Published in final edited form as:
*J Acoust Soc Am*. 1991 February ; 89(2): 874–886.

# Training Japanese listeners to identify English /r/ and /l/: A first report

**John S. Logan**, **Scott E. Lively**, and **David B. Pisoni**
Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington, Indiana 47405

## Abstract

Native speakers of Japanese learning English generally have difficulty differentiating the phonemes /r/ and /l/, even after years of experience with English. Previous research that attempted to train Japanese listeners to distinguish this contrast using synthetic stimuli reported little success, especially when transfer to natural tokens containing /r/ and /l/ was tested. In the present study, a different training procedure that emphasized variability among stimulus tokens was used. Japanese subjects were trained in a minimal pair identification paradigm using multiple natural exemplars contrasting /r/ and /l/ from a variety of phonetic environments as stimuli. A pretest–posttest design containing natural tokens was used to assess the effects of training. Results from six subjects showed that the new procedure was more robust than earlier training techniques. Small but reliable differences in performance were obtained between pretest and posttest scores. The results demonstrate the importance of stimulus variability and task-related factors in training nonnative speakers to perceive novel phonetic contrasts that are not distinctive in their native language.

## INTRODUCTION

When listeners are presented with speech stimuli from phonetic categories that are not used in their own language they typically show performance that is not as good as a native speaker of the language from which the phonemes were selected (e.g., Miyawaki *et al.,* 1975; Werker and Logan, 1985). This phenomenon has both practical and theoretical implications. In a practical sense, this means that an individual learning a second language may experience difficulty distinguishing certain phonetic contrasts in the second language. From a theoretical viewpoint, the phenomenon also poses several interesting questions: How did these language-specific linguistic categories arise?. How flexible is the adult perceptual system, with regard to novel phonetic categories? What conditions facilitate the development of novel phonetic categories in adults? The work described in the present paper focuses on several of these theoretical issues in the context of training Japanese listeners to identify the English phonemes /r/ and /l/. In particular, we examine several of the theoretical assumptions underlying previous efforts to train listeners to distinguish nonnative phonetic contrasts and consider some of the issues raised by this earlier work (see also Pisoni *et al.,* 1991).

A developmental account of the problems facing adult second language learners is a useful starting point for considering the points raised above because of the potential parallels between the acquisition of phonetic categories in the child's first language and acquisition of nonnative phonetic categories while learning a second language in adulthood. Werker (1989) has shown that within the first year of life the infant begins to move from language-universal abilities to the language-specific abilities that are characteristic of the adult.

Language-universal refers to how infants are able to discriminate virtually any phonetic contrast used in a language, regardless of the environment in which they are raised, while language-specific refers to the much more restricted abilities of mature adults to discriminate or identify stimuli from phonetic categories not used in their native language. The transition from language-universal to language-specific abilities appears to be a product of the interaction between innate perceptual mechanisms and early linguistic experience (Aslin and Pisoni, 1980). Early experience serves to modify the child's perceptual system so that only those phonetic contrasts that denote differences in meaning remain distinctive.

Jusczyk (1989) has recently suggested that attentional mechanisms underlie the modification of phonetic categories during development. Based on the recent work of Nosofsky (1986, 1987), Jusczyk has argued that attention to the stimulus dimensions differentiating phonetic categories can cause a change in the perceived similarity of stimuli varying along the dimensions in question. Nosofsky (1986) has shown that subjects presented with visual stimuli varying in several dimensions can selectively attend to a specific dimension, thereby affecting the perceived similarity of the stimuli. Stimuli varying along the attended dimension became more distinct from each other while stimuli varying along the unattended dimension became more similar to each other. According to Nosofsky, selective attention causes a "stretching" of the psychological distance along the attended dimension and a "shrinking" of distance along the unattended dimension. Jusczyk has taken Nosofsky's work on selective attention and applied it to how auditory dimensions become differentially weighted during perceptual development. The relative salience of dimensions that differentiate phonetic categories is modified as a consequence of the linguistic importance of the dimension to the formation of the categories during development. To the extent that Jusczyk's proposal is a plausible model of how adult phonetic categories are formed, it suggests that if a novel phonetic contrast is to be learned by adults, selective attention must be re-allocated to new acoustic-phonetic dimensions that were previously unattended.

Much of the work involving adult cross-language speech perception is based on findings obtained from experiments investigating the phenomenon of categorical perception (e.g., Liberman *et al.*, 1957), in which a listener's ability to discriminate a pair of stimuli appears to be limited to those stimuli the individual classifies as belonging to separate categories in an identification task (Studdert-Kennedy *et al.,* 1970). Under certain conditions, however, within-category discrimination is possible. Factors that facilitate the discrimination of within-category stimuli have generally relied on improving access to sensory memory (Pisoni, 1973; Pisoni and Lazarus, 1974), using minimal uncertainty procedures (Carney *et al.,* 1977), and extensive training of listeners (Samuel, 1977).

Some of the same procedures used to demonstrate within-category speech perception have also been utilized in cross-language investigations. For example, Werker and her colleagues (Werker and Tees, 1984; Werker and Logan, 1985) hypothesized that listeners could discriminate a non-native speech contrast if they had access to information in sensory memory. Werker employed a methodology developed by Pisoni (1973) in which the interval separating stimuli in an AX discrimination task was varied. She used natural CV stimuli from a place of articulation contrast occuring in Hindi and found that native speakers of English could discriminate these stimuli more accurately if the interval between the stimuli was reduced.

Strange and Dittmann (1984) also employed a procedure that had been used to demonstrate within-category perception. They attempted to modify Japanese listeners' perception of /r/ and /l/ using a psychophysical procedure developed by Carney *et al.* (1977). In this procedure, the first stimulus in an AX pair is always the same for a block of trials and only the second stimulus of the pair varies. Carney *et al.* showed that after several hundred training trials using this task, listeners could discriminate very small within-category differences between stimuli at various locations along a VOT (voice onset time) continuum. Strange and Dittmann used the same procedure to try and produce long-term changes in the phonological system of Japanese listeners. They trained Japanese subjects to discriminate stimuli from a synthesized "rock"–"lock" continuum for 14 to 18 sessions. The effectiveness of the training procedure was evaluated using a pretest–posttest design with natural speech tokens in which initial levels of performance were compared to performance after discrimination training. Strange and Dittmann found that although discrimination of the synthetic stimuli improved during training, the effects did not generalize to the natural speech stimuli used in the posttest. The authors summarized their findings by stating, "…we cannot conclude that this training experience generalized to perception of the phoneme contrast in real speech by a native AE (American English) speaker."

Before describing the goals of the present investigation, it is useful to consider what is known about the perception of /r/ and /l/ by Japanese listeners. The difficulty that Japanese listeners have with the /r/–/l/ contrast is well documented in the literature. Even after years of living in an English speaking environment, the performance of Japanese listeners presented with synthesized stimuli contrasting /r/ and /l/ in identification and discrimination tasks differs from native speakers of English when tested in the same tasks (MacKain *et al.,* 1981). The differences in performance between inexperienced Japanese subjects and native speakers of English are even larger; inexperienced Japanese generally have more poorly defined category boundaries and their discrimination functions are typically flat and close to chance performance (Miyawaki *et al.,* 1975; MacKain *et al.,* 1981; Mochizuki, 1981).

The problems that native speakers of Japanese have with the /r/–/l/ contrast are not confined to synthesized stimuli. Natural speech contrasting /r/ and /l/ is also poorly perceived by Japanese listeners, whether it is produced by a native English speaker or a native Japanese speaker (Goto, 1971; Sheldon and Strange, 1982). Several studies have shown that the performance of Japanese subjects presented with /r/ and /l/ is not uniform but instead is dependent on the phonetic context in which /r/ and /l/ appear (Gillette, 1980; Mochizuki, 1981; Sheldon and Strange, 1982). In general, identification performance is poorest for /r/ and /l/ in initial positions in either singleton and consonant cluster environments and intervocalic positions, while performance is much better for /r/ and /l/ in final position, in either singleton and cluster environments. According to Sheldon and Strange (1982), while there is no obvious *a priori* phonological explanation for these context effects, preliminary acoustic analyses of stimuli containing /r/ and /l/ suggests that there are systematic acoustic differences among the different phonetic environments (see also Dissosway-Huff *et al.,* 1982).

The overall goal of the present investigation was to develop a set of procedures that would be more likely to result in the learning of nonnative speech contrasts. We began by considering the methodology used by Strange and Dittmann (1984). Strange and Dittmann wanted to find out if subjects could transfer what they had learned about /r/ and /l/ from a low-level psychophysical discrimination task to other kinds of testing situations and if training subjects with a small number of synthetic stimuli containing /r/ and /l/ in initial position would be sufficient to form phonetic categories for /r/ and /l/ across a variety of other phonetic environments.

Strange and Dittmann's (1984) results showed that Japanese listeners failed to demonstrate a significant improvement in their ability to perceive /r/ and /l/ in a generalization task with naturally produced real words. There are several reasons why this may have happened. First, we consider the AX training task used by Strange and Dittmann. A fixed-standard AX discrimination task enables listeners to make use of low-level, sensory-based information in the speech signal. Unfortunately, training that emphasizes information contained in sensory memory has a high probability of failing to generalize to other conditions such as a minimal pair identification task, unless some more permanent memory code can be developed (Pisoni, 1973, 1975). Second, the choice of stimuli used by Strange and Dittman assumed that listeners would generalize what they learned about /r/ and /l/ in initial position and apply it to /r/ and /l/ in other positions. Although adult native speakers of English may treat /r/ or / l/ similarly regardless of phonetic environment, previous work (e.g., Mochizuki, 1981; Sheldon and Strange, 1982) has shown that the performance of Japanese listeners presented stimuli containing /r/ and /l/ varies systematically as a function of phonetic environment. As a consequence, training that assumes such an equivalence by presenting stimuli containing / r/ and /l/ from only one environment may not generalize to other phonetic contexts (cf. Jamieson and Morosan, 1986).

The choice of the training procedure and stimuli used in the present experiment was designed to circumvent some of the difficulties associated with Strange and Dittmann's study. As stated earlier, one goal of the present investigation was to determine under what conditions a group of native Japanese speakers could learn to identify the English phonemes /r/ and /l/. Moreover, we wanted to develop a training procedure that would promote transfer of knowledge acquired during training to novel stimuli that subjects had not been exposed to before. The present experiment utilized the same basic pretest–posttest design used by Strange and Dittmann to assess the effects of training. However, several changes were made in the training procedure.

First, a two-alternative forced-choice identification (ID) task was used during training. Listeners trained in an ID task are forced to develop and use phonetic memory codes in short-term memory rather than rely on information in sensory memory. An ID task also encourages classification of stimuli into categories instead of enhancing the perception of fine within-category acoustic differences that would be obtained using an AX discrimination task (Jamieson and Morosan, 1986), such as that employed by Strange and Dittmann. Furthermore, since an ID task was used during the training phase and in subsequent posttest and generalization phases of the experiment, transfer of the knowledge gained during training to each successive phase would be anticipated.

Second, the stimuli chosen for the present experiment were minimal pairs of English words contrasting /r/ and /l/ in several different phonetic environments produced by five different talkers. The motivation for selecting these stimuli involved a consideration of the role of stimulus variability in perceptual learning [or what Jamieson and Morosan (1986) call *acoustic context* and *uncertainty*]. As pointed out earlier, training that uses only a small number of stimuli varying in word-initial position assumes an equivalence of /r/ and /l/ across different phonetic contexts. Presenting stimuli containing /r/ and /l/ from various phonetic contexts exposes Japanese listeners to the full range of acoustic-phonetic cues that characterize /r/ and /l/ across different environments.

Natural speech tokens instead of synthesized speech were also used for similar reasons. Given that the acoustic cues signaling specific phonetic categories in synthetic speech may be considered impoverished compared to natural speech (Pisoni *et al,* 1985), subjects trained with synthetic speech may be exposed to misleading or incomplete information about the cues for the phonetic categories they are expected to learn, a vital concern when attempting to obtain stimulus generalization to novel tokens. The use of multiple talkers was a further attempt to form robust phonetic categories by increasing stimulus variability during learning. Different talkers produce widely varying acoustic output due to differences in vocal tract size and shape, glottal source function, dialect, and speaking rates. This additional stimulus variability may be important for developing stable and robust phonetic categories that show perceptual constancy across different environments (cf. Kuhl, 1983).

Although we considered the possibility that variability in talker and phonetic environment might pose initial difficulties for the Japanese listeners, the potential long-term benefits in terms of developing new phonetic categories outweighed these concerns. Support for our assumption that stimulus variability can benefit category learning comes from a well-known study by Posner and Keele (1968). They compared the performance of two groups of subjects in classifying visual stimuli. One group was trained on stimuli that had a high degree of variability, while the other group was trained on stimuli that had a low degree of variability. Posner and Keele found that when subjects from the two groups were presented with novel stimuli, the group that had received the high-variability stimuli during training performed much better than the group that had received low-variability stimuli during training.[1]

Finally, in addition to assessing the effectiveness of training in a pretest–posttest design, we also tested generalization by presenting subjects with novel words produced by both old and new talkers. Recently, Mullennix *et al.* (1989) found that talker variability has important perceptual consequences for the speed and accuracy of processing spoken words. In light of these findings, we were interested in assessing whether subjects would generalize to novel stimuli and the extent to which correct generalization would depend on the specific characteristics of the talkers used during training. We assumed that the variability of the talkers used during training would be sufficient to overcome talker-specific learning. That is, after training, subjects should be able to correctly identify novel words containing /r/ and /l/ produced by a novel talker. Such a test could be considered to be the most stringent measure for assessing whether the training procedure provided the Japanese listeners with sufficient information to form robust phonetic categories for /r/ and /l/ that would generalize to other contexts.

# I. METHOD

## A. Subjects

The subjects were six native speakers of Japanese living in the Bloomington, Indiana area. All were students at Indiana University and all had lived in the U.S. for periods ranging from 6 months to 3 years at the time of testing. All the subjects reported that they considered their proficiency with spoken English to be less than their ability to deal with written English. No subjects reported any history of a speech or hearing disorder. Subjects were paid $5.00 for each session.

## B. Stimuli

A computerized database containing approximately 20 000 words (*Webster's Seventh Collegiate Dictionary,* 1967) was searched to locate all minimal pairs 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 positions. Six talkers, four male and two female, recorded 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 recording booth. The words were low-pass filtered at 4.8 kHz and digitized at 10 kHz using a 12-bit analog-to-digital converter. The digitized

[1] The role of stimulus variability in the perceptual learning of /r/ and /l/ was acknowledged by Strange and Dittmann (1984) themselves when they suggested that future work should "include training of the contrast with more than one set of stimuli and in more than one phonetic context."

waveform files were edited and then equated for rms amplitude using a specialized signal processing package.

The stimuli were pretested with a separate group of native speakers of English to assess their intelligibility. An identification task was used in which listeners typed their response on a terminal after hearing each stimulus. The criteria for including a word in the experiment was that it have no more than a 15% error rate across all talkers and that no errors were due to misperception of /r/ or /l/. After pretesting, a set of 136 stimuli (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. A set of 96 additional stimuli (38 initial singleton words, 29 initial cluster words, 1 intervocalic word, 17 final singleton words, and 11 final cluster words) from a sixth talker (a male talker) were selected for use in one generalization test. A set of 98 additional stimuli (37 initial singleton words, 32 initial cluster words, 1 intervocalic word, 15 final singleton words, and 13 final cluster words) from one of the talkers used in the training set (Talker 4, a female talker) was selected for use in a second generalization test.[2]

Finally, the 32 words used by Strange and Dittmann (1984) in the pretest–posttest phase of their experiment were recorded by a male talker not included in either the training and generalization stimulus sets for use in the pretest and posttest phases of the present experiment. These stimuli were processed in the same way as the other stimuli used in the present experiment.

## C. Procedure

The experimental design employed a pretest–posttest procedure closely modeled after the procedure used by Strange and Dittmann (1984). In this design, the effects of training were assessed by comparing performance in a pretest and a posttest administered before and after a 3-week training period. Before training began, subjects were presented 16 minimal pairs contrasting /r/ and /l/, each presented twice (pretest phase). Subjects were required to identify the word presented from a minimal pair printed in an answer booklet by marking the correct response. The same test items were presented again after training (posttest phase). The words used in the pretest and posttest were the same as those used by Strange and Dittmann (1984). The pretest–posttest procedure required approximately 20 min to complete. The pretest was administered twice to three subjects prior to training in order to assess the extent to which mere exposure to the words used in the pretest might produce improvements in performance in the posttest. A 2-week period elapsed between the administration of the first and second pretest during which the subjects did not participate in the experiment.

The training phase also used a two-alternative identification task. Subjects were presented with a word from a minimal pair contrasting in /r/ and /l/. They were required to identify the stimulus presented from a minimal pair presented on a CRT screen by pressing a button on a response box. Feedback was given during the training task. If the subject made a correct response, the next trial began. If the subject made an incorrect response, the minimal pair remained on the CRT screen and a light on the response box corresponding to the correct response was illuminated followed by a second presentation of the stimulus, after which the

[2] Stimulus selection was motivated by our decision to provide as many stimuli from different phonetic environments as practical in both the training and generalization phases of the experiment. One consequence of this decision was that the number of stimuli from different phonetic contexts differed, providing listeners with varying amounts of exposure to /r/ and /l/ in different contexts. *A priori,* one might expect that the unequal distribution of stimuli according to context would influence training more than any of the other phases due to the much larger number of stimuli presented during the training phase. In turn, performance in the posttest and generalization phases might be expected to vary in relation to the distribution of stimuli used in training.

next trial began. 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 15 training sessions. Subjects were tested individually during training. Each session lasted approximately 40 min.

After the posttest phase, three of the subjects were tested again to assess the degree to which training generalized to novel stimuli. The first test of generalization (TG1) consisted of 96 novel words from minimal pairs contrasting /r/ and /l/ produced by a new talker (i.e., a talker not used in either the pretest–posttest phase or the training phase). A second test of generalization (TG2) consisted of 98 novel words from minimal pairs contrasting /r/ and /l/ produced by Talker 4, who 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 that used during training except that subjects did not receive any feedback. The tests of generalization were also administered individually.

Subjects were tested in a quiet sound-treated room containing individual cubicles. Each cubicle was equipped with a desk, a two-button response box, and a CRT monitor. Stimuli were presented over matched and calibrated TDH—39 headphones at 80 dB SPL. Presentation of stimuli and collection of responses was under the control of a laboratory computer (PDP-11/34). During training and tests of generalization, both identification responses and latencies were collected. Latencies were measured from the onset of the stimulus presentation.

## II. RESULTS

### A. Pretest–posttest

Results from the pretest–posttest phase of the experiment will be described first. The subset of subjects who were given the pretest twice prior to training showed no improvement in performance from the first administration (mean percentage of correct responses = 77.0%) to the second administration (mean percentage of correct responses = 76.5%), $t(2) = -0.83$. Thus mere repetition of the test vocabulary provided no reliable improvement in identification performance. No significant difference in pretest performance was found between the group of subjects administered the pretest twice (mean pretest 1 and pretest 2 = 76.8%) and the subjects administered the pretest only once (mean = 80.2%), $t(5) = 0.353$. Thus, in all subsequent analyses, the data for the two groups were combined.

Overall, there was a significant increase in the percentage of correct responses from the pretest (mean = 78.1%) to the posttest (mean = 85.9%), $F(1, 5) = 38.47$, $p < 0.005$. The overall pattern is also reflected in individual subject data (see the Appendix); all listeners without exception showed an improvement in performance from pretest and posttest. We conclude from this result that subjects were, in fact, able to transfer what they learned about /r/ and /l/ during training to the posttest stimuli.[3] This result demonstrates that native speakers of Japanese can learn to reliably identify English /r/ and /l/. Moreover, it contrasts with the null results reported earlier by Strange and Dittmann (1984).

The percentage of correct responses for each of the four phonetic environments in the pretest and posttest is plotted in Fig. 1.[4] An analysis of the four phonetic environments used in the

[3] For purposes of comparison, Strange and Dittmann (1984) obtained the following: In the pretest, overall performance for /r/ and /l/ was 69% while performance for /r/ and /l/ in initial position was 64.1%. In the posttest, performance for /r/ and /l/ in initial position was 69.5% (Strange and Dittmann did not give a value for the overall posttest performance).

[4] Since we used the same pretest and posttest words used by Strange and Dittmann (1984), our analysis of the pretest and posttest data was necessarily limited to the same four phonetic environments they examined.

pretest and posttest showed significantly better performance for words contrasting /r/ and /l/ in final position (mean percentage of correct responses = 96.9%) and intervocalic position (mean = 83.1%) than for words contrasting /r/ and /l/ in initial position (singleton [mean = 80.0%] and initial clusters [mean = 68.2%]), *F*(3,15) = 6.32, *p* <0.01. Moreover, there was a significant interaction between pretest–posttest performance and phonetic environments, *F*(3, 15) =3.1, *p* <0.05. Performance on words from initial clusters and intervocalic environments improved markedly from the pretest to the posttest. In contrast, performance on words from the other two environments improved only slightly from the pretest to the post-test, although it should be noted that even in the pretest, performance was close to ceiling for words contrasting /r/ and /l/ in final singleton position. With minor exceptions, these effects were consistently obtained across individual subjects as well (see the Appendix).

## B. Training

**1. Identification performance—**The results from the training phase of the experiment will be described next. An analysis of variance comparing week (weeks 1–3), talker (talkers 1–5) and phonetic environment (environments 1–5) was carried out; only statistically reliable (*p* < 0.05) main effects and interactions are reported. Figure 2 shows the percentage of correct responses as a function of week. A significant effect of week was obtained, *F*(2, 8) = 14.85, *p* <0.01. Identification accuracy improved significantly from week 1 to week 2, but the improvement from week 2 to week 3 was not statistically reliable. The overall improvement in identification accuracy from week 1 to week 3 during training was mirrored in the data from individual subjects (see the Appendix). In short, although presented a highly variable stimulus set, a significant change in subjects' perceptual mechanisms occurred during the course of training.

Figure 3 shows the percentage of correct responses as a function of the five talkers who produced the stimuli. An examination of Fig. 3 indicates some variability in identification performance among talkers. In the ANOVA, a significant effect of talker was obtained, *F*(4, 16) =21.88, *p*< 0.0001, confirming the trends shown in Fig. 3. Overall, talkers 4 and 5 were significantly more intelligible than talkers 1–3. Individual subjects consistently identified stimuli produced by talkers 4 and 5 more accurately than stimuli from talkers 1–3 (see the Appendix). Although all the stimuli were pretested with native speakers of English to ensure high intelligibility, the intelligibility of different talkers varied for the Japanese listeners. Preliminary acoustic analyses of the stimuli revealed that subject's identification performance for individual talkers was positively correlated with the duration of /r/ and /l/ in each token (Lively *et al.,* 1990). Further acoustic analysis of the tokens is currently underway.

Figure 4 shows the percentage of correct responses as a function of phonetic environment. For two of the environments, final singleton and final clusters, performance is close to ceiling whereas performance in the remaining three environments ranged from 70%–80% correct. The effect of phonetic environment was also significant, *F*(4, 16) = 16.96, *p* < 0.001. Performance during training was best for final singleton and final cluster positions. Performance was significantly lower for initial singleton and initial cluster positions, as well as for intervocalic positions. Individual subjects' performance on word-final singletons and clusters was uniformly high and close to ceiling. Identification accuracy for the two word-initial positions and the intervocalic position was consistently lower than for either of the two word-final positions and varied widely across subjects. In short, the group data parallel the performance of individual listeners (see the Appendix). Thus the effect of different phonetic environments found in the pretest–posttest data was also obtained in the training data. These results replicate previous work (Gillette, 1980; Mochizuki, 1981; Sheldon and

Strange, 1982) that obtained consistent differences in identification across phonetic environments.

Figure 5 shows the percentage of correct responses for each talker as a function of phonetic environment. For final singleton and final clusters, performance was uniformly good for all talkers whereas in the word-initial and intervocalic environments, performance was consistently lower and varied widely as a function of talker. The interaction between talker and phonetic environment was significant, $F(16, 64) = 3.01$, $p<0.001$. This result indicates that some talkers were much better than others in producing intelligible /r/'s and /l/'s in word-initial and intervocalic environments that, in general, were poorly perceived. For word-final singleton and cluster environments, where performance was close to ceiling, talker variability apparently made little difference in performance.

**2. Identification response times—**Response times were also collected during the training phase. An ANOVA comparing the mean response times for correct responses across week (weeks 1–3), talker (talkers 1–5), and phonetic environment (environments 1–5) was carried out. A significant effect of talker was obtained, $F(4, 16) =4.14$, $p <0.05$. Significantly faster response times were observed for Talker 1 and Talker 5 compared to Talker 3. There was no correlation between the mean latency for a talker and the identification data for that talker that can account for the pattern of response times. However, a systematic effect was found when the data were examined by week and phonetic environment, $F(8, 32) = 2.44$, $p < 0.05$. Figure 6 shows the mean response times for each week as a function of phonetic environment. These response times are for correct responses only. For those environments in which accuracy was relatively high at the outset of training; i.e., final singletons and final clusters, response times became faster each successive week, whereas for those environments in which accuracy was initially low, response times became much slower in week 2 than in week 1 but then reversed in week 3. Thus the changes in identification performance that occurred from week to week were paralleled by changes in the pattern of response latencies. Moreover, the pattern of response times from week to week varied systematically depending on whether the contrasts were from phonetic environments in which identification accuracy was initially high or from environments in which identification accuracy was initially low. This relationship between latencies and identification accuracy suggests that as subjects became more familiar with the cues for /r/ and /l/, the time required to identify the critical segment in each token was reduced. For word-initial singletons and clusters, an extra week of training was required before a significant reduction in response times occurred. Presumably, this was due to the difficulty subjects had in identifying the appropriate acoustic cues for /r/ and /l/ in word-initial and intervocalic environments.

## C. Generalization

The results of the two generalization tests will be described next. Recall that TG1 consisted of novel words produced by a novel talker and TG2 consisted of novel words produced by talker 4, an "old" talker used during training. An ANOVA comparing generalization test (TG1-TG2) and phonetic environment for the three subjects who were in both generalization tests yielded a nonsignificant trend for test ($p = 0.09$). Performance in TG2 (mean percentage of correct responses = 83.7%) was marginally better than performance in TG1 (mean = 79.5%). Subjects were more accurate in their identification of /r/ and /l/ when novel words were produced by an old talker they had heard during training than when novel words were produced by a new talker. Individual subject data were consistent with the group data (see the Appendix).

## III. DISCUSSION

The results of the present experiment demonstrate that laboratory training procedures can be used to modify Japanese listeners' perception of /r/ and /l/ in isolated English words. Compared to performance before training began, subjects' overall identification accuracy showed a significant, albeit small, improvement after training.[5] Moreover, subjects' performance depended on the phonetic context in which /r/ and /l/ were located. For word-final singleton and consonant cluster environments, performance was close to asymptotic levels even before training began. For word-initial singleton and cluster environments, and for intervocalic environments, performance improved after training relative to the levels obtained in the pretest, although these environments continued to be more poorly perceived than final positions.[6]

Phonetic context was also found to systematically affect response times. During training, subjects showed faster response times in each successive week for those phonetic environments in which identification performance was initially good. In contrast, for those phonetic environments in which identification performance was initially poor, response times showed an inverted U-shaped function. Response times were relatively fast in the first week, slower in the second week, and then faster again in the third week, suggesting that subjects required more exposure to the stimuli from difficult environments before the appropriate acoustic cues could be learned.

Identification of /r/ and /l/ was also found to depend on talker. During training, stimuli from some talkers were consistently identified more accurately than others. This effect was also found in the generalization tests. Subjects' performance in identifying /r/ and /l/ in novel words depended upon whether the talker had been heard before or not. Overall, these results suggest that, for the training of a nonnative phonetic contrast to be robust, the stimuli must be sufficiently variable and the training task must closely correspond to the task used in the testing phases of the experiment[7] (These results have recently been replicated and extended in Lively *et al.,* 1991.)

[5] The overall improvement in subjects' performance after training was reflected in the performance of individual subjects. Although subjects varied widely in their initial level of performance, the pattern of responses was consistent from subject to subject in all phases of the experiment, both as a function of talker and phonetic environment, indicating that the group data were representative of each subject's performance. The subjects used in the present experiment were also representative of the population of Japanese listeners living in the U.S. compared to the initial level of ID performance reported in other studies in which Japanese listeners were tested. In the present experiment, overall pretest identification accuracy was 78.1%. This compares favorably with Mochizuki (1981), 81.8%; Sheldon and Strange (1982), "good" listeners 89%, "poor" listeners 74%; and Strange and Dittmann (1984), 69%. Thus we would expect that the results obtained using the methodology employed in the present experiment would likely generalize to other groups of Japanese listeners.

[6] An examination of posttest performance as a function of the proportion of training stimuli from different phonetic contexts indicates that although the largest improvement in identification accuracy (20%) occurred in the most represented context in training, word-initial clusters, the second largest improvement (10%) occurred in the least represented context, intervocalic position. Thus the frequency with which stimuli from a particular phonetic context are presented in training does not have a proportional effect on posttest performance. However, this issue is also complicated by the fact that individual contexts vary widely in the initial level of performance as measured in the pretest, that some contexts may be intrinsically more difficult to learn, and that increments in identification accuracy may be nonlinear as a function of training when performance approaches asymptote. Further work is required to clarify these issues.

[7] A reviewer raised the possibility that the improvement between pretest and posttest could have been due to mere exposure to the training task rather than subjects learning something about the stimuli. We discount this possibility on the following grounds. In our laboratory, Schwab *et al.* (1985) carried out a study in which they assessed the effects of training on the perception of low-quality synthetic speech using a pretest–posttest design similar to the one employed in the present experiment. They compared performance in three groups of subjects: (1) subjects trained with synthetic speech, (2) subjects trained with the same procedure using natural speech, and (3) subjects that received no training. Schwab *et al.* were concerned that subjects might simply learn to do the experimental tasks better without necessarily learning anything about the stimuli. Their results showed that in the posttest, only the subjects that were specifically trained with synthetic speech showed any improvement. No differences were observed for the other two groups. Thus mere exposure to the training tasks and experimental procedures did not result in any improvement in performance.

The differences in identification performance for /r/ and /l/ across different phonetic environments found in the present experiment have been reported previously in the literature. Such a finding suggests the existence of systematic differences in the acoustic characteristics of /r/ and /l/ across phonetic contexts (Lehiste, 1964). Sheldon and Strange (1982) suggested that the temporal and spectral characteristics of /r/ and /l/ in initial environments, especially in initial consonant clusters, may differ from /r/ and /l/ in final positions. Specifically, they claim that when /r/ and /l/ "are coarticulated with stop consonants in prevocalic clusters [where performance is worst], their steady-state loci are often not reached or maintained." In final position, however, the acoustic characteristics of /r/ and /l/ tend to influence the formant structure of the preceding vowel, providing additional information about the identity of the liquid in final position. Thus, in the context of initial clusters, acoustic information differentiating /r/ and /l/ may be reduced, whereas, in final position, acoustic information differentiating /r/ and /l/ may actually be enhanced. According to Sheldon and Strange, these two contexts form the endpoints of a continuum and the "availability and duration" of acoustic features cuing /r/ and /l/ in other phonetic contexts lie between these two extremes.

Some support for Sheldon and Strange's hypothesis was obtained in a study carried out in our laboratory by Dissosway-Huff *et al.* (1982) who found that the duration of /r/ and /l/ in final position was longer than in other phonetic environments. As noted earlier, we have measured the durations of the stimuli used in the present experiment and found a pattern similar to that described in these earlier studies. Further acoustic analyses of the stimuli used in the present experiment are currently underway and will reported separately (Lively *et al,* 1990).

In a related study, Henly and Sheldon (1986) demonstrated that the phonological system of a language also can play a major role in determining the pattern of performance obtained across different phonetic environments. They found that for native speakers of Cantonese, identification of /r/ and /l/ in final singleton position and in initial consonant clusters was more difficult than identification of /r/ and /l/ in initial singleton position and intervocalic position. Henly and Sheldon argued that differences between the phonological systems of Cantonese and Japanese were responsible for the differences in performance observed between the two groups of listeners. In each case, the phonology of the listeners' native language acts to filter the acoustic characteristics of English /r/ and /l/. The effect of the filtering depends on the existence and /or distribution of /r/- or /l/- like phonemic categories in the listener's native language.

The consistent finding that phonetic context affects the perception of /r/ and /l/ by Japanese listeners suggests that for these listeners, /r/ and /l/ do not exist as abstract phonemic categories but instead may function as context-sensitive perceptual units. Indeed, we believe that one of the major reasons for the success of the present experiment was the use of training stimuli containing /r/ and /l/ in diverse phonetic contexts. In order for Japanese listeners to learn to identify /r/ and /l/ in different contexts, it appears that they must be exposed to stimuli containing /r/ and /l/ in these different contexts. As the results of Strange and Dittmann (1984) indicated, training on only one context is not likely to result in transfer to other contexts. With regard to developing a theory of phonological change in second language learning, this result implies that listeners may not necessarily proceed directly from the phonemic categories of their native language to the phonemic categories of the new language but instead may rely on intermediate, context-sensitive phonetic categories when initially learning a new phonetic contrast.

In the present experiment, we also found reliable effects associated with the use of different talkers. The use of multiple talkers was designed to increase the stimulus variability that subjects were exposed to during the training phase of the experiment. Our goal was to provide enough talker variability to enable the Japanese listeners to overcome idiosyncrasies in the realization of the acoustic cues for /r/ and /l/ that might be present in the stimuli produced by only one talker. We did not anticipate the degree to which subjects would become sensitive to the different talkers used during training, nor did we anticipate the extent to which generalization performance would depend on the relationship between the talkers used during training and the talkers used during generalization testing. To the best of our knowledge, there has been only one previous instance in the cross-language speech perception literature where the issue of talker variability in the perceptual learning of nonnative speech contrasts was even mentioned. In his 1971 paper, Goto reported that Japanese listeners' familiarity with an English talker was directly related to their perception of /r/ and /l/. For those talkers that the listeners had heard before, performance was better.

Recently, Mullennix *et al.* (1989) reported that talker variability can affect the perception of words by native English listeners. Subjects in their experiments were presented with isolated words under two conditions. In one condition, subjects were presented stimuli produced by a single talker. In the other condition, subjects were presented stimuli produced by multiple talkers that varied from trial to trial. Mullennix *et al.* found that in both naming and perceptual identification tasks, listeners assigned to the multiple talker condition were slower and less accurate than listeners assigned to the single talker condition. These results suggested the operation of a process in which talker variability is normalized in order for the physical stimulus to be mapped on to a more abstract phonetic representation. Mullennix *et al.*'s results demonstrated that this process operates at some cost to the perceptual system.

The results of the present investigation also suggest that nonnative listeners encode detailed talker-specific information and apparently store this information in long-term memory. Evidence for this effect was found not only in training but also in the generalization phase of the experiment in which listener's performance in identifying novel stimuli depended on whether the stimuli were produced by a talker they had heard before. If the novel stimuli were produced by an "old" talker, performance was better than if the novel stimuli were produced by a "new" talker. Goldinger *et al.* (in press) observed similar effects with native English listeners using a serial–ordered recall task. They found that although listeners typically recall lists of spoken words produced by a single talker more accurately than lists produced by multiple talkers (Martin *et al.,* 1989), the situation can be reversed if subjects are given sufficient time between successive list items to elaborate on these additional distinctive cues. Their results indicate that listeners encode talker-specific information as an integral component of the acoustic-phonetic representations of words in long-term memory (see also Mullennix and Pisoni, 1990).

The results of the present experiment are also consistent with recent accounts of the role of selective attention in perceptual learning. As noted earlier, Nosofsky (1986, 1987) has shown that selective attention plays an important role in the identification and categorization of multidimensional visual stimuli. Nosofsky's work demonstrated that selective attention to one stimulus dimension serves to maximize within-category similarity among exemplars sharing that dimension and minimize between-category similarity. Nosofsky has suggested that the role of attention in perceptual learning may be described as selectively distorting the psychological space corresponding to particular dimensions comprising the perceptual object in order to facilitate categorization. In the context of speech perception, the prior linguistic experience of listeners can be thought of as the means by which attention is allocated to specific acoustic-phonetic dimensions. For example, Terbeek (1977) showed that the distance between vowels in multidimensional psychological space depended on a listener's linguistic background. The perceptual distance between a pair of physically similar vowels was judged to be much larger if the members of the pair contrasted phonologically in the subject's native language. Similar findings on the role of linguistic experience in speech perception have been reported in cross-language investigations by Abramson and Lisker (1970), and Stevens *et al.* (1969), among others.

Jusczyk's (1989) recent elaboration of the role of selective attention in the development of phonetic categories in infancy has obvious parallels in adult cross-language speech perception work. Infants acquiring a first language must learn to weigh the acoustic cues in the speech they hear according to the salience of the cues in their linguistic environment. According to Jusczyk, the transition from the language-universal abilities of early infancy to the language-specific abilities of later infancy can be viewed as the allocation of selective attention to those acoustic cues that are relevant or appropriate to the specific language of the infant's environment (see also Strange, 1986; Werker, 1990). Similarly, adults who are acquiring a second language that contains a nonnative contrast must also learn to attend selectively to the acoustic dimensions that cue specific phonetic categories in the new language.

However, there are also important differences between infants and adults. One problem facing adults learning a nonnative phonetic contrast is that they have a pre-existing phonological system that can interfere with the allocation of attention to the novel phonetic categories. In the field of second language acquisition, this effect is known as "phonological filtering" (Flege, 1988). A classic example of this is the difficulty that Japanese listeners have with English /r/ and /l/. Since Japanese phonology contains a single liquid that has acoustic properties that make it similar to both /r/ and /l/, Japanese listeners must learn to allocate their attention to the acoustic cues that differentiate /r/ and /l/ in English. Phonological interference does not always occur when listeners are presented stimuli from a nonnative phonetic category, however.

Recently, Best *et al.* (1988) have proposed that the performance of subjects presented with nonnative speech sounds depends on the similarity of the nonnative sounds to the listeners own phonemic categories. Nonnative stimuli that are similar to native categories are assimilated to the native phonemic categories, causing poor performance in perceptual tasks. In contrast, nonnative stimuli that are distinct from native phonemic categories and thus not assimilable may be easily perceived. Evidence for this claim comes from work done by Best *et al.* in which they presented Zulu click stimuli to adult native speakers of English and infants raised in an English-speaking environment. They found that both adults and infants were able to reliably discriminate the clicks, despite the fact that these speech sounds are not used phonemically in English. Best *et al.* suggested that in those cases where stimuli from a nonnative speech category are not assimilated to a native phonemic category, such as when English listeners are presented Zulu clicks, psychophysical differences determine the accuracy of perception. However, the experience provided by exposure to allophonic variants of nonnative speech sounds within the native language of the listener may also determine how accurately nonnative speech sounds are perceived.

Since attention is intimately involved in the formation of phonetic categories it is useful to consider how attention can be most effectively modified, especially with regard to adults learning a nonnative contrast. In the present investigation several methodological factors were responsible for our success in training Japanese listeners to perceive nonnative phonetic categories. These factors included: (1) using a minimal pair ID task during training, (2) providing immediate feedback during training, (3) using the same ID task during testing and training, and (4) employing natural stimuli produced by several talkers that contained /r/ and /l/ in several phonetic contexts. Each of these factors will be considered below.

The first three factors can be grouped together as task-related variables. Since the ultimate goal of training listeners is to create novel phonetic categories, the training procedure should be designed to help subjects focus their attention on the critical attributes of the stimuli yet at the same time permit them to build up stable representations that allow for some stimulus variability. Basically, only two types of tasks are available: identification tasks and discrimination tasks. Jamieson and Morosan (1986) compared both types of tasks and concluded that training using an identification task was more likely to result in an improvement in the perception of nonnative phonetic categories than training using a discrimination task. Their reasoning was as follows: Identification tasks with immediate feedback during training require listeners to group stimuli from the same perceptual category together. In contrast, discrimination tasks tend to promote an increase in sensitivity to small within-category differences. Possible exceptions to this generalization are discrimination tasks in which listeners must ignore irrelevant stimulus variation while focusing on phonetic variation. For example, in the category-change procedure (Kuhl, 1983; Werker *et al.,* 1981), listeners are required to treat exemplars varying along a nonlinguistic dimension as belonging to the same phonetic category. Nonlinguistic dimensions include within-talker variability [e.g., multiple natural exemplars from the same talker (Werker *et al.,* 1981)] and between-talker variability [e.g., synthetic stimuli modeled after adult male, adult female, and children's voices (Kuhl, 1983)]. To the best of our knowledge, however, this type of discrimination procedure has not been applied to the formation of nonnative phonemic categories in laboratory training experiments. Immediate feedback during training is also important because it focuses subjects' attention on the criterial acoustic cues in a consistent manner from trial to trial (see Pisoni, 1977). The final task-related factor, the use of the same task during training and testing, emphasizes the importance of maintaining consistent mapping (Shiffrin and Schneider, 1977) between stimuli and responses across different phases of the experiment.

The modification of attention is also promoted by stimulus variability. In order for attention to be directed to the criterial acoustic cues across the range of stimuli possible for each category, the listener must be exposed to a set of stimuli with sufficient variability. Exposure to a broad range of stimuli is also necessary for the listener to learn about which acoustic cues are irrelevant to the categorization task. Thus the role of stimulus variability is to provide a representative sample of possible exemplars so that changes in the relative weightings of different acoustic cues appropriate to the novel categories can take place. In the present experiment, the use of stimuli from different phonetic environments produced by several different talkers provided a large number of different contexts in which the acoustic cues for /r/ and /l/ could be realized. The end result of providing listeners with stimulus variability is the formation of robust phonetic representations that sample a range of the stimulus variability possible in everyday settings (see Jamieson and Morosan, 1989).

The combination of task and stimuli used in the present experiment succeeded because of its similarity to real-world settings where listeners are typically faced with nonnative speech contrasts from a variety of phonetic contexts produced by many different talkers.[8] Since

[8] The success of the methodology employed in the present experiment suggests that a similar methodology could have practical applications in second language learning, where an important goal is to develop perceptual skills that are useful in conversational settings. However, there are several issues that need to be resolved before the procedure used here can be recommended unequivocally for use in second language learning. For example, the words used in our task were produced in citation form rather than in fluent connected speech. In citation form, the acoustic cues for /r/ and /l/ are more likely to be fully realized whereas in connected speech of the type found in conversational settings, the acoustic cues are likely to be imperfectly realized because of the operation of phonological rules. Viewed this way, identifying words presented in conversational settings would be a more difficult task than identifying words in isolation (unless the words were excised from a sentence). There is, however, at least one advantage to identifying words presented in conversational settings, namely, the contribution of semantic context. Semantic context is useful for resolving lexical ambiguity and for determining the identity of ambiguous utterances. Thus it remains an empirical question as to the best procedure for training listeners to perceive a nonnative phonetic category.

previous research has shown that even difficult-to-perceive categories can be learned over a long period of time if the listener is in an environment where the language is used on a regular basis (McKain *et al.,* 1981), the goal of a laboratory training task should be to simulate this experience in a concentrated form. Therefore, a useful training procedure must provide listeners the opportunity to develop representations that are robust with respect to the range of talkers and phonetic contexts that they will encounter in their everyday life. Using an ID task to present a large ensemble of stimuli produced by different talkers across a wide variety of phonetic contexts forces listeners to develop representations in LTM that will accommodate such a range of variation.

Aside from the specific characteristics of the method used to modify attention in the context of learning a nonnative contrast, the time necessary to develop usable nonnative phonetic categories is also an important consideration. For the Japanese subjects used in the present experiment, over 2500 identification trials were necessary to demonstrate a significant improvement in their identification performance. Thus we can conclude that learning to selectively attend to the relevant dimensions distinguishing /r/ and /l/ requires substantial practice before the mapping between category labels and stimulus input becomes an automatic process (Shiffrin, 1988).

In future work, we plan to address a number of the issues raised in the present experiment. First, the contribution of talker variability to the development of robust phonetic categories warrants further examination. It may be the case that using either a larger or smaller number of talkers during training may be more effective in facilitating the formation of phonetic categories for /r/ and /l/ in Japanese listeners. Another factor related to the effect of talker variability is determining what constitutes an intelligible talker. Whether listeners benefit more from training with “good” talkers or training with “bad” talkers is an open question. It may be the case that although training proceeds at a faster rate with good talkers, transfer to bad talkers is impaired if only good talkers are used during training. Second, additional work needs to be carried out to further assess the role of phonetic context. For example, eliminating those stimuli that are identified at asymptotic levels may be one way to make training more efficient (Atkinson, 1972). Finally, an additional direction for future research is to train subjects until they reach asymptotic levels of performance. Research along these lines would help answer questions regarding the time course of learning a phonetic contrast and the degree to which generalization performance would benefit from such training procedures.

In conclusion, the results of the present experiment demonstrate that the perception of /r/ and /l/ by Japanese listeners can be improved using a simple laboratory training task that requires identification of an item from a minimal pair of English words. However, performance was found to depend on the phonetic environment in which the contrast was located and the talkers used during training. The results of two generalization tests showed that listeners apparently learned characteristics of /r/ and /l/ that were not only conditioned by their phonetic environment but were also specific to the talker who produced the items during training. The present investigation raises many interesting and potentially important questions about the nature of stimulus variability in perceptual learning and its role in training nonnative listeners to perceive phonetic contrasts that are not distinctive in their language. Further work is currently underway in our laboratory to address these important issues.

## Acknowledgments

This research was supported, in part, by NIH Research Grant DC-00111-13 to Indiana University in Bloomington, IN, and, in part, by a Marie Z. Sinnott Fellowship from Indiana University to the first author. For their assistance in obtaining subjects, the authors thank Natsuko Tsujimura and N. John Castellan. The authors thank Pat Kuhl, Winifred Strange, and an anonymous reviewer for comments.

## References

Abramson, A.; Lisker, L. Discriminability along the voicing continuum: Cross-language tests. Proceedings of the Sixth International Congress of Phonetic Sciences; Academia, Prague. 1970. p. 569-573.

Aslin, R.; Pisoni, D. Some developmental processes in speech perception. In: Yeni-Komshian, G.; Kavanagh, J.; Ferguson, C., editors. Child Phonology Perception and Production. Academic; New York: 1980. p. 67-96.

Atkinson R. Ingredients for a theory of instruction. Am Psychol. 1972; 27:921–931.

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

Carney A, Widin G, Viemeister N. Noncategorical perception of stop consonants differing in VOT. J Acoust Soc Am. 1977; 62:961–970. [PubMed: 908791]

Dissosway-Huff, P.; Port, R.; Pisoni, D. Res Speech Percep, Prog Rep No 8. Speech Research Laboratory, Indiana University; Bloomington: 1982. Context effects in the perception of /r/ and /l/ by Japanese.

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

Gillette, S. Minn Papers Ling Philos Lang. Vol. 6. University of Minnesota; Minneapolis: 1980. Contextual variation in the perception of L and R by Japanese and Korean speakers; p. 59-72.

Goldinger S, Pisoni D, Logan J. On the nature of talker variability effects in the recall of spoken words lists. J Exp Psychol: Learn Memory Cognit. 1989

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

Henly E, Sheldon A. Duration and context effects on the perception of English /r/ and /l/: A comparison of Cantonese and Japanese speakers. Lang Learn. 1986; 36:505–521.

Jamieson D, Morosan D. Training non-native speech contrasts in adults: Acquistion of the English /δ/–/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]

Jusczyk, P. Developing phonological categories from the speech signal. Paper presented at International Conference on Phonological Development; Stanford University. 1989.

Kuhl P. Perception of auditory equivalence classes for speech in early infancy. Inf Behav Dev. 1983; 6:263–285.

Lehiste I. Acoustic characteristics of selected English consonants. Int J Am Ling. 1964; 30:10–115.

Liberman A, Harris K, Hoffman H, Griffith B. The discrimination of speech sounds within and across phoneme boundaries. J Exp Psychol. 1957; 54:358–368. [PubMed: 13481283]

Lively, S.; Logan, J.; Pisoni, D. Res Speech Percept, Prog Rep No 16. Speech Research Laboratory, Indiana University; Bloomington: 1990. An acoustic analyses of /r/ and /l/ across phonetic environments and talkers.

Lively, S.; Pisoni, D.; Logan, J. Some effects of training Japanese listeners to identify English /r/ and /l/. In: Tohkura, Yohichi, editor. Speech Perception, Production and Linguistic structure. OHM; Tokyo: 1991. (to be published)

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

Martin C, Mullennix J, Pisoni D, Summers W. Effects of talker variability on recall of spoken word lists. J Exp Psychol: Learn Memory Cognit. 1989; 15:676–684.

Miyawaki K, Strange W, Verbrugge R, Liberman A, Jenkins J, Fujimura O. An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. Percept Psychophys. 1975; 18:331–340.

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

Mullennix J, Pisoni D. Stimulus variability and processing dependencies in speech perception. Percept Psychophys. 1990; 47:379–390. [PubMed: 2345691]

Mullennix J, Pisoni D, Martin C. Some effects of talker variability on spoken word recognition. J Acoust Soc Am. 1989; 85:365–378. [PubMed: 2921419]

Nosofsky R. Attention, similarity, and the identification-categorization relationship. J Exp Psychol: General. 1986; 115:39–57.

Nosofsky R. Attention and learning processes in the identification and categorization of integral stimuli. J Exp Psychol: Learn Memory Cognit. 1987; 15:700–708.

Pisoni D. Auditory and phonetic codes in the discrimination of consonants and vowels. Percept Psychophys. 1973; 13:253–260.

Pisoni D. Auditory short-term memory and vowel perception. Mem Cognit. 1975; 3:7–18.

Pisoni D. Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. J Acoust Soc Am. 1977; 61:1352–1361. [PubMed: 881488]

Pisoni D, Lazarus J. Categorical and noncategorical modes of speech perception along the voicing continuum. J Acoust Soc Am. 1974; 55:328–333. [PubMed: 4821837]

Pisoni, D.; Logan, J.; Lively, S. Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception. In: Nusbaum, H.; Goodman, J., editors. Development of Speech Perception: The Transition from Recognizing Speech Sounds to Spoken Words. MIT; Cambridge. MA: 1991. (to be published)

Pisoni D, Nusbaum H, Greene B. Perception of synthetic speech generated by rule. Proc IEEE. 1985; 73:1665–1676.

Posner M, Keele S. On the genesis of abstract ideas. J Exp Psychol. 1968; 77:353–363. [PubMed: 5665566]

Samuel A. The effect of discrimination training on speech perception: Noncategorical perception. Percept Psychophys. 1977; 22:321–330.

Schwab E, Nusbaum H, Pisoni D. Some effects of training on the perception of synthetic speech. Hum Factors. 1985; 27:395–408. [PubMed: 2936671]

Shiffrin, R. Attention. In: Atkinson, R.; Herrnstein, R.; Lindzey, G.; Luce, R., editors. Steven's Handbook of Experimental Psychology. 2. Vol. 2. Wiley; New York: 1988. p. 739-811.Learning and Cognition

Shiffrin R, Schneider W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychol Rev. 1977; 84:127–190.

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

Stevens K, Liberman A, Ohman S, Studdert-Kennedy M. Cross-language study of vowel perception. Lang Speech. 1969; 12:1–23. [PubMed: 5789292]

Strange, W. Speech input and the development of speech perception. In: Kavanagh, J., editor. Otitis Media and Child Development. York; Parkton, MD: 1986. p. 12-26.

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]

Studdert-Kennedy M, Liberman A, Harris K, Cooper F. Motor theory of speech perception: A reply to Lane's critical review. Psychol Rev. 1970; 77:234–249. [PubMed: 5454133]

Terbeek, D. UCLA Work Pap Phon. Vol. 37. UCLA; Los Angeles: 1977. A. cross-language multidimensional scaling study of vowel perception.

Webster's Seventh Collegiate Dictionary. Library Reproduction Service; Los Angeles: 1967.

Werker J. Becoming a native listener. Am Sci. 1989; 77:54–59.

Werker, J. The ontogeny of speech perception. In: Mattingly, I.; Studdert-Kennedy, M., editors. Modularity and the Motor Theory of Speech Perception. Erlbaum; Hillsdale, NJ: 1990. (in press)

Werker J, Gilbert J, Humphrey K, Tees R. Developmental aspects of cross-language speech perception. Child Dev. 1981; 52:349–355. [PubMed: 7238150]

Werker J, Logan J. Cross-language evidence for three factors in speech perception. Percept Psychophys. 1985; 37:35–44. [PubMed: 3991316]

Werker J, Tees R. Phonemic and phonetic factors in adult cross-language speech perception. J Acoust Soc Am. 1984; 75:1866–1878. [PubMed: 6747097]

# APPENDIX

The Appendix contains individual subject data from the pretest–posttest (Table A1), training (Tables A2, A3, A4), and generalization phases (Table A5) of the experiment.

## TABLE A1

Pretest and Posttest performance (percent correct) as a function of phonetic context for individual subjects.

| | | Subjects | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Phonetic context** | **Test** | **YY** | **IS** | **MF** | **SK** | **HA** | **NY** | **Mean** |
| c r/l v… | Pre | 18.7 | 43.7 | 81.2 | 78.1 | 56.2 | 71.8 | 58.3 |
| | Post | 62.5 | 56.3 | 100.0 | 87.5 | 75.0 | 87.5 | 78.1 |
| r/l vc | Pre | 75.0 | 75.0 | 100.0 | 87.5 | 65.6 | 75.0 | 79.7 |
| | Post | 81.3 | 81.3 | 100.0 | 100.0 | 56.3 | 62.5 | 80.2 |
| …v r/l v… | Pre | 93.8 | 75.0 | 100.0 | 90.6 | 53.1 | 59.4 | 78.6 |
| | Post | 87.5 | 87.5 | 93.8 | 100.0 | 87.5 | 68.8 | 87.5 |
| …v r/l | Pre | 100.0 | 93.8 | 100.0 | 96.9 | 93.8 | 90.6 | 95.8 |
| | Post | 100.0 | 93.8 | 100.0 | 100.0 | 93.8 | 100.0 | 97.9 |
| Mean | Pre | 71.9 | 71.9 | 95.3 | 88.3 | 67.2 | 74.2 | 78.1 |
| | Post | 82.8 | 79.7 | 98.4 | 96.9 | 78.1 | 79.7 | 85.9 |

## TABLE A2

Training performance (percent correct) as a function of week for individual subjects.

| | Subjects | | | | | | |
|---|---|---|---|---|---|---|---|
| **Week** | **YY** | **IS** | **MF** | **SK** | **HA** | **NY** | **Mean** |
| 1 | 77.5 | 74.5 | 88.7 | 88.8 | 74.5 | 77.3 | 80.2 |
| 2 | 82.6 | 75.6 | 94.2 | 94.1 | 73.9 | 77.9 | 83.1 |
| 3 | 83.9 | 79.6 | 93.3 | 95.7 | 77.7 | 80.2 | 85.1 |
| Mean | 81.3 | 76.5 | 92.1 | 92.9 | 75.4 | 78.5 | 82.8 |

## TABLE A3

Training performance (percent correct) as a function of talker for individual subjects.

| | Subjects | | | | | | |
|---|---|---|---|---|---|---|---|
| **Talker** | **YY** | **IS** | **MF** | **SK** | **HA** | **NY** | **Mean** |
| 1 | 80.9 | 75.9 | 88.7 | 86.5 | 72.9 | 73.7 | 79.8 |
| 2 | 78.4 | 73.5 | 87.5 | 92.6 | 72.7 | 73.4 | 79.7 |
| 3 | 79.1 | 76.1 | 91.3 | 93.5 | 73.0 | 78.4 | 81.9 |

| | Subjects | | | | | | |
|---|---|---|---|---|---|---|---|
| **Talker** | **YY** | **IS** | **MF** | **SK** | **HA** | **NY** | **Mean** |
| 4 | 86.9 | 78.7 | 96.9 | 97.6 | 80.4 | 84.7 | 87.5 |
| 5 | 81.2 | 78.5 | 96.1 | 94.2 | 77.8 | 82.1 | 85.0 |
| Mean | 81.3 | 76.5 | 92.1 | 92.9 | 75.4 | 78.5 | 82.8 |

## TABLE A4

Training performance (percent correct) as a function of phonetic environment for individual subjects.

| | Subjects | | | | | | |
|---|---|---|---|---|---|---|---|
| **Phonetic context** | **YY** | **IS** | **MF** | **SK** | **HA** | **NY** | **Mean** |
| c r/l v… | 60.5 | 58.3 | 83.9 | 86.9 | 57.3 | 64.9 | 68.6 |
| r/l vc | 75.4 | 75.2 | 95.3 | 93.9 | 62.4 | 72.4 | 79.1 |
| …v r/l v… | 77.7 | 67.0 | 86.0 | 89.3 | 60.7 | 62.7 | 73.9 |
| cv r/l c | 95.3 | 85.6 | 96.8 | 95.8 | 97.4 | 94.9 | 94.3 |
| …v r/l | 97.7 | 96.6 | 98.4 | 98.4 | 99.0 | 97.5 | 97.9 |
| Mean | 81.3 | 76.5 | 92.1 | 92.9 | 75.4 | 78.5 | 82.8 |

## TABLE A5

Generalization performance (percent correct) for individual subjects. (Note: *TG1* refers to novel stimuli produced by a novel talker. *TG2* refers to new stimuli produced by Talker 4 whom subjects heard during training.)

| | Subjects | | | |
|---|---|---|---|---|
| **Test** | **SK** | **HA** | **NY** | **Mean** |
| TG1 | 90.6 | 71.9 | 76.0 | 79.5 |
| TG2 | 94.9 | 75.5 | 80.6 | 83.7 |

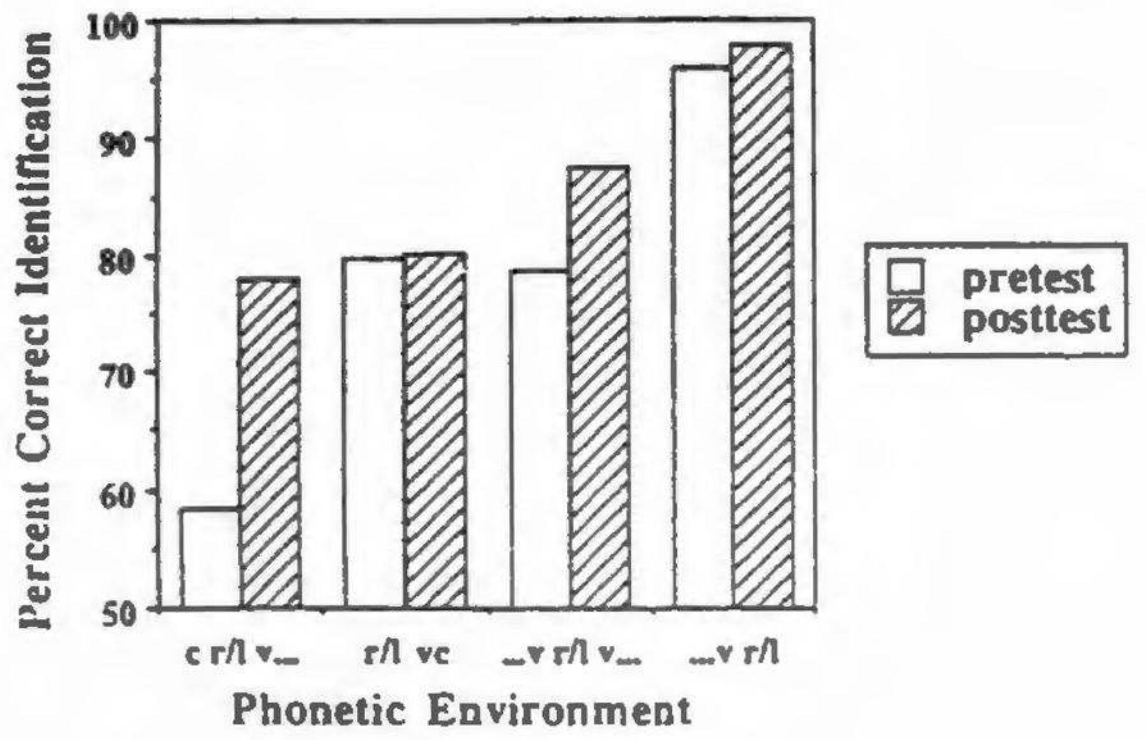

**FIG. 1.**

Mean percentage of correct response in the protest and posttest as a function of phonetic environment.

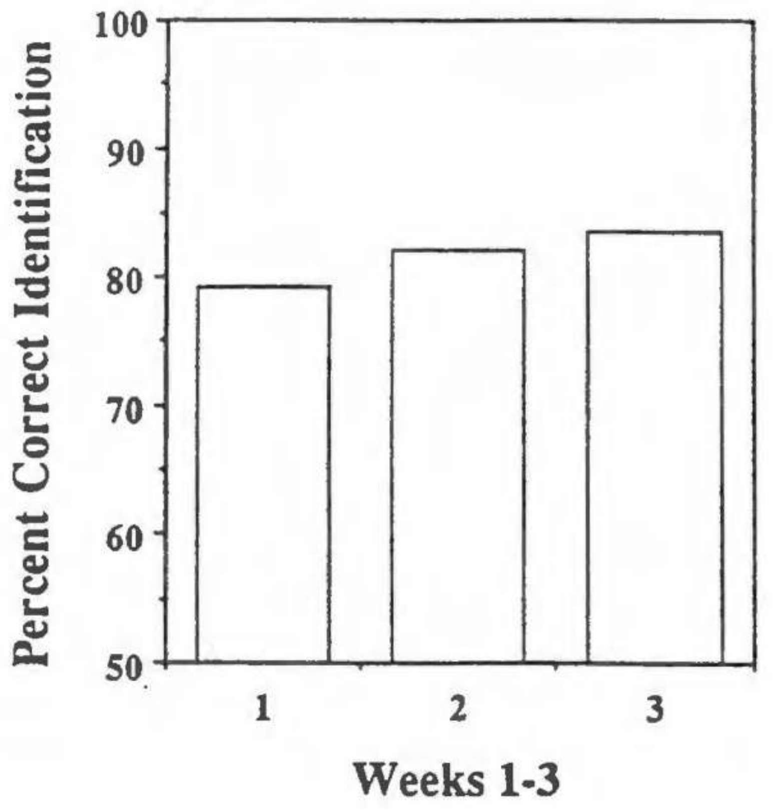

**FIG. 2.**

Mean percentage of correct responses during training as a function of week.

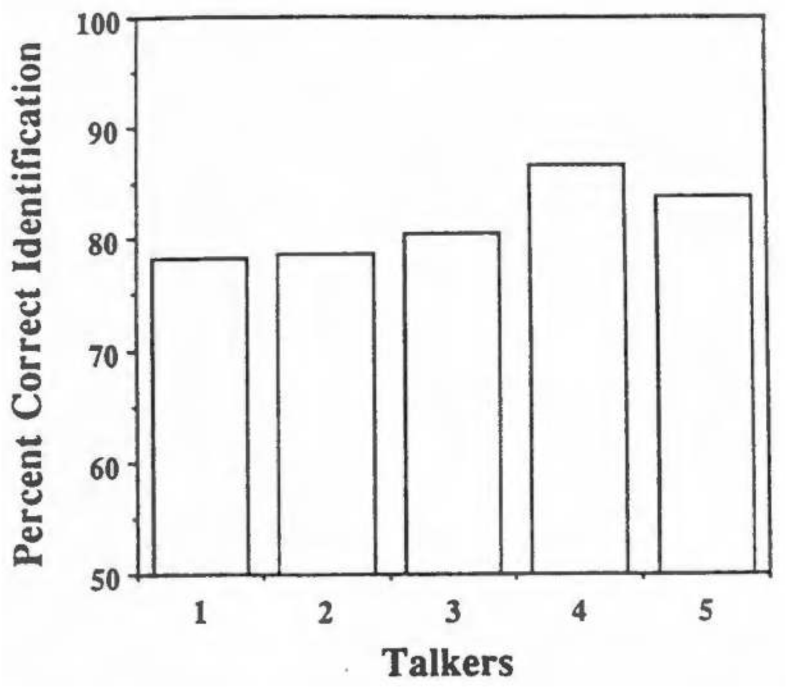

**FIG. 3.**

Mean percentage of correct responses during training as a function of talker.

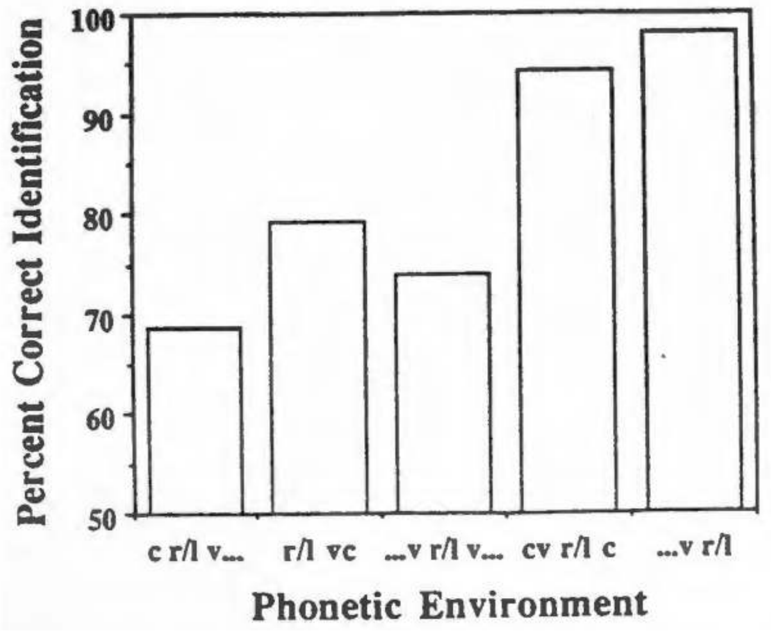

**FIG. 4.**
Mean percentage of correct responses during training as a function of phonetic environment.

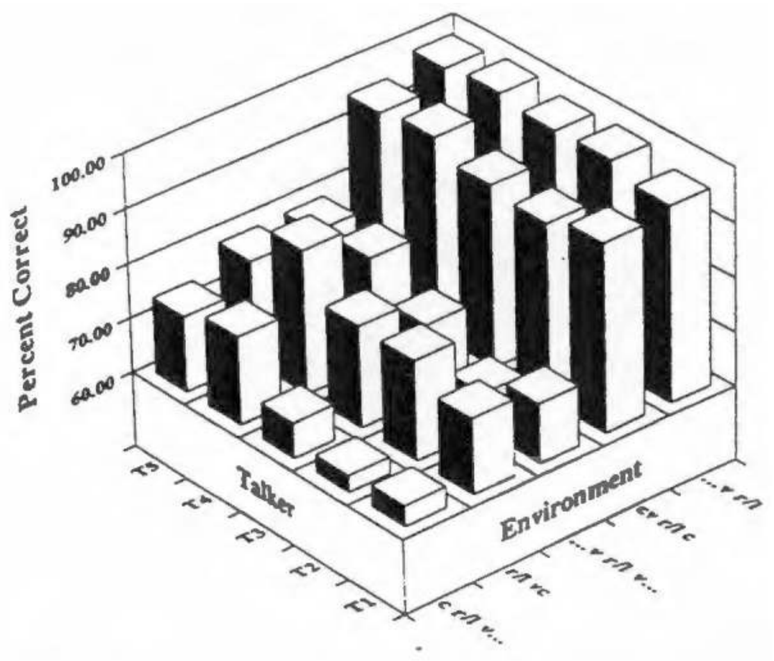

**FIG. 5.**
Mean percentage of correct responses during training as a function of talker and phonetic environment.

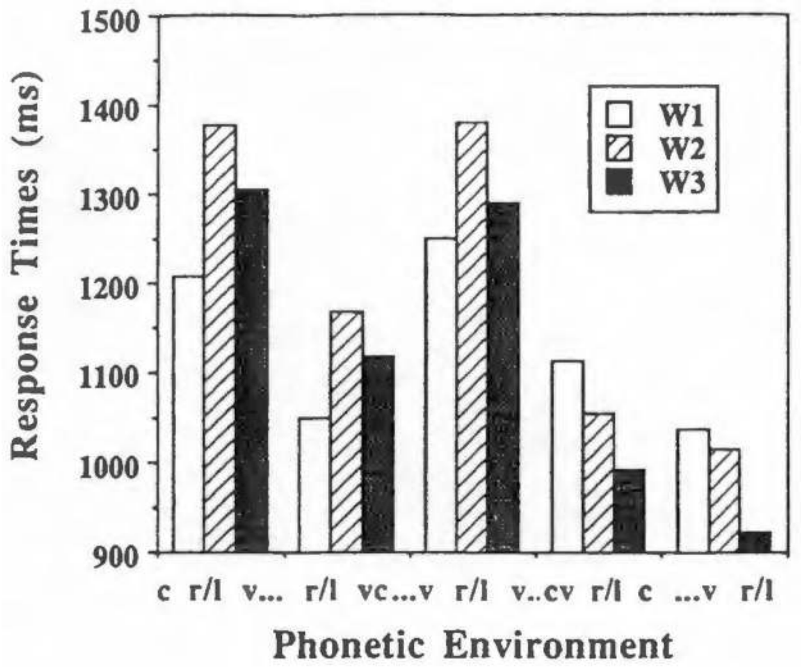

**FIG. 6.**

Mean latencies (in ms) for correct responses during training as a function of week and phonetic environment.