Published in final edited form as:
J Acoust Soc Am. 1997 April ; 101(4): 2299–2310.

Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production

Ann R. Bradlow,
Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington, Indiana 47405-1301

David B. Pisoni,
Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington, Indiana 47405-1301

Reiko Akahane-Yamada, and
ATR Human Information Processing Research Laboratories, 2-2 Hikaridai, Seika-cho Soraku-gun, Kyoto, 619-02 Japan

Yoh'ichi Tohkura
ATR Human Information Processing Research Laboratories, 2-2 Hikaridai, Seika-cho Soraku-gun, Kyoto, 619-02 Japan

Abstract

This study investigated the effects of training in /r/–/l/ perceptual identification on /r/–/l/ production by adult Japanese speakers. Subjects were recorded producing English words that contrast /r/ and /l/ before and after participating in an extended period of /r/–/l/ identification training using a high-variability presentation format. All subjects showed significant perceptual learning as a result of the training program, and this perceptual learning generalized to novel items spoken by new talkers. Improvement in the Japanese trainees' /r/–/l/ spoken utterances as a consequence of perceptual training was evaluated using two separate tests with native English listeners. First, a direct comparison of the pretest and post-test tokens showed significant improvement in the perceived rating of /r/ and /l/ productions as a consequence of perceptual learning. Second, the post-test productions were more accurately identified by English listeners than the pretest productions in a two-alternative minimal-pair identification procedure. These results indicate that the knowledge gained during perceptual learning of /r/ and /l/ transferred to the production domain, and thus provides novel information regarding the relationship between speech perception and production.

INTRODUCTION

The relationship between speech perception and speech production has been a long-standing issue in speech science and experimental phonetics. Some researchers have proposed a direct link between perception and production. For example, motor theorists (e.g., Liberman *et al*., 1967; Liberman and Mattingly, 1985; Liberman and Mattingly, 1989) claim that listeners perceive speech in terms of their own articulatory gestures that would produce the perceived sound. A central tenet of motor theory is that there is a specialized phonetic module that represents speech units in terms of articulatory gestures, and that this module mediates both speech perception and production. Thus, motor theory supposes a single, shared representation for speech perception and production. Other theorists have viewed the two processes of speech communication as much more autonomous. For example, proponents of acoustic-auditory theories of speech perception (e.g., Stevens and Blumstein, 1981; Diehl and Kluender, 1989) have argued that the processes of speech perception operate on the acoustic medium independently of the articulatory gestures that produced it. In other words, this approach takes the acoustic signal as the object of speech perception, and makes no explicit claims about the perception–production relationship. However, this approach does presuppose that speech perception and production are indirectly linked via common acoustically defined targets and auditory feedback mechanisms that operate during speech production. A third theoretical position, the direct-realist approach to speech perception (e.g., Fowler, 1986; Best, 1995), proposes that the listener directly perceives the articulatory gestures of the speaker in terms of the structure they impart to the acoustic medium. According to this view which is also known as event perception, the objects (events) of speech perception are the articulatory gestures, and speakers aim to achieve gesturally defined targets during speech production. Thus, the direct-realist approach proposes that speech perception and production are inextricably linked by virtue of their common communicative goal. In contrast to motor theory, however, the direct-realist approach does not propose a specialized phonetic module that mediates the direct perception–production link. Rather, direct-realism proposes that the direct speech perception–production link, which helps to ensure speaker–hearer parity, is a specific case of generally integrated event perception and action systems.

Also of long-standing interest is the acquisition of novel phonetic categories by non-native speakers. Previous research has shown that foreign accents persist even for highly proficient speakers of a non-native language (e.g., Tahta *et al.*, 1981; Flege and Hillenbrand, 1987), and that non-native speakers have extreme difficulty with both the perception and production of certain non-native phonetic contrasts (e.g., Flege, 1988; Goto, 1971). Second-language learners thus present cases where certain aspects of speaker–hearer parity break down; that is, where there is a mismatch between the phonetic system of the language-user and of the target language community. For this reason an investigation of speech perception and production by these subjects, and of the changes that occur as a result of second-language training, can provide a behavioral window into the mental representation that underlies the perception–production link.

To this end, the present study builds on earlier research from our laboratories that has established an effective procedure for training Japanese listeners to identify the English /r/–/l/ contrast, which is neutralized in Japanese (Logan *et al.*, 1991; Lively *et al.*, 1993, 1994). We wanted to know how the acquisition of a non-native perceptual contrast would affect control over production of that contrast. Thus, by directly examining the effect of perceptual learning on speech production, this project attempted to provide novel information regarding the relationship between speech perception and production in general. To accomplish this, we examined the effectiveness of perceptual identification training procedures for the acquisition of a non-native phonetic contrast in both perception and production. Any transfer of learning in perception to the production domain would provide important new evidence for a direct perception–production link. Furthermore, this outcome would suggest that perceptual identification training can facilitate the acquisition of production categories for second-language learners, as well as for other "special populations."

Previous studies that have investigated the relationship between perception and production of non-native phonetic contrasts, have generally focused on the subjects' performance in perception and production at a single point in time. For example, studies by Goto (1971) and by Sheldon and Strange (1982) showed that some Japanese subjects were able to produce identifiable /r/ and /l/ tokens even though they were unable to reliably identify native English /r/ and /l/ tokens. This finding led these researchers to conclude that production can precede perception in the acquisition of a non-native contrast. Similarly, in a study of /r/, /l/, and /w/ productions by a large number of Japanese speakers with varying degrees of exposure to English, Yamada *et al.* (1994) found that for some of the subjects, production abilities exceeded perception abilities, but not vice versa. Although these studies are informative about the relationship between perception and production in adult second-language learners, they do not provide quantitative information about how the changes in one domain (i.e., perception) affect performance in the other domain (i.e., production). Whereas the information provided by the studies discussed above is correlational, the major goal of the present study was to investigate the possibility of a functional perception–production link to the extent that success in perceptual learning leads directly to an improvement in speech production by adult second-language learners.

Two recent studies provide some indication that transfer of perceptual learning to speech production can occur. For example, Rochet (1995) reported that after perceptual identification training with a synthetic French /bu/–/pu/ continuum, Mandarin speakers displayed more French-like voice onset time (VOT) perceptual categorization. Furthermore, production data from the Mandarin subjects showed a change in VOT durations in the direction of native French VOT durations. In several recent studies with children who have articulation disorders, Jamieson and Rvachew (Jamieson and Rvachew, 1992; Rvachew, 1994; Jamieson and Rvachew, 1994) found that speech perception training can facilitate sound production learning in children who exhibit both perception and production deficits. For instance, Rvachew (1994) found that subjects who received perception training in conjunction with traditional speech production therapy showed greater improvement in /ʃ/ production than control subjects who did not receive perception training. This result indicated that perception training can enhance the effectiveness of speech production therapy for phonologically delayed children. In the present study, we examined this perception–production link further by investigating the effects of perceptual learning on production of the /r/–/l/ contrast by adult native speakers of Japanese in the absence of any explicit production training, and across a wide range of phonetic contexts.

The general design of the present study had four phases: a pretest phase, a perceptual training phase, a post-test phase, and a production assessment phase. During the pretest phase, both perception and production data were collected from a group of adult Japanese speakers. In the perceptual training phase, the subjects were trained to identify English /r/ and /l/ minimal pairs using the high-variability training technique developed in earlier work (see Logan *et al.*, 1991; Lively *et al.*, 1993, 1994; Yamada, 1993 for details regarding the principal motivation behind this procedure). In the post-test phase, both perception and production data were once again collected from the Japanese listeners. Finally, during the production assessment phase, the pre- and post-test utterances were evaluated by a group of native American English speakers. Thus, in this study, we investigated the effect of perceptual learning on subsequent performance in both perception and production.

I. PERCEPTUAL LEARNING

A. Method

1. Subjects—The subjects were 11 adult, native speakers of Japanese (5 females and 6 males), ranging in age from 19 to 22 years. None had lived abroad or had any special training in English conversation. However, as is typical in Japan, all of the subjects had studied English since junior high school (from about age 12). The subjects were recruited from Doshisha University, Kyoto prefecture, Japan. A comparable group of 12 Japanese speakers (6 females and 6 males) served as control subjects. These control subjects were also drawn from the same population as the experimental subjects. None of the subjects reported any history of a speech or hearing impairment at the time of testing. A hearing screening performed at 15 dB hearing level (HL) for the frequencies 250–8000 Hz showed all subjects to have normal bilateral acuity. All subjects were paid for their participation.

2. Procedure—The perceptual training program followed the high-variability procedures first developed by Logan *et al.* (1991), and later extended by Yamada (1993). This procedure consisted of a pretest phase, a training phase, and post-test phase. The pretest phase consisted of a minimal pair identification task with naturally produced English /r/–/l/ minimal word pairs produced by a native speaker of general American English. The perceptual learning phase involved 45 sessions (over a period of 3–4 weeks) of perceptual identification with feedback. The training stimuli consisted of a large number of naturally spoken /r/–/l/ minimal word pairs produced by five native speakers of general American English. Finally, the post-test phase included a perceptual identification post-test (identical to the pretest), and two tests of generalization. The tests of perceptual generalization consisted of a minimal word pair identification task with novel words spoken by a new speaker (test of generalization 1), and with novel words produced by one of the speakers used in creating the training stimuli (test of generalization 2). Control subjects performed the pretest, post-test, and two tests of generalization; however, these subjects did not go through the perceptual identification training program. For the control group, the time lag between the pretest and post-test phase was equal to the time it took for the trained subjects to participate in the entire 45-session training program (i.e., 3–4 weeks).

All perception training and testing was carried out at ATR Human Information Processing Research Laboratories in Kyoto, Japan. For all four perception tests (pretest, post-test, two tests of generalization) the same two-alternative forced choice minimal word pair identification procedure was used. Subjects were tested individually in a sound-treated room where they sat in a cubicle equipped with headphones (STAX-SR-Lambda Signature) and a NeXT workstation. Each trial began with a 500-ms presentation on the computer monitor of the standard English orthographies for an /r/–/l/ minimal pair. One member of the minimal pair appeared in the lower left corner of the screen; the other appeared in the lower right corner. The spoken test word was then presented at a comfortable listening level through the subjects' headphones. Subjects had 10 s to respond by pressing "1" to identify the spoken word as the orthographic word on the left of the screen, or "2" for the word on the right of the screen. For half the trials, a response of "1" corresponded to an /r/ identification label and a response of "2" corresponded to an /l/ identification label; for the other half of the trials, the order of identification labels was reversed. During the training trials, feedback was given in the form of a chime signaling a correct response and a buzzer signaling an incorrect response. After the buzzer for an incorrect response, the test word was repeated. As an additional motivation, each correct response received a 1 yen (approximately 1 cent) reward over and above the regular subject payment. There was no feedback in the pretest, post-test, or tests of generalization.

3. Stimuli—A large digital database of spoken words for the perception tests was originally recorded and compiled in the Speech Research Laboratory at Indiana University (see Logan *et al.*, 1991 for additional details). All stimuli were recorded in an IAC sound-attenuated booth. The utterances were low-pass filtered at 4.8 kHz and digitized at 10 kHz using a 12-bit analog-to-digital converter. The waveform files were then equated for rms amplitude using software developed in the Speech Research Laboratory. The files were then digitally transferred to ATR Human Information Processing Research Laboratories where they were upsampled to 22.05 kHz and rescaled to 16-bit resolution for presentation on the NeXT workstations.

The pretest and post-test stimuli were the same words as those used by Strange and Dittmann (1984). This set of stimuli consisted of 16 minimal pairs that contrasted /r/ and /l/ in four phonetic environments: initial singleton, initial cluster, intervocalic, and final cluster. There were also four minimal pairs that contrasted other English phonemes. These stimuli were recorded by a male speaker of general American English. The training stimuli consisted of 68 minimal pairs that contrasted /r/ and /l/ in five phonetic environments: 12 initial singleton pairs, 25 initial cluster pairs, 5 intervocalic pairs, 15 final singleton pairs, and 11 final cluster pairs. These stimuli were recorded by 5 speakers of general American English (3 males and 2 females). The stimuli for the first test of generalization consisted of an additional 96 words with /r/ or /l/ in five phonetic environments spoken by a new talker (i.e., new words by a new talker). The stimuli for the second test of generalization consisted of an additional 99 /r/–/l/ words (5 environments) spoken by one of the talkers in the training set (i.e., new words by an old talker).

B. Results of Perceptual Learning

Figure 1 shows the results of perceptual identification training for the experimental (left panel) and the control (right panel) groups. This figure displays the percentage of correct identifications for all four of the perceptual tests: the pretest, post-test, and the two tests of generalization. As shown in the left panel, the experimental (trained) group of subjects showed an improvement in their identification scores from pretest (65% correct identification) to post-test (81% correct identification), and this increase in performance was maintained for the two tests of generalization (83% and 80% correct identification for gen1 and gen2, respectively). Thus, on average, the trained subjects showed substantial gains in /r/–/l/ identification accuracy (16 percentage points). Nevertheless, it is important to note that this post-test level of identification accuracy is still substantially poorer than the near-perfect identification accuracy achieved by native English speakers.

A two-factor analysis of variance (ANOVA) with group (trained, control) and test (pre, post, gen1, gen2) as factors showed a significant main effect of group [$F(1,84) = 52.258$, $p < 0.0001$], and a significant group × test interaction [$F(3,84) = 5.136$, $p = 0.0026$]. *Post hoc* pairwise comparisons (Fisher's PLSD) showed no difference between the trained and control groups' pretest accuracies. However, the trained group performed significantly better than the control group ($p < 0.001$) on the post-test, as well as on both generalization tests. Furthermore, there was a significant improvement for the trained group from pretest to post-test ($p = 0.0074$), pretest to gen1 ($p = 0.0037$), and pretest to gen2 ($p = 0.0141$). There was no difference between the post-test and either of the tests of generalization scores for the trained group, and no difference between the scores on all four tests for the control group.

In order to gain some insight into the perceptual reorganization that resulted from the perceptual learning task, we examined the pre- and post-test identification accuracies for the /r/ and /l/ stimuli separately for the trained subjects (see Fig. 2). A three-factor repeated measures ANOVA was performed with test (pre or post) as the repeated measure, and phoneme (/r/ or /l/) and environment (four levels) as within-groups factors.

This analysis revealed three main findings. First, the main effect of test was highly significant [$F(1,80) = 40.369$, $p < 0.0001$], due to the overall improvement in identification accuracy from pretest to post-test. Second, the main effect of phoneme was also significant [$F(1,80) = 4.215$, $p = 0.043$], with /r/ being generally more accurately identified than /l/. Finally, there was a significant interaction between test and phoneme [$F(1,80) = 5.644$, $p = 0.012$]: /l/ tokens showed more overall improvement from pretest to post-test than /r/ tokens. This change from an asymmetrical distribution of /r/–/l/ identification accuracy at pretest to a more symmetrical distribution at post-test suggests that, after training, subjects show signs of developing perceptual categories that correspond more closely to the target English /r/ and /l/ categories.

There was also a main effect of phonetic environment [$F(3,80) = 4.603$, $p < 0.005$], indicating that the accuracy for the various phonetic environments increased from initial cluster position to medial to initial singleton to final. This dependency on phonetic context replicates earlier findings reported by Gillette (1980) and Mochizuki (1981), as well as Sheldon and Strange (1982). The interaction between phoneme and environment was significant [$F(3,80) = 4.647$, $p < 0.005$], due to the high identification accuracy of /l/ in initial cluster position relative to /r/ in that environment. Finally, the three-way interaction (test × phoneme × environment) [$F(3,80) = 7.272$, $p < 0.0002$] was also significant indicating that the perceptual learning of /r/ and /l/ was highly context-dependent.

In summary, these perceptual learning data provide a further replication of the findings of our earlier studies that used the high-variability perceptual training procedure to modify the acquisition of the English /r/–/l/ contrast by adult Japanese speakers (Logan *et al.*, 1991; Lively *et al.*, 1993; Lively *et al.*, 1994; Yamada, 1993). Having demonstrated significant perceptual learning for these Japanese adults, we now turn to the main concern of this study, that is, an assessment of the effects of perceptual learning on the production of English /r/–/l/ minimal word pairs by these subjects. Our goal here was to assess the extent to which the phonetic knowledge acquired during perceptual learning transferred to the production domain.

II. EFFECTS OF PERCEPTUAL LEARNING ON SPEECH PRODUCTION

A. Japanese /r/–/l/ productions

1. Procedure—During the pretest and post-test phases, audio recordings were made of the Japanese subjects producing English words that contrast /r/ and /l/. The pre- and post-test recordings were made directly before and after the perception pretest and post-test, respectively. The speech production task used a repetition procedure in which the subject read a set of English /r/–/l/ minimal pairs from a list of randomly ordered words. The subjects were provided with both visual and auditory prompts. The visual prompts consisted of a list of words written in standard English orthography. The auditory prompts consisted of a digital recording of the words spoken by a male speaker of general American English. This speaker was not one of the speakers that produced the stimuli for the perceptual identification tasks. The purpose of the auditory prompt was to provide the speakers with a model of how to pronounce the entire word aside from the target /r/ or /l/ segments. Because Japanese subjects have great difficulty with the /r/–/l/ perceptual contrast, we assumed that the subjects would not simply imitate the auditory model without relying on the printed word to inform them whether the word was the /r/ or /l/ member of the minimal pair. In this way, this task was not simply a direct imitation task, but rather was a repetition task that was mediated by linguistic intention. The printed prompts provided the subjects with the abstract linguistic-phonetic information they needed to guide their productions of the /r/ and /l/ phonemes, and the auditory prompts helped ensure that the rest of the words remained relatively stable across subjects. Since the main concern of this study was the change in /r/ and /l/ production due to perceptual identification training, this task was deemed appropriate for a pre- and post-test measure of production ability. Nevertheless, it is important to note that this production task was not a measure of spontaneous speech production. It remains for future research to determine the relationship between pre- and post-test performance on our repetition task and on more naturalistic measures of production ability.

The recordings were made in an anechoic chamber at ATR Human Information Processing Research Laboratories. The recordings were digitized at a sampling rate of 22.05 kHz with 16-bit resolution through DAT (Sony PCM-2500 or 2600) and a DAT interface, DAT-Link+ (Townshend Computer Tools., Inc.). The speech files were stored on the hard disk of a Sun Sparc workstation and were then digitally transferred to the Speech Research Laboratory at Indiana University where they were rescaled to 12-bit resolution for later presentation to native speakers of English using a PDP-11 laboratory computer.

2. Stimuli—The stimuli obtained from the pretest and post-test recordings consisted of 55 English words containing /r/ and /l/, giving a total of 110 words. These stimuli for the production tests included /r/–/l/ minimal pairs with the target phoneme in seven phonetic environments. The breakdown of the word-pairs by phonetic environment was as follows: 10 with initial singletons, 10 with initial clusters, 10 with medial singletons, 10 with final singletons, 10 with final clusters, 2 with medial clusters, and 3 with initial triple clusters (e.g., "splint-sprint"). Of these minimal pairs, half in each of the first five environments listed above came from the set of minimal pairs that were included in the earlier perceptual training stimuli, the other half were "new," that is, they were not used in any of the perception tests. None of the perceptual stimulus sets included minimal pairs with /r/ and /l/ in the last two environments listed above: these were all "new" word pairs for our subjects.

B. American English listeners' preference judgments

To assess the transfer of the Japanese trainees' perceptual learning to production, a group of native speakers of American English performed a paired comparison task using each Japanese trainee's pretest and post-test productions. The purpose of this procedure was to assess whether native speakers of American English could reliably discriminate between the trainees' pre- and post-test productions. If the perceptual learning procedure is effective in producing changes in control over speech production, then the native English listeners should display a consistent preference for the post-test tokens over the pretest tokens. This paired-comparison method of judging the trainees' improvement in production was selected as an initial test because it was expected to be sensitive to small differences in articulation. Our rationale was that if the Japanese trainees' post-test productions were indeed reliably preferred over the pretest productions, then we would have a reason to submit the pre- and post-test productions to additional tests of perceptual analysis. The initial paired-comparison task provides information about the degree and direction of change between the pretest and post-test tokens. The subsequent minimal pair identification task provides information about a change in speech intelligibility specifically related to improved /r/ and /l/ articulation.

1. Procedure—Each trial began with a visual presentation of the target word in standard English orthography centered on a CRT monitor. The listeners then heard a single Japanese trainee's pretest and post-test productions of this word over headphones. The two versions of the target word were separated by 500 ms of silence. The listeners then had to decide which version of the target word was "better," that is, which version was "a clearer and more intelligible pronunciation of the word shown on the screen." The judges responded on a seven-button response box which was labeled using a seven-point scale where "1" indicated that the first version was "much better" than the second version, "4" indicated no noticeable difference between the two versions, and "7" indicated that the second version was much better than the first. Each pair of utterances was presented twice: once with the pretest version first and the post-test version second, and once in the reverse order. There were 110 pre-post pairs in each of the two presentation orders, plus 10 practice trials at the start of the session, for a total of 230 trials per session. The initial practice trials were excluded from the final data analysis. Each listener judged the full set of pre- and post-test productions from a single Japanese speaker. Ten listeners were assigned to each of the 23 Japanese speakers. Because there were 11 Japanese trained subjects and 12 Japanese control subjects, a total of 230 native English speakers participated as subjects. No American listener judged more than one Japanese subjects' productions. In the final data analysis, the responses were recorded so that a response of "5" or higher always corresponded to a preference for the post-test version, and a response of "3" or lower always corresponded to a preference for the pretest version. This recoding simply takes into account the counterbalanced order of stimulus presentation.

2. Subjects—The American English listeners were all students at Indiana University. None reported any history of speech or hearing impairment at the time of testing, and all were monolingual native speakers of general American English. All received one hour of course credit for their participation.

3. Results—In the analysis of the data from the paired-comparison task, we examined the distribution of subjects' responses across the seven response categories for both the trained subjects (Fig. 3, left panel) and the control subjects (Fig. 3, right panel). The data shown in Fig. 3 are represented as proportions of the total number of responses from all listeners for each of the two groups of Japanese subjects.

As shown in Fig. 3, the distribution of ratings for the trained subjects' tokens (left panel) was skewed in favor of higher ratings, indicating a preference for the post-test tokens over the pretest tokens. In contrast, the native English speakers' ratings of the control subjects' tokens (right panel) were normally distributed across the seven response categories, indicating no systematic preference for either the pretest or the post-test tokens. The native English speakers' preference for the trained subjects' post-test productions over their pretest productions was confirmed by a highly significant chi square statistic, using the distribution of ratings for the control subjects' productions as the expected distribution [chi square = 1639.4, $p(6) < 0.001$]. This analysis was performed across all 24 200 trials for the trained subjects (220 trials × 11 trainees × 10 American listeners). Additionally, the Pearson coefficient of skewness for the trained subjects was negative (−0.527), indicating a greater median than mean; whereas for the control subjects, the mean and median were very close (Pearson coefficient of skewness = 0.090). Finally, the frequency of "post = pre" responses (response category 4) was lower for the trained subjects than for the control subjects [$t(21) = -7.297$, $p < 0.0001$], also indicating the increased discriminability of the trained subjects' pre- and post-test productions relative to those of the control subjects.

Taken together, these analyses of the preference data for the tokens produced by the trained and control Japanese subjects demonstrate reliable transfer of learning from perception to production for the trained subjects. Native speakers of English were able to reliably detect an improvement from the pretest to the post-test productions for the trained subjects; whereas, no reliable difference was observed for the tokens produced by the control subjects. Given that native speakers can discriminate the trained subjects' pre- and post-test tokens, our next step was to assess the extent to which the improvement in these utterances was due to improved /r/ and /l/ articulation; that is, whether American English listeners can identify the post-test productions more accurately than the pretest productions in a minimal pair identification task.

C. American English listeners' identification data

1. Procedure—The procedure for the minimal pair identification task was closely modeled after the task that the Japanese trainees performed during perceptual testing and training. In each experimental session, English listeners identified the full set of pre- and post-test productions from a single Japanese trainee. Each trial began with the two members of an English /r/–/l/ minimal pair appearing in standard English orthography on a CRT monitor in front of the subjects. One member of the minimal pair produced by a Japanese trainee was then presented over headphones. The listeners identified the word by pushing the left button on a two-button response box to select the word on the left of the CRT monitor, or the right button for the word on the right. Within a single experimental session, the complete set of pre- and post-test production tokens from a single Japanese trainee was presented in random order with each word presented twice, once with the correct response as a left button and once with the correct response as a right button. This arrangement resulted in a total of 440 experimental trials plus 10 practice trials at the beginning of the session, for a total of 450 trials. The productions of each of the 11 Japanese trainees and each of the 12 Japanese control subjects was identified in this manner by an independent panel of 10 English listeners, for a total of 230 listeners (10 different listeners for each of the 23 Japanese subjects). No American listener judged more than one Japanese subjects' productions.

2. Subjects—The English listeners were recruited from the university community in Bloomington, Indiana. None reported any history of speech or hearing impairment at the time of testing, and all were monolingual speakers of general American English. All were paid for their participation.

3. Results—Figure 4 shows percent correct identification of tokens from the trained (left panel) and control (right panel) Japanese subjects' productions as judged by the English listeners. Each panel shows the pretest level of performance along with the identification accuracy of the Japanese productions at the post-test phase for the "old" words (words that were included in the perceptual training stimulus set) and for the "new" words (novel words that the Japanese subjects had not been exposed to in any of the perceptual identification tests). The data shown here are averaged across the five phonetic environments that were included in the perceptual training stimulus set. The remaining two phonetic environments that were included in the production pre- and post-test set (medial clusters and initial triple clusters) were omitted from this "old" versus "new" analysis because there were no "old" stimuli for these two environments.

As shown in Fig. 4, utterances from the trained subjects displayed significant improvement in identification from pretest to post-test. Moreover, this improvement was consistent across both the "old" and the "new" items. A one-factor repeated measures ANOVA showed a significant effect of test [$F(2,20) = 8.857$, $p = 0.0018$]. Paired *t* tests established a significant difference between pretest and "old" post-test items [$t(10) = -3.321$, $p = 0.0077$], between pretest and "new" post-test items [$t(10) = -2.809$, $p = 0.0185$], but no difference between "old" post-test and "new" post-test items [$t(10) = 1.705$, $p = 0.1189$]. In contrast, for the control subjects there was no difference in identification across pretest, "new" post-test or "old" post-test items. These data demonstrate that the identifiability of the Japanese trainees' productions in a two-alternative forced-choice task improved as a result of the perceptual training program, and that this improvement generalized to both "old" and "new" tokens.

Table I gives the identification accuracy scores averaged across all 11 trained subjects at pretest and at post-test broken down by phonetic environment and by phoneme (/r/ or /l/). A three-factor repeated measures ANOVA was performed with test (pre or post) as the repeated measure, and phoneme and phonetic environment as the within-groups factors. In this analysis, we found a highly significant main effect of the repeated measure factor (i.e., test) [$F(1,140) = 21.850$, $p < 0.001$] indicating a strong overall improvement in performance from pre- to post-test. There was also a highly significant main effect of phoneme [$F(1,140) = 19.951$, $p < 0.001$], due to the generally higher identification accuracy of the /r/ tokens relative to the /l/ tokens at both pre- and post-test. There was no main effect of phonetic environment. There were also no significant interactions between test and either phoneme or environment, indicating that the degree of improvement in speech production was consistent across these factors. Thus, the results show that the Japanese trainees' post-test productions were more accurately identified by native English listeners than their pretest productions. Additionally, these minimal-pair identification data indicate that, at both pretest and post-test, the /r/ tokens were more accurately identified than the /l/ tokens.

The results of the two production evaluation tests (the paired comparison and the minimal pair identification task) clearly demonstrate significant improvements in the Japanese trainees' productions of /r/ and /l/ as a result of perceptual learning. The English listeners consistently judged the post-test utterances to be "better" tokens than the pre-test tokens, and they were more accurate in identifying the post-test tokens in an /r/–/l/ minimal-pair identification task. Furthermore, this improvement in production was robust in that it occurred across a variety of phonetic environments and it even generalized to novel words, i.e., words that the trainees had not been exposed to at all during perceptual learning. In contrast, the control subjects' productions showed no evidence of any change or improvement across any of these conditions. Having established that the perception training was effective in facilitating improvements in speech production, we now turn to an examination of the relationship between perception and production for individual subjects.

III. RELATION BETWEEN PERCEPTION AND PRODUCTION

Figure 5 displays the amount of learning observed in perception and production for each of the 11 Japanese trainees. This figure shows a "perception–production space" where the x axis represents each trainee's accuracy in perceptual identification of /r/ and /l/ minimal pairs, and the y axis represents accuracy in the identification of each trainee's productions by American English listeners. Thus, each trainee's performance is represented by a vector whose starting point corresponds to the trainee's pretest performance, and whose ending point corresponds to the trainee's post-test performance within this space. The group mean performance is indicated by the bold arrow, and the diagonal is the hypothesized vector that would indicate a perfect correlation between perception and production. The individual subjects' scores are given in Table II.

In perception, even though each subject showed some improvement from pretest to post-test, there was considerable individual variation across subjects in pretest accuracy, in post-test accuracy, as well as in the percentage change from pretest to post-test (see Table II). For instance, for the two subjects who performed well in the perception pretest (subjects 2 and 4), the training program was effective in enhancing their abilities to identify English /r/ and /l/ such that at post-test they approached native levels of performance. In contrast, for the poorest performer in the perception pretest (subject 7), the training program was only moderately effective. It is as if the two high performers used the training sessions to "fine-tune" an already well-defined preexisting two-way perceptual contrast. In contrast, even after 45 sessions of minimal-pair identification training, the poorest performer showed almost no evidence of learning to split a single perceptual category into two new categories. A striking individual difference that emerges from this perceptual identification data can be seen in the comparison between subjects 9 and 10. These two subjects performed at comparable levels at pretest; however, at post-test a difference of more than 20 percentage points was observed.

This wide range of individual performance is consistent with previous findings reported in other cross-language studies of /r/–/l/ perception (e.g., Goto, 1971; Mochizuki, 1981; MacKain *et al.*, 1981; Sheldon and Strange, 1982; Yamada *et al.*, 1994), nevertheless, is it still unclear what specific factors determine individual performance. Here we simply note the strong positive rank-order correlation (Spearman rho = 0.730, $p = 0.021$) between pretest level of performance and relative perceptual improvement, where relative perceptual improvement is defined as post-test accuracy minus pretest accuracy divided by 100 minus pretest accuracy. This is a measure of improvement as a proportion of the "room for improvement." This correlation indicates that pretest level of performance is a fairly good predictor of the effectiveness of the perceptual training program for individual subjects; however, as demonstrated by subjects 9 and 10, there are other factors at work here too.

We also observed considerable variation across individual subjects' production performance at pretest and post-test (Fig. 5 and Table II). Subjects 2 and 4 produced highly intelligible /r/'s and /l/'s at pretest, and therefore had very little room for any improvement to be observed in production. Several of the other subjects' pretest productions were identified at a level around 60% accuracy; however, there were also large differences in the degree of production improvement across these subjects. For example, the pretest productions of subjects 3 and 9 were identified at comparable levels of accuracy; however, subject 9's post-test productions were identified far more accurately than those of subject 3.

The results from the present investigation allowed us to extend these findings on individual variation in perception and production by looking at the relationship between changes in one domain (speech perception) and changes in the other domain (speech production). It is clear from the data shown in Fig. 5 and Table II that at the post-test phase, perception performance generally exceeded production performance. This is not surprising since the trainees had extensive training in speech perception, whereas there was no explicit training in speech production. Any improvement observed in speech production was a result of transfer of knowledge gained in perceptual learning to the production domain. By comparing the degrees of improvement in perception and production across the individual trainees, we can obtain additional insight into the underlying basis for the learning in the two domains. In this analysis, we examined the rank-order correlation between improvement in perception and production across all subjects to test the hypothesis that subjects who show the most perceptual learning also show the most improvement in production. However, as shown in the figure, there is no such correlation (Spearman rho = 0.202, $p = 0.522$). In other words, it is not the case that improvement in perception and production proceeded in parallel within individual subjects. Rather, it appears that, although perceptual learning generally transferred to improved production of this non-native contrast, as indicated by the positive slope of the mean vector, the two processes proceeded at different rates within individual subjects.

A more detailed examination of specific subjects' data revealed two kinds of situations that led to the lack of correlation between degrees of learning in perception and production that we observed in the present data. The first is illustrated by a comparison of subjects 9 and 10. These two subjects performed at a similar low level of performance in the perception pretest; however, at post-test, subject 10 performed considerably better than subject 9. However, in production, subject 9 showed a larger change than subject 10. A possible explanation for this discrepancy between learning in perception and production is that subject 9 continued to focus on perceptual cues that are not relevant for /r/–/l/ identification throughout the perceptual training program. However, in the production post-test, this subject was able to implement cues that were effective for improved /r/–/l/ identification in a two-alternative forced-choice identification task with native English listeners. For example, this subject may have focused on durational cues rather than spectral cues, and these attributes were sufficient to signal an /r/–/l/ contrast in production but were ineffective for the perceptual identification of /r/ and /l/ by native English speakers. In other words, for this subject, an apparent improvement in production was, in fact, a result of inappropriate, yet consistent, perception and production.

A second situation that could lead to a lack of correlation between learning in perception and production is illustrated by a comparison of subjects 6 and 3. These subjects show that production improvement can vary across individuals, even when initial performance and the degree of learning in perception are comparable. These two subjects performed at similar levels at the pretest phase in both perception and production. They also showed similar degrees of improvement in perceptual accuracy. Nevertheless, subject 6's post-test productions were identified more accurately by English listeners than subject 3's post-test productions. In other words, although subjects 3 and 6 showed comparable gain in perception, the transfer of perceptual learning to production was more effective for subject 6 than it was for subject 3 within the time frame of the present study. It is conceivable that, given more time to acquire the motor skills required for accurate /r/ and /l/ articulation, subject 3 would begin to show production improvement that is comparable to that of subject 6.

In summary, our investigation into the relationship between learning in perception and changes in production within individual subjects showed three main results. First, we found considerable variation across subjects in initial performance in both perception and production. Second, we observed a link between perception and production to the extent that perceptual learning generally transferred to improved production. This is seen by the positive slope of the mean vector in Fig. 5. Finally, we found little correlation between degrees of learning in perception and production after training in perception, due to the wide range of individual variation in learning strategies. This is seen by the deviation of the mean vector from the diagonal in Fig. 5. Taken together these findings support the hypothesis that learning in perception and production are closely linked, since perceptual learning generally transferred to improvement in production. However, learning in the perceptual domain is not a necessary or sufficient condition for learning in the production domain: the processes of learning in the two domains appear to be distinct within individual subjects.

IV. GENERAL DISCUSSION

The main goal of this study was to investigate the effects of perceptual learning on the production of non-native phonetic contrasts, and in so doing, to provide new data concerning the relationship between speech perception and production. First, we replicated earlier findings regarding the effectiveness of the high-variability perceptual training program for the acquisition of the English /r/–/l/ perceptual contrast by Japanese adults. Then, we showed that the knowledge gained about the non-native contrast from perceptual learning transferred to production of English /r/–/l/ words by the Japanese trainees. This improvement in production using a high-variability training procedure in perception was revealed by the results of two separate, but complementary, perceptual evaluation procedures using American English listeners as judges. The initial direct-comparison test demonstrated that the words produced by the Japanese trainees improved, in a general sense, from pretraining to post-training. The second two-alternative forced-choice identification test then showed, in a more specific sense, that the improvement in production resulted in better /r/–/l/ minimal pair intelligibility. This result was obtained for seven of the eleven individual subjects. Of the remaining subjects, the lack of improvement in production was due to a ceiling effect (subjects 2 and 4), or to consistently poor performance in perception (subject 7). Only one subject showed substantial improvement in perception, but consistently poor production (subject 5). Finally, a close examination of the degrees of improvement in perception and production within the individual Japanese trainees showed that, although initial performance in perception and production are well correlated, there is substantial individual variation in the degrees of learning in the two domains.

Having observed transfer of perceptual learning to aspects of speech production, we can now speculate as to the mechanisms that are responsible for this transfer, and what this tells us about the relationship between speech perception and production. A first possible account, along the lines of auditory-acoustic theories of speech perception (e.g., Stevens and Blumstein, 1981; Diehl and Kluender, 1989), would suppose that the learning in production involves a mechanism by which articulatory commands are tuned to internal acoustic representations. By this view, perceptual learning leads to more accurate internal acoustic representations of the target speech sounds, and these improved representations function as acoustic templates that play an important role in monitoring the articulatory output. Thus, the learning in production occurs during production *per se*, that is, there is no change in the articulatory commands until they are actually activated during articulation. An alternative possibility, along the lines of the motor theory (e.g., Liberman *et al.*, 1967; Liberman and Mattingly, 1985; Liberman and Mattingly, 1989) would suppose that it is the articulatory commands that are modified during perceptual training. Under this view, the changes that result from perceptual training constitute permanent changes in the internal representation that is common to both perception and production. Thus, according to this view, learning results in changes in a single, domain-independent, phonetic representation, and as such, accounts for the apparent transfer of learning in perception to improvement in production. A direct-realist approach (e.g., Fowler, 1986; Best, 1995) would posit that during the perceptually oriented training, the trainees become perceptually attuned to the invariant gestural features of English /r/ and /l/. The subsequent transfer of this perceptual learning to production occurs as a result of the post-test productions being guided by the now more accurate, gesturally defined /r/ and /l/ phonetic categories. According to this account, the /r/ and /l/ phonetic categories were modified on the basis of input from the auditory mode, but the effects of their modification are observed in both the perceptual and the production domains due to their integrated communicative function.

The present data do not provide conclusive evidence to support any one of these theoretical accounts regarding the underlying mechanisms that facilitate the transfer of perceptual learning to speech production. However, the finding that the transfer of perceptual learning to improvement in production occurred in the absence of any explicit instruction in /r/–/l/ production leads us to believe that there is a unified, common mental representation that underlies both speech perception and speech production. This claim is consistent with both the motor theory and direct-realist approach, which suppose that units of speech perception and production are integrally defined in terms of articulatory gestures, and therefore that changes in the one domain have concomitant changes in the other. This view is not inconsistent with the notion that the modified acoustic-perceptual representations function as “output monitors” during speech production. However, the fact that most subjects showed some improvement immediately after the perceptually oriented training program seems to provide evidence against the idea that the changes in production occurred *only* during production *per se*. Nevertheless, due to the lack of correlation between the degrees of learning in the two domains, it appears that the specific motor commands necessary for improved /r/–/l/ production may be acquired at different rates for different subjects. This suggests that modification of an underlying perceptuomotor, phonetic representation is not sufficient on its own to result in corresponding modifications in speech production.

From an applied point of view, this study provides very encouraging new data regarding the acquisition of non-native speech contrasts in laboratory settings. Our findings show very clearly that the high-variability perceptual training procedure is robust: it is not only effective in training Japanese adults to perceive the English /r/–/l/ contrast, as we have shown in several previous studies, but this training program is also effective in improving the pronunciation of these non-native speech sounds without any explicit training or feedback in speech production. This result is consistent with the recent findings of Rochet (1995) who reported that a change in VOT categorization was accompanied by a change in VOT production for Mandarin speakers exposed to a synthetic French VOT continuum. Additionally, the present results are compatible with the recent findings on phonologically delayed children reported by Jamieson and Rvachew (Jamieson and Rvachew, 1992; Rvachew, 1994; Jamieson and Rvachew, 1994), which showed clear benefits of perception training in conjunction with traditional speech production therapy for that population.

The research by Jamieson and Rvachew as well as the present study focused on cases where the observed pretraining phoneme inventory is reduced relative to the non-native target inventory, and where difficulties in both perception and production are known to occur. Similarly, the research by Rochet focused on a case where the trainees' native categories differed phonetically from the target categories in both perception and production. This situation, where there is a clear match between perception and production characteristics, is likely to be the case where transfer of perceptual learning to speech production will be observed. Recent models, such as Best's "Perceptual Assimilation Model" (PAM) (Best *et al*., 1988; Best, 1994; Best, 1995) and Flege's "Speech Learning Model" (SLM) (Flege, 1987, 1992, 1995), provide theoretical frameworks within which we can consider further which non-native contrasts are likely to show transfer of perceptual learning to production and which ones will not.

Both of these models propose that non-native phoneme perception abilities can be explained, at least in part, by reference to the native phonetic space. For example, in the case of the perception of English /r/ and /l/ by Japanese speakers, the observed difficulties can be explained in these models by the fact that Japanese has no such contrast in its native inventory. The most similar native Japanese phoneme is /ɾ/, which is described as an alveolar flap or stop depending on the phonetic context. Thus, with respect to the native Japanese phoneme inventory, English /r/ and /l/ are equally categorizable as this Japanese phoneme, and the contrast is therefore not supported by the native system in either perception or production. Whereas Best's PAM model makes predictions about the initial difficulty that a given non-native contrast proposes to listeners from a given native language background, Flege's SLM also makes several specific predictions about the persistence of foreign-accented production of a non-native contrast. SLM predicts that as long as native categories subsume non-native categories, accurate perception and production of the target categories will be blocked. Thus, in SLM, it is assumed that improvement in speech production as a consequence of perceptual learning is due to a reorganization of the auditory-acoustic phonetic space which is the underlying system used for both speech perception and production. Thus, SLM would predict that, with respect to adult second-language learning, changes in perception will transfer to changes in production, and these changes will proceed in parallel. Although PAM is not a model of speech learning, it would make similar predictions regarding the transfer of perceptual learning to speech production. PAM would predict that as the listener becomes more "attuned" to the gestural constellation that characterizes a non-native phoneme, he/she should learn to produce the required gestures for the target phonetic segment.

Although both models account for the main findings of the present study showing transfer of perceptual learning to speech production, neither model includes a mechanism to account for the observed lack of correlation between degrees of learning in the two domains. This reflects the main focus of these models on the relationship between the pretraining and the target phonetic categories. However, the present data indicate that more comprehensive models of second-language phonetic acquisition will need to address some of the specific characteristics of the relationship between learning in perception and production, as well as their variability across individual subjects.

In conclusion, we would like to emphasize that our goal in this research was to develop new techniques for the modification of the structure of the trainee's phonetic system, and in so doing, to investigate the nature of the relationship between learning in speech perception and production. In developing the perceptual training program, rather than explicitly focusing the trainee's attention on the detailed physical attributes of the perception and production of the target contrast, our approach has been to present the trainee with many exemplars of the target categories so that he or she can learn to integrate the exemplars into a linguistically meaningful phonetic space. The present study replicated earlier studies that showed the effectiveness of this "high-variability" approach to the acquisition of a non-native perceptual contrast. More importantly, however, this study has now extended these results by demonstrating that the changes produced by this approach to non-native phoneme acquisition occur at a level beyond the perceptual domain, that is, the modification of phonetic perception and the knowledge gained from this domain transferred to promote changes in speech production and motor control of these phonetic contrasts.

Acknowledgments

We are grateful to Luis Hernandez and Takahiro Adachi for technical support, and to Fernando Vanegas for subject running. We also thank Bernard Rochet, Catherine Best, and Winifred Strange for many helpful comments and suggestions. This work was supported by NIDCD Training Grant No. DC-00012 and by NIDCD Research Grant No. DC-00111 to Indiana University.

References

Best CT, McRoberts GW, Sithole NM. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. J. Exp. Psychol. Hum. Percept. Perform. 1988; 14:345–360. [PubMed: 2971765]

Best, CT. Learning to perceive the sound pattern of English. In: Rovee-Collier, C.; Lipsitt, L., editors. Advances in Infancy Research. Vol. SR-114. Norwood, NJ: Ablex; 1994. p. 31-80.Also in Haskins Laboratories Status Report on Speech Research

Best, CT. A direct-realist view of cross-language speech perception. In: Strange, W., editor. Speech Perception and Linguistic Experience: Issues in Cross-Language Speech Research. Timonium, MD: York; 1995. p. 171-206.

Diehl RL, Kluender KR. On the objects of speech perception. Ecol. Psychol. 1989; 1:1–45.

Flege JE. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. J. Phonetics. 1987; 15:47–65.

Flege, JE. The production and perception of foreign language speech sounds. In: Winitz, H., editor. Human Communication and its Disorders: A Review. Norwood, NJ: Ablex; 1988. p. 224-401.

Flege, JE. Speech learning in a second language. In: Ferguson, C.; Menn, L.; Stoel-Gammon, C., editors. Phonological Development: Models, Research, and Implications. Timonium, MD: York; 1992. p. 565-604.

Flege, JE. Second language speech learning: Theory, findings and problems. In: Strange, W., editor. Speech Perception and Linguistic Experience: Issues in Cross-Language Research. Timonium, MD: York; 1995. p. 233-272.

Flege, JE.; Hillenbrand, J. Limits on phonetic accuracy in foreign language speech production. In: Ioup, G.; Weinberger, S., editors. Interlanguage Phonology: The Acquisition of a Second Language Sound System. Cambridge: Newbury House; 1987. p. 176-201.

Fowler CA. An event approach to the study of speech perception from a direct-realist perspective. J. Phonetics. 1986; 14:3–28.

Gillette S. Contextual variation in the perception of *L* and *R* by Japanese and Korean speakers. Minn. Papers Ling. Philos. Lang. 1980; 6:59–72.

Goto H. Auditory perception by normal Japanese adults of the sounds 'l' and 'r'. Neuropsychologia. 1971; 9:317–323. [PubMed: 5149302]

Jamieson DG, Rvachew S. Remediating speech production errors with sound identification training. J. Speech-Lang. Pathol. Audiol. 1992; 16:201–210.

Jamieson, DG.; Rvachew, S. Perception, production and training of new consonant contrasts in children with articulation disorders. Proceedings of the International Conference on Spoken Language Processing; Yokohama: Acoustical Society of Japan; 1994. p. 199-1202.

Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. Psychol. Rev. 1967; 74:431–461. [PubMed: 4170865]

Liberman AM, Mattingly IG. The motor theory of speech perception revised. Cognition. 1985; 21:1–36. [PubMed: 4075760]

Liberman AM, Mattingly IG. A specialization for speech perception. Science. 1989; 245:489–494. [PubMed: 2643163]

Lively SE, Logan JD, Pisoni DB. Training Japanese listeners to identify English /r/ ad /l/: II. the role of phonetic environment and talker variability in learning new perceptual categories. J. Acoust. Soc. Am. 1993; 94:1242–1255. [PubMed: 8408964]

Lively SE, Pisoni DB, Yamada RA, Tohkura Y, Yamada T. Training Japanese listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories. J. Acoust. Soc. Am. 1994; 96:2076–2087. [PubMed: 7963022]

Logan JD, Lively SE, Pisoni DB. Training Japanese listeners to identify English /r/ and /l/: A first report. J. Acoust. Soc. Am. 1991; 89:874–886. [PubMed: 2016438]

MacKain KS, Best CT, Strange W. Categorical perception of English /r/ and /l/ by Japanese bilinguals. Appl. Psycholinguistics. 1981; 2:369–390.

Mochizuki M. The identification of /r/ and /l/ in natural and synthesized speech. J. Phonetics. 1981; 9:283–303.

Rochet, BL. Perception and production of second-language speech sounds by adults. In: Strange, W., editor. Speech Perception and Linguistic Experience: Issues in Cross-Language Research. Timonium, MD: York; 1995. p. 379-410.

Rvachew S. Speech perception training can facilitate sound production learning. J. Speech Hearing Res. 1994; 37:347–357. [PubMed: 8028316]

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

Stevens, KN.; Blumstein, SE. The search for invariant acoustic correlates of phonetic features. In: Eimas, PD.; Miller, JL., editors. Perspectives on the Study of Speech. Hillsdale, NJ: Erlbaum; 1981. p. 1-38.

Strange W, Dittmann S. Effects of discrimination training on the perception of /r/–/l/ by Japanese adults learning English. Percept. Psychophys. 1984; 36:131–145. [PubMed: 6514522]

Tahta S, Wood M, Loewenthal K. Foreign accents: Factors relating to transfer of accent from the first language to a second language. Lang. Speech. 1981; 24:265–272.

Yamada RA. Effects of extended training on /r/ and /l/ identification by native speakers of Japanese. J. Acoust. Soc. Am. 1993; 93:2391(A).

Yamada, RA.; Strange, W.; Magnuson, JS.; Pruitt, JS.; Clarke, WD, III. The intelligibility of Japanese speakers' productions of American English /r/, /l/, and /w/, as evaluated by native speakers of American English. Proceedings of the International Conference of Spoken Language Processing; Yokohama: Acoustical Society of Japan; 1994. p. 2023-2026.

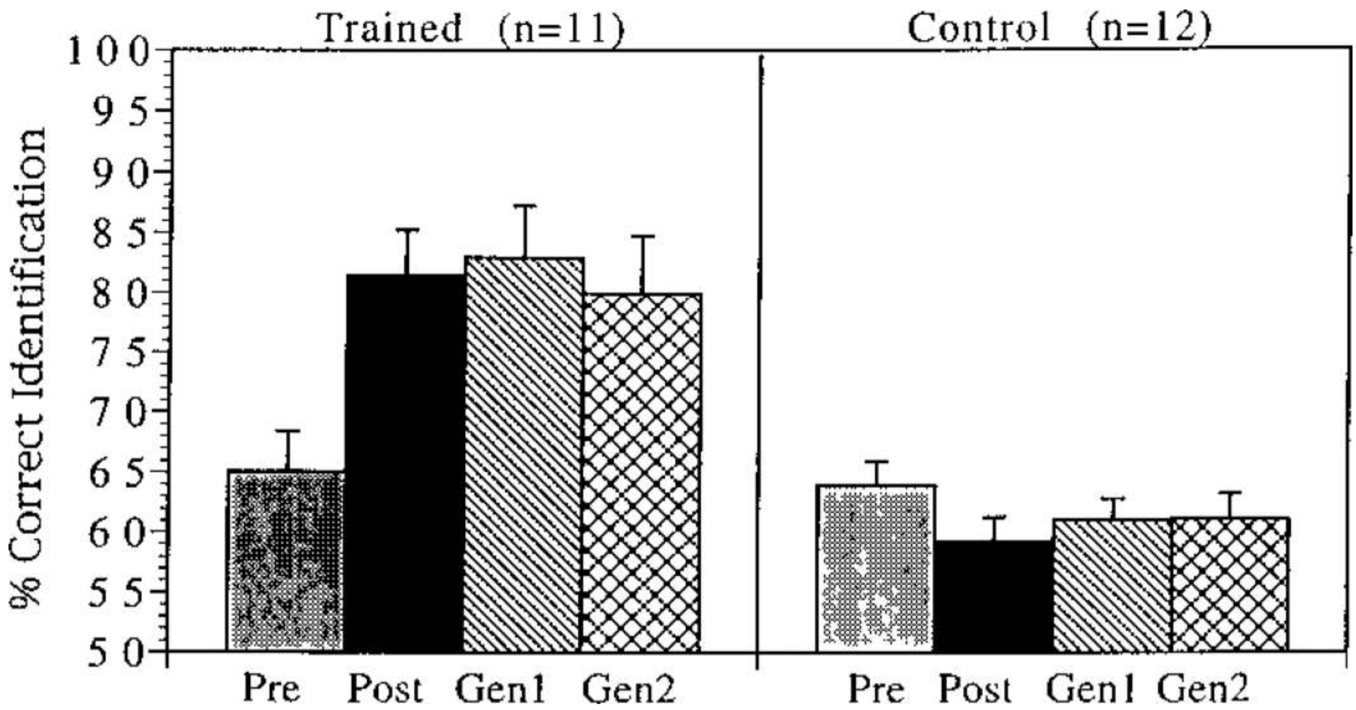

FIG. 1.
Percent correct perceptual identification performance for trained (left panel) and control (right panel) subjects at pretest, post-test, and the two tests of generalization. The error bars represent one standard error from the mean.

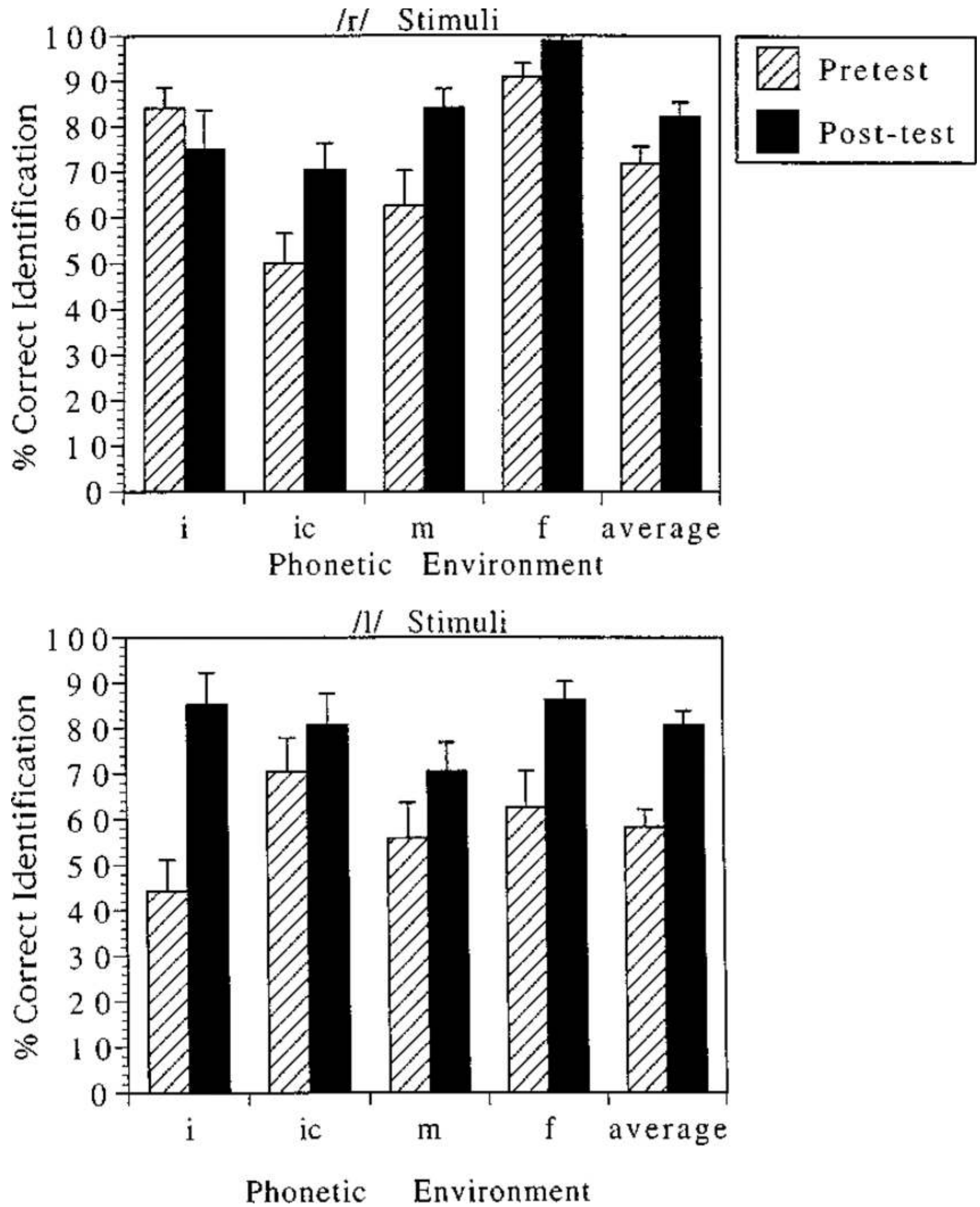

FIG. 2.

Distribution of the Japanese trainees' identification responses at pretest and at post-test by phonetic environment. The upper panel shows the /r/ stimuli; the lower panel shows the /l/ stimuli. The environments are: *i* = initial, *ic* = initial cluster, *m* = medial, *f* = final. The error bars represent one standard error from the mean.

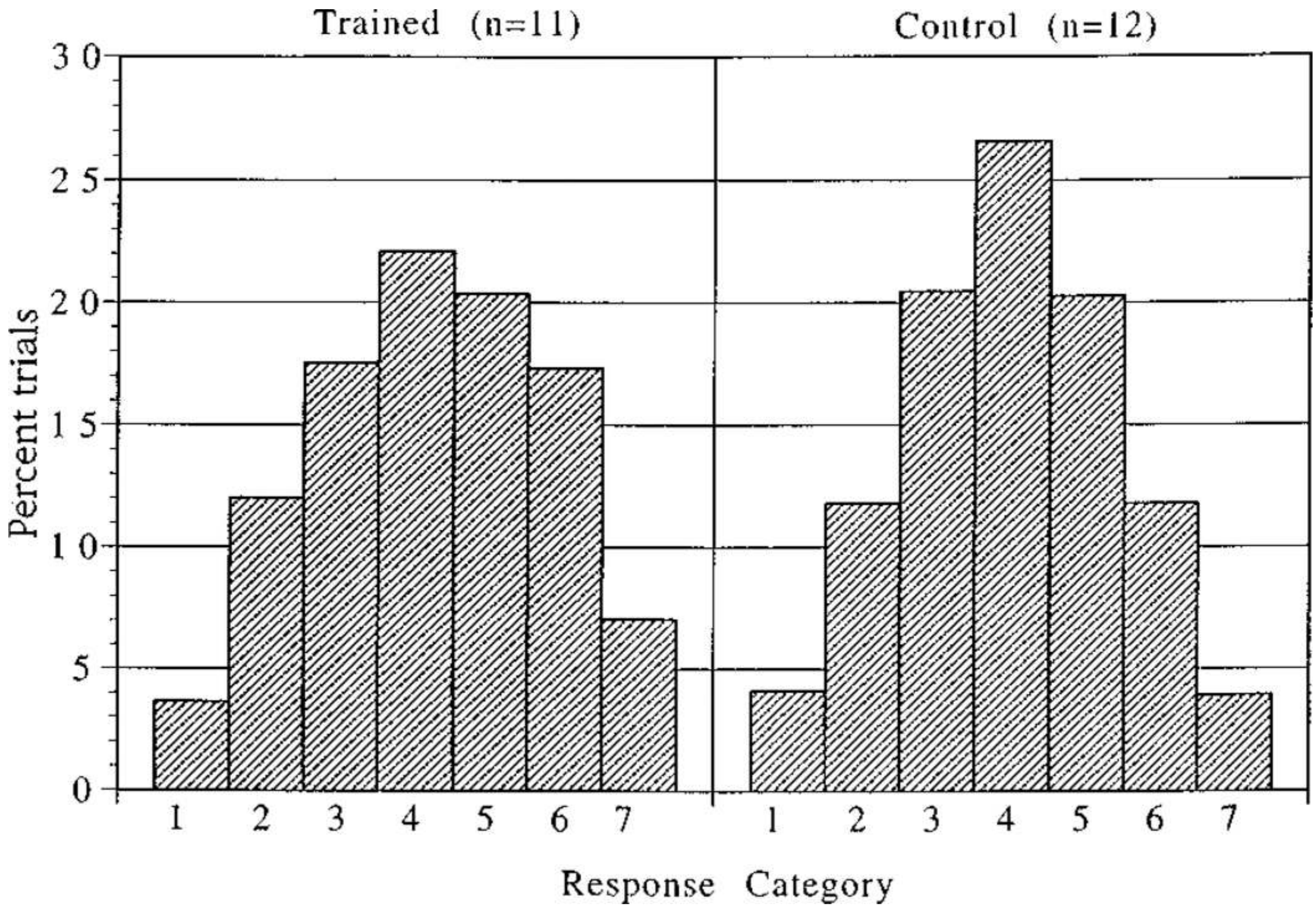

FIG. 3.

Distribution of preference ratings across all seven response categories by the American English listeners for the Japanese trained (left panel) and control (right panel) subjects' pre- and post-test productions. A response of "1" indicated that the pretest version was "much better" than the post-test version, "4" indicated no noticeable difference between the pre- and post-test versions, and "7" indicated that the post-test token was "much better" than the pretest version.

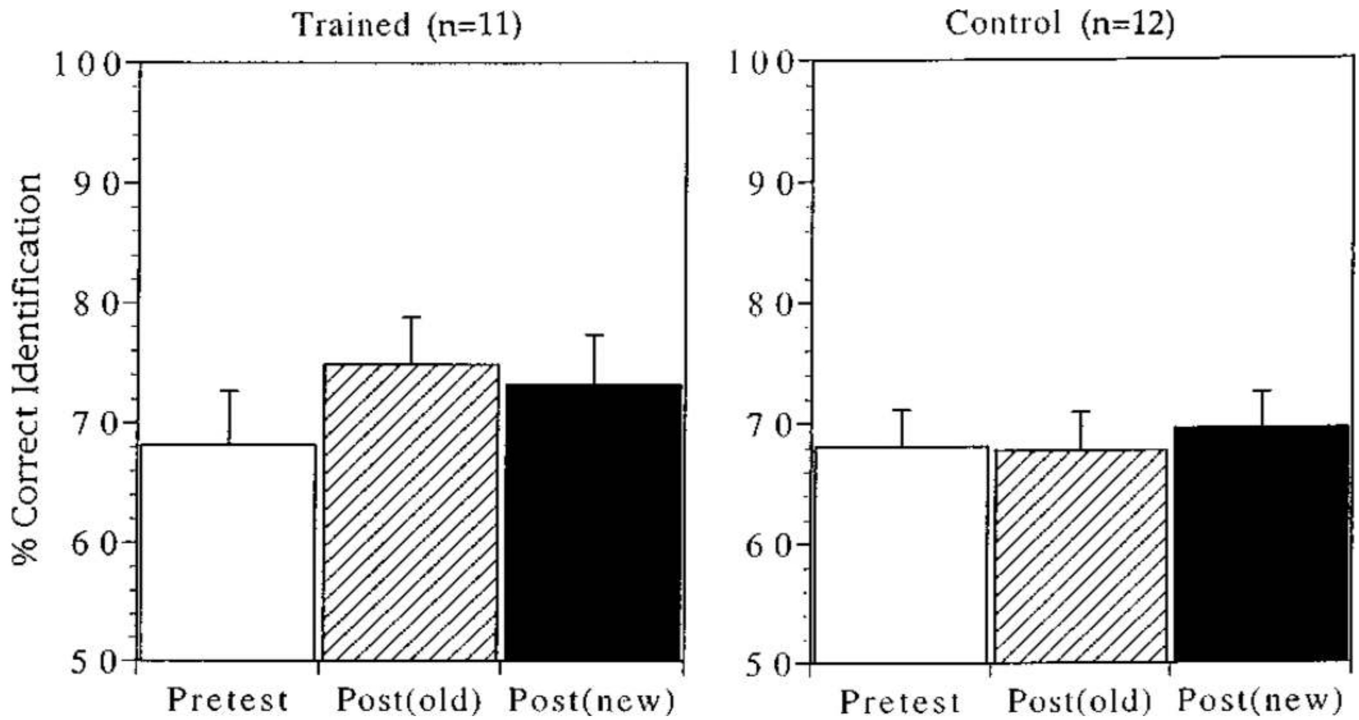

FIG. 4.

Percent correct performance for trained (left panel) and control (right panel) subjects' productions as judged by American English listeners in the minimal pair identification task. The open bar represents the full set of pretest tokens, the slashed bar represents the post-test tokens that were included in the perceptual training set, and the solid bar represents the post-tokens that were not included in the perceptual training set. The error bars represent one standard error from the mean.

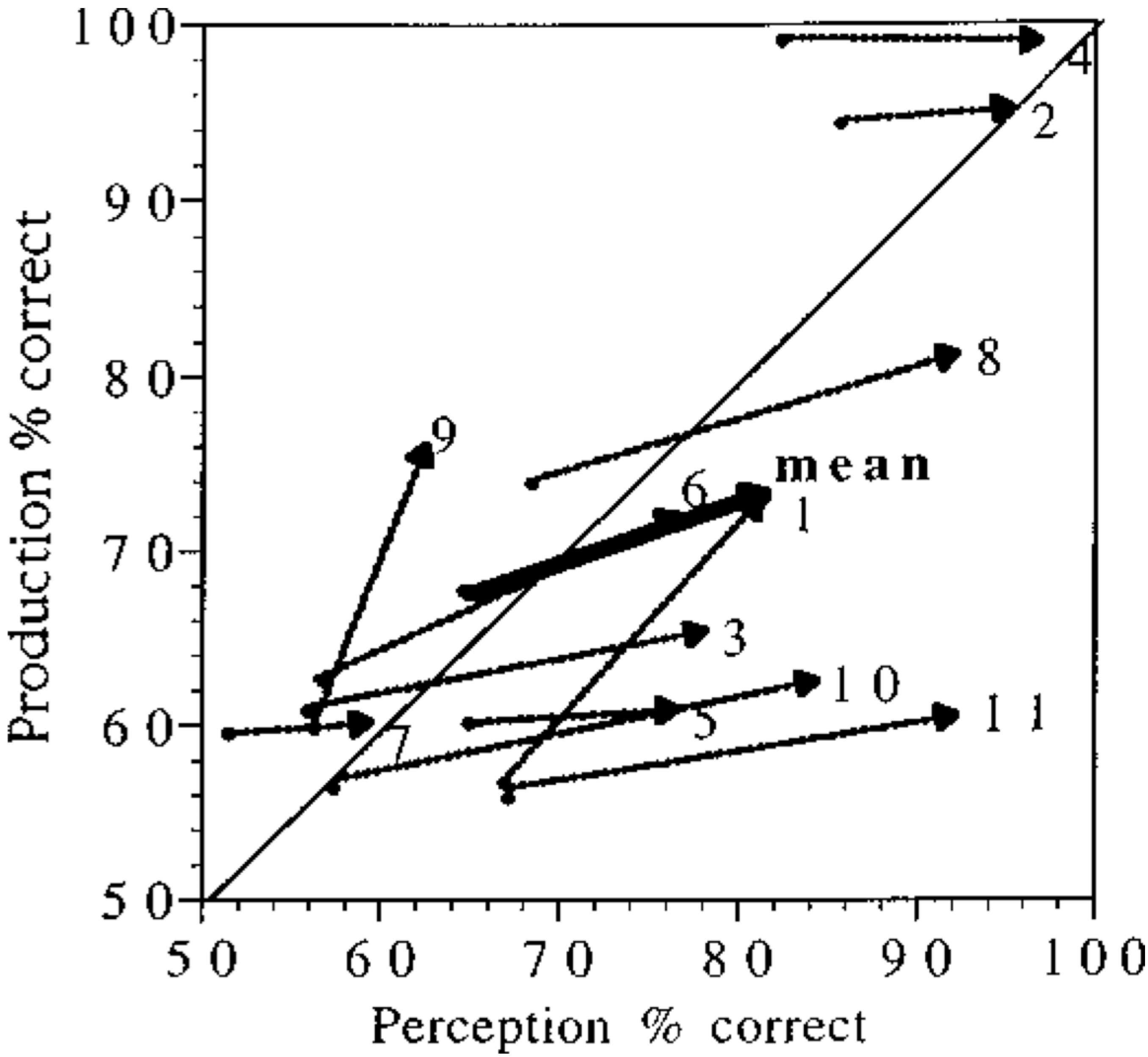

FIG. 5.
Vector plot of individual Japanese subjects' perceptual identification accuracy (*x* axis) and production identification accuracy (*y* axis) from pretest to post-test. Each individual subject's performance is indicated by a numbered vector. The mean performance is represented by the bold arrow. The diagonal represents the hypothetical vector location and orientation for a perfect correlation between perception and production.

TABLE I

Pretest and post-test identification accuracies by environment for the Japanese trainee productions as judged by American English listeners.

	/r/		/l/	
Environment	**Pretest**	**Post-test**	**Pretest**	**Post-test**
Initial	68.29	81.29	60.46	68.29
Initial cluster	72.83	77.08	49.33	54.92
Medial	74.83	77.50	48.00	57.96
Final	84.38	82.13	66.54	73.54
Final cluster	78.25	81.92	70.29	77.04
Initial triple cluster	76.26	84.03	43.47	49.18
Medial cluster	68.96	70.63	51.25	59.17
Totals	74.83	79.22	55.62	62.87

TABLE II

Individual Japanese trainee perception and production accuracy scores at pretest and at post-test. These data are averaged across /r/ and /l/, as well as across all phonetic environments.

	Perception			Production		
Trainee	**Pretest**	**Post-test**	**Difference**	**Pretest**	**Post-test**	**Difference**
1	67.19	81.25	14.06	57.18	73.00	15.82
2	85.94	95.31	9.37	94.59	95.18	0.59
3	56.25	78.13	21.88	61.18	65.41	4.23
4	82.81	96.88	14.07	99.18	98.95	−0.23
5	65.63	76.56	10.93	60.27	60.91	0.64
6	56.25	76.56	20.31	62.64	72.14	9.50
7	51.56	59.38	7.82	59.64	60.18	0.54
8	68.75	92.19	23.44	74.27	81.32	7.05
9	56.25	62.50	6.25	60.00	76.09	16.09
10	57.81	84.38	26.57	57.00	62.55	5.55
11	67.18	92.18	25.00	56.50	60.55	4.05
Totals	65.06	81.39	16.33	67.50	73.30	5.80