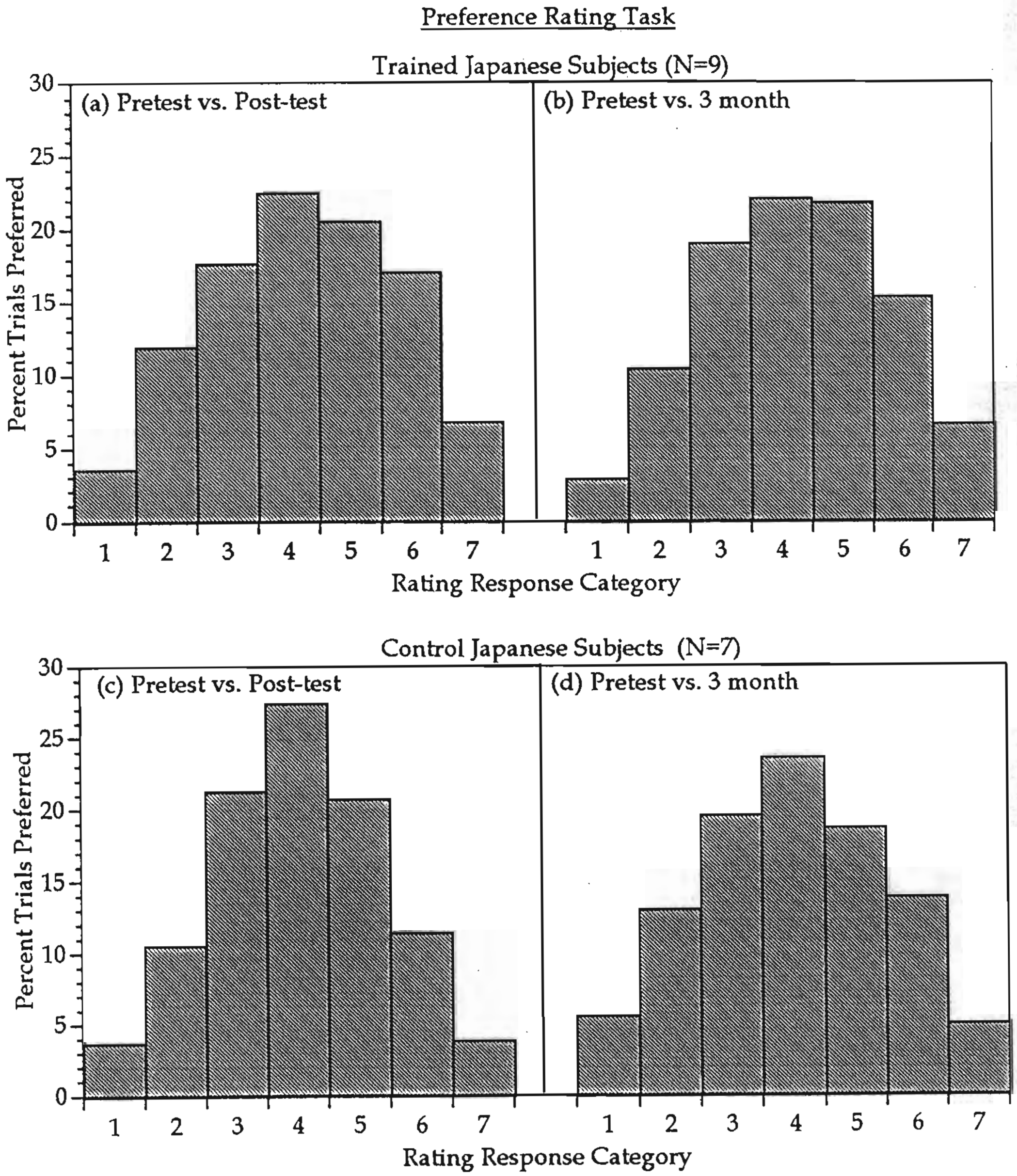

Figure 2. Distribution of responses across the seven response categories for the preference rating tests comparing the trained (top) and control (bottom) subjects' pretest versus post-test productions on the right (panels (a) and (c), respectively), and pretest versus three-month follow-up productions on the left (panels (b) and (d), respectively). A response of "1" indicated that the pretest version was preferred over the post-test or three-month token, "4" indicated no noticeable difference between the two tokens, and "7" indicated that the post-test or three-month follow-up token was preferred over the pretest token. Data are shown for only those subjects who participated in all three phases of the study, pretest, post-test and three-month follow-up.

Table II.

Individual trainee minimal-pair identification and open-set transcription scores as judged by American English listeners at pretest, post-test and three-month follow-up.

	Minimal-pair Identification			Open-set Transcription		
Trainee	pretest	post-test	3-month	pretest	post-test	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	---
mean	67.51	73.30	77.15	38.65	41.85	47.01

The control subjects' three-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 post-test, or from pretest to three-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 post-test productions (n=12), which showed no difference in the American English listeners' identification accuracies for the pretest and post-test productions. Thus, there was strong a priori evidence that the control subjects' productions did not change at all from pretest to post-test to three-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 post-test there would also be no differences at the three-month follow-up test.

The third, and final, production evaluation test allowed us to examine overall word intelligibility of the Japanese trainees' pretest, post-test and three-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 what they heard into the keyboard. Table II shows the percent correct transcription scores for the trainees' pretest, post-test and three-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 the minimal-pair 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 post-test, and this improved level of performance was maintained even three months after perceptual identification training was completed.

A one-factor repeated measures ANOVA with test (pretest, post-test, three-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 post-test ($t(8)=-2.356$, $p<.05$), and between pretest and three-month follow-up test ($t(8)=-4.155$, $p<.05$). We also found a significant difference between the post-test and three-month follow-up test ($t(8)=-2.583$, $p<.05$), such that the three-month follow-up productions were more accurately transcribed than the post-test productions. The reason for this increase is unclear at this time. However, the important finding for our purposes is that both the post-test and three-month follow-up productions were more accurately transcribed than the pretest productions and there was no decrease in the intelligibility scores after three-months.

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 resulted in higher overall word intelligibility that was retained even three months after training was 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 importance of stimulus variability for the acquisition and retention of fine phonetic details about 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 neuro-plasticity 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 three 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).

Third, the learning produced via the perceptual modality produces long-term modifications to both perception and production of the trained contrast, suggesting that changes to the underlying phonetic system occur at a level of representation that is common to both perception and production. In other words, perceptual training alone produces generalized changes that affect diverse speech processing operations, all of which are retained for several months after training is 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. 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 et al. (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 et al. (1994) found a positive correlation between perception and production of the English /r/-/l/ contrast in a large group of Japanese speakers. Thus, 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. 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 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 acquisition 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. *Proceedings of the Acoustical Society of Japan Fall meeting,* Honolulu, Hawaii.

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.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds 'l' and 'r'. *Neuropsychologia*, **9**, 317-323.

Jamieson, D. G. (1995). Techniques for training difficult non-native speech contrasts. *Proceedings of the International Congress of Phonetic Sciences,* **4,** 100-107.

Lively, S. E., Logan, J. D., & 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. D., 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.

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 and 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., Svirsky, M. A., Kirk, K. I., & Miyamoto, R.T. (1997). 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 City, NY, May 1-3, 1997.

Pisoni, D. B., Lively, S. E., & Logan, J. S. (1994). Perceptual learning of non-native speech contrasts: Implications for theories of speech perception. In H. C. Nusbaum and J. Goodman (Eds.). *Development of speech perception: The transition from speech sounds to spoken words*, (pp. 121-166). MIT Press: Cambridge, MA.

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.

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 and Hearing Research,* **39**, 676-686.

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

Yamada, R. A. (1993). Effects of extended training on /r/ and /l/ identification by native speakers of Japanese. *Journal of the Acoustical Society of America,* 93, Pt. 2, 2391.

Yamada, R. A. & Tohkura, Y. (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Perception and Psychophysics,* **52**, 376-392.

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 of Spoken Language Processing*. (pp. 2023-2026). Acoustical Society of Japan: Yokohama.