DOES TRAINING IN SPEECH PERCEPTION MODIFY SPEECH PRODUCTION?

Reiko Akahane-Yamada[1] *Yoh'ichi Tohkura*[1] *Ann R. Bradlow*[2] *David B. Pisoni*[2]

[1]ATR Human Information Processing Research Laboratories, Kyoto, Japan
[2]Indiana University, Bloomington, U.S.A.

ABSTRACT

To examine the relationship between speech perception and production in second language acquisition, this study investigated whether training in the perception domain transfers to improvement in the production domain. Native speakers of Japanese were trained to identify English /r/-/l/ minimal pairs. Recordings were made of the subjects' productions of minimal pairs before and after identification training. American-English listeners then perceptually evaluated these productions. The subjects showed significant improvements from pretest to post-test in perception as well as in production. Furthermore, the subjects retained these abilities in follow-up tests given three months and six months after the conclusion of training. These results demonstrate that training in the perception domain produces long-term modifications in both perception and production, implying a close link between speech perception and production.

1 Introduction

The relationship between speech perception and production has been a long-standing issue in second language (L2) acquisition as well as in native language (L1) acquisition. It is well-known that some phonetic contrasts in one language are difficult for speakers of another language to perceive and produce. For example, the English /r/-/l/ contrast is remarkably difficult for Japanese speakers to perceive and produce even after many years of education in English as an L2, or immersion in an English-speaking environment. Yamada et al. (1994 [6]) have shown a significant correlation between perception accuracy and production intelligibility of English /r/-/l/ tokens by Japanese speakers. This result implies a link between perception and production in L2 acquisition. However, few studies have examined this perception-production link directly. One way of addressing this issue is to investigate the effects of artificial changes in one domain, either perception or production, on the other domain.

In the past, laboratory training procedures in which Japanese speakers were trained to perceptually distinguish /r/ and /l/ using a discrimination task with synthesized stimuli, met with only limited success. The effect of discrimination training did not generalize to the identification of natural tokens (Strange & Dittmann,1984 [4]). Recently, however, a series of new studies (Logan et al., 1991 [3], Lively et al.,1994 [2]; Yamada, 1993 [5]) has demonstrated that laboratory training can be successful. For instance, when Japanese speakers were trained on /r/-/l/ minimal pairs using an identification task with natural tokens, the effects generalized to novel talkers and novel words. This result was obtained when the training stimulus set consisted of minimal pairs that contrasted /r/ and /l/ in various phonetic environments, and that were produced by multiple talkers. Furthermore, this ability was retained even three or six months after the conclusion of training.

Based on these initial findings, we have started examining effects of the changes in the perception domain on speech production (Bradlow et al., 1995 [1]). The present study investigated whether training in the perception domain transfers to an improvement in the production domain. In addition, we re-tested the Japanese subjects three months and six months after the conclusion of the perceptual training in order to assess the retention of this form of learning.

2 Experiment

2.1 Subjects

Twenty-three native speakers of Japanese served as subjects. All of the subjects had studied English since junior high school from about the age of 12, but none had lived abroad or had special training in English conversation. All of them reported no history of any speech or hearing disorder. A hearing screening confirmed all subjects to have normal bilateral hearing acuity. Eleven subjects (five females and six males) were randomly selected and assigned to the trained group, which received perception training. The remaining 12 subjects (six females and six males) were assigned to the control group, and participated only in the pre- and post-tests. The subjects ranged in age from 19 to 22 (average 20) for the trained group, and from 18 to 22 (average 20) for the control group.

2.2 Stimuli

The stimuli were identical to those used in the earlier series of studies training native speakers of Japanese to identify English /r/-/l/ minimal pairs (Logan et al., 1991 [3], Lively et al., 1994 [2], Yamada, 1993 [5]). English words contrasting /r/ and /l/ in various positions (e.g., word-initial singleton, word-initial consonant cluster, intervocalic, word-final singleton, and word-final consonant cluster) were used as the stimulus materials.

In the training, 136 words (68 /r/-/l/ minimal pairs) produced by five

talkers were used. In the pretest and post-test, the same 24 minimal pairs used by Strange and Dittmann (1984 [4]) were used. Sixteen pairs contrasted /r/ and /l/, and the other eight pairs were filler pairs which contrasted phonemes other than /r/ and /l/. These words were produced twice by a "new" male talker, i.e. a talker who was not one of the five training talkers. In one generalization test, novel words from /r/-/l/ minimal pairs produced by one of the training talkers were used. In a second generalization test, novel words produced by a new talker were used.

2.3 Procedure

The experiment employed the pretest/post-test design used by Strange and Dittmann (1984 [4]). Table 1 shows the experimental schedule. The perception pretest and post-test were administered before and after the training period, and generalization to the new words and new talker were tested only after the training period was completed. The trained group received 45 sessions of perception training over 15 days (three sessions per day). The training started on the day following the pretest. The subjects in the control group performed the pretest and then came back for the post-test phase approximately four weeks after the pretest. The method used in the perception tests and training followed the procedures first developed by Logan et al. (1991 [3]), and later modified by Yamada (1993 [5]). This method emphasized perceptual identification rather than discrimination. Recordings were made of the subjects' productions of /r/-/l/ minimal pairs in the pretest and post-test phases. Recordings were also made three months and six months after conclusion of the training, in order to assess the retention of the production ability. For the control group, only a three-month follow-up recording was administered. These productions were later evaluated by native speakers of American English (AE). The training and tests were administered at ATR HIP Lab; the perceptual evaluations of the productions were done in the Speech Research Lab at Indiana University.

Phase	Trained Group	Control Group
pretest phase	(day1) pretest recording	(day1) pretest recording
training	(day2-16) 45 training sessions	-
post-test Phase	(day17) post-test, gen-1,2 recording	(day2) post-test, gen-1,2 recording
3M phase	(3 months after) recording	(3 months after) recording
6M phase	(6 months after) recording	-

Table 1: Schedule of the experiment. Gen-1 and gen-2 indicate a generalization test for new items by a trained talker and by a new talker, respectively.

Test In the perception tests (pretest, post-test, and generalization tests), a two alternative forced choice (2AFC) task was used. On each trial, two members of a minimal pair were displayed on two buttons shown on a CRT monitor. One of the members was then played over headphones (Stax SR-Λ). The subjects chose one of the words by pressing the appropriate key. There was no feedback for the subjects' responses.

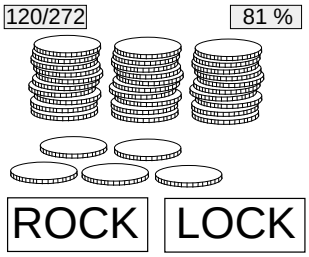

Figure 1: An example display on the CRT monitor during a training session. The response alternatives are shown on the two buttons at the bottom. The trial number (left-top corner), the accumulative correct response rate (right-top corner), and the graphical coins used as rewards are also presented.

Training The same task used in the tests was used during the training phase. However, feedback was provided for the subjects' responses. A chime (correct) or buzzer (incorrect) sounded after each response. The accumulative accuracy rate was always displayed on the CRT monitor. In addition, one graphical coin was added every time a subject made three correct responses (see Figure 1). Sixty-eight minimal pairs spoken by one talker were presented twice in one training session, yielding 272 trials per session. Each session lasted approximately 30-40 minutes. Five talkers cycled from T1 to T5 session-by-session.

Production Recordings In the production recording a repetition task was used in which a subject read a set of 55 English /r/-/l/ minimal pairs from a list of randomly ordered words. The subjects were provided with auditory prompts (words produced by a native speaker of General American English) as well as visual prompts (a list of words). Approximately half of the pairs were old pairs that appeared in the perception training; the remaining pairs were novel words. The aim of the auditory prompts was to provide the subjects with a model of how to pronounce the test words. Recordings were digitized at a sampling rate of 22.05kHz with 16-bit resolution at ATR HIP Lab, and digitally transferred to the Speech Research Lab at Indiana University where they were rescaled to 12-bit resolution.

Intelligibility A panel of 10 AE listeners for each Japanese subject evaluated the intelligibility of the productions from the pretest and post-test phases. A 2AFC task contrasting members of the /r/-/l/ minimal pairs was used for these perceptual evaluations.

Paired-Comparison Another panel of 10 native AE listeners for each Japanese subject provided preference ratings between the pretest and post-test versions of each test word using a paired-comparison

task. Listeners heard a single Japanese subject's pretest and post-test utterances of a word over headphones with an ISI of 500ms. They then indicated which version of the target word was better pronounced and to what degree with a rating between 1 and 7, where 1 indicated that the first version was much better than the second, 4 indicated no difference between the two versions, and 7 indicated that the second version was much better than the first. On half of the trials, the pretest version was presented first, and on the other half, the post-test version was presented first. All of the 110 pre-post pairs in each of the two presentation orders were presented in random order.

Retention In order to assess the retention of these novel production abilities, productions from the following pairs of experimental phases were compared using the paired-comparison procedure described above: pretest (pre) vs. three-month follow-up (3M), pre vs. six-month follow-up (6M), post-test (post) vs. 3M, and 3M vs. 6M. Only the pre vs. 3M and post vs. 3M were compared for the control group.

3 Results

3.1 Improvement in perception

Performance on the perception tests and training sessions averaged over the subjects in each group are shown in Figure 2. Subjects in the trained group improved in accuracy during the training. These perceptual learning effects generalized to untrained words by an untrained talker. The performance on each test was subjected to a one-factor ANOVA with test as the variable. Accuracy on the post-test was significantly higher than on the pretest [65% correct on the pretest vs. 81% correct on the post-test; $F(1,10)=54.5$, $p<0.0001$]. Performance in the two generalization tests also showed a high accuracy of 83% correct in a test of new words by one of the trained talkers, and 80% correct in a test of new words by a new talker.

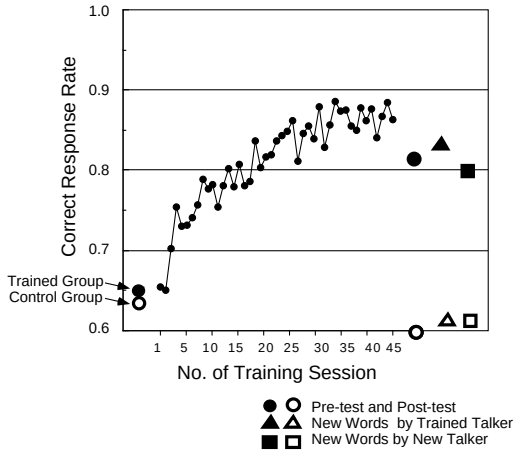

Figure 2: Accuracy in perception in the pretest, each training session, post-test and two generalization tests.

3.2 Intelligibility of productions

Intelligibility scores (in terms of how often the AE listeners' judgments matched the talkers' intended words in the 2AFC task between members in minimal pairs) were calculated for each subject's pretest and post-test productions. Figure 3 shows intelligibility scores for the pretest and post-test. An ANOVA showed a significant effect of the test (pre vs. post) for the trained group; the intelligibility improved significantly from pretest to post-test [$F(2,22)=12.946$, $p<0.001$]. Furthermore, both trained and un-trained words improved significantly from pretest to post-test, but there was no difference between trained and un-trained words within post-test productions. In contrast, there were no significant differences between the pretest and post-test productions for the control group.

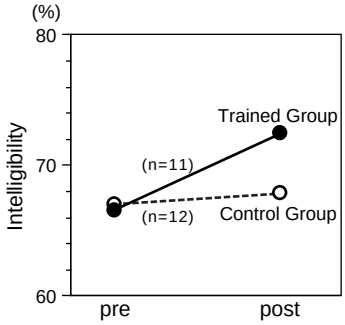

Figure 3: Intelligibility of subjects' pretest and post-test productions as judged by AE listeners.

3.3 Preference between pretest and post-test productions

To deal with the counter-balanced order of the stimulus presentation, the rating scores were recoded so that a response of "5" or higher corresponded to a preference for the post-test version, and a response of "3" or lower corresponded to a preference for the pre-test version. We compared the proportion of trials that received "1-3," indicating a preference for the pretest token, and the proportion that received "5-7," indicating a preference for the post-test token (circles in Figure 4). For subjects in the trained group, the proportion of trials in which the post-test version was preferred was significantly higher than the proportion of trials in which the pretest version was preferred. In contrast, there was no difference in preference for utterances produced by the control group.

3.4 Retention

Three-month follow-up Tokens from the pretest phase (pretest tokens) and from the 3-month follow-up (3M tokens) were compared using the same paired-comparison method as described above. The 3M tokens received a higher rate of "preferred" judgment than the pretest tokens for the trained group, but not for the control group (triangles in Figure 4). There was also no difference in preference between the post-test tokens and the 3M tokens for the trained group (squares in Figure 5).

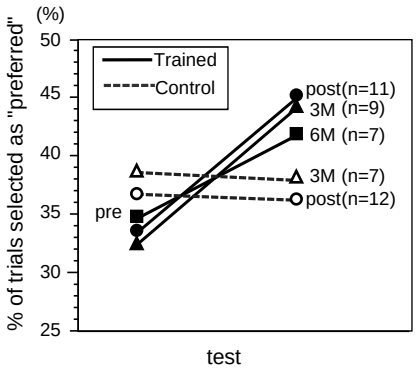

Figure 4: Preference comparison between tokens from two experimental phases. Pretest (pre) versions were compared to post-test (post) and follow-up test (3M and 6M) versions. The percentages of trials that received a "preferred" in the paired-comparison task as judged by AE listeners are shown.

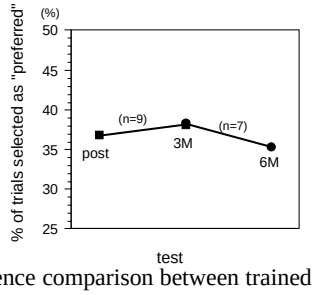

Figure 5: Preference comparison between trained subjects' tokens from post-test (post) vs. 3-month follow-up test (3M), and 3M vs. 6-month follow-up test (6M). The percentages of trials that received a "preferred" in the paired-comparison task as judged by AE listeners are shown.

Six-month follow-up The 6M tokens received a higher rate of "preferred" judgments than the pretest tokens for the trained group (squares in Figure 4). Also, no significant difference in preference was found between the 3M and 6M tokens for the trained group(circles in Figure 5).

4 General Discussion

The subjects in the trained group showed significant improvements from the pretest to post-test in perception, and these improvements generalized to novel talkers and novel items. This result replicated our previous findings (e.g. [2, 5]). In addition, the results obtained with the control subjects, who did not show significant improvements from the pretest to post-test, verified that the improvements for the trained group were due to the identification training.

More importantly, the production ability of the trained subjects also improved from the pretest to post-test: The intelligibility of the Japanese subjects' productions improved significantly from the pretest to post-test, and post-test productions were more frequently judged as better tokens of English /r/ and /l/ than pretest tokens. This pattern of results demonstrated that training in speech perception transfered to improvements in the production domain, suggesting a close link between speech perception and production in the development of L2 speech contrasts. In contrast, the pretest and post-test productions of the control group did not differ qualitatively, thus providing further support for the hypothesis that the improvements in /r/ and /l/ production for the trained group were due to the perceptual identification training.

The evaluations of the productions from the follow-up tests were somewhat surprising. The productions obtained from the three- and six-month follow-up tests were preferred over the pretest productions for the trained group, but not for the control group. There were no significant differences in preference between productions in the post-test and three-month follow-up, and between three- and six-month follow-up tests for the trained group. These results together with earlier findings of Lively et al. (1994 [2]) suggest that not only the perception improvement but also the production improvement was maintained at the post-test level even 3 or 6 months after completion of the perception training. We conclude here that the perception training produced long-term modifications in both perception and production.

The present results have several important implications. First, our findings provide evidence that phonetic categories are developed under close communication between the perception domain and the production domain. Second, the present results have a practical contribution: When considering the fact that computer-based training in production is more difficult than that in perception, we may suggest that "tuning the trainees' speech perception" will facilitate learning in production. Despite these contributions, further examination is necessary to understand the underlying mechanisms of the perception-production link in the development of new phonetic categories.

5 REFERENCES

1. A.R. Bradlow, D.B. Pisoni, R.A. Yamada, and Y. Tohkura. The effect of training in /r/-/l/ perception on /r/-/l/ production by Japanese speakers. In *Proc. 1995 International Congress on Phonetic Sciences*, pages 562–565, 1995.

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

3. J. S. Logan, S. E. Lively, and D. B. Pisoni. Training Japanese listeners to identify /r/ and /l/: A first report. *Journal of the Acoustical society of America*, 89:874–886, 1991.

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

5. R. A. Yamada. Effect of extended training on /r/ and /l/ identification by native speakers of Japanese. In *125th Meeting of the Acoustical Society of America*, volume 93, No.4, Pt.2, pages 4pSP8, p.2391, 1993.

6. R. A. Yamada, W. Strange, J.S. Magnuson, J. S. Pruitt, and W. D. Clarke III. The intelligibility of japanese speakers' production of American English /r/,/l/, and /w/, as evaluated by native speakers of American English. *Proceedings of the 1994 International Conference on Spoken Language Processing*, pages 2023–2026, 1994.