5pSC7

Testing the importance of talker variability in non-native speech contrast training

James S. Magnuson, Reiko A. Yamada, Yoh'ichi Tohkura

ATR Human Information Processing Research Laboratories
2-2 Hikaridai
Seika-cho, Soraku-gun,
Kyoto, 619-02 Japan
email: magnuson@hip.atr.co.jp, yamada@hip.atr.co.jp
Telephone: (+81)7749-5-1084
Fax: (+81)7749-5-1008

Ann R. Bradlow, Scott E. Lively

Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington, IN 47405-1301
email: abradlow@indiana.edu

This paper summarizes a poster presented at the 129th Meeting of the Acoustical Society of America, Washington, D.C., 3 June, 1995

ABSTRACT

In contrast to Japanese adults trained with stimuli produced by five talkers, Lively et al. [J. Acoust. Soc. Am., 94, 1242-1255 (1993)] reported that Japanese adults trained to perceive English /r/ and /l/ with stimuli produced by a single talker failed to improve from pretest to post-test, and did not generalize to novel stimuli. We extended that study by training five groups of subjects each with a different talker, and by examining the retention of learning after three and six months. The results reported previously were partially replicated. Although all five groups showed significant learning within the training sessions, only four of the five groups showed significant improvement from pretest to post-test; of a subset of subjects that returned for a 3-month follow-up test, only three of five groups were significantly more accurate in the post-test than in the pretest, and only one of those groups was significantly more accurate in the follow-up test. The results indicate that while multiple-talker training leads to consistently good results, training with stimuli produced by only one talker may fail to promote generalization to new stimuli and talkers.

INTRODUCTION

Adult learners of languages tend to have difficulty distinguishing foreign speech sounds which contrast in ways that are not meaningful in their native language. Among the best-known examples are English /T/ vs. /D/ for native speakers of French (e.g., Jamieson and Morosan, 1986), Hindi dental vs. retroflex stops for native speakers of English (e.g., Werker and Tees, 1984) and Japanese (Pruitt, 1994), and English /r/ vs. /l/ for native speakers of Japanese (Goto, 1971; Sheldon and Strange, 1982; see Strange, in press, for a review of non-native contrast research, and cross-language speech perception research in general). Training programs for the first two of these cases have been quite successful (e.g., Jamieson and Morosan, 1986, 1989; Pruitt, 1994). However, the case of /r/ and /l/ for Japanese speakers has proven more difficult. For example, Strange and Dittmann (1984) used synthetic words containing /r/s and /l/s using the fixed-standard discrimination task that Carney et al. (1977) had used to successfully train American English (AE) listeners to discriminate stops differing in voice-onset time (VOT). Subjects in this study met with some success: they showed significant improvement in discrimination performance on the training stimuli, and some generalization to new stimuli. However, only two of eight subjects improved in perceiving initial /r/-/l/ contrasts.

Logan et al. (1991) hypothesized that previous /r/-/l/ training efforts lacked sufficient variability in the training stimuli to prepare subjects for the amount and sort of variability that occurs in natural speech. By providing subjects with greater amounts of variability (both in the number of phonetic environments and the number of talkers used during training), using natural rather than synthesized speech, and by using an identification task rather than a discrimination task (as had Jamieson and Morosan, 1986, 1989), Logan et al. improved their subjects' identification performance significantly after approximately 15 hours of training. In addition, their subjects were able to generalize from the training stimuli to new tokens produced by a talker they had heard in training, and to tokens of novel words produced by a novel talker. In addition, the results have been replicated and extended (Lively et al., 1993, 1994). For example, Lively et al. (1994) replicated this study with a new, larger group of subjects, and extended it by testing some subjects' ability on the post-test and generalization materials in three- and six-month follow-up tests. Even after six months, with no additional training, subjects were 4.5% more accurate than they were in the pretest.

While these studies demonstrated that long-term modification of the phonetic categories of monolingual Japanese-speaking adults is possible using high-variability training materials, the relative contributions of phonetic-environment variability and talker variability were not clear (Logan et al., 1993; Pruitt, 1993). Lively et al. (1993) began examining this question by reducing the number of phonetic environments and the number of talkers used in training. In one experiment, only the three most difficult phonetic environments of the five used in other studies (Logan et al., 1991; Lively et al., 1994) were used in training. After training, these subjects showed improvement and generalization comparable to that of subjects trained with all five contexts. In a second experiment, subjects were trained with the productions of only one of the five talkers used in other studies. In contrast to the reduced-environment experiment, subjects showed some improvement but did not improve significantly from pretest to post-test, and they were not able to generalize to a new talker.

The current experiment was designed to test further the importance of talker variability. The experiment was motivated by talker-specific effects observed in previous training and perceptual studies. In previous /r/-/l/ training and perception experiments, we have found significant talker differences intelligibility. In multiple talker training experiments, accuracy has been consistently higher on some talkers (Logan et al., 1991; Lively et al., 1993, 1994). In perception experiments, we have found that some talkers' productions of /r/ and /l/ are "R"- or "L"-like relative to those of other talkers. When subjects listen to stimuli produced by a single talker, they respond "R" approximately 50% of the time. However, when the same stimuli are mixed with other talkers' productions, rate of "R" response changes significantly for some talkers' productions (Magnuson and Yamada, 1994). These talker effects suggest that learning in single-talker training might also vary between talkers. A related motivation was to ensure that the previous finding that single-talker training is not effective was due to the lack of talker variability in training, rather than specific characteristics of the one talker used in that study (Lively et al., 1993).

Here we report a replication and extension of the single-talker study: five groups of subjects were trained with stimuli produced by a different one of five talkers. Each group participated in pretest -- post-test, training and generalization phases. Subsets of the subjects returned for three- and six-month follow-up tests.

I. METHOD

A. Subjects

50 native speakers of Japanese living in Kyoto prefecture, Japan, participated in the training experiment. 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 hearing or speech disorders.

B. Stimuli

The training and pretest--post-test materials were those employed by Logan et al. (1991) and Lively et al. (1994; Lively et al., 1993, used subsets of these materials). The materials used for training were compiled as follows. All words beginning with /r/ and /l/ were retrieved from a computerized database containing approximately 20,000 words (Webster's Seventh Collegiate Dictionary, 1967). A total of 207 minimal pairs contrasting /r/ and /l/ in word-initial, word-final, and intervocalic positions were found. Five talkers, three male (talkers T1, T3 and T5) and two female (T2 and T4), 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 using a 12-bit analog-to-digital converter at Indiana University. The digitized wave form files were then edited and equated for RMS amplitude using a specialized signal processing package.

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 transcription task was used in which listeners typed what they heard at 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/. 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) The stimuli used in the tests of generalization were also those used by Logan et al. (1991) and Lively et al. (1994). Two sets of tokens were recorded. The first set consisted of 95 novel words produced by whichever talker subjects heard in training. In this set of stimuli, 37 words had /r/ or /l/ in initial singleton position, 32 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.

Files were digitally transferred to ATR laboratories, where they were upsampled to a rate of 44.1 kHz and rescaled to a resolution of 16 bits.

C. Procedure

The experimental design employed a pretest--post-test procedure closely modeled after the methods used by Strange and Dittmann (1984) and in our previous reports on /r/-/l/ training (Logan et al., 1991; Lively et al., 1993; Lively et al., 1994). In this design, the effects of training were assessed by comparing performance on a pretest and a post-test administered before and after a 15-session training period. In previous studies, the training occurred over a period of three weeks (5 daily sessions per week). However, Yamada (1993) has demonstrated that the training is equally effective whether subjects are trained with 15 daily sessions or with 3 sessions per day over a period of 5 days. To reduce the number of times subjects had to travel to the laboratory, subjects were trained in five days, with three sessions per day. Generalization to new words and a new talker was measured after the post-test.

All testing and training was carried out in Japan at the Advanced Telecommunications Research (ATR) Human Information Processing (HIP) Research Laboratories in Kyoto prefecture. All sessions were administered in quiet, sound-treated rooms. With the exception of 11 subjects’ pretests and post-tests (which were conducted with a paper-and-pencil test; see below), the procedure was as follows: subjects sat at desks equipped with a keyboard, and a CRT monitor. Stimuli were binaurally presented over headphones (STAX SR-Lambda Signature) at a comfortable listening level. Presentation of stimuli, feed back, and collection of response was controlled by a computer workstation (NeXT cube). During training, tests, and generalization sessions, both identification responses and latencies were collected. Latencies were measured from the onset of the stimulus presentation to the subject's response (however, response times are not reported in this presentation; detailed analyses of the results will be presented in a later paper).

Before training began, subjects were given a pretest. Eleven subjects participated in a paper-and-pencil version of the pretest. 24 minimal pairs of words were recorded onto digital audio tape (DAT) using a Sony DTC-1500ES DAT recorder. 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 rest of the subjects' pretest and post-test data was collected via a computer controlled two-alternative identification task. This change was made to allow us to collect response times. The same stimuli were used and presented twice, but different random orders of the stimuli were generated for each subject.

On each trial, the two members of a minimal pair contrasting /r/ and /l/ were displayed on the left and right sides of a "window" displayed on the CRT. 500 ms later, the audio stimulus was presented through the headphones. Subjects responded by pressing keys on the keyboard: if subjects thought the stimulus played through the headphones matched the word displayed on the left, they pressed the "1" key; if they thought the stimulus matched the word displayed on the right, they pressed the "2" key. On each trial, the experimental control program randomly assigned the /r/ and /l/ members of each minimal pair to the right and left sides of the window. The same test items were presented after training in the post-test, and again for the 3- and 6-month follow-up tests. All 3- and 6-month follow-up tests were computer controlled. These tests each took approximately 20 minutes to complete. Training began on day 2.

Subjects were randomly assigned to one of 5 single-talker training groups. Each single-talker group was trained with stimuli produced by a different single talker. Single-talker group and talker numbers were matched (i.e., group 1 trained on T1, etc.). In the training phase, the same two-alternative identification paradigm described for the pretest was used, with the addition of feedback. During training, if the listener responded correctly, a chime sounded, and the next trial began 2 s later. To motivate subjects to pay attention during training, a monetary incentive was offered. For every 3 correct responses, 1 yen was added to the subject's payment. After every third correct response, a coin was displayed on the computer screen. If the subject made an error, a buzzer sounded, the stimulus word was repeated, and the subject had to respond again. The orthographic forms of the minimal pair were randomly reassigned to left and right sides, in order to require subjects to identify the word, not just respond with a different key. This procedure was repeated until subjects responded correctly. However, only the first response to each stimulus was used in the analyses.

In each of the 15 training sessions, a set of 68 minimal pairs were each presented twice, yielding a total of 272 trials in each session. Subjects heard only one talker throughout the training. Subjects were trained individually in three sessions per day. The sessions each took approximately 40 minutes to complete.

Following training (on day 7), subjects participated in a post-test (identical to the pretest, described above), and two tests of generalization. The first generalization test consisted of 93 novel words from minimal pairs contrasting /r/ and /l/ produced by a new talker (i.e., a male talker not used in either the pretest--post-test or training phases; see the stimuli section, above). The second generalization test consisted of 95 novel words from minimal pairs contrasting /r/ and /l/ produced by an "old" talker; whichever talker subjects in the single-talker training groups had heard during training (e.g., T1 for group 1). The test stimuli consisted of words subjects had not heard before in the experiment. In both tests of generalization, the task was identical to that used in the pretest--posttest and training phases. As in the pretest--post-test phase, no feedback was given.

II. RESULTS

Accuracy was measured in every session. Separate ANOVAs were conducted on each group's data from training, pretest, post-test, generalization and follow-up sessions. All post-hoc tests were conducted using Tukey's Honestly Significant Difference test at the .05 level.

A. Training

Although data were collected by day, phonetic context, and response ("R" or "L"), for the purposes of this presentation, we will only consider the main effect of day. Full results will be presented in a later paper.

As can be seen in Figure 1, all groups improved significantly during identification training.[1] Post-hoc tests comparing accuracy on the first day of training with accuracy on the last day of training were significant for every group. The results of the F-tests on the main effect of day and the post-hoc tests for each group are presented in Table 1. We now turn to the results from the pretest -- post-test comparisons to see if the training benefits generalized to the test talker.

		Main effect of Day		**Accuracy**		**Post-hoc:**
Group	**n**	**F**	**p**	**Day 1**	**Day 5**	**Day 5 > Day 1**
G1	10	(4,36)=32.60	**<.001**	74.40	87.33	*
G2	10	(4,36)=69.13	**<.001**	70.40	86.71	*
G3	10	(4,36)=56.35	**<.001**	63.50	75.70	*
G4	10	(4,36)=39.40	**<.001**	75.34	91.00	*
G5	10	(4,36)=48.55	**<.001**	68.76	84.29	*

Table 1: Training accuracy by group and day.

B. Pretest--post-test

Although data were collected by test, phonetic context, and response (“R” or “L”), for the purposes of this presentation, we will only consider the main effect of test. Full results will be presented in a later paper.

---

[1] Figures follow the References section. For comparison, the figures contain data from a group of subjects trained with all five talkers (Group M). The data is from Yamada et al. (in preparation).

In contrast to the single-talker experiment reported by Lively et al. (1993), four of the five groups were significantly more accurate in the post-test than in the pretest (every group but G3; see Table 2 and Figure 2). We now turn to the results from the tests of generalization to see if the training benefits generalized to novel stimuli and new talkers.

		Main effect of Test		**Accuracy**	
Group	**n**	**F**	**p**	**Pretest**	**Post-test**
G1	10	(1,9)=12.90	**<.01**	72.80	81.40
G2	10	(1,9)=12.14	**<.01**	68.65	77.00
G3	10	(1,9)=.97	.350	59.24	63.44
G4	10	(1,9)=20.48	**<.005**	62.74	75.13
G5	10	(1,9)=5.31	**<.05**	65.98	76.56

Table 2: Pretest and post-test accuracy by group.

3. Generalization

Although data were collected by talker, phonetic context, and response ("R" or "L"), for the purposes of this presentation, we will only consider the main effect of talker, and compare accuracy with pretest levels of accuracy. Full results will be presented in a later paper.

The effects of talker on accuracy and response time were not significant for any group. However, accuracy by talker is compared with accuracy on the pretest in Figure 3. In the figure, you can see that compared with performance on the pretest, all groups showed some ability to generalize to new stimuli produced by their training talker (the "old" talker), and that subjects, especially those in group 3, tended to generalize less well to the new talker.

D. Three-month follow-up test

Twenty-one subjects returned for the 3 month follow-up test. The test consisted of two parts: a test on the pretest--post-test materials (identical to the pretest and post-test, aside from order of stimulus presentation), and a repetition of the test of generalization. Separate ANOVAs were performed on both tests for each group. Please note that the number of subjects who returned is small for some groups. Hence, our interpretations of this data are tentative.

1. Post-test materials

Again, although data were collected by test, phonetic context, and response, for the purposes of this presentation, we will only consider the main effect of test. Full results will be presented in a later paper.

Post-hoc tests showed that of the subset of subjects who returned for the 3-month follow-up test, three groups were significantly more accurate in the pretest than in the post-test (G1, G4 and G5). However, only Group 5 was also significantly more accurate in the 3-month test than in the pretest (see Table 3 and Figure 4).

		Main effect of Test		Accuracy			Post-hoc:	
Group	**n**	**F**	**p**	**Pre**	**Post**	**3 mo.**	**Post > Pre**	**3 mo. > Pre**
G1	3	(2,4)=3.48	.133	80.73	89.58	86.46	*	**ns**
G2	4	(2,6)=1.20	.364	70.75	73.49	73.05	**ns**	**ns**
G3	5	(2,8)=.78	.492	60.35	63.42	66.25	**ns**	**ns**
G4	4	(2,6)=10.95	**<.01**	60.97	79.84	71.48	*	**ns**
G5	5	(2,8)=5.73	**<.05**	61.87	79.38	72.50	*	*

Table 3: Accuracy by test for subjects who returned for the 3-month follow-up session.

2. Generalization materials

Although data were collected by test (post and 3-month), talker, phonetic context, and response, for the purposes of this presentation, we will only consider the main effects of talker and test, and compare accuracy with pretest levels of accuracy. Full results will be presented in a later paper.

The effects of test and talker and the interaction of test and talker in accuracy and response time were not significant for any group, and post-hoc tests showed that there were no significant differences between old and new talkers in either the generalization test given at the time of the post-test or the one given at the time of the 3-month follow-up test. Accuracy by talker is compared with accuracy on the pretest and accuracy by talker at the time of the post-test in Figure 5. In the figure, you can see that compared with performance on the pretest and the generalization test given at the time of the post-test, subjects in groups G1, G4 and G5 showed substantial generalization ability at post-test time and substantial retention after 3-months.

E. Six-month tests

Nine subjects returned for the 6 month follow-up test. The test consisted of two parts: a test on the pretest--post-test materials (identical to the pretest and post-test, aside from order of stimulus presentation), and a repetition of the test of generalization. ANOVAs were performed for both tests for each group. Please note that the number of subjects is very small; we can only make tentative interpretations of these data.

		Main effect of Test		Accuracy			Post-hoc:	
Group	**n**	**F**	**p**	**Pre**	**Post**	**6 mo.**	**Post < Pre**	**6 mo. < Pre**
G1	2	(2,2)=14.62	.064	71.09	84.38	85.16	*	*
G2	*1*	*(2,17)=.17*	*.849*	*62.50*	*62.50*	*58.69*	***ns***	***ns***
G3	*1*	*(2,17)=.51*	*.610*	*64.06*	*56.25*	*53.13*	***ns***	***ns***
G4	3	(2,4)=15.08	**<.05**	64.06	78.65	78.65	*	*
G5	2	(2,2)=2.48	.288	74.22	83.59	86.72	**ns**	**ns**

Table 4: Accuracy by test for subjects who returned for the 6-month follow-up session. Note that the statistics for G2 and G3 are based on a different model than that used for the other groups. See the text for details.

1. Post-test materials

Again, although data were collected by test, phonetic context, and response, for the purposes of this presentation, we will only consider the main effect of test. Full results will be presented in a later paper. Only 1 subject returned for each of two groups (G2 and G3). For these groups, the effect of test was calculated using the error term for the entire ANOVA model.

Post-hoc tests showed that of the subset of subjects who returned for the 6-month follow-up test, two groups were significantly more accurate in the post-and 6-month tests than in the pretest (G1 and G4). The results are summarized in Table 4 and Figure 6.

2. Generalization materials

Although data were collected by test (post and 6-month), talker, phonetic context, and response, for the purposes of this presentation, we will only consider the main effects of talker and test, and compare accuracy with pretest levels of accuracy. Full results will be presented in a later paper.

The effects of test and talker and the interaction of test and talker in accuracy and response time were not significant for any group. Accuracy by talker is compared with accuracy on the pretest and accuracy by talker at the time of the post-test in Figure 7. In the figure, you can see that compared with performance on the pretest and the generalization test given at the time of the post-test, all subjects in groups G1, G4 and, to a lesser extent, G5, showed substantial generalization ability at post-test time and substantial retention after 6-months.

III. CONCLUSIONS

The results of the second experiment reported by Lively et al. (1993), were partially replicated: one group failed to improve in the pretest--post-test comparison (G3), and subjects in groups G2, G3 and G5 failed to meet at least one of the following criteria (i.e., they failed to perform comparably to subjects in previous multiple-talker training studies): good generalization to new stimuli and new talkers; and substantial retention in three-month follow-up tests. However, in contrast to the result found by Lively et al. (1993), subjects in some groups (G1, G4, and G5) showed significant pretest--post-test improvement, and performed significantly better on at least one of the follow-up tests administered three and six months after training than they did on the pretest administered prior to training.

Thus, while multiple talker training leads to consistently good results, single-talker training with some talker may fail to promote learning and generalization ability. Although some talkers may provide enough variability to promote learning and generalization, further research is required to quantify the advantages of multiple-talker training, and the talker-specific advantages observed in the current and previous studies.

REFERENCES

Carney, A. E., Widin, G. P., and Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing VOT. Journal of the Acoustical Society of America, 62, 961-970.

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

Jamieson, D. G., and Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /T/ - /D/ contrast by francophones. Perception and Psychophysics, 40, 205-215.

Jamieson, D. G., and Morosan, D. E. (1989). Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. Canadian Journal of Psychology, 43, 88-96.

Lively, S. E., Logan, J. S., and Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. Journal of the Acoustical Society of America, 94, 1242-1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., and Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories. Journal of the Acoustical Society of America, 96, 2076-2087.

Logan, J. S., Lively, S. E., and Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. Journal of the Acoustical Society of America, 89, 874-886.

Logan, J. S., Lively, S. E., and Pisoni, D. B. (1993). Training listeners to perceive novel phonetic categories: How do we know what is learned? Journal of the Acoustical Society of America, 94, 1148-1151.

Magnuson, J. S., and Yamada, R. A. (1994). Talker variability and the identification of American English /r/ and /l/ by Japanese listeners. Journal of the Acoustical Society of America, 95, 2872 (abstract).

Pruitt, J. S. (1993). 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)], Journal of the Acoustical Society of America, 94, 1146-1147.

Pruitt, J. S. (1994). Identification of Hindi dental and retroflex consonants by native English and Japanese speakers. Journal of the Acoustical Society of America, 95, 3011 (abstract).

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

Strange, W. (in press). Cross-Language studies of speech perception: An historical review. To be published in W. Strange (Ed.), Speech Perception and Linguistic Experience: Issues in Cross-Language Speech Research. Timonium, MD: York Press.

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

Werker, J. F., and Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. Journal of the Acoustical Society of America, 75, 1866-1878.

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

Yamada, R. A., Tohkura, Y., Bradlow, A. R., and Pisoni, D. B. (in preparation). The effect of extended training on the perception of /r and /l/ by native speakers of Japanese.

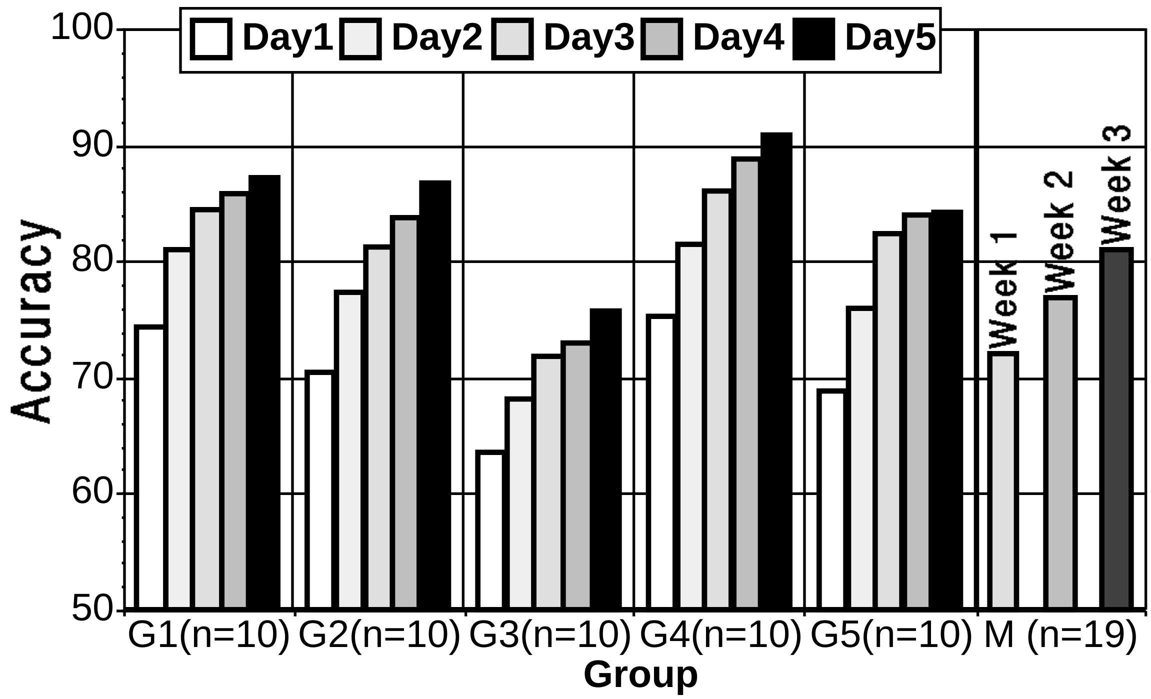

Figure 1: *Training accuracy by group.*

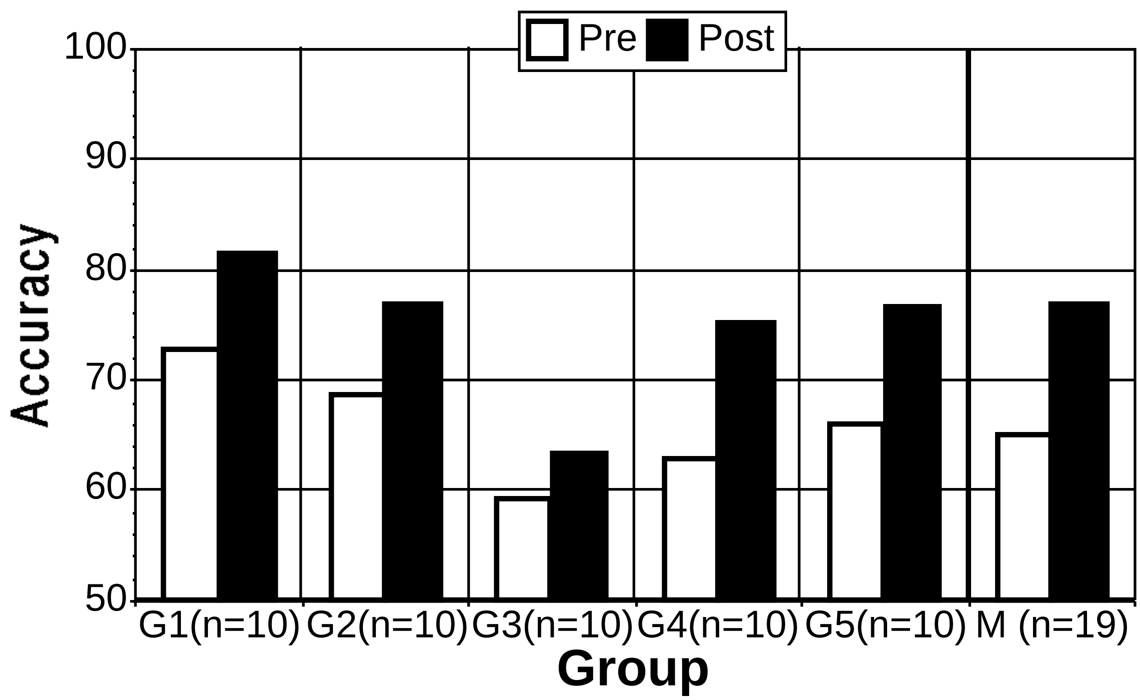

Figure 2: *Pretest/post-test accuracy by group.*

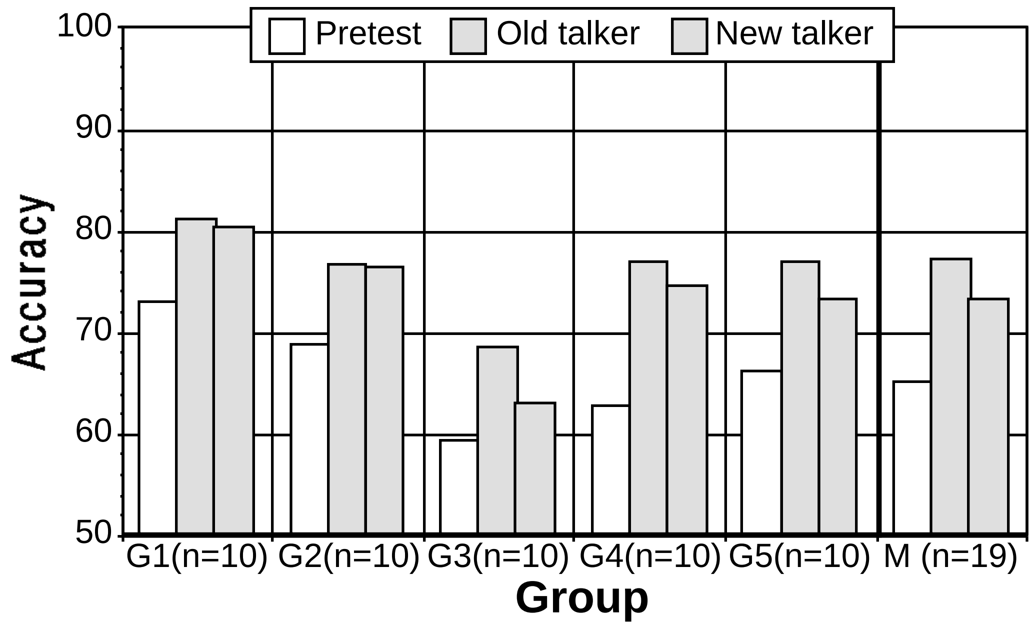

Figure 3: *Generalization accuracy at post-test.*

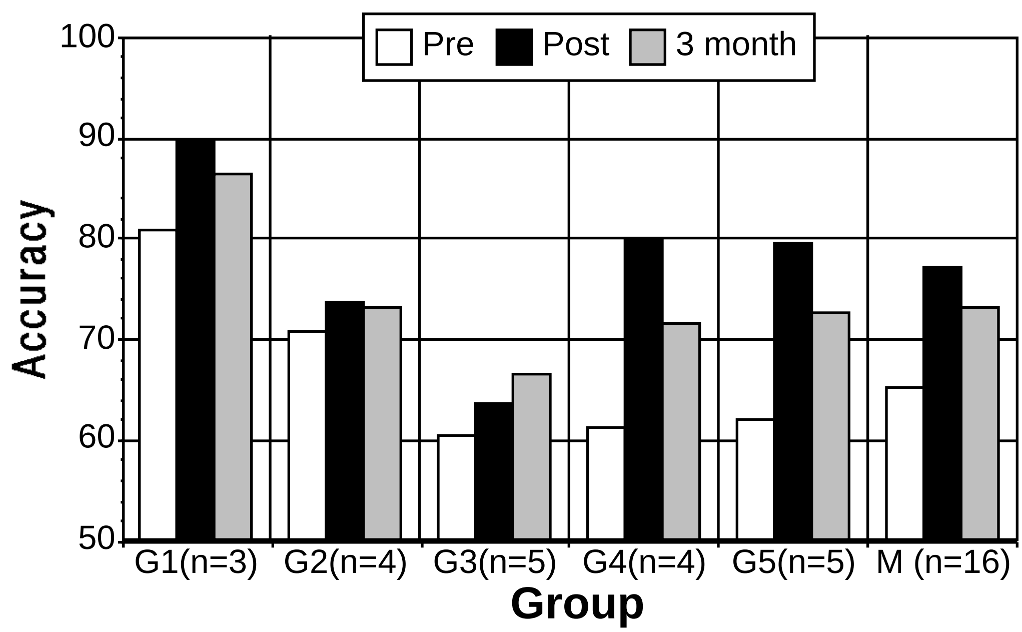

Figure 4: *Accuracy on the 3-month follow-up test on the post-test materials by group.*

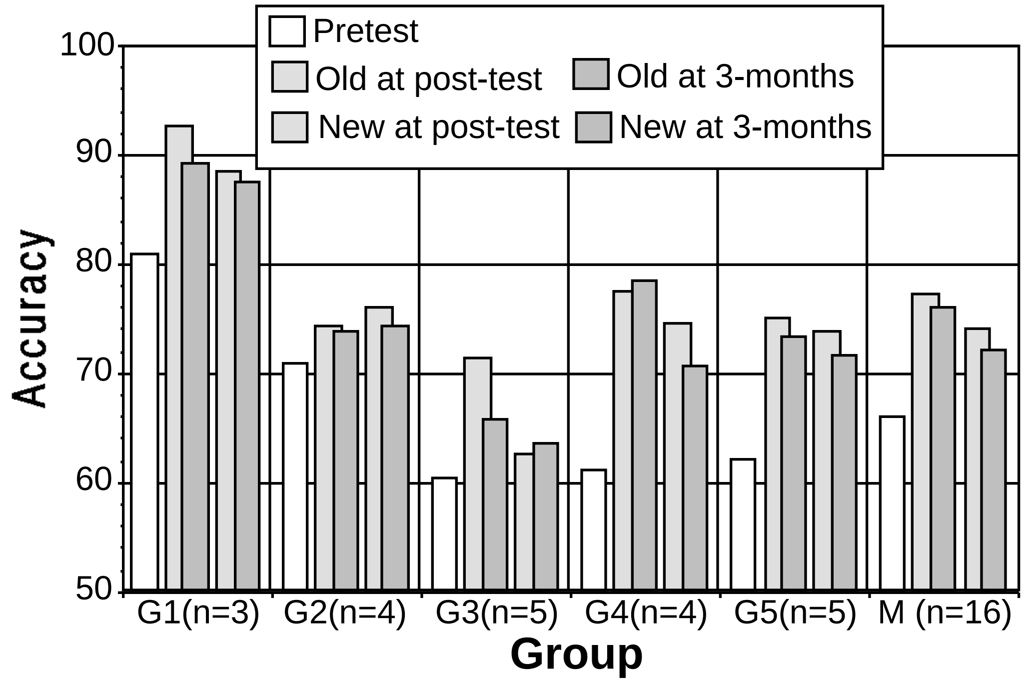

Figure 5: *Generalization accuracy in the 3-month follow-up test.*

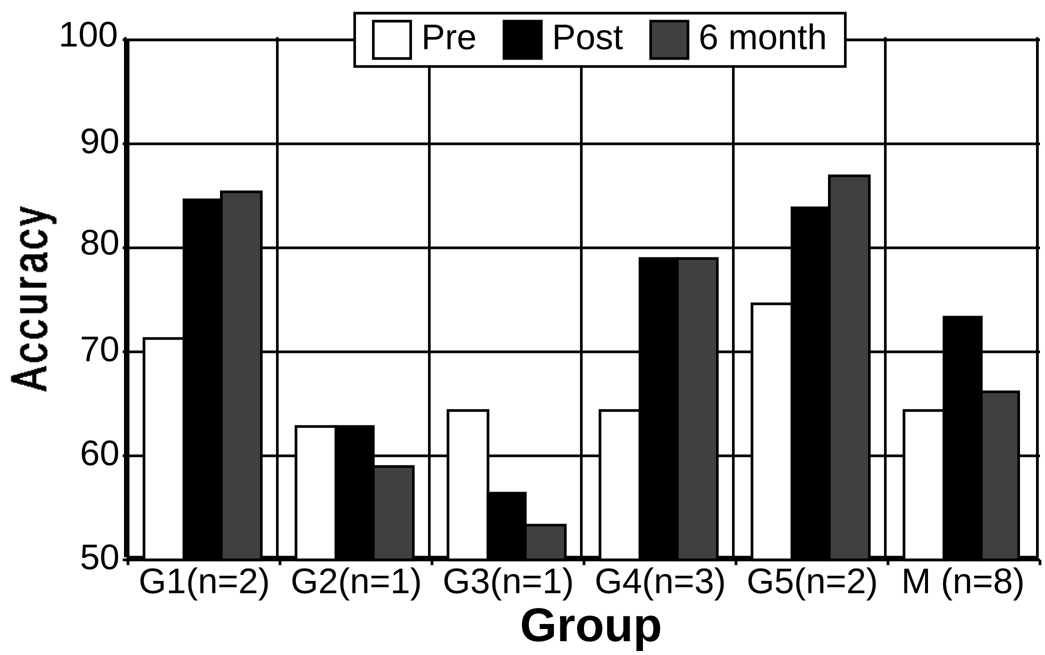

Figure 6: *Accuracy on the 6-month follow-up test on the post-test materials by group.*

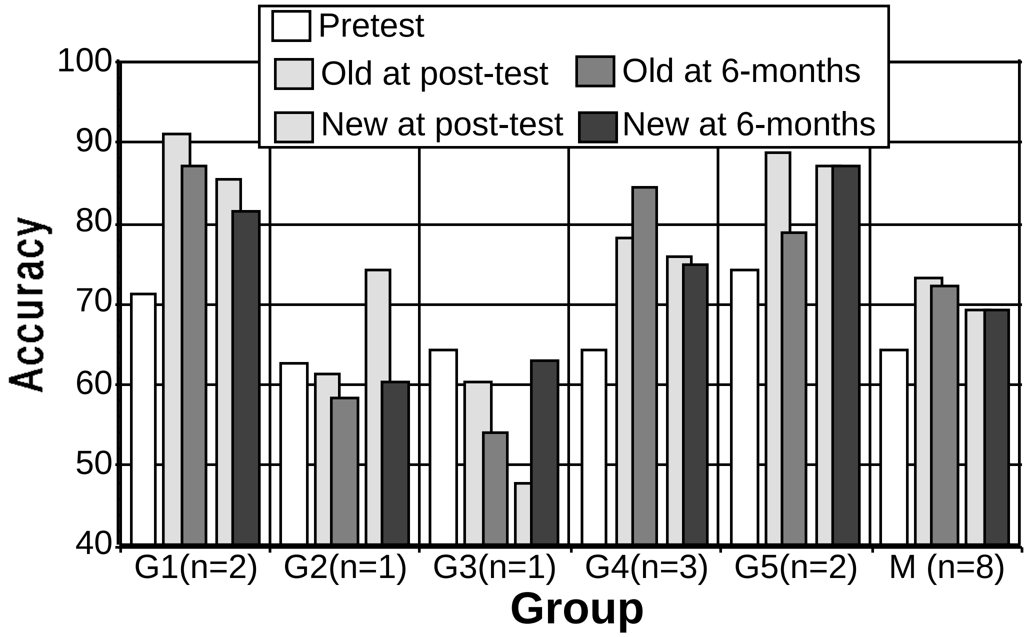

Figure 7: *Generalization accuracy in the 6-month follow-up test.*