The Role of Talker Variability in Non-Native Phoneme Training *

◎ James S. Magnuson, Reiko A. Yamada, Yoh'ichi Tohkura
(ATR Human Information Processing Laboratories)
David B. Pisoni, Scott E. Lively, Ann R. Bradlow
(Indiana University)

In previous reports on a high-variability /r/-/l/ identification training paradigm, we have stressed the importance of variability in the training stimuli [3, 1, 2], especially in phonetic context and number of talkers. This training method has proven effective for modifying the phonetic categories of adult, monolingual speakers of Japanese: it leads to significant improvement in a test given before and after training, generalization to new stimuli and new talkers, and retention of the modified phonetic categories for at least six months [2].

In one previous experiment, a group of subjects was trained with stimuli produced by only one talker [1]. Although these subjects showed improvement in training, they did not perform significantly more accurately in a post-test than in a pretest – unlike subjects trained with stimuli produced by five talkers. In a test of generalization, they showed less generalization to new words produced by the training talker than subjects trained with several talkers, and failed to generalize to stimuli produced by a new talker. The results indicated that training with a single talker did not expose subjects to sufficient variability to prepare them to generalize to new stimuli.

The current experiment was designed to test this conclusion further. In order to examine the possibility that single-talker training may be more effective with some talkers than others, five groups of subjects were trained with stimuli produced by different single talkers. Furthermore, we examined retention of learning in three- and six-month follow-up tests.

1 Method

1.1 Subjects

44 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 studied English for approximately six years in Japanese schools, although conversation skills were not emphasized. No subjects reported any history of hearing or speech disorders.

1.2 Stimuli

The training materials consisted of the 68 minimal pairs of real English words contrasting /r/ and /l/ produced by three male and two female talkers used in our previous training studies (see [3] for details).

Two sets of stimuli were used in the tests of generalization. The first set consisted of 99 novel words produced by each of the five talkers heard in training. The second set of 96 novel items was produced by a new male native speaker of English.

The pretest–post-test materials consisted of forty-eight items used in previous training studies [3, 1, 2], produced by a male talker not used in training or generalization. These consisted of sixteen minimal pairs contrasting /r/ and /l/, and eight pairs contrasting phonemes other than /r/ and /l/. Only the /r/-/l/ pairs were evaluated in the analyses.

1.3 Procedure

Subjects were randomly assigned to one of five training groups. Each group was trained using stimuli produced by a different one of the five talkers used in previous training studies. In the training phase, a two-alternative identification paradigm was used. Subjects were seated at computer terminals and wore headphones. On each trial, subjects saw orthographic forms of a minimal pair of real English words contrasting /r/ and /l/ displayed on the CRT. Subjects then heard one of the words played through the headphones. If subjects thought the audio stimulus matched the word on the left side of the screen, they pressed the "1" key. If subjects thought the audio stimulus matched the word on the right side of the screen, they pressed the "2" key. Subjects were given immediate feedback. The subjects participated in fifteen training sessions (three per day for five days). In each session, both members of each of the 68 minimal pairs were presented twice, for a total of 272 trials. This was the same number of sessions and trials used in the previous training experiments using five talkers [3, 1, 2].

The effects of the training sessions were assessed by comparing performance on a pretest and a post-test administered before and after the training period. Generalization to new words and a new talker was tested after the post-test. Some subjects returned three and six months after the post-test to be tested on the test items again. The test and generalization sessions employed the same two-alternative identification paradigm used for training, although no feedback was given.

2 Results

2.1 Pretest–Post-test

Mean accuracy scores were submitted to an analysis of variance (ANOVA). Training group (G1, G2, G3, G4 and G5) was a between-subjects variable. Test (pre- vs. post-test) was a within-subjects variable.

Accuracy on the post-test was significantly higher than accuracy on the pretest (pretest = .647, post-test = .726; $F(1,4)=29.657$, $p<.001$). However, not all groups improved, as can be seen in Figure 1. Tukey-Kramer HSD post-hoc comparisons showed

*外国語音知覚訓練における話者変動の役割 J.S. マグナソン、山田玲子、東倉洋一 (ATR 人間情報通信研究所) D.B. ピゾーニ、S.E. ライブリィ、A.R. ブラドロウ (インディアナ大学)

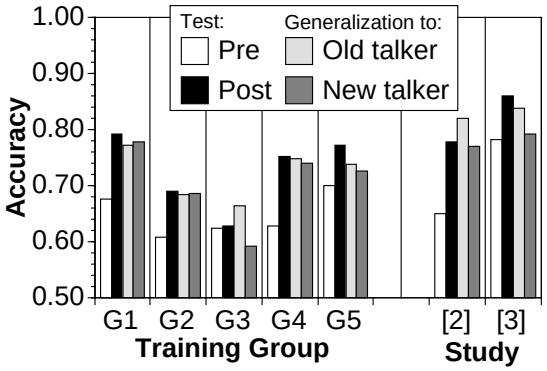

Figure 1. Accuracy by group on test and generalization materials. Data from this experiment is on the left, data from previous studies ([2] and [3]) is presented for comparison.

that only groups 1 and 4 improved significantly ($p<.05$) between pretest and post-test.

2.2 Training

The main effect of training day was highly significant, with accuracy increasing substantially from .678 on day 1 to .831 on day 5 ($F(4,16)=235.835$, $p<.001$). There was also a significant main effect of training group ($F(4,39)=3.447$, $p<.05$). Post-hoc tests showed that the only pair-wise comparison of groups to reach significance was group 4 ($M=.831$) vs. group 3 ($M=.700$). However, the interaction between day and training group was not significant. In Figure 2, we show the interaction of day and training group. All groups improved: post-hoc tests comparing accuracy on day 1 with accuracy on day 5 were significant for all groups.

2.3 Generalization

The results of the tests of generalization are also shown in Figure 1. In order to quantify the generalization scores, paired t-tests were performed comparing generalization performance on old and new talkers with pretest performance and performance in the first training session for each group. According to this comparison, groups 1 and 4 were significantly more accurate in generalization tests than in the pretest and first training session.

2.4 Follow-up tests

Twenty subjects returned for the three month follow-up test. An ANOVA showed that the between-subjects effect of training group was not significant. The effect of test was significant ($F(2,30)=7.56$, $p<.01$), with accuracy higher in both the post-test and three-month test than in the pretest for the twenty subjects. However, pair-wise post-hoc comparisons showed that only group 4 was significantly more accurate in the three-month test than in the pretest. The other groups were more accurate in the three-month test than in the pretest, but not significantly so, and were not significantly less accurate in the three-month test than in the post-test.

Ten subjects returned for a six-month follow-up test with the pretest–post-test materials. Of the subjects that returned, only subjects in groups 1 and 4 were substantially more accurate in the six-month test than in the pretest. Thus, the two groups that showed significant pretest–post-test gains also showed substantial retention after six months.

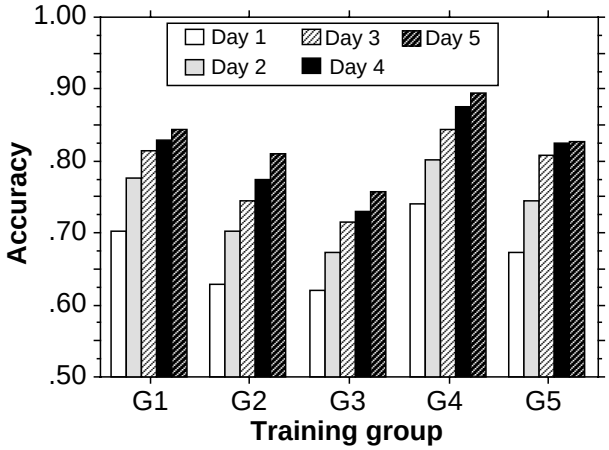

Figure 2. Training accuracy by day and group.

3 Discussion

The previous one-talker training results [2] were partially replicated (see Table 1): although all groups showed significant improvement within the training sessions, subjects in three of five groups did not show significant improvement from pretest to post-test, did not generalize well to new stimuli, and did not show good retention in three- and six-month follow-up tests. However, subjects in two of the five groups improved significantly from pretest to post-test, generalized well to new stimuli, and showed retention comparable to that of subjects trained with multiple talkers, even six months after the training sessions. 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 under certain conditions.

Table 1. Summary of results.

Group	Post-test	General-ization	Follow-up test
G1	*	*	*
G2			
G3			
G4	*	*	*
G5			

* Significantly better than pretest

REFERENCES

[1] 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. J. Acoust. Soc. Am., 94, 1242-1255.

[2] 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. J. Acoust. Soc. Am., 96, 2076-2087.

[3] Logan, J. S., Lively, S. E., and Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. J. Acoust. Soc. Am., 89, 874-886.