



Published in final edited form as:

J Acoust Soc Am. 1994 October ; 96(4): 2076–2087.

Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories

Scott E. Lively,

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 A. Yamada,

ATR Human Information Processing Research Laboratories, Kyoto, Japan

Yoh'ichi Tohkura, and

ATR Human Information Processing Research Laboratories, Kyoto, Japan

Tsuneo Yamada

Department of Behavioral Engineering, Faculty of Human Sciences, Osaka University, Osaka, Japan

Abstract

Monolingual speakers of Japanese were trained to identify English /r/ and /l/ using Logan *et al.*'s [J. Acoust. Soc. Am. **89**, 874–886 (1991)] high-variability training procedure. Subjects' performance improved from the pretest to the post-test and during the 3 weeks of training. Performance during training varied as a function of talker and phonetic environment.

Generalization accuracy to new words depended on the voice of the talker producing the /r/–/l/ contrast: Subjects were significantly more accurate when new words were produced by a familiar talker than when new words were produced by an unfamiliar talker. This difference could not be attributed to differences in intelligibility of the stimuli. Three and six months after the conclusion of training, subjects returned to the laboratory and were given the post-test and tests of generalization again. Performance was surprisingly good on each test after 3 months without any further training: Accuracy decreased only 2% from the post-test given at the end of training to the post-test given 3 months later. Similarly, no significant decrease in accuracy was observed for the tests of generalization. After 6 months without training, subjects' accuracy was still 4.5% above pretest levels. Performance on the tests of generalization did not decrease and significant differences were still observed between talkers. The present results suggest that the high-variability training paradigm encourages a long-term modification of listeners' phonetic perception. Changes in perception are brought about by shifts in selective attention to the acoustic cues that signal phonetic contrasts. These modifications in attention appear to be retrained over time, despite the fact that listeners are not exposed to the /r/–/l/ contrast in their native language environment.

INTRODUCTION

Recent studies on the perception of English /r/ and /l/ by native speakers of Japanese have demonstrated that simple laboratory procedures can be used to modify phonetic perception in adult listeners (Lively *et al.*, 1991, 1993; Logan *et al.*, 1991; Pisoni *et al.*, 1994). We chose to examine the /r/-/l/ contrast because it has been argued that contrasts based on spectral cues, such as /r/-/l/, may be much more difficult for listeners to acquire than contrasts based on temporal cues, such as distinctions using voice onset time (Strange and Dittmann, 1984; Strange and Jenkins, 1978). Thus, experiments that examine the acquisition of /r/ and /l/ address one of the more difficult test cases for modifying phonetic perception using laboratory-based training procedures. In the present investigation, we wanted to determine if the high-variability identification training procedure first used by Logan *et al.* would be effective with monolingual speakers of Japanese. Furthermore, we wanted to know if laboratory-induced modifications in phonetic perception were retained over time.

Because this study attempts to extend the findings of Logan *et al.* (1991), we briefly review the assumptions, methodology, results, and limitations of their earlier study. Logan *et al.* trained and tested six Japanese listeners who were living in an English-speaking environment using a pretest–post-test design. Generalization to new words and a new talker was also assessed immediately after the conclusion of training. Subjects responded in a two-alternative forced-choice identification task throughout the experiment. Immediate feedback was given to subjects only during the training phase. The minimal uncertainty of the two-alternative forced-choice procedure, combined with the use of immediate feedback, was assumed to promote the formation of new phonetic categories that are robust across non-contrastive sources of stimulus variability, such as changes in voice, speaking rate, and ambient listening conditions (Jamieson and Morosan, 1986, 1989). The identification training procedure encouraged subjects to group similar objects into the same category and different objects into different categories (see Lane, 1965, 1969). During training, listeners heard tokens from five talkers who produced English words with /r/ and /l/ in five phonetic environments. The high variability of the stimulus set was assumed to be important for developing perceptual constancy (Kuhl, 1983) and was thought to provide a broad base for generalization to new items and new talkers (Posner and Keele, 1968, 1970).

Several aspects of the Logan *et al.* results motivated the present investigation. First, significant increases in identification accuracy were observed from the pretest to the post-test. It should be noted that Logan *et al.* used the same words in the pretest and post-test that Strange and Dittmann (1984) employed in their earlier study. In Strange and Dittmann's study, Japanese subjects were trained in an AX fixed-standard discrimination paradigm. Although listeners improved during training in their ability to *discriminate* synthetically produced tokens that contrasted /r/ and /l/, Strange and Dittmann failed to find any change in performance when generalization was tested with naturally produced speech stimuli. The success of Logan *et al.* in training subjects to perceive /r/ and /l/ in English words and Strange and Dittmann's failure to find any changes in pretest–post-test performance suggest that the high-variability identification paradigm using natural speech was more effective than the low-variability discrimination paradigm using synthetic speech.

Second, significant increases in accuracy and decreases in response time were also obtained during training. Listeners' accuracy increased approximately 5% and their response latencies decreased by 125 ms for /r/'s and /l/'s in final singleton position and final consonant clusters over the course of the 3-week training period. These results indicate that the high variability of the training set, combined with the identification procedure, was effective in modifying non-native listeners' phonetic perception in a short period of time. However, given that the improvements were modest and that subjects had some exposure to

English because they were living in an English-speaking environment at the time of testing, we were interested in testing a monolingual group of subjects in the present experiment in order to generalize our earlier findings to a different population of listeners.

Third, we found that performance varied significantly as a function of phonetic environment. Tokens that contained /r/ and /l/ in final singleton position were identified most accurately, while /r/'s and /l/'s in initial consonant clusters were identified least accurately. Several other studies have reported a similar pattern of findings (Gillette, 1980; Goto, 1971; Lively *et al.*, 1993; Mochizuki, 1981; Sheldon and Strange, 1982; Strange and Dittmann, 1984). Durational and coarticulatory cues have been proposed as possible factors responsible for listeners' differential sensitivity to /r/ and /l/ in different phonetic environments (Dalston, 1975; Dissosway-Huff *et al.*, 1982; Henly and Sheldon, 1986). The durations of /r/ and /l/ are longest in final singleton position and shortest in initial consonant clusters (Lehiste, 1960), where accuracy is highest and lowest, respectively. Furthermore, /r/ and /l/ in final position tend to co-occur with the preceding vowel, while /r/ and /l/ in consonant clusters may be coarticulated with preceding consonants and may not reach their steady-state targets (Sheldon and Strange, 1982).¹

Fourth, Logan *et al.* found that performance during training varied as a function of the talker who produced the stimulus words. Tokens produced by one of the female talkers were identified significantly better than tokens produced by the other talkers. This finding was surprising because all of the training and testing materials were pretested with native speakers of English and all talkers were found to be equally intelligible. Thus, one of the goals of the present study was to replicate the finding of talker-specific effects in perceptual learning in order to determine if reliable differences do exist among our training talkers for a group of monolingual Japanese listeners.

Finally, Logan *et al.* found only a marginal difference in identification accuracy during the tests of generalization in which subjects had to respond to novel tokens produced by an old and new talker (old talker=83.7%, new talker =79.5%, $p<0.09$). One of the goals of a successful training paradigm would be to show that listeners transfer learning from items presented during training to new words produced by new voices. However, evidence of transfer can be obscured by differences in intelligibility between the talkers used to test generalization. In the present investigation, we dissociated possible differences in base line intelligibility among our talkers from transfer of training effects by pretesting the tokens used in the generalization tests with a control group of untrained monolingual Japanese listeners. If we can demonstrate that the talkers used in the transfer phase are equally intelligible, then we can rule out intelligibility differences as the source of the effects observed during generalization (Logan *et al.*, 1993; Pruitt, 1993).

On the basis of the findings reviewed above, Logan *et al.* argued that the minimal uncertainty of the closed-set identification task, combined with the high variability of the stimulus items, promoted the acquisition of the English phonemes /r/ and /l/ by Japanese listeners. Subjects showed modest increases in identification accuracy from the pretest to the post-test and during training. Decreases in response times during training were also

¹The similarity of the non-native contrast to phonetic distinctions found in the native language as well as the phonotactic constraints in the native language may also contribute to the difficulty of different phonetic environments (Best, 1994; Best and Strange, 1992; Flege, 1989, 1990; Flege and Wang, 1989; Mann, 1986; Polka, 1991, 1992). Japanese has an /r/ that is similar to English /d/ or /l/. It is produced either as a stop consonant or as a flap, depending on the vowel environment (Price, 1981; Yamada and Tohkura, 1992). In terms of phonotactic constraints, Japanese does not generally allow consonant clusters. Thus, in learning to perceive English /r/ and /l/ Japanese listeners have two disadvantages: first, they are being asked to discriminate between sounds that are not contrastive in their native language and are not similar to any known contrast in their language. Second, the /r/'s and /l/'s in initial consonant clusters occur in an unfamiliar phonotactic construction.

observed. These findings demonstrate that simple, laboratory-based training procedures can be effective in modifying listeners' phonetic perception in a short period of time. However, these changes appear to be specific to the voice of the talker producing the contrast and the phonetic environment in which it occurs.

Several limitations of the Logan *et al.* study can be identified (Logan *et al.*, 1993; Pruitt, 1993). First, Logan *et al.* trained and tested only six subjects. Thus, some of the effects that they observed may have been due to the small sample size used in the original study. Second, because of subject attrition, only three listeners were able to participate in the final tests of generalization to new tokens and to a new talker. This severely limited the possibility of observing significant differences between the two talkers used in the generalization test (see, however, Lively *et al.*, 1993). Finally, Logan *et al.* used subjects who had been living in the United States for several months. We assume they received some exposure to English outside of the laboratory.

Because of the theoretical importance of the Logan *et al.* results to issues in perceptual learning and development and the limitations of the original study noted above, we decided to carry out an extension of their training study with a larger number of monolingual Japanese listeners. We predicted that subjects would improve in their identification accuracy from the pretest to the post-test and during training. These findings would confirm the effectiveness of the high-variability training procedure. But more importantly, a replication would also establish once again the importance of methodological factors in speech perception studies and demonstrate that adult listeners have flexible perceptual capabilities that may not be shown by specific experimental procedures (Jenkins, 1979; Pisoni *et al.*, 1994; Pisoni and Lively, in press).

In addition to these considerations, we were also interested in retesting subjects several months after the completion of training to assess retention. If the changes observed during training are short term, then listeners' performance should return to pretest levels in the absence of any further training or feedback in their linguistic environment. This outcome might be expected because subjects in the present study were living in a monolingual Japanese speaking environment and would not be expected to have had much exposure to spoken English outside of the laboratory. However, if our training procedure produces long-term modifications in listeners' perception of /r/ and /l/, then we would predict very little change in performance from the post-test given at the end of training to follow-up tests given 3 and 6 months later. Thus, the results of the follow-up tests provide a way to measure the retention of the changes that we observed during training.

I. METHOD

A. Subjects

The subjects were 19 native speakers of Japanese (11 female, 8 male) living in Kyoto, Japan. Listeners ranged in age from 18 to 34. All of the subjects reported that they were monolingual speakers of Japanese and had never lived abroad. They had received some training in English grammar. However, conversation skills were not emphasized. No subjects reported any history of a speech or hearing disorder at the time of testing. A hearing screening performed at 25 dB HL for the frequencies 250 through 6000 Hz showed all subjects to have normal bilateral hearing acuity.

An additional group of 23 monolingual speakers of Japanese was recruited to serve in a control condition that assessed changes in performance from the pretest to the post-test without training. These subjects met the same selection criteria as the subjects who received

training. All subjects in the control condition had some training in English grammar, but had little or no production skills.

B. Stimuli

The stimulus materials were identical to those employed by Logan *et al.* (1991). A computerized database containing approximately 20 000 words (*Webster's Seventh Collegiate Dictionary*, 1967) was searched to locate all minimal pairs of words contrasting /r/ and /l/. A total of 207 minimal pairs were found. These words contrasted /r/ and /l/ in word-initial and final positions, in singleton and cluster environments, and in intervocalic position. Five talkers, three male and two female, recorded tokens of the words in an IAC sound-attenuated booth using an Electro-Voice D054 microphone. Talkers were given no special instructions concerning pronunciation of the words, which were presented individually in random order on a CRT monitor located inside the booth. The utterances were low-pass filtered at 4.8 kHz and digitized at 10 kHz using a 12-bit analog-to-digital converter at Indiana University. The digitized waveform files were then edited and equated for rms amplitude using a specialized signal processing package. Files were then digitally transferred to ATR laboratories, where they were upsampled at a sampling rate of 44.1 kHz and rescaled to a resolution of 16 bits.

The stimuli were originally pretested at Indiana University with a separate group of native speakers of English to assess their intelligibility (Logan *et al.*, 1991). An identification task was used in which listeners typed their responses on a computer terminal after hearing each stimulus. The criteria for including a word in the experiment were that it have no more than a 15% error rate across all listeners and that no errors were due to misperception of /r/ or /l/ (see footnote 2). After pretesting, a subset of 136 words (68 minimal pairs—12 initial singleton pairs, 25 initial cluster pairs, 5 intervocalic pairs, 15 final singleton pairs, and 11 final cluster pairs) from five talkers was selected for use in the training phase of the experiment.

The stimuli used in the tests of generalization were also identical to those used by Logan *et al.* (1991). Two sets of tokens were recorded. The first set consisted of 95 novel words produced by one of the female training talkers. In this set of stimuli, 37 words had /r/ or /l/ in initial singleton position, 32 had /r/ or /l/ in initial consonant clusters, 15 had /r/ or /l/ in final singleton position, and 11 had /r/ or /l/ in final consonant clusters. The second set of 93 novel items was produced by a new male native speaker of English. In this set of stimuli, 38 words had /r/ or /l/ in initial position, 29 had /r/ or /l/ in initial consonant clusters, 18 had /r/ or /l/ in final singleton position, and 8 had /r/ or /l/ in final consonant clusters.

Finally, the 24 minimal pairs used by Strange and Dittmann (1984) in the pretest–post-test phase of their experiment were recorded by a new male talker. Sixteen minimal pairs contrasted /r/ and /l/ in one of four phonetic environments (initial singleton, initial consonant cluster, intervocalic, final singleton). The remaining eight pairs contrasted phonemes other than /r/ and /l/. These items were processed in the same way as the other stimuli used in the present experiment.

C. Procedure

The experimental design employed a pretest–post-test procedure closely modeled after the methods used by Strange and Dittmann (1984) and Logan *et al.* (1991). In this design, the effects of training were assessed by comparing performance on a pretest and a post-test administered before and after a 3 week training period. We assessed retention of the new

²These data were reported earlier by Logan *et al.* (1991).

phonetic categories 3 and 6 months after the conclusion of training. Generalization to new words and a new voice was measured after training and during the two follow-up tests. A control group of native Japanese subjects who took only the pretest and post-test over the same 3-week interval was also included to insure that any changes observed in the performance of the experimental group could not be accounted for by repeated testing on the same items.

All testing and training was carried out at the ATR Human Information Processing Research Laboratories in Kyoto, Japan, and was administered individually in a quiet sound-treated room. Subjects sat at a cubicle that was equipped with a desk, a keyboard, and a CRT monitor. Stimuli were binaurally presented over headphones (STAX-SR-Lambda Signature) at a comfortable listening level. Presentation of stimuli, feedback, and collection of responses was controlled by a microcomputer (NeXT cube). During training and tests generalization, both identification responses and latencies were collected. Latencies were measured from the onset of the stimulus presentation to the subject's response. Feedback was given only during the training phase.

Before training began, subjects were given a pretest. This test consisted of 24 minimal pairs of words that were recorded onto digital audio tape using a DAT recorder (SONY DTC-1500ES). Two randomizations of the 48 words were recorded for a total of 96 trials. On each trial of the pretest, subjects were presented with an isolated word and were required to identify the stimulus by circling their response in an answer booklet. The same test items were presented after training (post-test phase) and again 3 and 6 months after the conclusion of training.³ The pretest–post-test procedure required approximately 20 min to complete.

The training phase also used a two-alternative identification task. On each trial, the two members of a minimal pair contrasting /r/ and /l/ were displayed in the lower left and right corners of the CRT for 500 ms prior to stimulus presentation. Subjects then heard a test word presented over their headphones. Responses were made by pressing a button on the keyboard: Subjects identified stimuli corresponding to words on the left side of the CRT by pressing “1” and words on the right side of the screen by pressing “2.” The position of the word containing /r/ appeared on the left side of the CRT; on the remaining trials, the word containing /r/ was on the right side. Listeners had a maximum of 10 s to respond. If no response was made, an error was scored.

Feedback was given on each trial during the training phase. If the listener responded correctly, a chime sounded and the next trial was presented 2 s later. If the subject made an error, a buzzer sounded and the stimulus word was repeated. Responses were made during the repetition phase and this procedure continued until the listener made a correct response. Correct responses and reaction times for responses made during repetitions were not included in the analysis of the data. Subjects were also shown a graphical representation of a coin on the CRT every time they made three correct responses. At the end of a training session, listeners were paid an additional bonus based on the level of their performance.

Stimuli from a set of 68 minimal pairs were each presented twice during a training session, yielding a total of 272 trials in each session. During each training session, stimuli from only one talker were presented. Subjects cycled through the set of five talkers used during training three times for a total of fifteen training sessions. Subjects were tested individually during training. Each session lasted approximately 40 min per day.

³A main effect for phonetic environment was also observed for subjects in the control condition [$F(3,66) = 58.95, p < 0.01$]. /r/'s and /l/'s in final position were identified most accurately, while /r/'s and /l/'s in initial consonant clusters were identified least accurately.

After the post-test, subjects were tested again to assess the degree to which training transferred to novel stimuli. The first test of generalization consisted of 93 novel words from minimal pairs contrasting /r/ and /l/ produced by a new talker (i.e., a talker not used in either the pretest–post-test phase or the training phase). A second test of generalization consisted of 95 novel words from minimal pairs contrasting /r/ and /l/ produced by talker 4, whom subjects had heard during training. The test stimuli consisted of new words that the subjects had not heard before. In both tests of generalization, the task was identical to the procedures used during training except that subjects did not receive any feedback. The tests of generalization were also administered individually. Both tests of generalization were repeated again 3 and 6 months after the conclusion of training.

II. RESULTS

A. Pretest–post-test

Mean accuracy scores for each subject from each test were submitted to an analysis of variance (ANOVA). Separate ANOVAs were conducted for subjects from the experimental and control groups. Phonetic environment and pretest versus post-test were within-subjects variables. *Post hoc* tests were conducted using Tukey's HSD procedure.

No main effect for time of test was observed for subjects in the control group [pretest: 66%, post-test: 63%, $F(1,22) = 2.01$, $p > 0.05$]. The interaction between time of test and phonetic environment did not approach significance [$F(3,66) < 1$]. This result demonstrates that any changes in performance from the pretest to the post-test observed in the experimental group cannot be accounted for by repeated testing on the same items. Because no differences in performance were observed for the control subjects over the 3-week interval between the pretest and the post-test, it was not necessary to retest these subjects at 3- and 6-month intervals. As a consequence, no further attention will be given to the subjects in this control condition.⁴

Figure 1 displays accuracy in the pretest and post-test as a function of phonetic environment for the trained listeners. Overall, subjects improved significantly in their ability to identify /r/ and /l/ from the pretest to the post-test [pretest: 65%, post-test: 77%, $F(1,18) = 92.99$, $p < 0.01$]. Individually, 18 of 19 subjects were more accurate during the post-test than during the pretest. Fourteen of the subjects showed an improvement of 8% or more. Accuracy varied widely as a function of phonetic environment [$F(3,54) = 37.13$, $p < 0.01$]: /r/'s and /l/'s in final singleton position were identified more accurately than /r/'s and /l/'s in initial singleton position and initial consonant clusters ($p < 0.05$). Target phonemes in initial singleton position were identified more accurately than phonemes in initial consonant clusters ($p < 0.05$).

B. Training

Separate ANOVAs were conducted on subjects' mean accuracy and response latency scores from each day of training. Week of training, talker, and phonetic environment were treated as within-subjects variables.

The main effects for week of training on response accuracy and latency demonstrate the effectiveness of the training procedures. As shown in Fig. 2, subjects' responses became significantly more accurate and faster during training [$F_{pc}(2,36) = 31.67$, $p < 0.01$; $F_{\pi}(2,36) = 23.57$, $p < 0.01$]. Increases in accuracy were localized between weeks 1 and 2 of

⁴The pretest–post-test procedure was under computer control for the 3-month and 6-month follow-up tests. On each trial of the follow-up test, subjects saw each member of a minimal pair presented on a CRT monitor and responded by pressing a button that corresponded to the word that they heard over their headphones.

training, although response latencies decreased significantly during all three weeks of training. All subjects showed at least a small increase in identification accuracy during training. The smallest increase was 3%, while the largest increase was 21%. Fourteen of the nineteen subjects improved by 8%–16% over the course of the 3-week training interval. Consistent decreases in response latencies between 180 and 1930 ms were also obtained. Only one subject failed to show a decrease in latency during training.

Figure 3 shows the effects of talker on response accuracy and latency. The effects of talker variability obtained earlier by Logan *et al.* (1991) and Lively *et al.* (1993) were replicated in the present experiment [$F_{pc}(4,72)=32.41, p<0.01$; $F_{rt}(4,72)=3.77, p<0.01$]. Tukey's HSD tests revealed that subjects responded more accurately to talker 4 than to any other talker used in training, except talker 5. Subjects also responded more accurately to talker 5 than to talker 1 or talker 2. Responses to stimulus tokens produced by talker 4 and talker 5 were significantly faster than responses to items produced by talker 1.

A main effect for phonetic environment was also observed in the accuracy and latency analyses [$F_{pc}(4,72)=153.60, p<0.01$; $f_{rt}(4,72)=22.22, p<0.01$]. As shown in Fig. 4, subjects' responses to targets in final singleton position were faster and more accurate than to targets in all other positions, /r/'s and /l/'s in final consonant clusters were identified more slowly and accurately than /r/'s and /l/'s in initial consonant clusters and intervocalic position. Finally, targets in initial singleton position were responded to more accurately than targets in initial consonant clusters. However, responses were faster to targets in initial consonant clusters than to phonemes in initial singleton position.

In addition to the main effects for week, talker, and phonetic environment, several interactions were also significant in the analyses of the accuracy and latency data. First, an interaction between talker and week was observed in the accuracy scores [$F(8,144) = 4.46, p<0.01$]. These findings are displayed in Fig. 5. Performance was significantly higher during week 3 of training than during week 1 for all talkers. However, significant increases were also obtained between weeks 1 and 2 for talkers 1, 2, and 3. In terms of differences among talkers as a function of week of training, during week 1, responses to words produced by talker 4 were significantly more accurate than to words produced by any of the remaining training talkers. In addition, tokens produced by talkers 3 and 5 were identified more accurately than tokens produced by talker 1 and 2. During week 2 of training, accuracy was higher to items produced by talker 4 than to any other talker. Finally, during week 3 of training, words produced by talker 4 were identified significantly more accurately than words produced by talkers 1 and 2. None of the remaining differences among talkers reached significance.

Second, a significant interaction between week of training and phonetic environment was also observed in the accuracy scores [$F_{pc}(16,288)=8.22, p>0.01$]. These data are displayed in Fig. 6. Accuracy increased in all phonetic environments from week 1 to week 2. However, significant increases were obtained only for /r/'s and /l/'s in initial singleton, initial consonant clusters, and intervocalic positions from week 2 to week 3.

Third, interactions between talker and phonetic environment were observed in both the accuracy and latency analyses [$F_{pc}(16,288)=8.22, p>0.01$; $F_{rt}(16,288) = 2.64, p<0.01$]. The top panel of Fig. 7 shows accuracy as a function of phonetic environment and talker, while the bottom panel displays the corresponding response times. In general, talker 4 was responded to more accurately than all other talkers when /r/ and /l/ occurred in initial singleton, initial consonant clusters, and intervocalic positions. No significant differences were observed among any of the talkers for targets in final consonant clusters and final singleton position. Talker 4 was also responded to faster than talker 1 when /r/ and /l/

occurred in initial singleton, initial consonant clusters, intervocalic, and final consonant cluster positions. Responses to talker 5 were also faster than to talker 1 when targets occurred in final consonant clusters and final singleton position.

Finally, a three-way interaction among week, talker, and phonetic environment was obtained in the response time data [$F(32,576) = 1.76, p < 0.01$]. Responses to talker 1 tended to be slowest during week 1 of training for /r/'s and /l/'s in final consonant clusters and final singleton position. By week 3, however, no significant differences were observed among any talkers for targets in final consonant clusters and final singleton position.

Overall, subjects' performance improved over the course of training: Accuracy increased while response times decreased. Performance varied both as a function of talker and phonetic environment. Accuracy was highest when subjects listened to one of the female talkers (talker 4). Listeners also identified /r/ and /l/ in word final singleton position most accurately. These findings replicate the major results obtained by Logan *et al.* (1991) during the training phase of their experiment.

C. Tests of generalization

Separate ANOVAs were conducted on the mean accuracy scores and response latencies collected during the tests of generalization. Talker and phonetic environment were treated as within-subjects variables in each analysis. A main effect for talker was obtained in the accuracy analyses [$F(1,18) = 12.51, p < 0.01$]. These results are shown in Fig. 8. Responses to items produced by the familiar talker were significantly more accurate than responses to tokens produced by the unfamiliar talker (82% vs 77%, respectively). The main effect for talker did not approach significance in the analysis of the response time data [$F(1,18) < 1$].

Main effects for phonetic environment were also observed in both the accuracy and latency analysis [$F_{pc}(3,54) = 48.03, p < 0.01$; $F_{rt}(3,54) = 9.05, p < 0.01$]. Targets in final consonant clusters and final singleton position were identified more accurately than targets in initial singleton position and initial consonant clusters. In addition, /r/'s and /l/'s in final singleton position were identified more accurately than /r/'s and /l/'s in final consonant clusters. Targets in initial singleton and final singleton positions were identified faster than targets in initial consonant clusters and final consonant clusters.

Because we observed significant differences in accuracy as a function of talker during the tests of generalization, it is important to determine if these results are due to differences in intelligibility or whether they are due to listeners encoding talker-specific information during training (Nygaard *et al.*, 1994). We tested these two possibilities by having 14 additional untrained Japanese listeners perform the two tests of generalization. These subjects were also monolingual native speakers of Japanese living in Kyoto who ranged in age from 24 to 37 years old. The same two-alternative forced-choice identification procedure was used and the order of presentation was counterbalanced across subjects. Mean accuracy and latency scores from each subject were submitted to separate ANOVAs. Talker and phonetic environment were within-subjects variables in each analysis. No main effect for talker was obtained in either the analysis of the accuracy data or the latency data [old talker: 71%, new talker: 70%, $F_{pc}(1,13) < 1$; old talker: 2971 ms, new talker: 2746 ms, $F_{rt}(1,13) < 1$]. The interaction of talker with phonetic environment was not significant in either analysis [$F_{pc}(3,39) = 1.32, p < 0.3$; $F_{rt}(3,39) < 1$]. Thus, these results, obtained with a control group of untrained listeners, rule out any baseline differences in intelligibility as the source of the differences in performance between the talkers in the tests of generalization and support the proposal that listeners have encoded talker-specific information during training.

D. Three-month tests

Three months after the conclusion of training, 16 of the original 19 subjects returned to the laboratory and performed the post-test and two tests of generalization again. A separate ANOVA was conducted on the accuracy data from these subjects to compare performance on the pretest and post-test and to assess retention after 3 months. Test and phonetic environment were within-subjects variables. Mean accuracy scores are presented as a function of test and phonetic environment for these subjects in Fig. 9. A significant main effect for test was obtained [$F(2,30) = 53.93, p < 0.01$]. Subjects' accuracy increases significantly from the pretest to the post-test, but did not decrease significantly from the post-test to the 3 month follow-up. All subjects' accuracy on the 3-month follow-up test was above pretest levels of performance and 12 subjects were at least 8% above their pretest levels. The main effect for phonetic environment was also significant [$F(3,45) = 38.14, p < 0.01$]. /r/'s and /l/'s in final singleton position were identified more accurately than targets in any other environment. In addition, targets in initial position were identified more accurately than targets in initial consonant clusters.

Several additional ANOVAs were conducted on the mean accuracy and latency scores collected from the 16 subjects who participated in the original tests of generalization and the follow-up tests of generalization | given 3 months after the conclusion of training. Time of test, talker, and phonetic environment were within-subjects variables. Mean accuracy scores are plotted as a function of each of these variables in Fig. 10. Original test scores for these 16 subjects are displayed in the left panel, while 3-month follow-up scores are displayed in the right panel.

Several results are noteworthy: First, the main effect for time of test was not significant in the analysis of the accuracy data or the response time data [$F_{pc}(1,15) = 1.60, p < 0.2$; $F_{rt}(1,15) < 1$]. Mean accuracy for the original tests of generalization was 79.4%, while mean accuracy for the generalization tests given 3 months later was 77.9%. Second, a main effect for talker was obtained in the analysis of the accuracy data [$F(1,15) = 11.80, p < 0.01$]. Subjects were more accurate in identifying /r/'s and /l/'s produced by a familiar talker than by an unfamiliar talker (81.0% vs 76.3%, respectively). Separate ANOVAs conducted on the data from the original tests of generalization and the follow-up tests given after 3 months replicated this pattern [$F_{orig}(1,15) = 9.08, p < 0.01$; $F_{3-month}(1,15) = 10.04, p < 0.01$]. Finally, main effects for phonetic environment were also obtained in the analysis of the accuracy and latency data [$F_{pc}(3,45) = 37.83, p < 0.01$; $F_{rt}(3,45) = 5.40, p < 0.01$]. Targets in final singleton position were identified more accurately than /r/'s and /l/'s in any other phonetic environment. Words containing /r/'s and /l/'s/ in final consonant clusters were also identified more accurately than words with /r/'s and /l/'s in initial singleton position or initial consonant clusters. Responses to phonemes in initial and final singleton position were significantly faster than responses to targets in initial consonant clusters and final consonant clusters.

E. Six-month tests

Six months after the conclusion of training, only eight subjects were able to return to the laboratory again to participate in the post-test and the tests of generalization. These subjects were a subset of the 16 who returned for the 3-month tests. Separate ANOVAs were conducted on the pretest-post-test data and the generalization data in order to assess changes in performance over time.

An ANOVA was conducted on mean pretest–post-test follow-up data from each subject. Time of test and phonetic environment were within-subjects variables. A main effect was observed for time of test [$F(3,21) = 6.04, p < 0.01$]. Accuracy on the original posttest and the

posttest given 3 months after the conclusion of training was significantly higher than accuracy on the pretest. After 6 months, accuracy was still 4.5% greater than pretest levels. Fisher's LSD test indicated that this difference was neither significantly greater than pretest levels of performance nor significantly less than post-test levels of performance. These results are displayed in Fig. 11. In terms of individual subjects, six subjects were more accurate on the follow-up test given after 6 months than on the pretest. A main effect for phonetic environment was also observed [$F(3,21) = 19.60, p < 0.01$]. Subjects were more accurate at identifying /r/ and /l/ when they occurred in final singleton position than in any other phonetic environment.

Additional ANOVAs were conducted on mean accuracy and latency scores from the eight subjects who participated in the original tests of generalization and the two follow-up tests. Time of test, talker, and phonetic environment were within-subjects variables. Mean accuracy scores are plotted as a function of these variables in Fig. 12. Original generalization scores are shown in the left panel, 3-month scores are displayed in the middle panel, and 6-month scores are presented in the right panel.

The main effect for time of test was not significant in the analysis of the accuracy data or the response times [$F_{pc}(2,14) < 1; F_{rt}(2,14) < 1$]. Performance for these eight subjects did not decrease significantly during the 6 months after the conclusion of training. This replicates the finding observed after 3 months without training.

The overall ANOVA revealed a main effect for talker in the accuracy data [$F(1,7) = 7.32, p < 0.05$]. Words produced by the familiar talker were identified more accurately than tokens produced by the unfamiliar talker. Separate ANOVAs were conducted on the accuracy data collected at the end of training, during the 3-month follow-up, and during the 6-month follow-up. In general, the results of these ANOVAs confirmed the overall pattern. During the original test, listeners were significantly more accurate with the familiar voice [$F(1,7) = 8.06, p < 0.05$]. At 3 months, the results showed a marginal advantage for the familiar talker [$F(1,7) = 5.89, p < 0.11$]. Finally, after 6 months without any additional training, a significant advantage was still observed for the familiar talker [$F(1,7) = 5.89, p < 0.05$]. These results suggest that familiarity with a talker's voice acquired during initial training facilitates the later recognition of new words spoken by the same talker (Nygaard *et al.*, 1994). This familiarity effect appears to extend over a period of at least 6 months.

In addition to the overall main effect for talker, a main effect for phonetic environment was also obtained in the analysis of the accuracy data [$F(3,21) = 20.88, p < 0.01$]. /r/'s and /l/'s in final singleton position and final consonant clusters were identified significantly more accurately than phonemes in any other environment. In addition, targets in final singleton position were identified more accurately than /r/'s and /l/'s in final consonant clusters.

III. DISCUSSION

One of the long-term goals of cross-language training experiments is to develop new perceptual categories for non-native phonemic contrasts that are robust and permanent. These new perceptual categories must be applicable to highly variable stimuli, as listeners routinely encounter the non-native phonetic contrast in many different phonetic environments. Furthermore, the categories must also be applicable to many different talkers and must be highly discriminable even under degraded listening conditions. The goal of the present investigation was to examine the effectiveness of a high-variability identification training procedure in helping monolingual speakers of Japanese to acquire and retain the English /r-/l/ contrast.

Several results from the present experiment and the earlier Logan *et al.* (1991) study demonstrate the effectiveness of the high-variability identification training paradigm. First, listeners' accuracy increased by 12% from the pretest to the post-test and by 11% during the 3 weeks of training. Logan *et al.* observed a more modest gain of 5%–7% from the pretest to the post-test and during training. In the present investigation, subjects' accuracy increased significantly during each week of training for the contrasts that occurred in the three most difficult phonetic environments (initial singleton, initial consonant clusters, and intervocalic positions).⁵ Mean levels of performance in the Logan *et al.* study were comparable to those reported here. These findings demonstrate that the present training procedure is effective in modifying listeners' phonetic perception and that these methods can be used with monolingual subjects. The magnitude of the training effect appears to be modulated by the amount of exposure subjects have to the contrast in the natural language-learning environment (see MacKain *et al.*, 1982).

Second, listeners retained information about the new phonetic contrast for a period of at least 6 months. Performance on the tests of generalization did not decrease significantly over the 6-month interval during which subjects received no further training or exposure to /r/ and /l/ and post-test performance remained 4.5% above baseline accuracy established during the pretest. These results were initially somewhat surprising to us, given that subjects were living in a monolingual Japanese-speaking environment and were not exposed to much spoken English during this time. The present findings demonstrate that our high-variability identification training procedure was effective in changing the long-term representations of these perceptual categories. However, knowledge gained during training was not completely retained over time, as shown by the decrease in accuracy on the post-test after 6 months without additional training (see also Werker and Tees, 1984).

In addition to acquiring and retaining information about the new phonetic categories, listeners also become attuned to the voices used during the training phase (Goto, 1971). Initially, large differences in accuracy were observed among the five training talkers. However, subjects' identification accuracy increased significantly during the training phase when responding to each of the talkers. By the end of training, performance was almost equivalent among all talkers. Only talker 4 was responded to more accurately than talker 1 and talker 2. Taken together, the present findings show that listeners not only acquire information about the acoustic-phonetic cues for phonetic categorization of the novel items, but they also acquire information about the “indexical” or “personal” characteristics of the talker's voice during perceptual learning (Laver, 1980; Laver and Trudgill, 1979; Nygaard *et al.* 1994).

The results from the tests of generalization indicate that talker-specific information acquired during training was used to facilitate the identification of new words spoken by a familiar talker, relative to the identification of new words produced by an unfamiliar talker. During the first tests of generalization, mean response accuracy was 77% for tokens from the unfamiliar talker versus 82% for the familiar talker. Furthermore, the results of the follow-up tests of generalization suggest that talker-specific details were retained over time: After 6 months without training, subjects were still more accurate at identifying /r/ and /l/ produced by an familiar talker than by an unfamiliar talker. The present results, taken together with our previous findings (Lively *et al.*, 1993; Logan *et al.*, 1991), suggests that accuracy during

⁵Lively *et al.* (1993) reported a similar interaction in their second experiment. In that experiment listeners were trained using the high-variability identification paradigm with tokens from only a single talker. Accuracy increased from week 1 to week 2 for /r/'s and /l/'s in an initial singleton and intervocalic positions and initial consonant clusters. Significant increases were also observed from week 2 to week 3 for /r/'s and /l/'s in initial consonant clusters.

the tests of generalization is dependent on previous experience and the amount and type of variability available in the stimulus set (see Goggin *et al.*, 1991).

Although our results indicate that listeners retained knowledge about both the new phonetic categories and the voices used in training, it is necessary to qualify these observations. When subjects were retested 3 and 6 months after the conclusion of training, they were given the same tests that they had previously received. Thus they listened to the pretest–post-test stimuli for the fourth time during the 6-month post-test and heard the stimuli from the tests of generalization for the third time. It is possible that subjects learned something about the specific tests themselves, even in the absence of feedback, and were therefore responding on the basis of stimulus-specific knowledge, rather than more abstract knowledge about the new phonetic contrast (Jacoby and Brooks, 1984).

Several observations about the results of the present experiment tend to discount this possibility. First, consider the methodology of the tests. Listeners were given feedback only during the training phase of the experiment and they would have no way to judge the accuracy of their responses from trial to trial during the pretest, post-test, or tests of generalization. Second, subjects who only received the pretest and post-test as a control condition actually showed a nonsignificant decrease in identification accuracy. This result would not be anticipated if listeners had encoded stimulus-specific information about the test items. Third if listeners had memorized individual stimulus items during the tests, performance should not have decreased during the post-test given 6 months after the conclusion of training. In future studies, new test items should be used whenever possible. However, it is worth emphasizing here that the words used in the pretest–post-test, training, and tests of generalization virtually exhaust all of the minimal pairs of words in English containing /r/ and /l/.

Taken together, the results of the present experiment suggest that the high-variability identification procedure is effective in modifying non-native listeners' phonetic perception. More importantly, however, this perceptual information was retained over time for at least a 6-month period. The results from this experiment raise two broad theoretical questions: First, what is the nature of the perceptual mechanisms that are responsible for this change in phonetic perception? Second, what factors contribute to listeners' long-term retention of the new categories, given that exposure to the contrast outside of the laboratory is minimal?

Several issues concerning the acquisition and retention of new phonetic categories can be addressed by considering recent proposals about the role of selective attention in categorization and perceptual learning (see Jusczyk, 1989, 1993, 1994; Nosofsky, 1986, 1987). According to Nosofsky, selective attention “stretches” and “shrinks” the listener's perceptual space during category acquisition. Representations are stretched along contrastive psychological dimensions so that items from different categories are made *less similar* to each other. Similarly, representations are shrunk along noncontrastive dimensions so that items from the same category are made more similar to each other.

Jusczyk (1993, 1994) has recently extended this basic framework to address specific issues of phonetic category acquisition in infants. He has proposed that acoustic dimensions are extracted from the incoming signal by a bank of auditory processing filters (Sawusch, 1986) and that these perceptually relevant dimensions are then weighed automatically in terms of their importance in determining linguistic contrasts (see also Kruschke, 1992). These “attention weights” are then used to generate a precompiled interpretative scheme that can be applied automatically to incoming fluent speech (Jusczyk, 1989, 1993; Klatt, 1979). A similar process may occur when Japanese listeners are trained to identify English /r/ and /l/ using the high-variability identification procedures. Over the course of training, selective

attention weights are changed to favor the new phonetic categories, which then causes a modification of listeners' phonological spaces or filters (Flege, 1989, 1990; Terbeek, 1977).

In addition to accounting for the acquisition of new phonetic contrasts, an explanation must also be provided for the results obtained on the retention tests of the present investigation. To accommodate these results, it is important to consider the nature of adult listeners' phonological systems. Within the framework of Jusczyk's model, adult listeners have well-developed selective attention weights that serve to maximize contrasts in their native language. As training progresses, we assume selective attention weights are modified to incorporate the new phonetic contrast (Kruschke, 1992). Simultaneously, however, the weights remains stable enough so that contrasts in the native language are not disrupted. When training ends, the newly established attention weights may decay slowly toward their initial states.

Two aspects of the linguistic environment suggest this possibility. First, the listeners in the present experiment were monolingual adults who already had well-established phonological systems capable of perceiving the contrasts of the native language. In contrast, infants who are in the process of acquiring their native language would not be expected to have a stable set of attention weights. Rather, their phonological systems are dynamically adapting to sound patterns in their native language environment (Aslin and Pisoni, 1980). Only after approximately 12 months of age would infants be expected to converge on a fixed set of attention weights (Werker, 1989; see, however Kuhl *et al.*, 1992).

Second, because our subjects were living in a monolingual Japanese-speaking environment, little exposure to the English /r/-/l/ contrast would be expected. Thus, these listeners would have little experience or feedback (either explicit or implicit) in perceiving the /r/-/l/ contrast. Because feedback is unavailable after the end of training, selective attention weights would have little impetus to remain stable or increase. Indeed, the lack of feedback may lead to a decay of the weights. When subjects are tested after 3 months and then 6 months later without training, they may try to employ the same interpretative scheme they had used during testing at the end of training. However, because attention weights decay over time, only a corrupted set of weights can be applied to the later test items. Thus, decreases in performance would be expected over time without additional training and feedback on the new contrast.

Two basic assumptions underlie this selective attention-based account of category acquisition and retention. First, we assume that selective attention weights were changed during training to favor the acquisition of a new phonetic contrast. Second, we assume that the new weights decay slowly over time because little input or feedback from the environment was available to keep them stable. This account makes a testable set of predictions for future research. If subjects were trained during the 6-month interval on a new contrast that required a different weighting of the same stimulus dimensions, interference would be anticipated at the time of retest. The magnitude of the interference should be modulated by the degree to which the selective attention weights were modified. These changes would be determined by the similarity of the two contrasts on which subjects were trained (Best, 1994; Best and Strange, 1992).

IV. CONCLUSIONS

In summary, we have replicated our previous findings using a larger group of monolingual subjects and therefore, we are confident about our earlier conclusions about the effects of laboratory training on the acquisition of new phonetic contrasts (Lively *et al.*, 1993; Logan *et al.*, 1991). First, the high-variability identification training procedure provides an effective means for quickly modifying adult non-native listeners' phonetic perception. Monolingual

native speakers of Japanese can learn to identify English /r/ and /l/. Identification accuracy increased 11% during 3 weeks of training. Second, listeners retained some information about the new categories over a 6-month interval, without any further training. Performance on the post-test and the tests of generalization did not decrease significantly over a 3-month retention interval. At 6 months, accuracy on the post-test was still 4.5% above pretest levels. Third, listeners demonstrated talker-specific generalization. Identification accuracy was higher when listeners heard new words produced by a familiar talker than when they heard new words produced by an unfamiliar talker. To account for the pattern of these results, we suggest that a selective attention mechanism differentiates items along contrastive perceptual dimensions. Over the course of training, this mechanism stretches and shrinks, the underlying psychological space in order to make the contrastive phonemes more dissimilar from each other. Taken together, the present results indicate listeners acquire and retain information about both the new phonetic contrast and the fine perceptual details related to a talker's voice during the process of perceptual learning.

Acknowledgments

This work was supported, in part, by NIDCD Research Training Grant No. DC-00012-14 and NIDCD Research Grant No. DC-00111-16 to Indiana University in Bloomington, IN. The authors wish to thank John Logan for his helpful suggestions and discussions on this work and Rie Kawakami for her assistance in collecting the data.

References

- Aslin, RN.; Pisoni, DB. Some developmental processes in speech perception. In: Yeni-Komshian, G.; Kavanagh, JF.; Ferguson, CA., editors. *Child Phonology: Perception and Production*. Academic; New York: 1980. p. 67-96.
- Best, CT. The emergence of language-specific influences in infant speech perception. In: Goodman, J.; Nusbaum, HC., editors. *Development of Speech Perception: The Transition From Recognizing Speech Sounds To Spoken Words*. MIT; Cambridge, MA: 1994. p. 167-224.
- Best CT, Strange W. Effects of phonological and phonetic factors on cross-language perception of approximants. *J Phon*. 1992; 20:305-320.
- Dalston RM. Acoustic characteristics of English /w,r,l/ spoken correctly by young children and adults. *J Acoust Soc Am*. 1975; 57:462-469. [PubMed: 1117100]
- Dissosway-Huff, P.; Port, R.; Pisoni, DB. *Research on Speech Perception, Prog Rep No 8*. Speech Research Laboratory, Indiana University; Bloomington: 1982. Context effects in the perception of /r/ and /l/ by Japanese.
- Flege, J. Perception and production: The relevance of phonetic input to L2 phonological learning. In: Ferguson, C.; Huebner, T., editors. *Crosscurrents in Second Language Acquisition and Linguistic Theories*. John Benjamins; Philadelphia, PA: 1990.
- Flege J. Chinese subjects' perception of the word-final English /t-/d/ contrast: Performance before and after training. *J Acoust Soc Am*. 1989; 86:1684-1697. [PubMed: 2808918]
- Flege J, Wang C. Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t-/d/ contrast. *J Phon*. 1989; 17:299-315.
- Gillette, S. *Minnesota Papers on Linguistics and the Philosophy of Language*. Vol. 6. University of Minnesota; Minneapolis: 1980. Contextual variation in the perception of L and R by Japanese and Korean speakers; p. 59-72.
- Goggin JP, Thompson CP, Strube G, Simental LR. The role of language familiarity in voice identification. *Mem Cognit*. 1991; 19:448-458.
- Goto H. Auditory perception by normal Japanese adults of the sounds 'L' and 'R'. *Neuropsychologia*. 1971; 9:317-323. [PubMed: 5149302]
- Henly E, Sheldon E. Duration and context effects on the perception of English /r/ and /l/: A comparison of cantonese and Japanese speakers. *Language Learning*. 1986; 36:505-521.

- Jacoby, L.; Brooks, L. Nonanalytic Cognition: Memory, perception, and concept learning. In: Bower, G., editor. *The Psychology of Learning and Motivation*. Vol. 18. Academic; New York: 1984. p. 1-47.
- Jamieson D, Morosan D. Training non-native speech contrasts in adults: Acquisition of the English /-/θ/ contrast by francophones. *Percept Psychophys*. 1986; 40:205–215. [PubMed: 3580034]
- Jamieson D, Morosan D. Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. *Can J Psychol*. 1989; 43:88–96. [PubMed: 2819599]
- Jenkins, JJ. Four points to remember: A tetrahedral model of memory experiments. In: Cermak, LS.; Craik, FIM., editors. *Levels of Processing in Human Memory*. LEA; Hillsdale, NJ: 1979. p. 429-466.
- Jusczyk, P. Developing phonological categories from the speech signal. paper presented at The International Conference on Phonological Development; Stanford University. 1989. (unpublished)
- Jusczyk P. From general to language-specific capacities: The WRAPSA model of how speech perception develops. *J Phon*. 1993; 21:3–28.
- Jusczyk, P. Infant speech perception and the development of the mental lexicon. In: Goodman, J.; Nusbaum, HC., editors. *The Transition from Speech Sounds to Spoken Words: The Development of Speech Perception*. MIT; Cambridge, MA: 1994. p. 227-270.
- Klatt DH. Speech perception: A model of acoustic-phonetic analysis and lexical access. *J Phon*. 1979; 7:279–312.
- Kuhl PK. Perception of auditory equivalence classes for speech in early infancy. *Infant Behav Dev*. 1983; 6:263–285.
- Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*. 1992; 255:606–608. [PubMed: 1736364]
- Kruschke J. ALCOVE: An exemplar-based connectionist model of category learning. *Psychol Rev*. 1992; 90:22–44. [PubMed: 1546117]
- Lane H. The motor theory of speech perception: A critical review. *Psychol Rev*. 1965; 7:275–309. [PubMed: 14348425]
- Lane, H. A behavioral basis for the polarity principle in linguistics. In: Salzinger, K.; Salzinger, S., editors. *Research in Verbal Behavior and Some Neurological Implications*. Academic; New York: 1969. p. 79-98.
- Laver, J. *The Phonetic Description of Voice Quality*. Cambridge U.P; Cambridge: 1980.
- Laver, J.; Trudgill, P. Phonetic and linguistic markers in speech. In: Scherer, KR.; Giles, H., editors. *Social Markers in Speech*. Cambridge U. P; Cambridge: 1979. p. 1-32.
- Lehiste I. Acoustic characteristics of selected English consonants. *Int J Am Linguistics*. 1960; 30:10–115.
- Lively SE, Logan JS, Pisoni DB. Training Japanese listeners to identify English /r/ and /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.; Logan, JS. Some effects of training Japanese listeners to identify English /r/ and /l/. In: Tohkura, Y.; Vatikiotis-Bateson, E.; Sagisaka, Y., editors. *Speech Perception, Production and Linguistic Structure*. OHM; Tokyo: 1991. p. 175-196.
- Logan JS, 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]
- Logan JS, Lively SE, Pisoni DB. Training listeners to perceive novel phonetic categories: How do we know what is learned? *J Acoust Soc Am*. 1993; 94:1148–1151. [PubMed: 8370868]
- Mann V. Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English R and L. *Cognition*. 1986; 24:169–196. [PubMed: 3816123]
- MacKain K, Best C, Strange W. Categorical perception of /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 Phon*. 1981; 9:283–303.

- Nosofsky RM. Attention, similarity, and the identification categorization relationship. *J Exp Psychol: General*. 1986; 115:39–57.
- Nosofsky RM. Attention and learning processes in the identification and categorization of integral stimuli. *J Exp Psychol: Learning, Memory and Cognition*. 1987; 15:700–708.
- Nygaard LC, Sommers MS, Pisoni DB. Speech perception as a talker-contingent process. *Psychol Sci*. 1994; 5:42–46. [PubMed: 21526138]
- Pisoni, DB.; Lively, SE. Variability and invariance in speech perception: A new look at some old problems in perceptual learning. In: Strange, W., editor. *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-language Speech Research*. York; Timonium, MD: (in press)
- Pisoni, DB.; Lively, SE.; Logan, JS. Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception. In: Goodman, J.; Nusbaum, H., editors. *Development of Speech Perception: The Transition From Recognizing Speech Sounds to Spoken Words*. MIT; Cambridge, MA: 1994. p. 121-166.
- Polka L. Cross-language speech perception in adults: Phonemic, phonetic and acoustic contributions. *J Acoust Soc Am*. 1991; 89:2961–2977. [PubMed: 1918634]
- Polka L. Characterizing the influence of native language experience on adult speech perception. *Percept Psychophys*. 1992; 52:37–52. [PubMed: 1635856]
- Posner M, Keele S. On the genesis of abstract ideas. *J Exp Psychol*. 1968; 77:353–363. [PubMed: 5665566]
- Posner M, Keele S. Retention of abstract ideas. *J Exp Psychol*. 1970; 83:304–308.
- Price, JP. unpublished doctoral dissertation. University of Pennsylvania; 1981. A cross-linguistic study of flaps in Japanese and in American English.
- Pruitt JS. Comments on ‘Training Japanese listeners to identify /r/ and /l/: A first report’ [J. S. Logan, S. E. Lively, and D. B. Pisoni, *J. Acoust. Soc. Am.* 89, 874–886 (1991) [71], “. *J Acoust Soc Am*. 1993; 94:1146–1147. [PubMed: 8240541]
- Sawusch, JR. Auditory and phonetic coding of speech. In: Schwab, EC.; Nusbaum, HC., editors. *Pattern Recognition By Humans and Machines: Vol. 1, Speech Perception*. Academic; New York: 1986. p. 51-88.
- Sheldon A, Strange W. The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech perception can precede speech production. *Appl Psycholinguist*. 1982; 3:243–261.
- 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]
- Strange, W.; Jenkins, J. Role of linguistic experience in perception of speech. In: Walk, RD.; Pick, HL., editors. *Perception and Experience*. Plenum; New York: 1978. p. 125-169.
- Terbeek, D. *Working Papers in Phonetics*. Vol. 37. University of California; Los Angeles: 1977. A cross-language multidimensional scaling study of vowel perception.
- Werker J. On becoming a native listener. *Am Sci*. 1989; 77:54–59.
- Werker J, Tees R. Phonemic and phonetic factors in adult cross-language speech perception. *J Acoust Soc Am*. 1984; 75:1866–1878. [PubMed: 6747097]
- Yamada R, Tohkura Y. The effects of experimental variables in the perception of American English /r,l/ by Japanese listeners. *Percept Psychophys*. 1992; 52:376–392. [PubMed: 1437471]

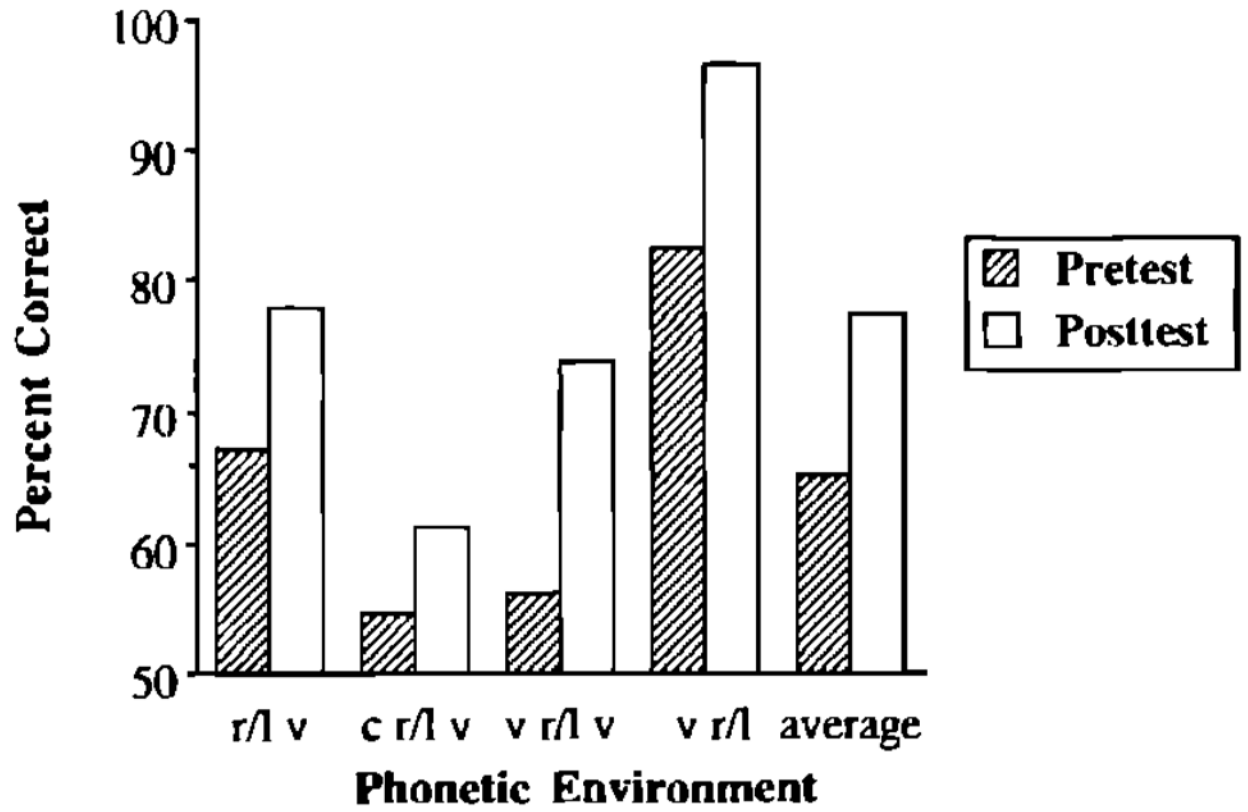


FIG. 1.
Subjects' accuracy in the pretest and post-test as a function of phonetic environment.

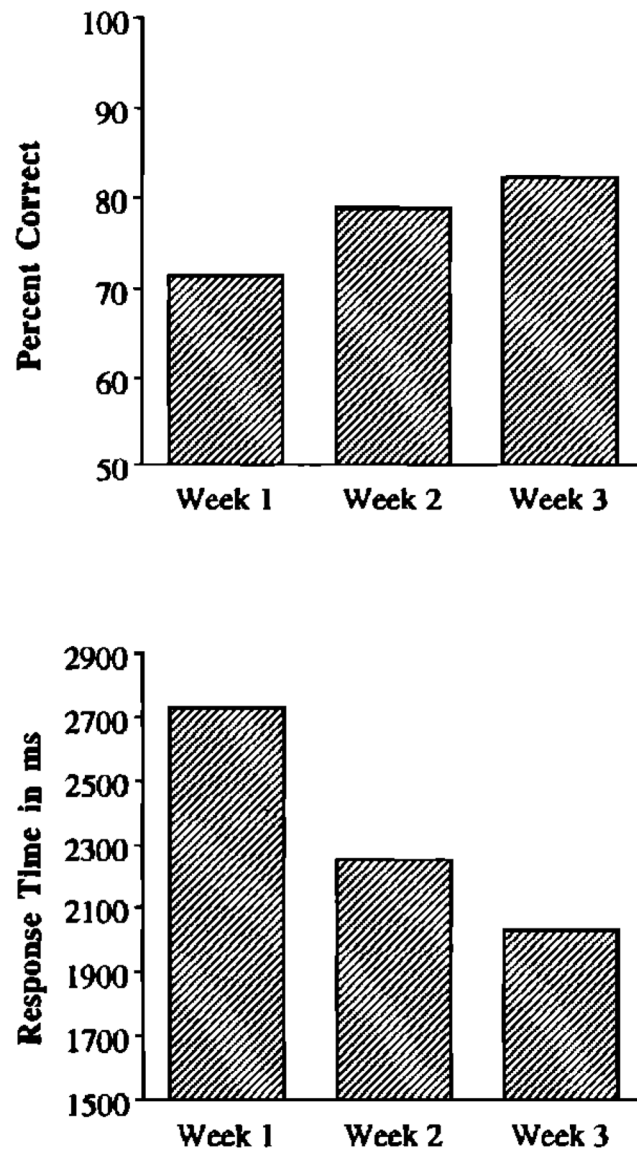


FIG. 2. The top panel shows subjects' accuracy during training as a function of week of training. The lower panel shows response latencies.

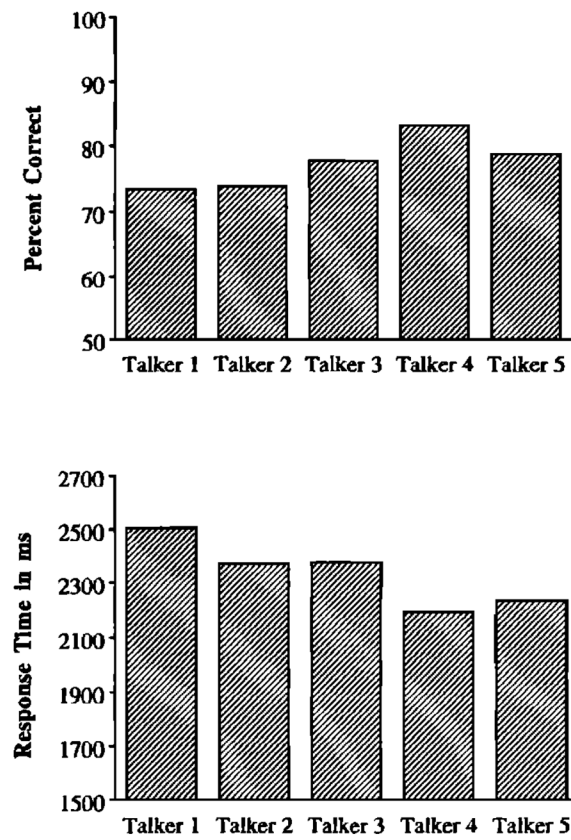


FIG. 3. The top panel shows accuracy during training as a function of talker. The lower panel shows response latencies.

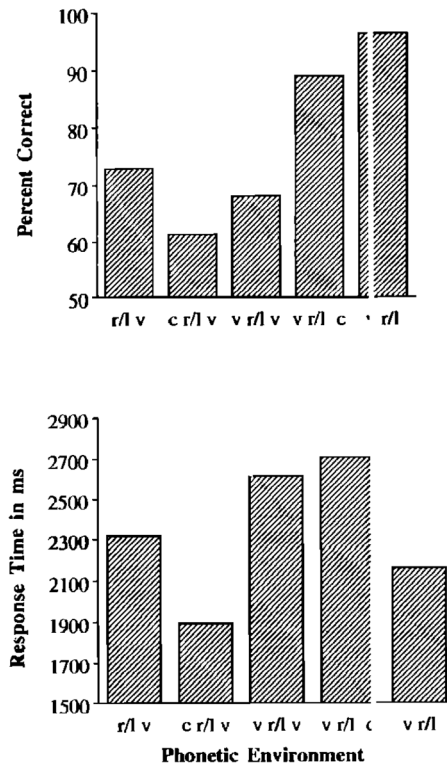


FIG. 4. The top panel shows accuracy during training as a function of phonetic environment. The lower panel shows response latencies.

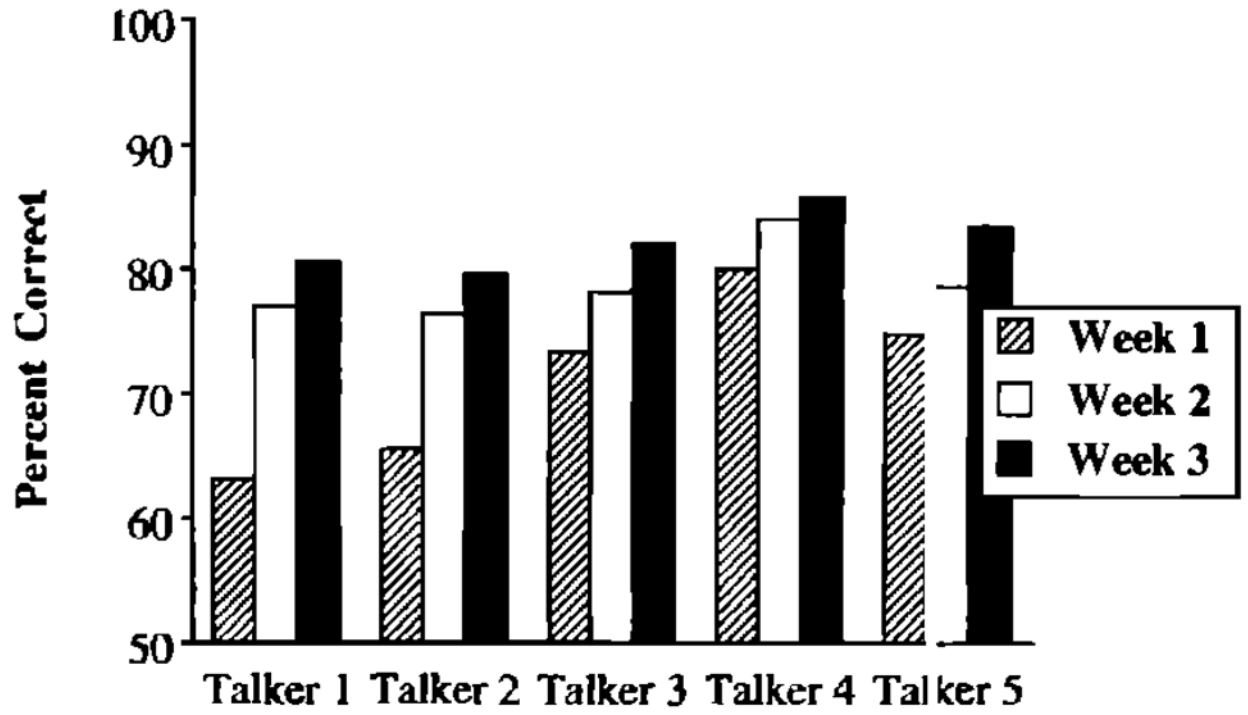


FIG. 5.
The interaction of talker and week in the accuracy data during the training phase.

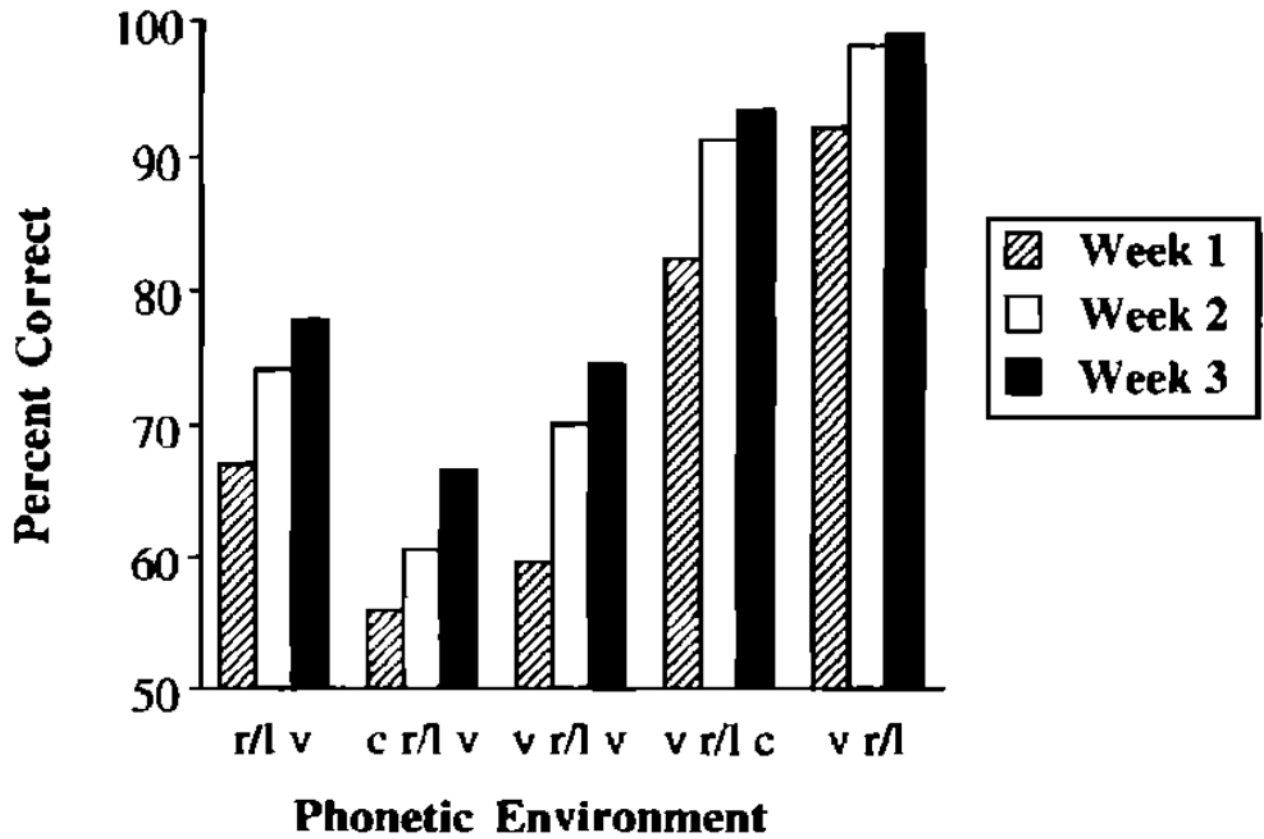


FIG. 6.
The interaction of phonetic environment and week in the accuracy data during the training phase.

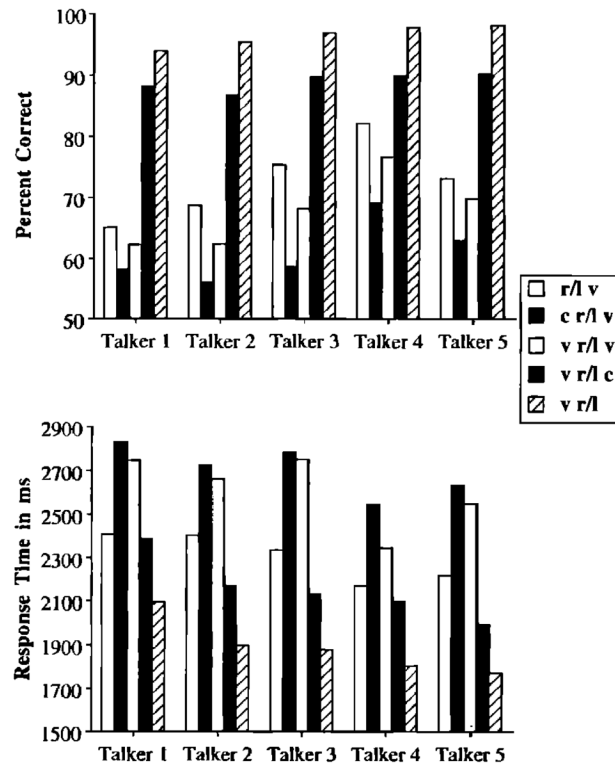


FIG. 7. The top panel shows the interaction of talker with phonetic environment during training in the accuracy data. The lower panel shows a similar interaction in the response time data.

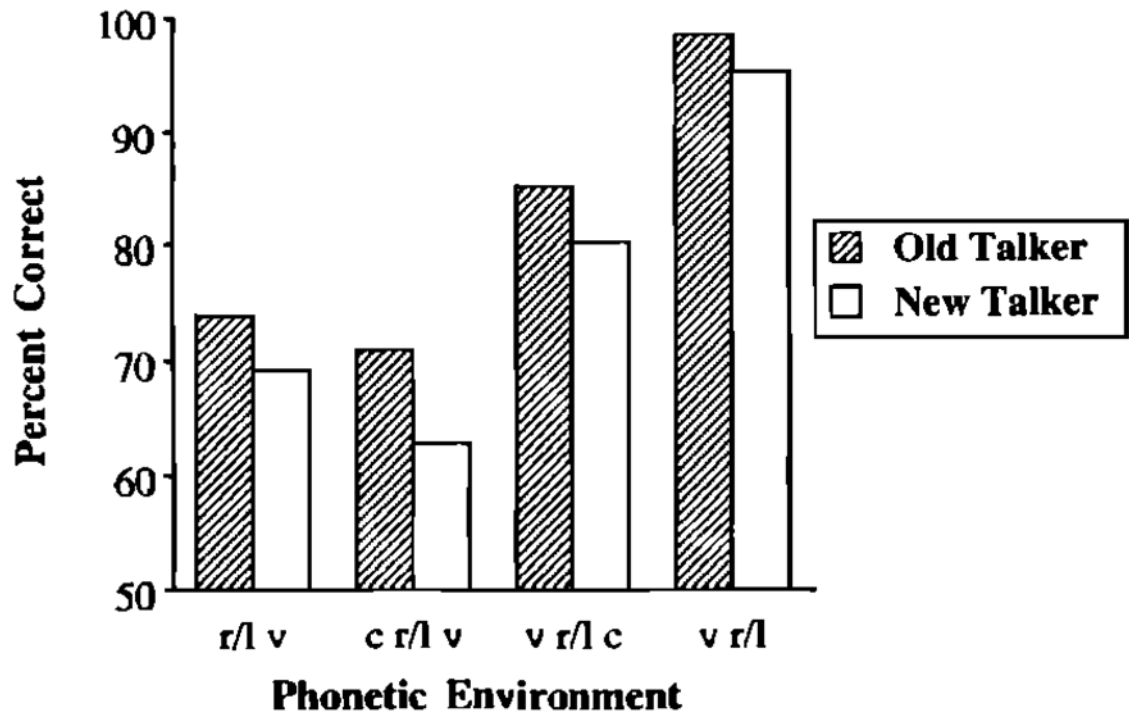


FIG. 8. Subjects' accuracy in the tests of generalization as a function of talker.

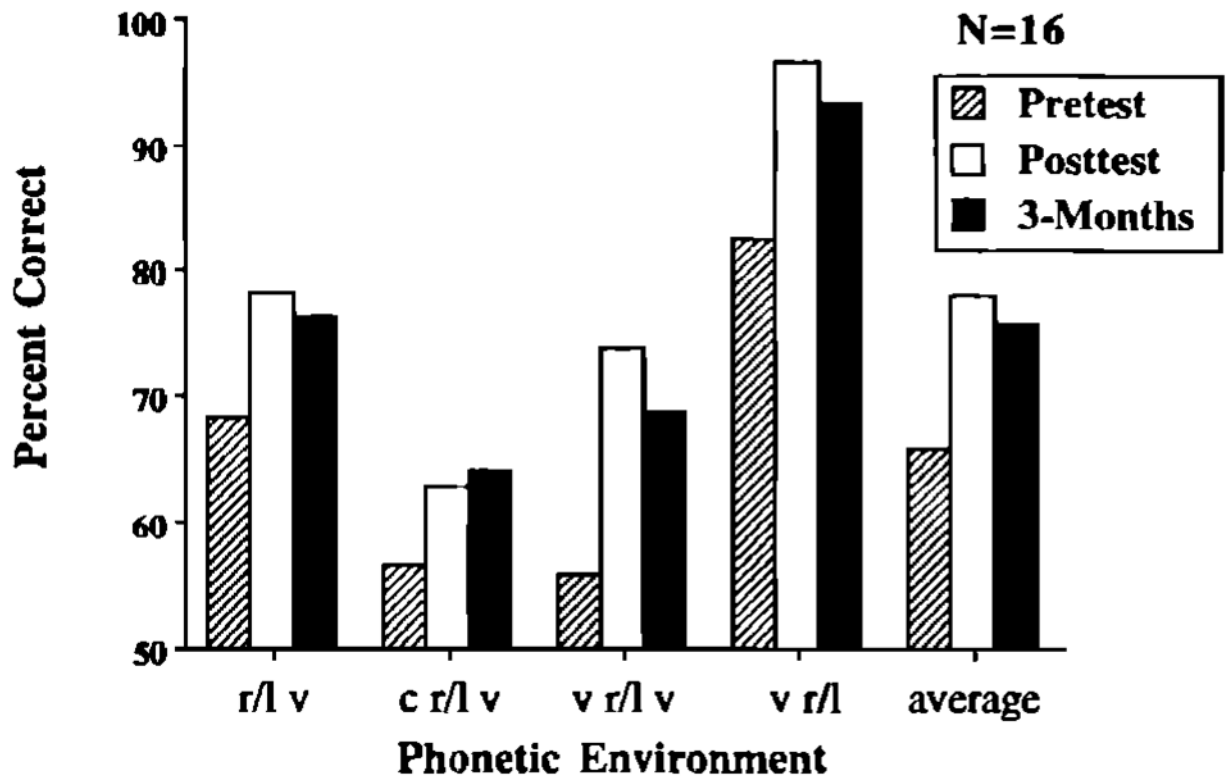


FIG. 9. Accuracy scores for the pretest, post-test, and 3-month follow-up tests of 16 subjects.

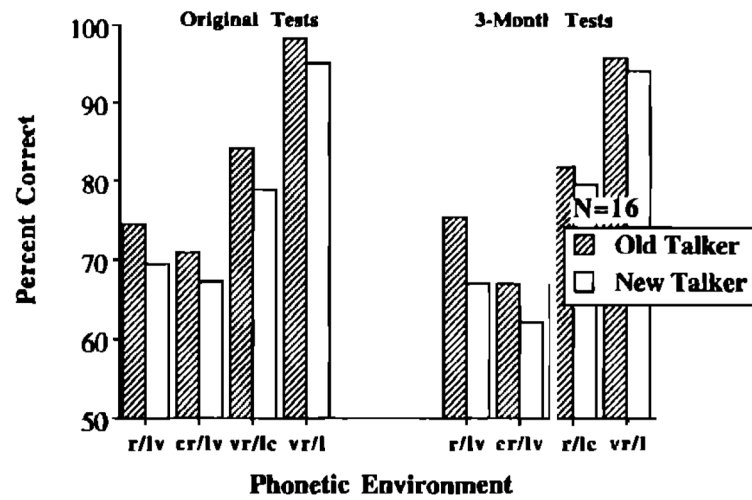


FIG. 10. The left panel of the figure shows accuracy scores during the first tests of generalization for 16 subjects. The right panel shows accuracy scores from the same subjects during 3-month follow-up tests of generalization.

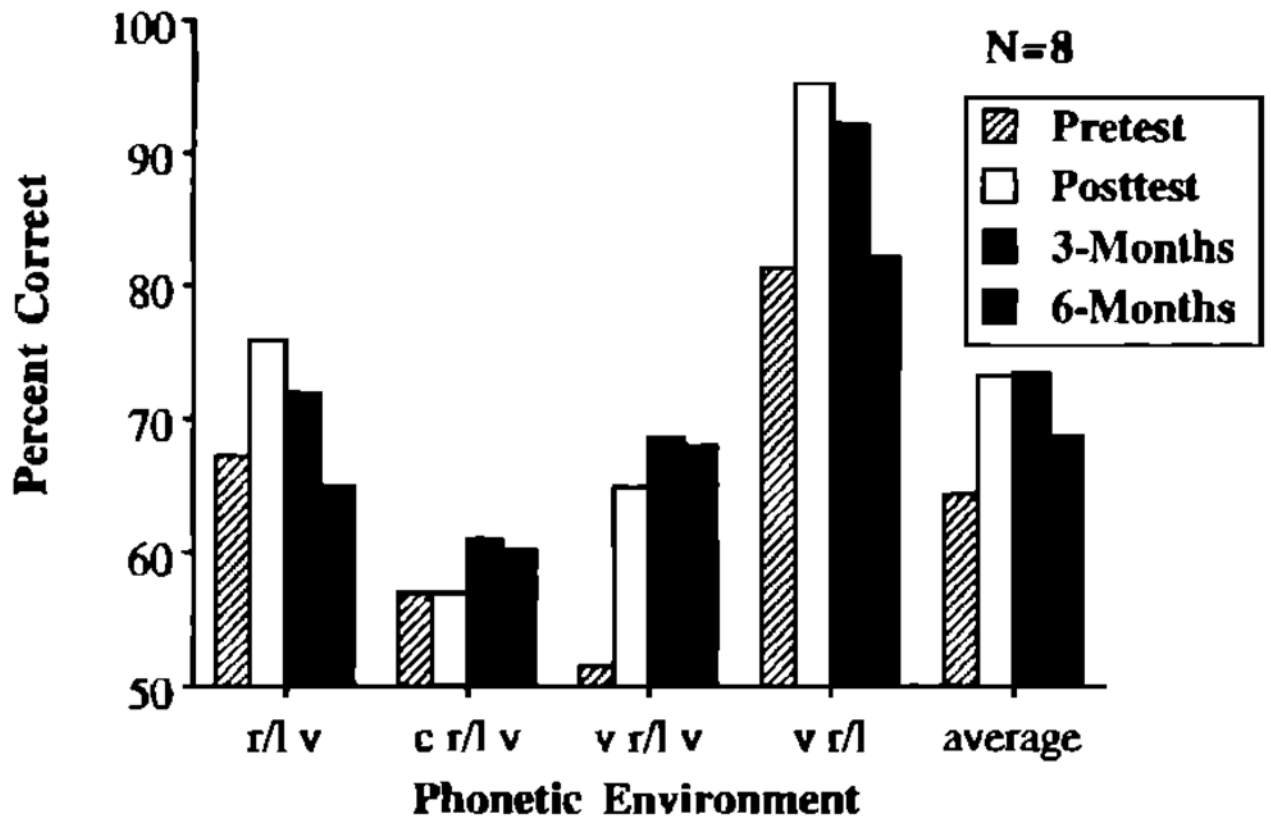


FIG. 11. Accuracy scores for the pretest, post-test, 3- and 6-month follow-up tests of eight subjects.

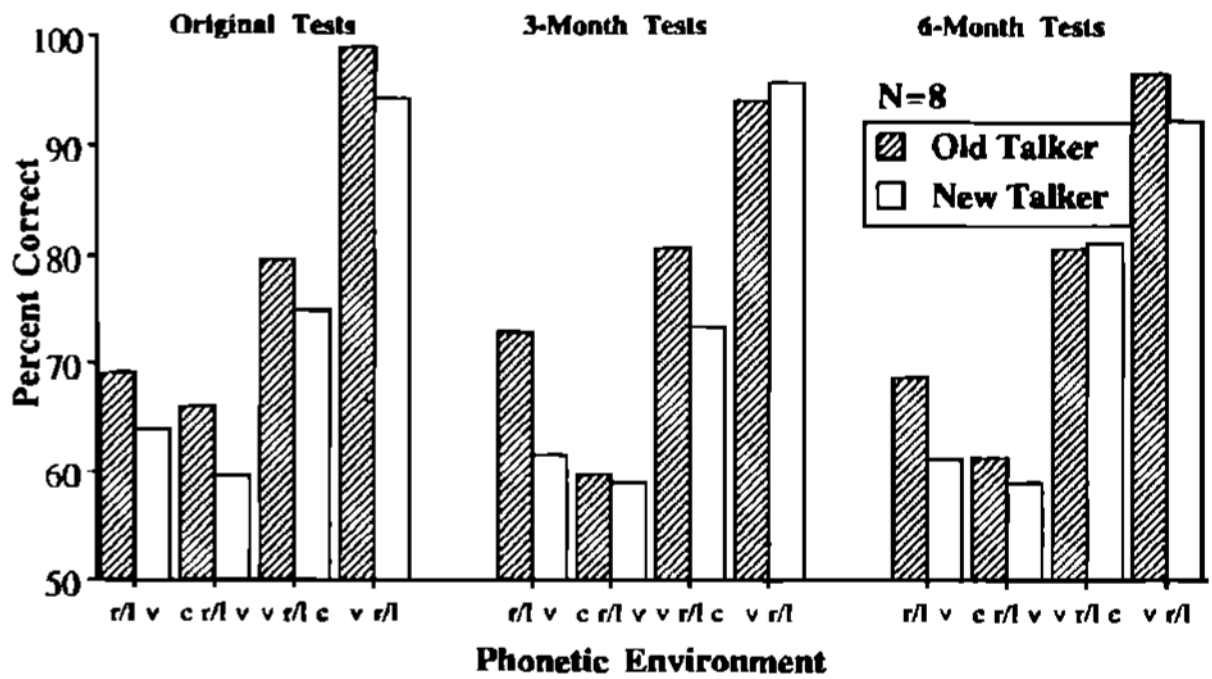


FIG. 12.

The left panel shows accuracy scores during the first tests of generalization for eight subjects. The middle panel shows accuracy scores from the same subjects during the 3-month follow-up tests of generalization. The right panel shows accuracy scores during the 6-month follow-up tests of generalization.