

Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults

Paul Iverson, Valerie Hazan, and Kerry Bannister

Department of Phonetics and Linguistics, University College London, 4 Stephenson Way, London NW1 2HE, United Kingdom

(Received 5 May 2005; revised 5 August 2005; accepted 17 August 2005)

Recent work [Iverson *et al.* (2003) *Cognition*, **87**, B47-57] has suggested that Japanese adults have difficulty learning English /r/ and /l/ because they are overly sensitive to acoustic cues that are not reliable for /r/-/l/ categorization (e.g., F2 frequency). This study investigated whether cue weightings are altered by auditory training, and compared the effectiveness of different training techniques. Separate groups of subjects received High Variability Phonetic Training (natural words from multiple talkers), and 3 techniques in which the natural recordings were altered via signal processing (All Enhancement, with F3 contrast maximized and closure duration lengthened; Perceptual Fading, with F3 enhancement reduced during training; and Secondary Cue Variability, with variation in F2 and durations increased during training). The results demonstrated that all of the training techniques improved /r/-/l/ identification by Japanese listeners, but there were no differences between the techniques. Training also altered the use of secondary acoustic cues; listeners became biased to identify stimuli as English /l/ when the cues made them similar to the Japanese /r/ category, and reduced their use of secondary acoustic cues for stimuli that were dissimilar to Japanese /r/. The results suggest that both category assimilation and perceptual interference affect English /r/ and /l/ acquisition. © 2005 Acoustical Society of America. [DOI: 10.1121/1.2062307]

PACS number(s): 43.71.Hw, 43.71.Es [ARB]

Pages: 3267–3278

I. INTRODUCTION

Infants are born with an ability to tune their perceptual processes to the sounds of their first language (L1), but the perceptual processes of adults are much less plastic during second language (L2) learning. Although some scientists have argued that this change in plasticity is a result of a biologically delimited critical period that ends at puberty [e.g., Lenneberg (1967); Patkowski (1990)], the current evidence suggests that this change is a gradual consequence of learning one's L1. For example, Flege (1999) has shown that there is no discrete point where learning switches from "easy" to "hard;" rather, it becomes linearly harder to learn an L2 without an accent as one gets older. Learning does not become globally harder for all L2 phonemes; adults have the greatest difficulty learning L2 phonemes that are strongly assimilated into L1 categories, and are better able to produce and perceive L2 phonemes that are dissimilar from any existing L1 phonemes [e.g., Best *et al.* (1988); Flege (1995); Flege *et al.* (2003); Guion *et al.* (2000)]. It appears that plasticity for L2 speech perception progressively declines as individuals become neurally committed to processing their native language [e.g., Kuhl (2000)].

The case of Japanese adults learning the English /r/-/l/ distinction has become the canonical example of "hard" L2 phoneme learning. Japanese adults tend to be very poor at distinguishing English /r/-/l/ [e.g., Goto (1971); Miyawaki *et al.* (1975)]. Although the perception and production of English /r/ and /l/ can improve with experience and training [e.g., Bradlow and Pisoni (1999); Bradlow *et al.* (1999); Hazan *et al.* (in press); MacKain *et al.* (1981); Logan *et al.* (1991)], it can take decades of English-language experience

before individuals achieve native levels of performance [Flege *et al.* (1995)]. Best and Strange (1992) hypothesized that the English /r/-/l/ distinction is particularly hard for Japanese adults because they are both assimilated into a single Japanese /r/ category. The Japanese /r/ is a lateral flap, which is much more rapid than the English /r/ or /l/, but it has a range of F2 and F3 frequencies that overlap with those of English /r/ and /l/ [Lotto *et al.* (2004)]. Best and Strange argued that English /r/ and /l/ may sound the same to Japanese adults because they both are the same in respect to the Japanese phonological system (i.e., they both are poor exemplars of the Japanese /r/). Aoyama *et al.* (2004) have further suggested that assimilation patterns can account for the finding that Japanese adults are somewhat better at learning English /r/ than /l/; they argue that English /l/ is more similar to the Japanese /r/ category than is English /r/, and learning to produce and perceive English /r/ is thus easier because it is subject to less L1 interference.

Although Best's (1994) Perceptual Assimilation Model can account for some patterns of L2 phoneme perception [e.g., Best *et al.* (2000); Harnsberger (2001)], our recent work [Iverson *et al.* (2003)] has suggested that it cannot explain the perception of English /r/ and /l/ by Japanese adults. Iverson *et al.* (2003) replicated the common finding that Japanese adults are much poorer, compared to English speakers, at discriminating /r/-/l/ differences near the category boundary. However, we found that Japanese adults were actually better than English speakers at discriminating within-category acoustic variation, and were more sensitive to acoustic variation in F2 frequency. It is thus not the case that English /r/ and /l/ stimuli all sound the same to Japanese listeners. Instead, the problem may be that Japanese adults

are particularly sensitive to acoustic differences that are irrelevant to the English /r/-/l/ categorization. Iverson *et al.* (2003) hypothesized that these patterns of perceptual sensitivities interfere with /r/-/l/ learning by making it harder for Japanese adults to focus attention on more critical acoustic cues (i.e., F3 differences near the category boundary) and more likely to form category representations based on cues that are not critical to native listeners [e.g., F2 frequency or duration differences; see Gordon *et al.* (2001); Yamada (1995)]. Learning English /r/ and /l/ may be hard because it requires Japanese listeners to alter their perceptual space for these phonemes in order to reduce these perceptual interference effects. That is, they must learn to become less sensitive to acoustic cues that are not important for the /r/-/l/ distinction.

The aims of the present study were to test whether the reliance on secondary cues (i.e., acoustic differences that are not critical for distinguishing /r/ and /l/, such as F2 and duration) is reduced during learning, and to compare the effectiveness of different training methods. The baseline method was High Variability Phonetic Training [HVPT; Logan *et al.* (1991)], which involves having subjects give identification judgments with feedback for natural recordings of words produced by multiple talkers, with target phonemes in multiple syllable positions. Pisoni and colleagues [e.g., Logan *et al.* (1991); Lively *et al.* (1993)] have argued that exposing listeners to a wide range of natural stimuli is better than training with a small number of stimuli because the distributions of natural stimuli teach individuals which cues are most reliable; listeners are thought to store individual exemplars that they hear in training, and the multidimensional categorization space for these stimuli gets stretched along dimensions where /r/ and /l/ differ and shrunk along dimensions that do not distinguish /r/ and /l/ [see Nosofsky (1986) (1987)]. This shrinking/stretching account is compatible with the notion of perceptual interference [Iverson *et al.* (2003)]; such a process is exactly what Japanese listeners would need to alter their perceptual space so that they can better attend to the F3 differences that are critical to the /r/-/l/ distinction.

Although HVPT has emphasized the importance of natural variability, it is possible to experimentally manipulate stimuli to specifically target the perceptual interference problems of Japanese listeners. For example, the Perceptual Fading technique [Jamieson and Morosan (1986)] has been used to help listeners focus on critical acoustic cues; listeners are initially trained on stimuli that are maximally contrastive (i.e., enhanced) on the primary acoustic cues used for a phonetic contrast, and the degree of enhancement is decreased as training progresses. This approach has been used to train English /r/ and /l/ for Japanese listeners [Doeleman *et al.* (2000); McCandliss *et al.* (2002); McClelland *et al.* (2002); Protopapas and Calhoun (2000)], but it has not been directly compared to HVPT. The present study used a version of Perceptual Fading that was designed to parallel HVPT; the natural stimuli used in HVPT were signal-processed to increase the difference in F3 for /r/ and /l/ at early stages of training, and decrease this difference at later stages. The study also tested a similar training technique, All Enhanced, in which subjects received enhanced F3 differences at every

stage (i.e., no fading); this tested whether a lack of exposure to stimuli with natural variability in F3 would affect the learning process and/or generalization to natural stimuli.

Our work [Iverson *et al.* (2003)] suggested that too much sensitivity to secondary cues was as much a problem for learning as too little attention to the primary acoustic cues. The present study thus constructed a Secondary Cue Variability training technique that was essentially the complement of Perceptual Fading; individuals started training on stimuli that had been signal-processed to equate secondary acoustic cues (i.e., no variability in F2, closure duration, or transition duration) and the amount of random variability in these cues was increased throughout training. The intention was to keep subjects from being distracted by secondary cues at early stages of training (i.e., making it easier to pay attention to F3 differences) and then progressively teach them to ignore this kind of variation.

The study thus compared 4 training techniques (HVPT, All Enhanced, Perceptual Fading, and Secondary Cue Variability). The stimuli from all conditions were based on natural recordings from 10 talkers speaking 100 initial-position /r/-/l/ minimal pair words. The positional variability normally used in HVPT (i.e., stimuli with /r/ and /l/ in multiple syllable positions) was not used here because little is known about which acoustic cues are most important for distinguishing /r/ and /l/ in medial position and consonant clusters; to compensate, more variability was introduced by using more talkers and words than had been used in earlier studies [e.g., Logan *et al.* (1991)]. Subjects were given a battery of tests before and after training to examine how well their training generalized to new words, talkers, and syllable positions. In addition, they were tested on signal-processed stimuli to examine whether training altered their use of secondary acoustic cues.

II. METHOD

A. Subjects

A total of 73 subjects completed testing (87 began the training program but 14 did not finish all of the sessions). Of the subjects that completed, 5 were dropped from the data analysis because of computer problems (i.e., missing data), and 6 were dropped because their pre-training identification of initial-position /r/-/l/ was greater than 90% correct. There was thus a total of 62 subjects included in the data analysis (16 each in the Secondary Cue Variability and All Enhanced conditions; 15 each in HVPT and Perceptual Fading). Forty-six of these subjects participated in all pre/post tests; the other 16 subjects were given the pre/post tests using natural stimuli but not those using cue-manipulated stimuli (see Procedure) due to testing time limitations.

All subjects were native speakers of Japanese with no known hearing or language impairments. Their ages ranged from 18 to 40 years (median 20 years), and the age at which they began learning English ranged from 4 to 23 years (median 12 years). Forty-one subjects were tested in Japan; they were students taking a course in English language at Kochi University, and all but one of these participants had never lived in an English speaking country. Twenty-one subjects

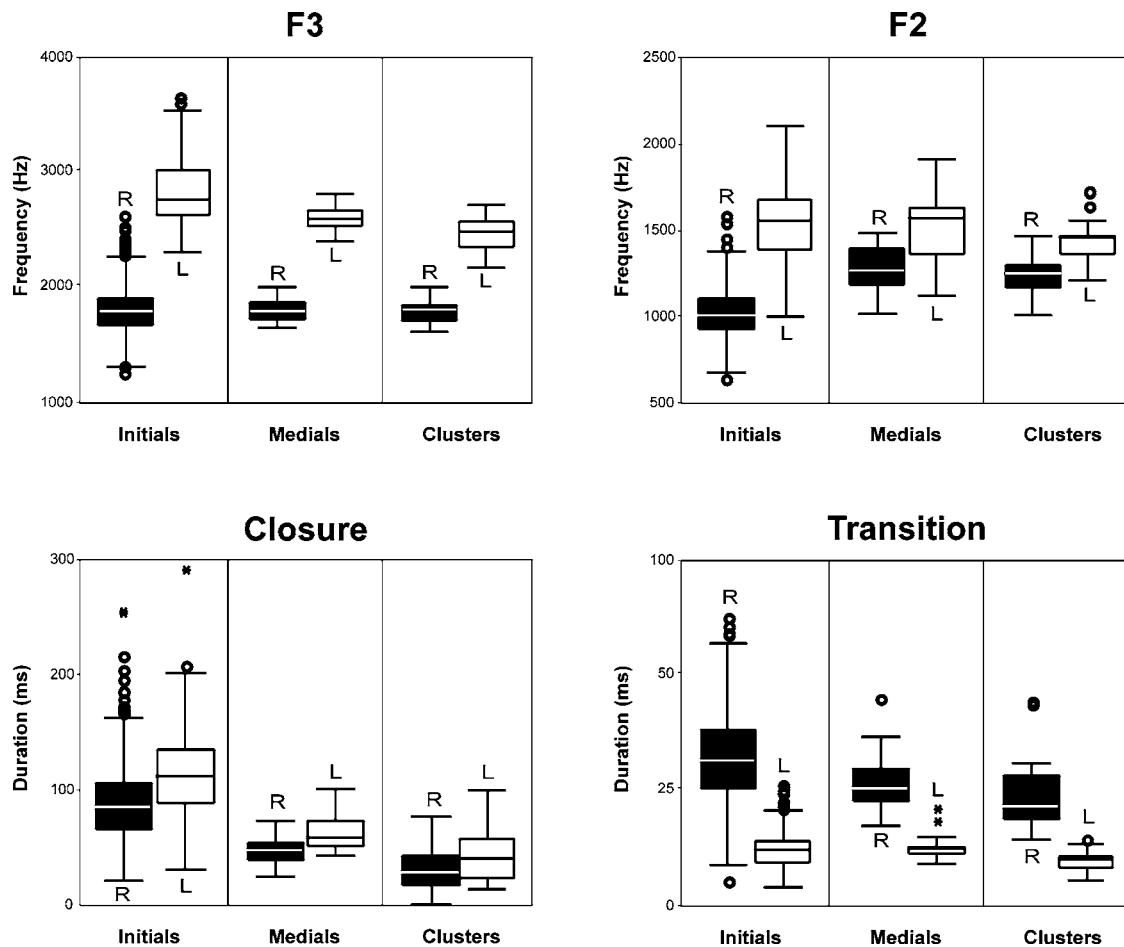


FIG. 1. Distributions of acoustic cues in natural stimuli. The boxplots represent the quartile ranges of each acoustic measure. The distributions of initials were based on 1000 stimuli (100 words each from 10 talkers). The distributions of medials and clusters were based on 40 stimuli (20 words each from 2 talkers). The formant frequency measurements (F3 and F2) were normalized to each speaker's vowel formant frequencies.

were tested in England; they were students who were attending English language or phonetics courses in London, and 5 had lived for more than 12 months in an English-speaking country (the longest for 5 years). The subjects in the two locations had very similar levels of /r/-/l/ identification accuracy (e.g., the pre-test percentage correct across all conditions averaged 60.5% in Kochi and 60.3% in London).

B. Apparatus

Subjects were tested and trained using either laboratory PCs or their own laptops and headphones. When subjects used their own laptops, a research assistant supervised the installation and checking of the software, and the subjects were able to borrow laboratory headphones if they did not own any of sufficient quality. Subjects were allowed to adjust the amplitude of the stimuli to a level that they found comfortable. Testing was completed in a quiet room under supervision of a research assistant. Training was completed by the subjects on their own (e.g., at home or in the laboratory), with the details of each session (e.g., time and date completed) automatically logged in a password-protected file that the subjects could not read or change.

C. Stimuli

1. Natural recordings

Twelve adult native speakers of British English (6 male and 6 female) were digitally recorded in an anechoic chamber with a calibrated microphone. The stimuli were recorded with 44 100 16-bit samples per second, and downsampled to 22 050 samples per second at a later stage. The words were spoken in isolation, but were presented to the talkers on a computer screen one at a time to avoid list intonation. The words were presented in a random order, mixed with words that did not contain /r/ or /l/. The training corpus was recorded by 10 speakers and consisted of 100 initial-position /r/-/l/ minimal-pair words (e.g., *rock* and *lock*). The testing corpus was recorded by two additional speakers and included 40 initial-position /r/-/l/ minimal-pair words from the training corpus, 40 initial-position /r/-/l/ minimal-pair words that were not used in training, 40 medial-position /r/-/l/ minimal-pair words (e.g., *arrive* and *alive*), and 40 consonant-cluster /r/-/l/ minimal pair words (e.g., *crash* and *clash*). The word lists are in Table A1 of the Appendix.

Acoustic measurements of the natural stimuli are displayed in Fig. 1. Formant frequency measurements were made using hand-corrected LPC analyses. For Fig. 1 (not for later signal processing), the formant frequency measure-

ments were normalized for each talker by a multiplicative factor that equated the F2 frequency that each speaker used in the vowel /i/ [which is stable between speakers; Evans (2005)]; this normalization was similar to methods that have been used to equate vowel spaces between speakers [see Adank *et al.* (2004) for a review]. The closures (i.e., the beginning of the consonant in which the articulators are held in a relatively static position) and transitions (i.e., between the consonant and the following vowel) were hand marked by inspecting F1 transitions and the amplitude envelope changes.

As expected, there was little overlap between the distributions of F3 frequencies for /r/ and /l/, although the frequency difference between /r/ and /l/ was moderately lower in medials and clusters. F2 also differed between /r/ and /l/ for initials, but there was substantial overlap between the distributions; the F2 difference was reduced further for medials and clusters. The F2 cue thus has some utility for distinguishing initial-position /r/ and /l/, and has less value in other positions [F2 may be even less useful for American English; see Lotto *et al.* (2004)]. There was substantial overlap between the distribution of closure durations for /r/ and /l/ at all syllable positions, indicating that this would be an unreliable acoustic cue for listeners. Transition duration was longer for /r/ than for /l/, with some overlap between distributions; the differences between transition durations was reduced for medials and clusters compared to initials, but the overlap between the distributions was not greater. Transition duration was thus similar to F2 frequency; both had some utility as secondary cues for distinguishing /r/ and /l/.

2. Signal processing

All of the signal processing was conducted only on the initial-position stimuli, and the processing combined changes in duration with changes in formant frequencies. The duration changes were made using the PSOLA function in Praat [Boersma and Weenink (2004)], and the duration of the closure and transition intervals were independently manipulated. The formant frequency changes were made via LPC analysis and resynthesis within Praat; the LPC parameters (e.g., prediction order and frequency cutoff) were hand selected for each stimulus so that the analysis correctly tracked the formants in the spectrogram, an LPC residual was created by inverse filtering the stimulus, a new LPC filter was created by manipulating the formant frequencies in the LPC analysis, and the final stimulus was created by filtering the LPC residual with the new LPC parameters. In order to improve the naturalness of the stimuli, the high-frequency energy that was removed by LPC (i.e., energy that was above the cut-off frequency) was added back into the signal following the LPC manipulations.

Prior to the construction of the final versions of the stimuli, the signal processing was pilot tested to make sure that it did not reduce the identifiability of these stimuli. A group of 9 native British English speaking listeners gave forced-choice /r/-/l/ identification responses for stimuli from multiple talkers and words, with the signal-processing dimensions (F3 enhancement, F2, closure duration, and transition duration) varying independently. Listeners were correct

on 98.7% of the trials. Given this high level of accuracy, the final versions of the stimuli were simply screened by a research assistant to ensure that they were intelligible.

a. All Enhancement condition. Throughout training, F3 was set to extreme values during the closure (enhancing the difference between /r/ and /l/) and the duration of the closure was increased by 100 ms (ensuring that all stimuli would have a closure that was long enough to be audible; see Fig. 2 for example spectrograms, and Table A2 of the Appendix for specific values). For /r/, F3 was set to be 100 Hz higher than the median F2 frequency during the closure. For /l/, F3 was set to be 100 Hz lower than the median F4 frequency during the closure. The distribution of F3 across all stimuli was thus bimodal, with the values for /r/ and /l/ being further apart than in natural stimuli. To prevent the formants from crossing, F2 and F4 were flattened by setting them to their median values throughout the closure. During the transition, the degree of F3 enhancement was reduced linearly so that there was no enhancement at the end of the transition (i.e., the F3 frequency was the same as in the original recording by the time that the transition was over).

b. Perceptual Fading condition. On the first day of training, the stimuli were fully enhanced (i.e., the same as in the All Enhancement condition) and the amount of enhancement was linearly decreased each day until, by Day 10, the difference between /r/ and /l/ was less distinctive than normal (see Fig. 2 and Table A2 of the Appendix). The F3 values were based on the differences between the fully enhanced and normal values, such that there was 100% enhancement of F3 values on Day 1 (i.e., 100 Hz greater than F2 for /r/ and 100 Hz less than F4 for /l/), there was 50% enhancement on Day 4 (i.e., values were the average of the fully enhanced and normal), there was 0% enhancement on Day 7 (i.e., normal values), and -50% enhancement on Day 10 (e.g., for /r/, F3 was higher than normal, by an amount equal to half the difference between the fully enhanced and normal values). The “negative enhancement” in Days 8–10 lead to overlap of the F3 distributions for /r/ and /l/, but F3 remained a cue to the contrast because of vowel coarticulation (i.e., F3 in the vowel tends to be lowered following an /r/ and the vowel formant frequencies were unaffected by this manipulation). The amount of closure duration lengthening was 100 ms on Day 1, and decreased linearly to 0 ms (i.e., no lengthening) on Day 7; closure duration remained at its normal values on Days 8–10 (i.e., it was not shortened to match the negative F3 enhancement).

c. Secondary Cue Variability condition. On the first day of training, F2 during the closure, closure duration, and transition duration were set to the median values for all stimuli from the speaker; the stimuli thus had no variability and no differences between /r/ and /l/ along these acoustic dimensions. By Day 10, the stimuli randomly varied between the maximum and minimum F2 frequency, closure duration, and transition duration used by that speaker for all /r/ and /l/ stimuli (see values in Table A2 of the Appendix). That is, the stimuli had random combinations of short and long closures and transitions, and high and low F2 frequencies, for both /r/ and /l/. The distributions of values were bimodal (i.e., the values were either set to the minimum or maximum). The

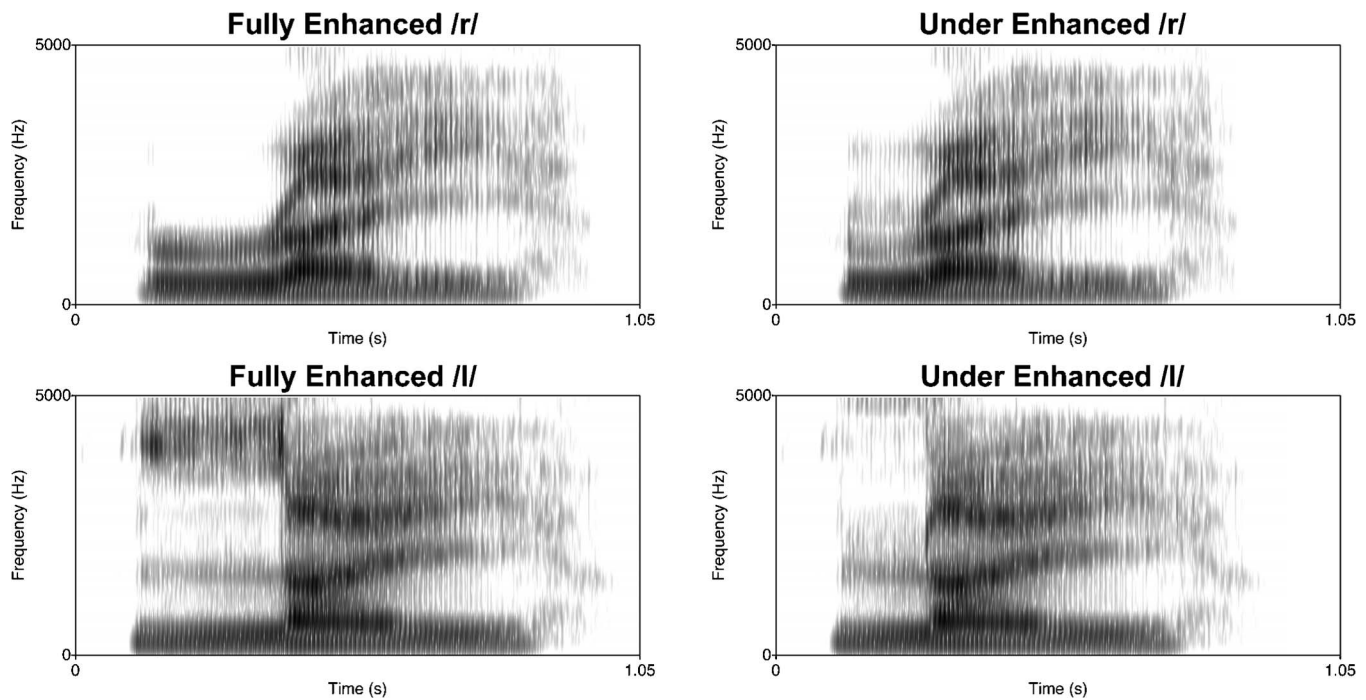


FIG. 2. Spectrographic examples of stimuli in Perceptual Fading, for the words *ray* and *lay*. The fully enhanced versions (used on the first day of training in Perceptual Fading, and on all days in All Enhanced) set F3 to extreme values and increased the duration of the initial closure by 100 ms. The under-enhanced versions (used on the last day of training in Perceptual Fading) decreased the contrast in F3 frequencies and had natural closure durations.

values of F2 were limited for each stimulus so that they were at least 100 Hz greater than F1 and 100 Hz less than F3 (e.g., if the maximum F2 frequency across all stimuli was greater than the F3 frequency for a particular stimulus, as occurred sometimes for /t/, F2 was set to be 100 Hz less than F3). The variability increased from Day 1 to Day 10 by increasing the range of the values. That is, Day 1 had 0% range (all values set to medians for that speaker), Day 10 had 100% range (all values set to the maximum or minimum), and Day 2, for example, had 11% range (all values set to 11% of the difference between the median and the maximum or minimum).

D. Procedure

1. Training

Each subject was randomly assigned to one of the 4 training conditions: HVPT, All Enhanced, Perceptual Fading, and Secondary Cue Variability. Except for the stimulus differences between the conditions, the training procedures were identical. The training comprised 10 sessions, each taking approximately 1/2 hour to complete. The subjects ran no more than one session per day, and completed the training over a 2 to 3 week period. There was a different talker each day [as in previous HVPT studies, e.g., Logan *et al.* (1991); Lively *et al.* (1993)], and each subject received the same talker order regardless of condition.

At the beginning of each session, subjects heard a greeting from the talker (e.g., “Hello, my name is Ian. You’re going to hear my voice in the training today. Let’s get started.”) that was synchronized with an animated face. They then completed 300 trials (3 repetitions of the 100 stimuli, presented in a random order) of forced-choice identification with feedback. On each trial, subjects saw minimal pair

words on the computer screen (e.g., *rock* and *lock*; the words varied for each stimulus), heard one of the words, and then clicked on the “R” or “L” button to indicate which of those words they thought they heard. If they answered correctly, they saw a “Correct!” message on the computer screen, heard a cash register sound, and heard the stimulus played again. If they answered incorrectly, they saw a “Wrong.” message on the computer screen, heard two beeps with descending pitch, and then heard the stimulus played twice again. The screen displayed a running tally of the percentage of correct responses during the training session.

After each training session, subjects completed a short identification test without feedback or display of the percentage correct. This test tracked their performance as training progressed, and consisted of 20 words that were randomly selected from the training corpus. In order to directly compare the different conditions, the HVPT, Perceptual Fading, and Secondary Cue Variability conditions all used natural speech (i.e., unprocessed) from the talker that had been used in the training session. The All Enhancement condition used enhanced speech, in order to test whether subjects were able to improve when they had not heard *any* natural speech during the course of the training program.

2. Pre/Post identification testing

Before and after completing the period of training, subjects were tested in terms of their identification of natural and cue-manipulated stimuli. The format of each trial was the same as in the training (i.e., forced-choice identification of /t-/l/ minimal pairs), except that subjects did not receive feedback.

a. Natural stimuli. Subjects first completed a practice

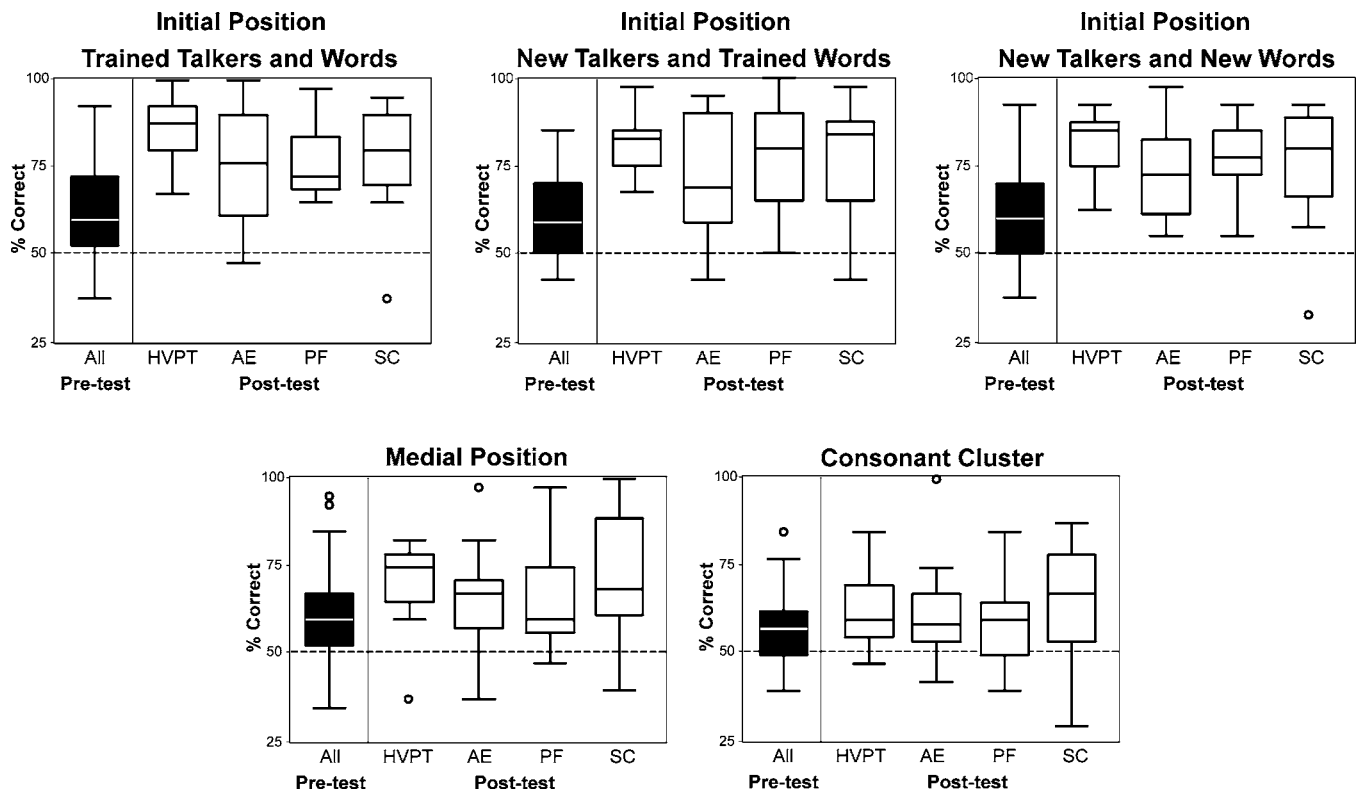


FIG. 3. Identification accuracy for natural /r/ and /l/ stimuli before and after training, for High Variability Phonetic Training (HVPT), All Enhanced (AE), Perceptual Fading (PF), and Secondary Cue Variability (SC). Pre-test scores are aggregated across training techniques, because there were no significant differences between the subjects assigned to the different techniques. Although there was high between-subject variability in pretest scores, there were reliable within-subject improvements after training. Improvement was greatest for initial-position stimuli, and this training generalized to new talkers and initial-position words. Generalization was weaker for medials and clusters. There were no significant differences between training techniques.

block of 10 trials, comprising initial-position /r/-/l/ minimal pairs from the training set. Subjects then completed 10 experimental blocks (5 conditions, with 2 talkers for each condition) of 20 trials each. The 5 conditions were: (1) trained talkers and words, (2) new talkers and trained words, (3) new talkers and new initial-position words, (4) new talkers and new medial-position words, and (5) new talkers and new consonant-cluster words. Each of the 10 blocks had a different word list (i.e., words were not repeated). All subjects received the same 10 lists of words, and all subjects were tested on the same talkers. The word lists are in Table A1 of the Appendix.

b. Cue-manipulated stimuli. Following identification testing for natural stimuli, subjects completed the same forced-choice identification task with stimuli that had been signal-processed to alter the acoustic cues. There were two talkers and 9 stimulus conditions, and the talkers and words were drawn from the training corpus. The conditions were: (1) Short Closure (i.e., closure duration set to a speaker's minimum, as in Day 10 of Secondary Cue Variability), (2) Long Closure (i.e., closure duration set to a speaker's maximum, as in Day 10 of Secondary Cue Variability), (3) Short Transition (i.e., transition duration set to a speaker's minimum, as in Day 10 of Secondary Cue Variability), (4) Long Transition (i.e., transition duration set to a speaker's maximum, as in Day 10 of Secondary Cue Variability), (5) Low F2 (i.e., F2 set to a speaker's minimum, as in Day 10 of Secondary Cue Variability), (6) High F2 (i.e., F2 set to a

speaker's maximum, as in Day 10 of Secondary Cue Variability), (7) Enhanced F3 (i.e., F3 the same as Day 1 of Perceptual Fading, with no duration lengthening), (8) Negative Enhancement of F3 (i.e., F3 the same as Day 10 of Perceptual Fading), and (9) Natural stimuli (i.e., no acoustic manipulation). There were 2 blocks (1 for each talker), with 180 trials per block (20 trials per stimulus condition, with the conditions mixed randomly within each block).

III. RESULTS

A. Pre/Post identification of natural stimuli

The pre- and post-training results for initial-position words (see Fig. 3) were analyzed by MANOVA; the RAU-transformed percentages correct [Rationalized Arcsine Units; Studebaker (1985)] were analyzed with pre/post (i.e., before and after training) and stimulus condition (trained talkers and words; new talkers and trained words; and new talkers and new words) coded as within-subject factors (i.e., as a repeated measure) and training condition coded as a between-subject factor. There was a significant effect of pre/post, $F(1, 58) = 102.01, p < 0.001$, demonstrating that identification performance improved after training (mean improvement of 18 percentage points). There was a significant effect of stimulus condition, $F(2, 57) = 57.00, p = 0.008$; on average, accuracy for the trained talkers and words was 2.3 percentage points higher than for the other two conditions. However, there was no significant interaction of stimulus condition and

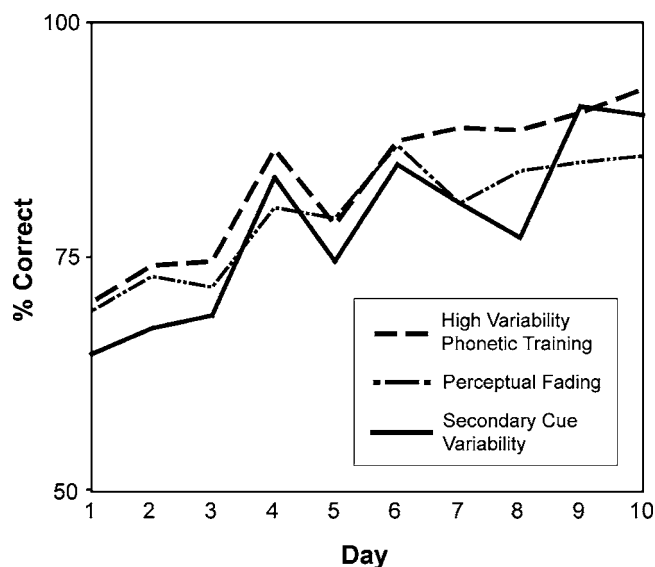


FIG. 4. Changes in identification accuracy for /r/ and /l/ over time. Subjects gradually improved in identification accuracy during the 10 training sessions. There were no significant differences between training techniques.

pre/post, $p > 0.05$, which indicated that the magnitude of improvement with training was similar across the stimulus conditions (i.e., the training generalized to new words and talkers). There was no significant effect of training condition and no significant interactions, $p < 0.05$, demonstrating that all four training conditions yielded similar levels of improvement in identification accuracy.

To further examine generalization, the pre- and post-training results for the different syllable positions (initial, medial, cluster) were also analyzed by MANOVA. There was a significant effect of pre/post, $F(1, 58) = 99.65, p < 0.001$, demonstrating that identification performance improved after training. There was a significant effect of position, $F(2, 57) = 42.29, p < 0.001$, demonstrating that initials were more accurately identified than were medials and clusters. There was also a significant interaction of position and pre/post, $F(2, 56) = 12.61, p < 0.001$, demonstrating that the improvement in training on initials did not fully generalize to the other syllable positions (median improvement in accuracy was 18.0 percentage points for initials, 8.7 percentage points for medials, and 5.5 percentage points for clusters). There was no significant effect of training condition and no significant interactions, $p < 0.05$, demonstrating that all four training conditions yielded similar levels of improvement in identification accuracy.

B. Changes in identification performance during training

Although the pre/post analysis demonstrated that there were no differences in improvement between training methods, the results for each training session were analyzed to examine whether identification improved at different rates among the training conditions. The results for the tracking test at the end of each training session, for all conditions except All Enhanced (which did not have a tracking test with natural stimuli), are displayed in Fig. 4. Identification performance appeared to improve steadily for all conditions, with

out reaching asymptotic levels. There was some variation between conditions in Days 4–6 that may have been caused by differences in the intelligibility of speakers. A MANOVA analysis of the RAU-transformed percentages correct revealed that there was a significant main effect of day, $F(9, 29) = 9.62, p < 0.001$, demonstrating that training improved recognition. There was no significant effect of training condition, $p > 0.05$, but there was a marginally significant interaction between day and training condition, $F(18, 58) = 1.67, p = 0.072$. It is thus possible that there were some differences in the rate of learning in the different conditions, but the mean data in Fig. 4 suggests that the differences, if reliable, were small.

C. Pre/Post assessment of cue-manipulated stimuli

Figure 5 displays the percentage correct for /r/ and /l/ when the secondary cues were altered. A preliminary examination revealed that there were substantial differences in response bias between conditions (e.g., listeners gave more /l/ responses when transition durations were short than when they were long), so Detection Theory [Macmillan and Creelman (1991)] was used to calculate the sensitivity (d') and bias (c) for each condition. The bias statistic provides a way of measuring cue weighting. For example, if listeners are biased to identify stimuli with long transitions as /r/ and short transitions as /l/, this would demonstrate that the transition duration affects whether they identify the stimulus as /r/ or /l/ and thus indicate that transition duration had high weighting in the categorization decision. If listeners had zero bias for a cue, this would indicate that the cue does not affect /r/-/l/ identification, and thus had low weighting.

For each stimulus condition, a MANOVA analyzed the bias with pre/post (i.e., before and after training) as a within-subjects variable and training condition as a between-subjects variable. Although d' was also analyzed, the results are not reported here because they simply corresponded with the natural identification results (i.e., d' increased with training, but there were no interactions with training condition).

For closure duration, listeners had a strong bias before training to label long closures as /r/ and short closures as /l/, demonstrating that they had a high weighting for this cue in their categorization decision. After training, subjects had significantly reduced response bias for long closures, $F(1, 42) = 2.51, p < 0.001$, but there was no change for short closures, $p > 0.05$; training thus modified their cue weightings somewhat (particularly for long closures), but they continued to use this cue. There was no main effect of training condition or significant interaction with pre/post, $p > 0.05$.

For transition duration, listeners had a strong bias before training to label long transitions as /r/ and short transitions as /l/, demonstrating that they had a high weighting for this cue. After training, subjects had significantly reduced bias for long transitions, $F(1, 42) = 21.0, p < 0.001$, but had significantly increased bias for identifying short-transition stimuli as /l/, $F(1, 42) = 18.0, p < 0.001$; training thus changed how they used this cue, but they continued to give high weight to this cue overall. For long-transition stimuli, there was a significant main effect of training condition, $F(3, 42) = 4.4, p$

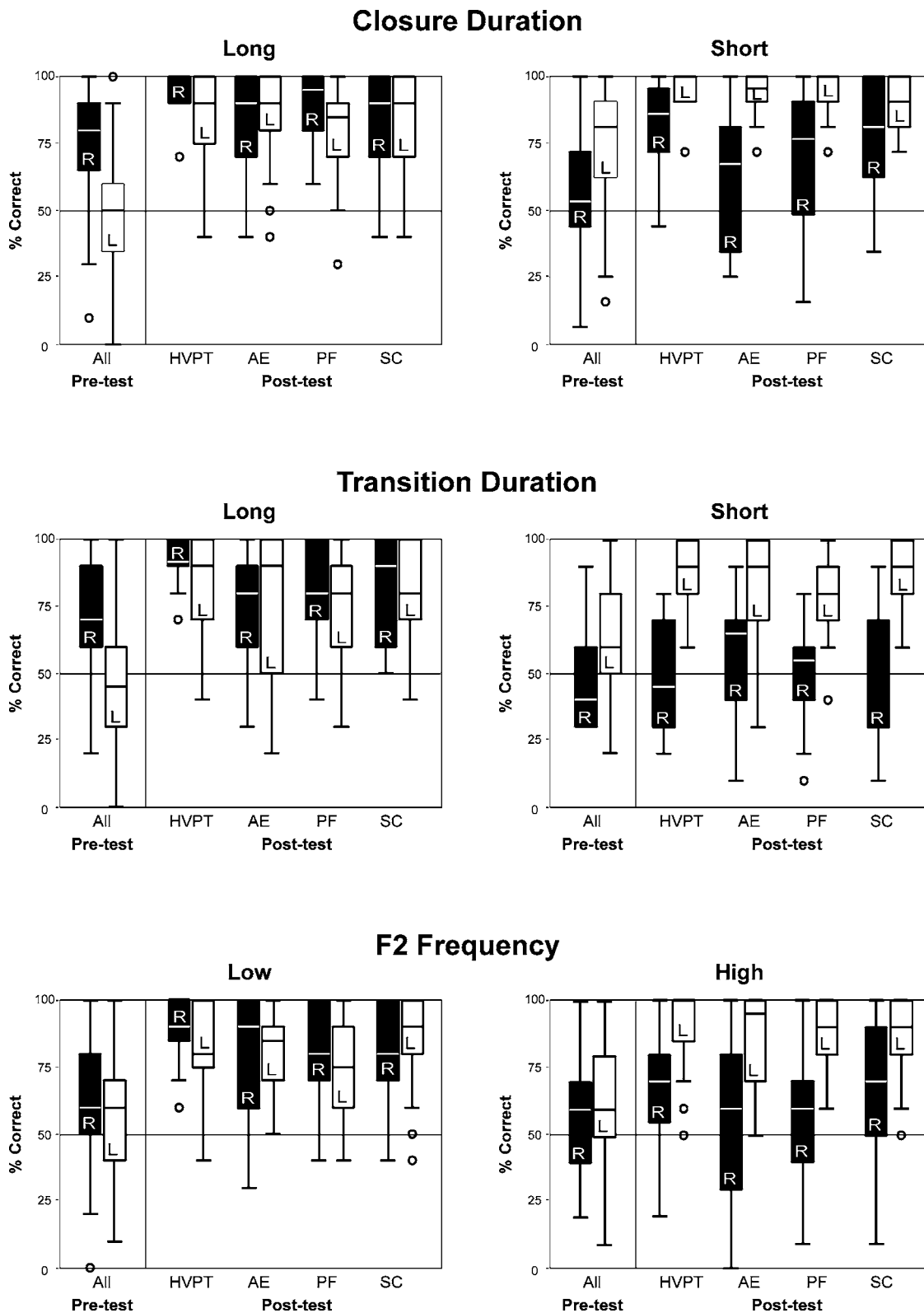


FIG. 5. Identification accuracy for cue-manipulated stimuli before and after training, for High Variability Phonetic Training (HVPT), All Enhanced (AE), Perceptual Fading (PF), and Secondary Cue Variability (SC). Subjects were biased to identify stimuli with long closures and transitions as /r/ before training, but this bias was reduced after training. Subjects were biased to identify stimuli with short closures and transitions as /l/ before training; this /l/ bias increased after training for short transitions and for high F2 frequencies.

=0.009, but no significant interaction with pre/post, $p > 0.05$; there were some differences in long-transition biases between subject groups (i.e., All Enhanced subjects had less bias to identify /r/) but this difference seemed more a result

of the subject assignment to conditions rather than a result of training. For short-transition stimuli, there was no main effect of training condition or significant interaction with pre/post, $p > 0.05$.

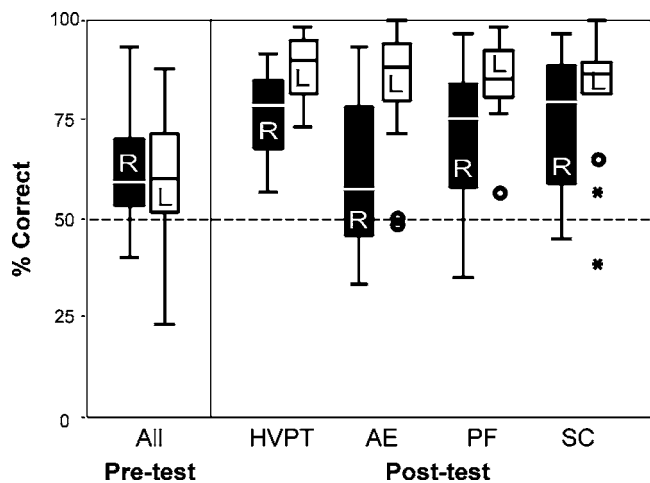


FIG. 6. Identification accuracy for natural stimuli before and after training, divided by /r/ and /l/, for High Variability Phonetic Training (HVPT), All Enhanced (AE), Perceptual Fading (PF), and Secondary Cue Variability (SC). Subjects became biased to identify stimuli as /l/ after training, leading to more accurate identification of /l/ than /r/.

For F2 frequency, listeners began with no strong biases, demonstrating low cue weighting. After training, they did not significantly change their biases for low-F2 frequency stimuli, $p > 0.05$, but they increased their bias to identify high-F2 frequency stimuli as /l/, $F(1, 42) = 25.8, p < 0.001$, demonstrating an increased weighting for this cue. There were no main effects of training condition or significant interactions with pre/post, $p > 0.05$.

For the other conditions in the cue-manipulated identification tests, there was low bias before training, but a small bias to identify stimuli as /l/ after training (natural stimuli, $F(1, 42) = 31.6, p < 0.001$; enhanced F3, $F(1, 42) = 11.8, p = 0.001$; negatively enhanced F3, $F(1, 42) = 9.7, p = 0.003$). Although there were no significant interactions between training condition and pre/post, $p > 0.05$, there were significant main effects of training condition for enhanced F3, $F(3, 42) = 3.2, p = 0.033$, and negatively enhanced F3, $F(3, 42) = 4.7, p = 0.006$; subjects in the All Enhanced condition had a slightly stronger bias overall to identify stimuli as /l/.

To further explore the apparent increase in /l/-bias after training, the identification of natural stimuli in the main pre/post test was reanalyzed in terms of response bias. As displayed in Fig. 6, there was a significant increase in bias to identify stimuli as /l/ after training, $F(1, 58) = 46.7, p < 0.001$. There was no main effect of training condition or significant interaction, $p > 0.05$. The increase in /l/-bias can also be interpreted as a differential increase in identification accuracy. That is, individuals had greater improvement in identification accuracy for /l/ than for /r/.

IV. DISCUSSION

The results demonstrated that there was significant improvement of /r/ and /l/ identification by Japanese adults; the identification of initials improved by an average of 18 percentage points, which is at least as large as in previous studies that have used HVPT [Hazan *et al.* (in press); Logan *et al.* (1991); Lively *et al.* (1993); Bradlow and Pisoni (1999);

Bradlow *et al.* (1999)]. There were no significant differences in how well the methods improved perception, both in terms of correct identification and in the use of secondary acoustic cues. Regarding the applied goal of aiding L2 phoneme learning, it thus appears that training with natural speech is currently the best method, because the signal processing techniques used here are more labor intensive and offer no additional gains in performance. However, the lack of significant differences between training methods also demonstrates that there is nothing particularly special about having fully natural variability; Pisoni and colleagues [e.g., Logan *et al.* (1991); Lively *et al.* (1993)] have emphasized the importance of exposing listeners to natural speech from multiple talkers in order to teach listeners how individual talkers covary their acoustic cues, but training was just as successful here using methods in which the F3 or secondary cues were varied unnaturally. Even though the specific signal-processing methods used here did not improve upon natural speech, the general approach of training with manipulated speech is supported by our results; listeners clearly could learn under these conditions. It remains to be seen whether differences between the methods would emerge if long-term retention or production were measured, if other phonetic contrasts or cues were trained that might benefit more from enhancement, or if more natural enhancement (i.e., clear speech) was used.

The changes in secondary-cue biases support the general view that secondary cues are important to L2 phoneme learning, but the changes were not as predicted by the perceptual interference account [Iverson *et al.* (2003)]. That is, there was not a general reduction in the salience of secondary acoustic cues as identification performance improved. Rather, there was a decrease in response bias for some types of secondary cues (long closures and transitions) and an increase in others (short transitions and high F2 frequencies).¹ The patterns of secondary cue weightings after training also did not correspond to their validity in natural stimuli, contrary to the patterns of shrinking and stretching predicted by exemplar models [e.g., Logan *et al.* (1991); Lively *et al.* (1993)]. For example, /r/ and /l/ have substantial overlap in terms of closure duration (Fig. 1), with /r/ having shorter closures on average, but listeners persisted in being biased to label short-closure stimuli as /l/. Moreover, the acoustic measurements revealed that F2 frequency and transition duration had some cue validity, but their natural distributions did not predict the asymmetries in subjects' cue weightings. For example, the acoustic measurements indicated that the presence of a long transition was a more reliable cue for /r/ than a short transition was for /l/ (i.e., there was more overlap between the distributions for durations < 25 ms), but the subjects changed their cue weightings in an opposite way, reducing /r/-bias for long transitions and increasing /l/-bias for short transitions.

On the surface, the results are also at odds with category assimilation accounts of /r/-/l/ learning. Aoyama *et al.* (2004) claimed that English /l/ is more strongly assimilated into the Japanese /r/ category than is English /r/, and this makes English /r/ easier to learn. Our results showed the opposite pattern of learning, with /l/ identification improving more with

training than /r/. However, it is notable that all of the unpredictable cue biases described above involved stimuli that would be expected to be strongly assimilated into the Japanese /r/ category. That is, assimilation has been shown to be stronger for English /r/ and /l/ stimuli that have high F2 frequencies [Iverson *et al.* (2003)], and the short duration of the Japanese /r/ would likely produce stronger assimilation effects for English stimuli with short closures and transitions; all of these types of stimuli were biased to be identified as English /l/ after training. It is thus plausible that training caused subjects to learn to systematically label a stimulus as /l/ whenever it was strongly assimilated into the Japanese /r/ category [Japanese students are sometimes taught this strategy when learning English; Lotto *et al.* (2004)]. For stimuli that were probably not strongly assimilated (low F2 frequencies, and long closures and transitions), subjects had low biases for secondary acoustic cues after training, in accord with the predictions of perceptual interference [Iverson *et al.* (2003)] and exemplar models [e.g., Logan *et al.* (1991); Lively *et al.* (1993)]. Category assimilation and perceptual interference may therefore combine to affect how English /r/ and /l/ are learned.

This account may help explain why there were no differences between training conditions. The Secondary Cue Variability technique was designed to eliminate the validity of secondary acoustic cues, such that, for example, the distributions of transition duration would be identical for /r/ and /l/ rather than as in natural stimuli (i.e., /r/ longer than /l/, with some overlap between distributions). This technique successfully reduced /r/-bias for stimuli with long transitions, but /r/-bias was reduced by all of the other training techniques too. It seems that natural variability in transition durations was sufficient to change biases, and that the Secondary Cue Variability technique offered no additional improvement. For the stimuli that were strongly assimilated into the Japanese /r/ category, the variability in the stimuli may not have mattered very much (e.g., stimuli with short closures were biased to be labeled as /l/ even though this was unmotivated by the acoustic distribution of stimuli in any of the conditions), and hence there were no differences between training conditions.

The poor generalization of training to medials and clusters may simply have occurred because they were too dissimilar to the initial-position stimuli that were used during training. That is, they may not have mapped onto the same categorization space [cf., Lively *et al.* (1993)]. The acoustic measurements, though, suggest that category assimilation may also have played a role. That is, medials and clusters had shorter closure durations, and /r/ in those positions had higher F2 frequencies, all of which promote assimilation into the Japanese /r/ category. That being said, training did not cause subjects to identify all medials and clusters as /l/ (there was an /l/-bias, as in the other conditions), and these positions have not been shown to be particularly resistant to training in previous studies that included medials and clusters in the training set [e.g., Logan *et al.* (1991); Lively *et al.* (1993)].

In summary, the results demonstrate that listeners modify their use of secondary cues during L2 phoneme

learning, and suggest that both perceptual interference and assimilation affect the learning process. Although the signal-processing methods used here did not improve the effectiveness of training, the results demonstrate that there is still room for improvement in existing methods; identification performance for most individuals did not reach ceiling after training, and category assimilation remained a barrier to learning. What may be needed are new techniques that are specifically targeted to reduce assimilation effects.

APPENDIX

TABLE A1. Words used in the experiment.

Trained words					
lack	rack	leer	rear	loaves	roves
lad	rad	lent	rent	lob	rob
lag	rag	lice	rice	lobe	robe
laid	raid	lick	Rick	lock	rock
lake	rake	lid	rid	long	wrong
lamb	ram	lies	rise	look	rook
lane	rain	life	rife	loom	room
lank	rank	lift	rift	loss	Ross
late	rate	light	right	lot	rot
laughed	raft	limb	rim	loud	rowed
laws	roars	lime	rhyme	lout	rout
lay	ray	line	Rhine	low	row
laze	raise	lined	rind	lows	rose
leach	reach	link	rink	lump	rump
leaf	reef	lip	rip	lush	rush
leak	reek	lit	writ	lust	rust
led	red	loan	roan		
New initial-position words (pre/post test)					
lace	race	lest	rest	loot	root
lamp	ramp	lewd	rude	lope	rope
lap	wrap	lied	ride	lord	roared
lapse	raps	list	wrist	lose	ruse
law	raw	load	road	lug	rug
leap	reap	loam	roam	lung	rung
lens	wrens	loon	rune		
New medial-position words (pre/post test)					
alive	arrive	elect	erect	palling	poring
allows	arouse	fairly	fairy	pilot	pirate
bawling	boring	fallow	farrow	starling	starring
believe	bereave	holler	horror	tally	tarry
bellies	berries	mallow	marrow	teller	terror
calling	coring	miller	mirror	whirling	whirring
collect	correct	palate	parrot		
New cluster words (pre/post test)					
bland	brand	flame	frame	glue	grew
bloom	broom	flesh	fresh	plank	prank
blunt	brunt	flows	froze	plays	praise
blush	brush	flute	fruit	plod	prod
clamp	cramp	glass	grass	splay	spray
climb	crime	glaze	graze	splint	sprint
cloud	crowd	glow	grow		

TABLE A2. **Ranges of acoustic cue values for each training condition and day.** The values in each box represent the minimum and maximum values for /r/ and /l/ (i.e., min-max [r]/min-max [l]); the SC condition had identical acoustic distributions for /r/ and /l/ in terms of F2, closure duration, and transition duration, so only one range of values is listed. HVPT (High Variability Phonetic Training) used natural recordings with a different talker on each day. AE (All Enhanced), PF (Perceptual Fading), and SC (Secondary Cue Variability) used signal-processed versions of these natural recordings.

Day	Condition	F2 (Hz)	F3 (Hz)	Closure (ms)	Transition (ms)
1	HVPT (natural)	719–1377/894–1538	1378–2016/2144–2810	17–92/31–108	6–44/7–20
	AE	same as natural	819–1477/2981–3423	117–192/131–208	same as natural
	PF	same as natural	819–1477/2981–3423	117–192/131–208	same as natural
	SC	1140–1140	same as natural	64–64	17–17
2	HVPT (natural)	716–1382/1638–2275	1892–2722/2954–3994	41–132/72–191	28–64/10–33
	AE	same as natural	816–1482/3470–4586	141–232/172–291	same as natural
	PF	same as natural	1040–1647/3434–4419	124–215/155–274	same as natural
	SC	1581–1754	same as natural	98–115	27–33
3	HVPT (natural)	945–1370/1616–2346	1932–2498/2656–3458	22–78/25–103	20–58/9–24
	AE	same as natural	1054–1470/4054–4686	122–178/125–203	same as natural
	PF	same as natural	1395–1760/3634–4281	89–145/92–170	same as natural
	SC	1467–1778	same as natural	43–61	19–30
4	HVPT (natural)	620–1114/1077–1692	1298–1724/2424–2998	49–186/106–167	26–78/8–30
	AE	same as natural	755–1214/3270–3967	149–286/206–267	same as natural
	PF	same as natural	1135–1408/2922–3371	99–236/156–217	same as natural
	SC	961–1318	same as natural	96–141	20–43
5	HVPT (natural)	956–1613/1317–1881	1551–2699/2402–3768	30–93/30–118	13–63/9–30
	AE	same as natural	1056–1713/3722–4389	130–193/130–218	same as natural
	PF	same as natural	1403–2287/2837–3871	63–126/63–151	same as natural
	SC	1214–1625	same as natural	54–93	15–39
6	HVPT (natural)	735–1346/1645–2151	1813–2267/2881–3925	29–91/70–137	36–69/9–29
	AE	same as natural	1019–1446/3757–4391	129–191/170–237	same as natural
	PF	same as natural	1701–2098/3056–3959	46–108/87–154	same as natural
	SC	1139–1927	same as natural	51–111	18–52
7	HVPT (natural)	619–1122/1041–1596	1216–1694/2432–2826	42–113/49–118	19–50/7–28
	AE	same as natural	719–1222/2975–3601	142–213/149–218	same as natural
	PF	same as natural	same as natural	same as natural	same as natural
	SC	777–1429	same as natural	54–104	12–44
8	HVPT (natural)	1017–1324/1415–2150	1749–2640/2924–3965	41–102/68–138	21–69/7–33
	AE	same as natural	1160–1424/3621–4658	141–202/168–238	same as natural
	PF	same as natural	1840–2874/2774–3855	same as natural	same as natural
	SC	1098–1979	same as natural	50–126	12–60
9	HVPT (natural)	602–1211/1208–2960	1507–2586/2759–3336	34–117/33–148	30–89/9–35
	AE	same as natural	862–1311/3837–4470	134–217/133–248	same as natural
	PF	same as natural	1682–3060/2277–3036	same as natural	same as natural
	SC	682–2778	same as natural	38–140	12–83
10	HVPT (natural)	835–1132/1044–1423	1105–1797/2208–2659	52–212/71–242	26–75/5–30
	AE	same as natural	995–1232/3352–3799	152–312/171–342	same as natural
	PF	same as natural	1117–2144/1507–2184	same as natural	same as natural
	SC	835–1423	same as natural	52–242	5–75

ACKNOWLEDGMENTS

We are grateful to Professor Masaki Taniguchi (Kochi University) for his help in organizing the testing sessions in Japan. This research was funded by Grant No. RES-000-22-0445 from the Economic and Social Research Council of Great Britain.

¹The pattern of bias changes could also be interpreted as being a result of a global increase in /l/ bias, because most conditions changed bias in the direction of more /l/ responses (i.e., reductions in /r/ bias and increases in /l/ bias both result from increases in the proportion of /l/ responses). However, the /l/ bias did not change to the same extent across all conditions. For example, there was no bias change for low-F2 frequencies, but a significant increase in /l/ bias for high-F2 frequencies. This suggests that there were cue-specific changes to biases rather than a simple global increase in the

proportion of /l/ responses or an overall increase in /l/ identification accuracy.

- Adank, P., Smits, R., and van Hout, R. (2004). "A comparison of vowel normalization procedures for language variation research," *J. Acoust. Soc. Am.* **116**, 3099–3107.
- Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., and Yamada, T. (2004). "Perceived phonetic distance and L2 learning: The case of Japanese /r/ and English /l/ and /r/," *J. Phonetics* **32**, 233–250.
- Best, C. T. (1994). "The emergence of native-language phonological influences in infants: A perceptual assimilation model." in *The development of speech perception: The transition from speech sounds to spoken words*, edited by J. C. Goodman and H. C. Nusbaum (MIT Press, Cambridge, MA), pp. 167–224.
- Best, C. T., McRoberts, G. W., and Goodell, E. (2001). "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system," *J. Acoust. Soc. Am.* **109**, 775–794.
- Best, C. T., McRoberts, G. W., and Sithole, N. M. (1988). "Examination of

- perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants," *J. Exp. Psychol. Hum. Percept. Perform.* **14**, 345–360.
- Best, C. T., and Strange, W. (1992). "Effects of language-specific phonological and phonetic factors on cross-language perception of approximants," *J. Phonetics* **20**, 305–330.
- Boersma, P., and Weenink, D. (2004). "Praat: doing phonetics by computer" [Computer program]. Retrieved from <http://www.praat.org/>
- Bradlow, A. R., and Pisoni, D. B. (1999). "Recognition of spoken words by native and non-native listeners: Talker-, listener- and item-related factors," *J. Acoust. Soc. Am.* **106**, 2074–2085.
- Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., and Tohkura, Y. (1999). "Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production," *Percept. Psychophys.* **61**, 977–985.
- Doeleman, T. J., Conley, R. J., Pruitt, J. S., Iverson, P., Kuhl, P. K., and Stevens, E. B. (2000). "Perceptual identification training of American English /r/ and /l/ by Japanese speakers generalizes to novel stimuli and tasks," *J. Acoust. Soc. Am.* **108**, 2652.
- Evans, B. G. (2005). "Plasticity in speech perception and production: A study of accent change in young adults," Unpublished PhD dissertation, University of London, UK.
- Flege, J. E. (1995). "Second language speech learning: Theory, findings, and problems," in *Speech perception and language experience: Issues in cross-language research*, edited by W. Strange (York Press, Baltimore), pp. 233–277.
- Flege, J. E. (1999). "Age of learning and second language speech." In Birdsong (Ed.), *Second Language Learning and the Critical Period Hypothesis* (pp. 101–131). London: Erlbaum.
- Flege, J. E., Schirru, C., and MacKay, I. R. A. (2003). "Interaction between native and second language phonetic subsystems," *Speech Commun.* **40**, 467–491.
- Flege, J. E., Takagi, N., and Mann, V. (1995). "Japanese adults can learn to produce English /r/ and /l/ accurately," *Lang Speech* **38**, 25–55.
- Gordon, P. C., Keyes, L., and Yung, Y. F. (2001). "Ability in perceiving nonnative contrasts: Performance on natural and synthetic speech stimuli," *Percept. Psychophys.* **63**, 746–758.
- Goto, H. (1971). "Auditory perception by normal Japanese adults of the sounds "L" and "R"," *Neuropsychologia* **9**, 317–323.
- Guion, S. G., Flege, J. E., Akahane-Yamada, R., and Pruitt, J. (2000). "An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants," *J. Acoust. Soc. Am.* **107**, 2711–2724.
- Harnsberger, J. D. (2001). "On the relationship between identification and discrimination of non-native nasal consonants," *J. Acoust. Soc. Am.* **110**, 489–503.
- Hazan, V., Sennema, A., Iba, M., Faulkner, A. (in press). "Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English," *Speech Commun.*
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (2003). "A perceptual interference account of acquisition difficulties for non-native phonemes," *Cognition* **87**, B47–B57.
- Jamieson, D. G., and Morosan, D. E. (1986). "Training non-native speech contrasts in adults: Acquisition of the English /ð/-/θ/ contrast by francophones," *Percept. Psychophys.* **40**, 205–215.
- Kuhl, P. K. (2000). "A new view of language acquisition," *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11850–11857.
- Lenneberg, E. (1967). *Biological Foundations of Language* (Wiley, New York).
- Lively, S. E., Logan, J. S., and Pisoni, D. B. (1993). "Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories," *J. Acoust. Soc. Am.* **94**, 1242–1255.
- Logan, J. S., Lively, S. E., and Pisoni, D. B. (1991). "Training Japanese listeners to identify English /r/ and /l/: a first report," *J. Acoust. Soc. Am.* **89**, 874–886.
- Lotto, A. J., Sato, M., and Diehl, R. L. (2004). "Mapping the task for the second language learner: The case of Japanese acquisition of /r/ and /l/," in J. Slifka, S. Manuel, and M. Matthies (Eds.) *From Sound to Sense: 50 + Years of Discoveries in Speech Communication* (MIT Research Laboratory in Electronics, Cambridge, MA), pp. C-181–C-186.
- MacKain, K. S., Best, C. T., and Strange, W. (1981). "Categorical perception of English /r/ and /l/ by Japanese bilinguals," *Applied Psycholinguistics* **2**, 369–390.
- Macmillan, N. A., and Creelman, C. D. (1991). *Detection theory: a user's guide*. (Cambridge University Press, New York).
- McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., and McClelland, J. L. (2002). "Success and failure in teaching the r-l contrast to Japanese adults: predictions of a hebbian model of plasticity and stabilization spoken language perception," *Cognitive, Affective, and Behavioral Neuroscience* **2**, 89–108.
- McClelland, J. L., Fiez, J. A., and McCandliss, B. D. (2002). "Teaching the /r/-/l/ discrimination to Japanese adults: behavioral and neural aspects," *Physiol. Behav.* **77**, 657–662.
- Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J., and Fujimura, O. (1975). "An effect of language experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English," *Percept. Psychophys.* **18**, 331–340.
- Nosofsky, R. (1986). "Attention, similarity, and the identification-categorization relationship," *J. Exp. Psychol. Gen.* **115**, 39–57.
- Nosofsky, R. (1987). "Attention and learning processes in the identification and categorization of integral stimuli," *J. Exp. Psychol.* **15**, 87–108.
- Patkowski, M. (1990). "Age and accent in second language learning: A reply to James Emil Flege," *Applied Psycholinguistics* **11**, 73–89.
- Protopapas, A., and Calhoun, B. (2000). "Adaptive phonetic training for second language learners," 2nd International Workshop on Integrating Speech Technology in Language Learning (InSTIL). Dundee, Scotland.
- Studebaker, G. A. (1985). "A "rationalized" arcsine transform," *J. Speech Hear. Res.* **28**, 455–462.
- Yamada, R. A. (1995). "Age and acquisition of second language speech sounds: Perception of American English /r/ and /l/ by Native Speakers of Japanese," in *Speech perception and language experience: Issues in cross-language research*, edited by W. Strange (York Press, Baltimore), pp. 305–320.