Training Japanese Listeners to Perceive American English Vowels: Influence of Training Sets

Kanae Nishi
Diane Kewley-Port
Indiana University

Purpose: Studies on speech perception training have shown that adult 2nd language learners can learn to perceive non-native consonant contrasts through laboratory training. However, research on perception training for non-native vowels is still scarce, and none of the previous vowel studies trained more than 5 vowels. In the present study, the influence of training set sizes was investigated by training native Japanese listeners to identify American English (AE) vowels.
Method: Twelve Japanese learners of English were trained 9 days either on 9 AE monophthongs (fullset training group) or on the 3 more difficult vowels (subset training group). Five listeners served as controls and received no training. Performance of listeners was assessed before and after training as well as 3 months after training was completed.
Results: Results indicated that (a) fullset training using 9 vowels in the stimulus set improved average identification by 25%; (b) listeners in both training groups generalized improvement to untrained words and tokens spoken by novel speakers; and (c) both groups maintained improvement after 3 months. However, the subset group never improved on untrained vowels.
Conclusions: Training protocols for learning non-native vowels should present a full set of vowels and should not focus only on the more difficult vowels.

KEY WORDS: bilingualism, Japanese, speech perception, English language learners

It is well-known that phonology of a learner's native language (L1) and the target language interact in a complicated manner. In order to account for difficulty that second language (L2) learners may face, several models have been proposed that describe the process of non-native speech perception. The perceptual assimilation model (PAM; Best, 1995) and speech learning model (SLM; Flege, 1995) are the two major models. Both assume that naïve second language learners evaluate L2 sounds using their L1 system (*assimilation*). The assimilation of an L2 sound depends on two factors: (a) whether there exists an equivalent L1 category and (b) its associated goodness of fit (ranging from marginal to perfect) in the equivalent L1 category. The accuracy of discrimination between contrasting L2 sounds depends on the similarity of their assimilation patterns into L1 categories.

Strange, Akahane-Yamada, Kubo, Trent, and Nishi (2001) described a study that was motivated by these models. They assessed the perceptual assimilation of 11 American English (AE) vowels (/iː, ɪ, eᶦ, ɛ, æː, ɑː, ʌ, ɔː, oᶷ, ʊ, uː/) by Japanese listeners in six consonantal contexts (/b_b, b_p, d_d, d_t, g_g, g_k/). Their results showed that in the initial stage of learning, none of the AE vowels were perfectly assimilated

1092-4388/07/5006-1496

into a single Japanese category and that the assimilation patterns changed depending on the consonantal contexts.

Briefly, Japanese has five spectrally distinctive long–short vowel pairs (/i-ii, e-ee, a-aa, o-oo, ɯ-ɯɯ/) in which vocalic duration is a primary phonemic cue (Shibatani, 1990). The average temporal ratio between long and short Japanese vowels ranges from 2.2 to 3.2, and the spectral differences between the five long–short pairs are very small (Hirata & Tsukada, 2004). In contrast, AE has 10 monophthongs (/iː, ɪ, ɛ, æː, ɑː, ʌ, ə, ɔː, ʊ, uː/) and six diphthongs (/eᶦ, aᶦ, aᶷ, oᶷ, ɔᶦ, ʲu/; Ladefoged, 1993). Many of these AE vowels are distinguished primarily by spectral properties, and the average duration ratio between inherently long (/iː, eᶦ, æː, ɑː, ɔː, oᶷ, uː/) and short (/ɪ, ɛ, ʌ, ʊ/) vowels is 1.3 (Strange, Bohn, Nishi, & Trent, 2005). Therefore, when the smaller temporal difference between long and short AE vowels is not detected, Japanese listeners are expected to assimilate more than one AE vowel to a Japanese category using primarily spectral cues.

In fact, Strange et al. (2001) reported that Japanese listeners did differentiate between AE /iː/ and /ɪ/. Because of their duration differences, they were perceived as either exemplars of a single spectrally equivalent Japanese category (/ii/) but with different goodness of fit or equivalent to different Japanese categories (/ii/ and /i/, respectively) in all six consonantal contexts. On the other hand, the distinction between /uː/ and /ʊ/ in alveolar contexts was difficult, as expected, because their spectral and temporal differences were very small, and they were perceived as equally good exemplars of a Japanese /ɯɯ/ in alveolar contexts. More interestingly, AE vowels /æː, ɑː, ʌ/ were all assimilated to Japanese /aa, a/, again, each with varying goodness of fit across consonantal contexts. These and other related results raise the question of whether these contrasts are equally trainable and what training method should be used.

Many training studies for difficult L2 contrasts have shown that structured, intensive laboratory training successfully helps L2 learners improve their perception on such difficult distinctions (Iverson, Hazan, & Bannister, 2005; Lively, Logan, & Pisoni, 1993; Logan, Lively, & Pisoni, 1991; Pruitt, Jenkins, & Strange, 2006; also see Bradlow, in press, for a detailed review). Results of these studies also suggested that phonetic identification tasks using highly variable, naturally produced (HVNP) stimulus materials yield the most improvement. Generalization of improvement to novel stimuli and to tokens produced by novel speakers was also reported (Lively et al., 1993; Logan et al., 1991; Pruitt et al., 2006). Based on these results, in the present study we used a phonetic identification task on HVNP stimuli in tests and training. Training materials were nonsense words spoken by 2 speakers (trained speakers). In addition to the trained nonsense word stimuli, nonsense words spoken by 2 additional speakers (new speakers) and real-word stimuli spoken by both trained and new speakers were included in the tests. In this way, generalization was assessed for novel speakers and novel material.

Although they are informative in designing a training protocol, many of the training studies reported thus far have focused mainly on consonant training. They have included, at most, three consonants that contrast by a single feature such as voicing, manner, and place. However, the number of vowel categories in languages is usually more than three, and any two vowels tend to contrast by more than one feature (e.g., various combinations of tongue height, tongue advancement, diphthongization, duration, lip rounding, rhoticity, and so forth; for more detail, see Ladefoged, 1993, 2001). To make the problem with vowels more complicated, the acoustic properties of vowels vary depending on speakers' gender, age, and dialect (Hillenbrand, Getty, Clark, & Wheeler, 1995; Peterson & Barney, 1952) as well as speaking styles (Ferguson & Kewley-Port, 2002; Krause & Braida, 2002, 2004). It is well-known that consonants influence neighboring vowels (*allophonic variation*), but the realization of allophonic variation for the "equivalent" vowels in two languages may not be similar (Hay, Sato, Coren, Moran, & Diehl, 2006; Strange, Weber, Levy, Shafiro, Hisagi, & Nishi, 2007). All of these differences between vowels and consonants indicate that vowel training may require somewhat different protocols from those frequently used in consonant training.

Among the few vowel training studies using HVNP stimuli are those by Akahane-Yamada, Strange, and Kubo (1997) and Sperbeck, Strange, and Ito (2005), who trained Japanese listeners on three AE vowels, /æː, ɑː, ʌ/, and Lambacher, Martens, Kakehi, Marasinghe, and Molholt (2005), who trained Japanese listeners on five AE vowels, /æː, ɑː, ʌ, ɔː, ɝ/. These vowels were chosen based on the common confusions observed for Japanese learners of English. They reported improved perception for trained stimuli as well as tokens spoken by novel speakers, indicating that listeners could concurrently learn up to five L2 vowel categories through training, and improvement was not limited to the specific voice or words presented during training.

However, an interesting result was reported by both Akahane-Yamada et al. (1997) and Sperbeck et al. (2005). They included two additional vowels, /ɛ, ɔː/, in the pre- and posttests and found no improvement for these untrained vowels. In contrast, McClaskey, Pisoni, and Carrell (1983) showed that training on the voicing contrast in stop consonants generalized to other stops with different places of articulation. These results imply that the generalization of training may be contrast-cue specific and that a training protocol using only the subset of vowels may not help in learning a complete set of L2 vowels. In other words, unlike the consonant training, the particular vowel set might have trained learners to focus their attention on the cues relevant only for the trained vowels and to ignore other cues required for other AE vowels. However, it is not known whether the manageability of set size becomes an issue with training of large vowel sets. Thus, in the present study, two stimulus sets were introduced to examine the overall training efficacy: one smaller set (/ɑː, ʌ, ʊ/ [subset]) and another that covers the entire vowel space (/iː, ɪ, ɛ, æː, ɑː, ʌ, ɔː, ʊ, uː/ [fullset]). All AE vowels were embedded in six consonantal contexts (/b_bə/, /b_pə/, /d_də/, /d_tə/, /g_gə/, /g_kə/) that formed 54 nonsense words. These particular contexts were chosen based on the previous cross-language acoustic studies (e.g., Strange et al., 2001) and to allow Japanese listeners to perceive ranges of perceptual similarity between Japanese and AE vowels. Subset vowels in the present study were chosen based on the pilot data that indicated that percent correct identification was lower than for the other vowels examined at tests and during training (i.e., more difficult). As a result, the set used in the present study was slightly different from the set (/æː, ɑː, ʌ/) used by Akahane-Yamada et al. (1997) and Sperbeck et al. (2005).

Two predictions were made concerning the set size. First, if no differences were found between the fullset and the subset training groups' performance at posttest, then vowel training using only the subset of vowels would be effective and efficient when the set is chosen based on training difficulty rather than on commonly observed confusion. Second, if fullset training improved perception of all vowels, whereas no improvement was found for the untrained vowels of the subset training, then vowel training protocols should include more vowels than only the more difficult ones.

In addition to the use of two training sets, the present study examined long-term retention of training. Long-term retention has been reported for consonant training (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999), but apparently there are no reports for vowel training. The present study examined whether or not the improvement observed at the posttest, if any, was retained after 3 months. Based on the results of previous consonant training studies, it was predicted that both groups would retain improved perception of trained vowels, although no prediction was made for whether the subset training would show any delayed improvement on untrained vowels after 3 months.

Besides the differences of vowel sets used and the examination of long-term retention, there were some notable differences between the present study and that of Akahane-Yamada et al. (1997). First, due to the number of vowels used, real-word stimuli no longer formed minimal contrasts that differed only by vowels. Instead, real words were used only for assessment and were selected to be common words that are familiar to beginning L2 learners. These real-word stimuli examined vowel perception in addition to the nonsense words. Second, the listeners recruited for the present study were Japanese learners of English who had recently arrived in the United States. They represented more diverse proficiency levels than those who participated in the previous studies. This diversity of levels allowed us to extend previous vowel training results to a slightly different population. Thus, the main inclusion criterion was the length of residency in the United States, and vocabulary size and English proficiency were not controlled. However, screening ensured that all listeners had enough room to improve on the training vowels.

Method

Speakers

Five native speakers of AE (2 female [F1, F2], 3 male [M1, M2, M3]; ages 20–27 years) recorded stimulus materials. All grew up in northern Indiana, which coincides with the North Midland dialect region (Labov, Ash, & Boberg, 2006). They had no special training in speaking. Speaker M1's tokens were used as stimuli for task familiarization only.

All speakers were recorded in a sound-treated room in the Speech Psychophysics Laboratory at Indiana University. Speakers were given a reading list and were instructed to read at their normal speaking rate but to enunciate each sound clearly without exaggeration.

Stimulus Materials

There were two categories of stimulus materials: Thirty-six monosyllabic consonant–vowel–consonant (C_1VC_2) real words (RW) and 54 disyllabic nonsense words (NSW; /C_1VC_2ə/). The NSW were used both in training and tests, but the RW were used only in the test to examine generalization in more varied consonantal contexts. Each of the stimulus words included one of the nine AE monophthongs, /iː, ɪ, ɛ, æː, ɑː, ʌ, ɔː, ʊ, uː/. All stimulus words were produced in a carrier sentence, "The first word is ___, isn't it?" with a falling intonation before the tag question. Each speaker recorded two tokens of each stimulus word.

The NSW were /C_1VC_2ə/, where C_1–C_2 combinations were /b-b, b-p, d-d, d-t, g-g, g-k/. The consonants in the RW stimuli were /b, p, d, t, k, h, s, z, ʃ, tʃ, dʒ, m, n, l, w/ for C_1 and /b, p, d, t, g, k, s, z, ʃ, m, n/ for C_2. All of these consonants have comparable categories in Japanese and were assumed to include allophonic variations of the vowel without requiring listeners to learn new consonants.[1]

Digital recording was made directly to a computer at a sampling rate of 48 kHz (16-bit) using a headset microphone (Shure SM10A) connected to a mixer (Shure M267 Professional Microphone Mixer) and a real-time processor (Tucker-Davis Technologies [TDT] RP2). A custom-programmed MATLAB (The MathWorks, 2004) script controlled the recording. Sentences were blocked by the target vowel, and each block containing 5–10 sentences was saved as a separate mother file. Prior to saving a mother file, the MATLAB script adjusted the maximum amplitude of the mother file to 99% of the available amplitude range.[2] All files were later downsampled to 24.414 kHz. The stimulus words were excised from the downsampled mother files.

To ensure the intelligibility of the vowels in the stimulus materials, excised stimulus words were presented to 9 native AE listeners from the North Midland dialect region. On average, vowels were correctly identified 91% in RWs and 83% in NSWs. The greatest confusions were observed between /ɑː/ and /ɔː/. When the responses between these two vowels were collapsed, correct identification was 94% for RWs and 89% for NSWs. The 4 AE speakers were then rank-ordered by their overall NSW intelligibility. Tokens produced by the speakers with the first (F2: 85.4%) and third (M3: 81.4%) ranks were used as stimuli in training and testing (trained speakers). Tokens produced by speakers with the second (M2: 85.1%) and the fourth (F1: 79.6%) ranks were used only in testing (new speakers).

There were two sets of training stimuli. The first set included all nine vowels and was used for the fullset training. The other set, subset, included only the three more difficult vowels, /ɑː, ʌ, ʊ/, which were chosen based on the results of a pilot study using the fullset training protocol. In short, mutual confusion was observed for these vowels at pre-test, and their identification remained less accurate throughout the training.

Japanese Listeners

There were 17 Japanese listeners (20;8 [years; months]–43;10; M = 27 years). All were native speakers of Japanese. The listeners were randomly assigned to one of the three experimental groups. Six listeners were assigned to the fullset training group (J91–J96), another 6 were assigned to the subset training group (J31–J36), and the remaining 5 were assigned to a group that did not receive any training (control: JC1–JC5). Listeners in the two training groups were students in the Intensive English Program at Indiana University or family members of Japanese graduate students. The listeners in the control group were graduate students who had recently arrived in the United States, a student in the diploma program in the music school, or a young spouse of a native AE speaker who had recently moved to the United States from Japan.[3] All were paid for their participation. The first inclusion criterion was that a listener had never lived outside Japan for more than 1 year. The second criterion was that they had enough room (30%) for improvement on vowels in NSW tokens as determined by the pre-test results. One listener was excluded for not meeting the second criterion.

Procedures

All Japanese listeners were individually tested in a sound-treated room. Stimulus presentation was controlled by MATLAB software. Stimulus files were low-pass filtered at 5 kHz using a built-in 10th-order Butterworth filter (PF1) on the TDT system II. All stimuli were presented to the listener's right ear through earphones (TDH-39P) at a fixed listening level determined to be comfortable during pilot testing.

Familiarization for response alternatives. Prior to the pre-test, all listeners were familiarized with the response alternatives and software used in all sessions. First, the listeners' familiarity with the 18 key words (team/meet, hit/kids, set/ten, hat/map, hot/mop, gum/luck, sauce/walk, foot/look, two/who) to be shown on the computer interface was confirmed. Then, using the same interface as in the tests and training, speech samples for the key words recorded by Speaker M1 were presented. The interface displayed International Phonetic Alphabet (IPA) symbols for the nine target vowels and two key words below each symbol. The experimenter reminded the listeners that their task during familiarization was not to identify the vowels in key words but to memorize the relationship between each IPA symbol and key words. Speech samples for key words were presented 4 times—twice in a fixed order first, then two more times in a

[1]It would have been ideal to use real words that cover several common consonantal contexts for all vowels, but it was virtually impossible because of the large number of vowels in the present study in relation to English vocabulary structure. Therefore, the real words were chosen so that they are common words and are familiar to beginning second language learners and were not shown to the listeners prior to the experiment.

[2]This amplitude adjustment technique was used to equalize levels across the reading blocks. Casual observation during the recording indicated that the maximum amplitude occurred on the word "first," not at the target vowel in the carrier sentence. Therefore, it was assumed that the natural amplitude difference among the vowels was preserved.

[3]The listeners who attended the Intensive English Program (IEP) included those who were preparing to enter the graduate school and those who were attending to improve their English skills. The spouses of the graduate students also attended the IEP part time and attended a weekly English conversational group especially offered for them at an on-campus residential hall. All of the listeners who attended IEP were enrolled in classes at intermediate or higher levels.

random order. The listeners were asked to indicate the key word that they heard by clicking on a button located under each IPA symbol. After this response familiarization, a brief written confirmation test was given. Each page on the test form listed the 18 key words in different random orders. The listeners were asked to identify the IPA symbol associated with a key word by circling it on a test form (18 words × 5 randomizations = 90 questions). The passing score was 90% correct. All listeners passed the confirmation test. However, if a listener scored less than 100%, a brief review was provided for the keyword–IPA relationship.

Tests. The same set of three listening tasks was given to all listeners before training (pre-test), after training (posttest), and 3 months after training (3mo-test). The three listening tasks were (a) a similarity rating of vowels in a pair of nonsense syllable tokens (part of a separate study); (b) vowel identification in RW; and (c) vowel identification in NSW. In addition, a speech sample recording of the nine AE monophthongs was made (for separate analysis). These tasks were given over 2 days to maintain a session length within 2 hours. The similarity rating task was given on the first day after response familiarization. The other two listening tasks were given on the second day. None of the tasks for pre- or posttests were given on the same day as training.

In the two vowel identification tasks, listeners were asked to identify an AE vowel in a word (the first vowel, in the case of a /C_1VC_2ə/ NSW) by choosing one of the nine response alternatives on the computer screen. The 18 key words used in the response familiarization were always shown below the nine IPA vowel symbols to guide the listeners. Each listener was tested on four 72-trial blocks (4 words × 9 vowels × 2 tokens) of RW tokens and four 108-trial blocks (6 consonantal contexts × 9 vowels × 2 tokens) of NSW tokens. All listeners were tested on the RW first. Stimulus materials were blocked by speaker, and the presentation order of speakers was randomized among the listeners. All listeners except J34 (who relocated out of state) completed the 3mo-tests approximately 3 months after posttest.

Training. The listeners in the two training groups had training sessions for 9 days between the pre- and posttests. One session lasted an average of 90 min. A session consisted of six blocks of 108 trials (fullset: 6 consonantal contexts × 9 vowels × 2 tokens; subset: 6 consonantal contexts × 3 vowels × 2 tokens × 3 repetitions). Among the six blocks, tokens spoken by a female speaker (F2) were presented in three blocks, and the other three blocks contained the tokens produced by the male speaker (M3). Half of the listeners began training with Speaker F2, and the other listeners began training with M3. Listeners alternated the blocks by the 2 speakers.

The procedures for the training were similar to the vowel identification task given in the tests, except that interactive feedback (adopted and modified from Miller, Dalby, Watson, & Burleson, 2004) was provided for each trial. When a listener identified a target vowel correctly, the text feedback “Correct” appeared on the computer screen, and the next trial began. When the answer was wrong, a subwindow appeared on the screen with two response buttons for the correct and incorrect vowels. The listener then listened to the sound of the correct answer (stimulus) and the incorrect answer (randomly chosen from the two tokens by the same talker for each button press) up to 10 times in any combination, with an option to stop and proceed to the next trial at any time. Listeners also could choose not to use feedback listening by clicking on “stop” before the other two buttons.[4] Listeners completed all sessions, including pre- and posttests, within 1 month. The listeners in the control group did not receive any training, but the time between pre- and posttests for this group was comparable to the other two groups and varied from 2 weeks to 1 month among individual listeners.

Results

Pretraining Confusions

The responses collected for all 17 Japanese listeners at pre-test are presented in Table 1. The first column shows the stimulus vowels. The next two columns list the modal response for a stimulus vowel along with the percentage of frequency out of 816 total opportunities (Column 3). Columns 4 and 5 present the data for the second modal responses. The last two columns are for the other responses that occurred more than 5% of time.

First, notice that none of the nine AE vowels were perfectly identified. The modal responses for most of the vowels were correct responses, but all were also confused with some other vowel(s) at a rather high rate. For example, one of the three vowels, /ɑː/, chosen for the subset training was confused with four vowels with rather equally distributed confusion rates among them. These results indicate that none of the AE vowels were perceived as perfect exemplars of Japanese vowels and that allophonic variation influenced their assimilation into Japanese categories as predicted by Strange et al. (2001). This, in turn, influenced identification of AE vowels. As can be seen in the table, similar results can be found also for vowels /ɛ, æː, ʌ, ɔː, ʊ/.

Next, notice that the three vowels (/ɑː, ʌ, ʊ/, 3V) chosen for the subset training were the least accurately

[4] In the present study, the use of feedback was not recorded. However, feedback was recorded in the follow-up study with Korean learners of English (Nishi & Kewley-Port, 2005).

Table 1. Vowel confusion patterns at pre-test for all Japanese listeners (N = 17).

	Modal		Second modal		Other (>5%)	
Stimulus	Response	%	Response	%	Response	%
iː	iː	69	ɪ	30		
ɪ	ɛ	51	ɪ	41		
ɛ	ɛ	50	æː	22	ʌ	19
					ɑː	8
æː	æː	70	ɛ	14	ʌ	9
					ɑː	6
ɑː	ɑː	34	ʌ	27	ɔː	20
					æː	18
ʌ	ʌ	44	ɑː	27	æː	12
					ɔː	9
ɔː	ɔː	62	ɑː	23	æː	8
					ʌ	6
ʊ	ʊ	45	uː	23	ɑː	11
					ʌ	10
uː	uː	62	ʊ	35		

Note. Total number of observations per stimulus vowel is 816.

identified vowels. In addition, although the vowels /ɑː, ʌ/ were not confused with /ʊ/, the vowel /ʊ/ was confused with both /ɑː/ and /ʌ/. The results presented in the next section also confirm that these vowels were always more difficult than other vowels during training, as well.

Time Course of Training

Vowels. The time course of the 9-day training (Tr-1 to Tr-9) was summarized as daily scores averaged across the tokens produced by the trained speakers. Figure 1 presents the time course of training for vowels. Individual lines represent vowels, and points indicate daily averages for each trained vowel across the 6 training blocks and 6 listeners. The 3V /ɑː, ʌ, ʊ/ chosen for the subset training were represented by solid lines, and the other six vowels (6V), /iː, ɪ, ɛ, æː, ɔː, uː/, are shown as dotted lines. Thus, in the lower panel, nine vowels are shown only for the tests. In the upper panel (fullset), it can be readily noticed that the 3V chosen for the subset training were, in fact, among the least accurately identified in almost all training sessions. This result, together with the initial confusion pattern (see Table 1), confirms that the 3V were indeed more difficult for Japanese listeners to learn than the other vowels. However, the subset group showed remarkably greater improvement on the 3V compared with the fullset group, and all three vowels reached ceiling performance during training. In addition, in the presence of the 6V at both post- and 3mo-test, and although the performance was at a lower level than the last training day (Tr-9), the 3V were still among the most accurately perceived vowels by the subset group.

Individual listeners. The two panels in Figure 2 summarize the time course of training for individual listeners in the fullset group (upper panel) and the subset group (lower panel). Each line represents a listener, and points indicate daily averages across six training blocks and trained vowels for the nine training sessions. Even though tests presented all nine vowels, in order to make comparison between training and test performance easier, averages were calculated only for the trained vowels (i.e., nine for fullset, three for subset).

All listeners in both training groups showed improvement over the 9-day training. As for the rate of learning, all listeners showed higher performance on the first training day (Tr-1) than at pre-test (11.7% for fullset; 40.4% for subset) and then gradually improved between Tr-1 and Tr-9 at a similar rate (15.4% for fullset; 15.6% for subset). Comparison of individual listeners' learning curves within the training group indicated that the differences among the listeners in the subset group reduced as training proceeded (SDs: Tr-1 = 7.82, Tr-9 = 2.59), whereas those for the fullset group did not (SDs: Tr-1 = 9.31, Tr-9 = 9.59). This difference between the two groups is probably caused by the fewer number of training categories for the subset training.

Pre-, Post-, and 3mo-Test Performance

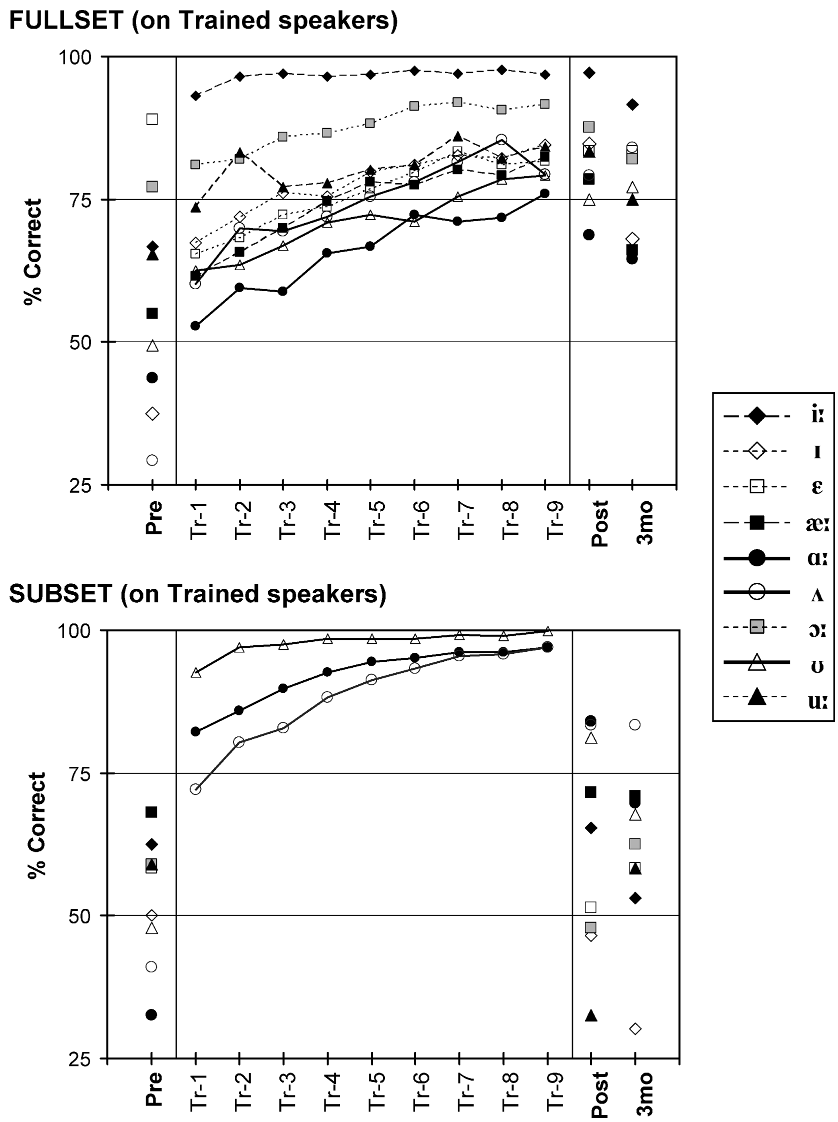
This section reports the listeners' performance at the three tests (pre-, post-, and 3mo-tests). All the results presented in the figures are percent correct scores, but scores were converted into rationalized arcsine units (Studebaker, 1985) for statistical analyses in order to allow score distributions at the extremes (under 15% and over 85%) to be more linear. In the statistical analyses presented below, listener group was a between-subjects variable, whereas test, speaker, and training set were treated as within-subjects variables.

Figure 3 presents the summary for the NSW stimuli produced by the trained speakers (left side) and new speakers (right side) at three tests. Scores were calculated as averages across nine vowels and the 2 speakers. In the figure, data are organized so that the three test scores for a listener group are presented together. Error bars indicate the within-group SD.

Trained tokens. Although the control group's pre-test performance (61%) appeared slightly higher than that of the other two groups (fullset = 57%, subset = 53%), the result of an analysis of variance (ANOVA) on individual listeners' pre-test scores indicated no difference among the three groups, $F(2, 14) = 1.64$, $p = .23$.

The next analysis was performed only for the control group in order to gauge the effect of exposure to the

Figure 1. Percent correct identification scores for fullset (upper panel) and subset (lower panel) training groups at pre-, post-, and 3mo-tests as well as during training (Tr-1 to Tr-9), summarized separately for each of the trained vowels in the nonsense word tokens.

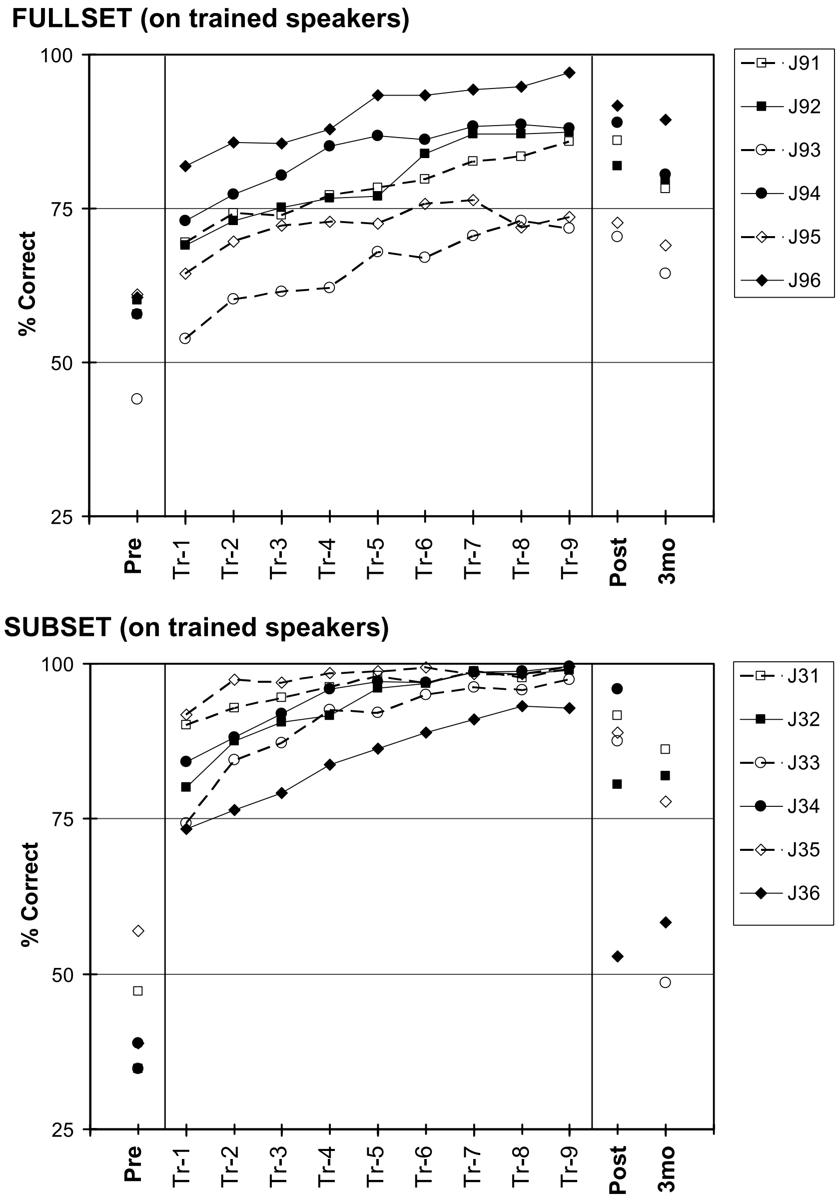


same stimulus materials at the three tests as well as the effects of daily exposure to spoken English. A repeated measures ANOVA on individual listeners' scores at three tests revealed no significant difference among the tests, $F(2, 8) = 0.08$, $p = .92$, suggesting that neither repeated exposure to the stimulus materials nor the daily exposure to the spoken English over 3 months helped listeners to improve AE vowel perception.

Turning to the two training groups, a mixed-design ANOVA indicated significant main effects of listener group, $F(1, 9) = 11.98$, $p < .01$, and test, $F(2, 18) = 52.25$, $p < .001$. The Group × Test interaction was also significant, $F(2, 18) = 8.98$, $p < .005$, because the overall improvement was greater for the fullset group than for the subset group. Follow-up tests using Tukey's HSD revealed that the subset group's posttest performance was significantly higher than pre-test ($p < .02$), but their 3mo-test was not ($p = .08$). For the fullset group, both posttest and 3mo-test performance were considerably higher than pre-test ($p < .001$ for both). It was also found that the fullset group scored higher than the subset group at both post- ($p < .01$) and 3mo-tests ($p < .05$).

Generalization of training effects: New speakers. As was described earlier, NSW tokens produced by the new speakers were presented only in the tests so that generalization to new voices could be evaluated. First,

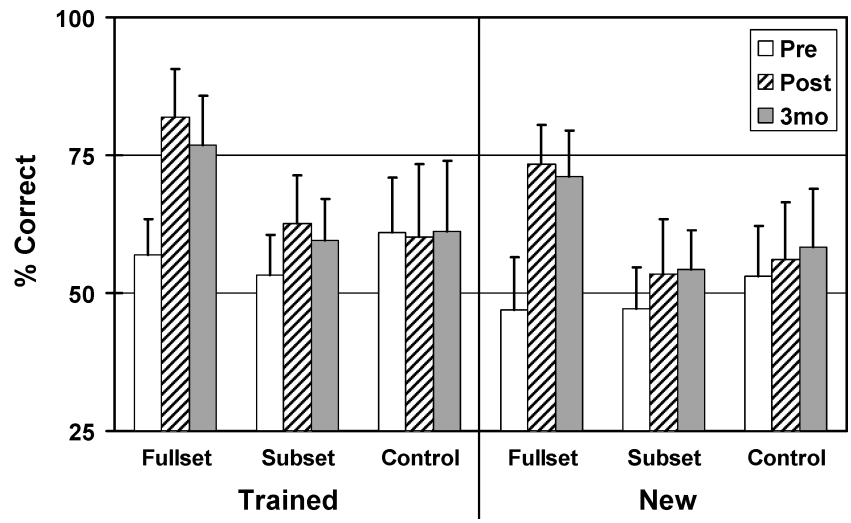
Figure 2. Percent correct identification scores at pre-, post-, and 3mo-tests as well as during training (Tr-1 to Tr-9), summarized separately for individual listeners in fullset (upper panel) and subset (lower panel) training groups for the nonsense word tokens spoken by the two trained speakers.



consider the rank order of intelligibility among the four AE speakers observed at pre-test. The result obtained from Japanese listeners was not the same as that from native AE listeners (reported previously in the *Stimulus Materials* section). The trained speakers were ranked first (F2) and third (M3) intelligible for native AE listeners, but they were ranked first and second for Japanese listeners. Therefore, the 2 new speakers were unexpectedly less intelligible (48.8%) for Japanese listeners than the trained speakers (56.8%) at pre-test, $F(1, 14) = 32.20$, $p < .001$, but no difference was found among the groups, $F(2, 14) = 1.10$, $p = .36$ (see also Figure 3).

To examine speaker generalization after training, a mixed-design ANOVA (Training Group × Test) was performed only on the scores for the new speakers. The results were similar to those for the trained speakers. The significant effects were observed for training group, $F(1, 9) = 9.89$, $p < .05$; test, $F(2, 18) = 25.75$, $p < .001$; and interaction, $F(2, 18) = 5.85$, $p < .05$. Tukey's HSD on the interaction revealed that only the fullset group showed considerably improved performance at both posttest

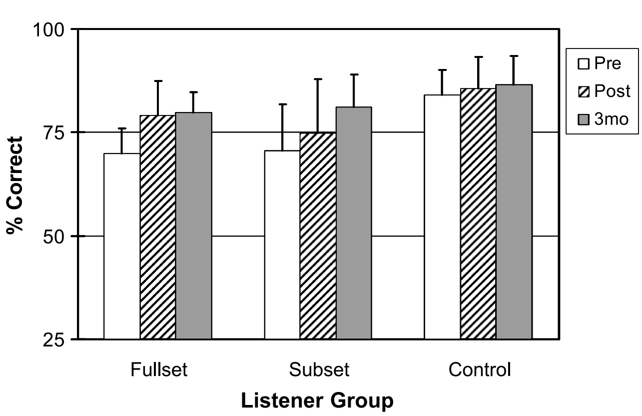
Figure 3. Percent correct identification scores obtained for the nonsense word tokens spoken by the two trained speakers (left side) and new speakers (right side) at pre-, post-, and 3mo-tests for the three listener groups. Error bars indicate standard deviations.



(73%) and 3mo-test (71%), compared with pre-test (47%) ($ps < .001$). The results of the subset group revealed no significant differences between pre- and posttests ($p = .24$), but their performance after 3 months was considerably higher than that at pre-test ($p < .005$).

Generalization of training effects: Real words. The next analysis was performed for the RW tokens (as opposed to the NSW tokens used in training). The percent correct scores for the three groups at pre-, post-, and 3mo-tests were averaged across the 4 speakers and are presented in Figure 4. In the figure, data are organized in terms of the listener groups to show the changes in performance between the tests.

Figure 4. Percent correct identification scores for the real-word tokens spoken by all 4 speakers obtained at pre-, post-, and 3mo-tests for the three listener groups. Error bars indicate standard deviations.



A Listener Group × Test ANOVA indicated significant effects of test, $F(2, 26) = 24.55$, $p < .001$, and interaction, $F(4, 26) = 2.75$, $p < .05$, but not group, $F(2, 13) = 2.16$, $p = .15$. Analyses on the interaction using Tukey's HSD revealed that a significant posttraining change was observed only for the fullset group ($p < .001$ for both pre–post and pre–3mo comparisons).

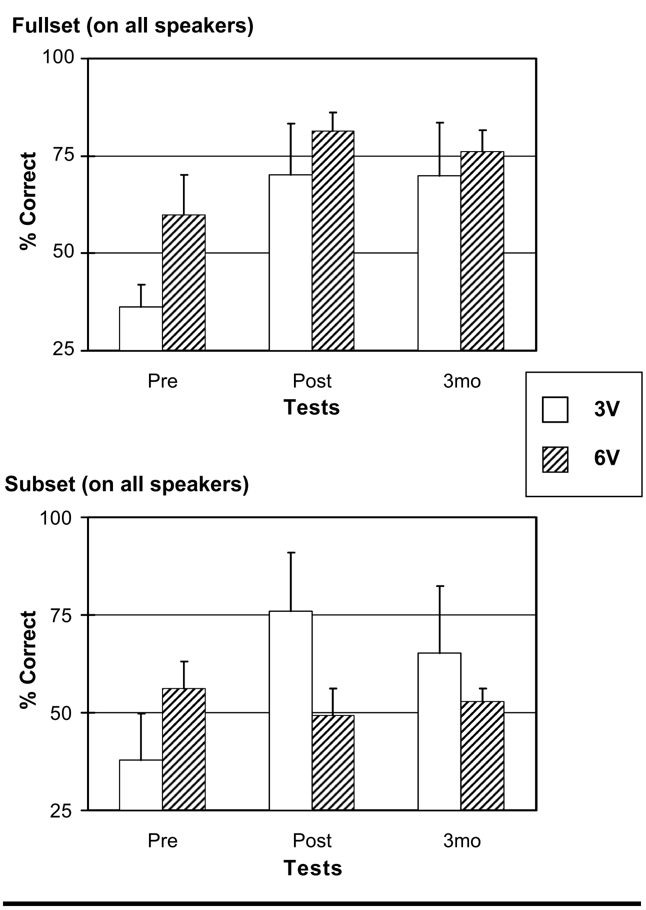
For the RW results, because overall scores for all three groups at pre-test were already higher than NSW, one might suspect that lexical information largely influenced listeners' performance on RW. Indeed, although it was not significant, the control group scored somewhat higher than the other groups on the RW tokens (see Figure 4), and the subset group showed improvement between post- and 3mo-tests without any training, possibly because they became more efficient in lexical access after living in the United States for 3 months. Recall that vowels in the RW were presented in more varied consonantal contexts than the NSW used for training. Considering that the significant change was observed only for the fullset group, these results indicate that a training method using HVNP stimulus materials that present a large number of vowels in sufficient consonantal variations can help L2 listeners accurately identify nonnative vowel categories even in novel consonantal contexts.

Difference Between Trained and Untrained Vowels

To confirm that the 3V (/ɑː, ʌ, ʊ/) were actually more difficult than the other six vowels (/iː, ɪ, ɛ, æː, ɔː, uː/, 6V) for all Japanese listener groups at the pre-test, average scores across 4 speakers were calculated separately for 3V and 6V. Individual listeners' average pre-test scores for the two vowel sets for each group were then submitted to an ANOVA. The only significant effect was found between vowel sets, $F(1, 14) = 51.16$, $p < .001$, indicating that 3V (41%) were more difficult than 6V (59%) for all Japanese listeners at pre-test. Therefore, the differences found between the listener groups at post- and 3mo-tests should be attributed to training. For the following analyses concerning the effect of vowel set, only the two training groups were compared because it was shown above (Figure 3) that the control group did not improve through exposure.

Figure 5 presents the average test scores across 4 speakers for 3V and 6V for the fullset group (upper panel) and for the subset group (lower panel). The results of a three-way mixed-design ANOVA indicated that the main effects of training group, $F(1, 9) = 5.18$, $p < .05$, and test, $F(2, 18) = 47.58$, $p < .001$, as well as all interactions, were significant: Training Group × Vowel Set, $F(1, 9) = 21.26$, $p < .005$; Training Group × Test, $F(2, 18) = 3.97$, $p < .05$; Vowel Set × Test, $F(2, 18) = 24.15$, $p < .001$; Training Group × Vowel Set × Test, $F(2, 18) = 8.79$, $p < .005$. Follow-up analyses on the three-way interaction using Tukey's HSD indicated that the fullset group's post- and 3mo-test scores on 3V (70% for both tests) as well as 6V (81% and 76%, respectively) were significantly higher than the pre-test scores (3V = 36%; 6V = 60%; $ps < .002$). However, for the subset group, significant differences were observed only for the 3V (pre-test = 38%; posttest = 76%; 3mo-test = 65%, $p < .001$ for both pre–post and pre–3mo comparisons) but not for the 6V (pre-test = 56%; posttest = 50%; 3mo-test = 53%, $p = 1.00$ for both pre–post and pre–3mo comparisons). Although the greater decline of performance on the 3V after 3 months was observed for the subset group (10.65%; fullset = 0.23%), the post–3mo difference was not significant for either of them.

Figure 5. Percent correct identification scores for 3V (/ɑː, ʌ, ʊ/) and 6V (/iː, ɪ, ɛ, æː, ɔː, uː/) obtained at pre-, post-, and 3mo-tests across four speakers for fullset (upper panel) and subset (lower panel) training groups. Error bars indicate standard deviations.



Discussion

The present study followed for 4 months three groups of Japanese L2 learners recently immersed in an English-speaking environment. One of the groups was perceptually trained on nine AE monophthongs covering the entire vowel space (/iː, ɪ, ɛ, æː, ɑː, ʌ, ɔː, ʊ, uː/; fullset training protocol), the second group was trained only on the three more difficult vowels (/ɑː, ʌ, ʊ/; subset training protocol), and the third group was provided no training (control). The two training groups were tested before (pre-test), after (posttest), and 3 months after (3mo-test) training. The control group was given the same tests at similar intervals as the training groups. No statistical difference was found among the three groups at pre-test.

Results showed that both training protocols successfully improved Japanese listeners' perception of AE vowels presented during training. Performance of the control group did not change over 3 months, indicating that the naturalistic exposure to spoken English, as well as learning by means of the three tests alone, can be excluded as possible causes for the improved performance of the trained groups. The fullset protocol was more effective than the subset protocol. Differences between the two training protocols are discussed in the next three sections in terms of the effects of training set, long-term retention, and the effects of feedback.

Effects of Training Set

The present study showed for the first time that training using a large set of sound contrasts—in this case, nine vowels—did not make learning extremely difficult. Rather, the fullset protocol resulted in rather uniform improvement in perception for all nine vowel sounds. Results also demonstrated that vowel training should include more than just "difficult" vowels. Specifically, the present study used three difficult vowels (/ɑː, ʌ, ʊ/) and found no improvement on the six untrained vowels. Similar results were reported by Akahane-Yamada et al. (1997) and Sperbeck et al. (2005), although they selected their vowels, /æː, ɑː, ʌ/, on the basis of a different principle. Altogether, the results imply that efficient learning of nonnative vowels requires exposure to a full set of vowel categories, both easy and difficult, in the target language.

The comparison of time course of training showed that the subset training improved on the three difficult vowels more rapidly than the fullset training (see Figure 1). However, considering that the generalization of improvement to RW tokens was not significant for the subset group, subset training may not be as effective as the fullset training. These results suggest that training protocols that use a small number of categories are less effective than training protocols that use a larger set of categories.

Long-Term Retention of Training Effects and the Nature of Changes in Performance

In our review of the literature, the present study was the first demonstration of long-term retention for vowel training. As expected from the results of the previous consonant training studies, both training groups maintained their improved performance on the trained vowels for 3 months after the completion of training, whereas the control group did not change. It was also of interest whether the subset group would generalize training to the untrained vowels after 3 months. Unfortunately, no cross-vowel generalization occurred (see Figure 5).

Recall that the procedures used for training were the same between the two protocols except for the vowel sets. Because the listener groups were not different at pre-test, it is suggested that the two training protocols elicited different learning mechanisms. Goldstone (1998) summarized four underlying mechanisms of perceptual change that include attentional weighting, imprinting, differentiation, and unitization. These underlying mechanisms, by themselves, are stated to be less long-lasting than perceptual learning. Applying this theoretical approach to the present results, the changes that occurred for the two training groups may be different in nature. Given that the learning observed for the subset group was limited only to the trained vowels and was apparently less long-lasting than that of the fullset group (see Figure 5), their improvement may not represent perceptual learning but rather a strategy shift in one of the underlying mechanisms to attend only to the task-relevant cues that resulted in higher performance. These listeners had not merely memorized the acoustic characteristics of each token because improved performance was also observed for the tokens spoken by untrained speakers. Does this mean that intentional strategy shift may limit perceptual learning? To achieve an effective training protocol, future research should consider the mechanisms of perceptual learning in the design and evaluation of specific protocols.

Effects of Feedback

Recall that the feedback system used in the present study was interactive and provided listeners with opportunities to choose and listen to multiple repetitions of the sound of correct and wrong answers when they misidentified the target vowel. This feedback system has no known precedents in L2 training. Although detailed records of the use of feedback were not kept in this experiment, listeners reported using the listening options especially during the early training sessions. Note the following relevant observations. Even though the same interactive feedback system was used in both fullset and subset training protocols, because of the size of the training set, the possible feedback pairs in the fullset protocol were 36, whereas the subset group's feedback was limited to only three possible pairs of vowels (/ɑː-ʌ/, /ʌ-ʊ/, or /ʊ-ɑː/). Although the 36 pairs may not have been presented as feedback with equal frequency, the fullset protocol probably owes its success not only to the specific vowels included in the nine-vowel set but also to the quality of feedback listeners received. More specifically, in the presence of eight other vowels, the vowel /ʊ/ was confused most with /uː/ by all listeners (see Table 1 and Figure 1), but in the subset protocol, listeners did not receive any contrasting feedback other than /ɑː, ʌ/ for /ʊ/. Thus, the feedback in the fullset protocol was not restricted to such a small number of contrasts and, therefore, could provide multiway comparison feedback for the target vowel. This is consistent with the prediction by Schmidt's (1975) schema theory of discrete motor learning even though the present domain is speech perception learning, not motor learning. Schema theory hypothesizes that people learn by exploring the parameters (e.g., direction, duration, strength, etc.) involved in the realization of the target movement and approximate outcomes to the goal by adjusting the parameters through trial and error. Thus, the theory predicts that people learn the relationship between the parameters and outcomes more quickly if they attempt a wide variety of parameter combinations and if they experience errors. Similarly, the present results provided opportunities for listeners to experience a wide range of auditory parameter combinations (e.g., spectral and temporal characteristics as well as allophonic variations due to consonantal contexts and speakers) and to learn from the actual errors. However, as discussed previously, the errors experienced by the fullset group were more varied than those experienced by the subset group, resulting ultimately in the fullset group having a richer learning opportunity than the subset group.

The present results, therefore, have some potentially strong implications for training methods that focus on specific minimal pairs, in which contrasting sounds and response alternatives used during training are limited to a predefined set of two. Barlow and Gierut (2002) reviewed different selection criteria for the minimal pairs in relation to clinical research on intervention for phonologically delayed children. They recommended the use of minimal pairs that represent two phonemes that are both absent in a child's phonemic inventory and are maximally different from each other by major class properties (consonants/vowels, glides/consonants, obstruents/sonorants) as well as nonmajor class distinctions (place, manner, and voice). However, because all reviewed studies were on production of consonants and for children with phonological delay in their native language, it is not easy to determine whether the same method would be effective for L2 vowel perception training. Even if minimal pairs were chosen based on Barlow and Gierut's well-evaluated criteria, it seems unclear how to efficiently expand the pairs to include other sounds because errors made in a minimal pair may not reflect those observed in real life. The underlying assumption for the subset training, as well as for minimal-pair methods, is to increase learners' awareness of differences between contrasts that are presumed to be the primary cause of errors. However, the results of the present study caution against the use of a smaller set of categories for vowel perception training because it might allow learners to ignore alternative cues important for the accurate categorization of an entire vowel set. As a result, the learners may show remarkably fast and substantial improvement on the basis of strategy learning that may fail to generalize to the genuine perceptual learning that is needed to classify the complete set of vowels in the target language.

One might wonder, then, whether it is possible to avoid the "no improvement" on the untrained vowels observed for the subset protocol by using multiple minimal pairs that cover the entire vowel space. However, not only should such a protocol be evaluated first, but given that feedback is still limited within a specific pair, selecting effective pairs may be challenging. Alternatively, a training protocol that does not require minimal pairs to be determined in advance but provides contrasting feedback based on actual observed errors should be more efficient and less likely to introduce unexpected complications in the learning process.

Another possible training protocol is to start with a smaller set of vowels that are more difficult (e.g., 3V set here) and expand it to a larger set of training vowels. The results of the subset training illustrated the possible strengths of the training using a problem-focused set—namely, early and rapid improvement on the more difficult vowels. To take advantage of these strengths and correct for the weakness—specifically, "no improvement" on the untrained categories—the two training protocols used in the present study can be combined. Research is in progress on the effectiveness of two hybrid training protocols that combine the fullset and subset protocols (Nishi & Kewley-Port, 2005).

Limitations and Future Directions

The results of the present study demonstrate that the training protocol using up to nine categories is not only effective but also produces higher overall improvement compared with the training using only three vowels. However, there may still be room for improving this nine-vowel training protocol. For example, a set of nine monophthongal vowels were chosen for the present study, but it is not known whether nine is the optimal number of vowels. AE has diphthongs and rhotic vowels in addition to the monophthongs used in the present study. Previous cross-language perceptual assimilation studies predicted that Japanese listeners would confuse /eɪ-ɛ/ and /oʊ-ɔː/ pairs (Strange et al., 2001). The results of the present study suggest that perception of diphthongs and rhotic vowels may not improve as a result of training only on the monophthongs and that they may also need to be included in vowel training.

The results of the present study were limited to vowels. Therefore, it is not known whether a similar training protocol incorporating many consonant categories would also be effective. However, because the number of consonants in AE, including clusters and varying positions in a word, is quite large, a protocol presenting all possible consonants/clusters (items) in a single training session is unrealistic. Research is needed to examine possible ways to reduce the number of items into tractable group sizes appropriate for training.

It is important to note that the present study trained only native Japanese listeners. Therefore, it is not known whether the fullset protocol is also effective for learners with different L1 backgrounds. As shown by Pruitt et al. (2006), even though the same protocol is used, L2 speech perception training may not produce similar results for listeners with different native languages. For this reason, the efficacy of using a large set of vowels in a training protocol needs to be investigated with different languages. As a first step to address this issue, a training study with Korean learners of English is in progress (Nishi & Kewley-Port, 2005).

Conclusion

Results of this study imply that even when care is taken to choose contrasts, perceptual training only on the more difficult vowels is not an effective vowel training protocol. Although it was small, generalization of learning to novel speakers and different consonantal contexts was observed for the subset training protocol using highly variable naturally produced stimuli, but it was limited only to *trained* vowels. Furthermore, the long-term retention was observed only for the protocol using nine vowels, compared with the protocol using three vowels. Thus, it can be concluded that an effective vowel training protocol should present a large set of vowels.

The present study was the first attempt to simultaneously train L2 learners on many vowel categories. It compared the efficacy of such a training protocol to the training protocol using a smaller, problem-focused set. Previous laboratory training on consonant contrasts has used minimal pairs to train nonnative listeners. However, the results of the present study caution against the use of sets with a small number of categories, at least for vowel training, because learners may learn to ignore cues that are related to other categories and may therefore fail to learn the complete set in the target language. To avoid such uneven learning and the risk of introducing unexpected complications in the learning process, we recommend including a large number of categories in a perceptual training protocol for L2 learners.

Acknowledgments

A portion of this article was presented at the 149th meeting of the Acoustical Society of America in Vancouver, Canada, in May 2005. This work was supported by National Institutes of Health (NIH) Postdoctoral Fellowship F32 DC-006313 (awarded to the first author) and NIH Research Grant R01 DC-02229 (awarded to the second author). Manuscript preparation was supported, in part, by NIH Research Grant R01 DC-04300 (awarded to Patricia G. Stelmachowicz). William B. Mills wrote the training program used here. James J. Jenkins and Winifred Strange made significant contributions to the development of the initial research design. The consultation with Charles S. Watson and James D. Miller regarding the feedback system was invaluable. Finally, we wish to thank the study participants.

References

Akahane-Yamada, R., Strange, W., & Kubo, R. (1997). Training Japanese listeners to identify American English vowels. *Proceedings of Fall Meeting of the Acoustical Society of Japan,* 379–380.

Barlow, J. A., & Gierut, J. A. (2002). Minimal pair approaches to phonological remediation. *Seminars in Speech and Language, 23,* 57–67.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Baltimore: York Press.

Bradlow, A. R. (in press). Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/-/l/ contrast. In J. Hansen & M. Zampini (Eds.), *State-of-the-art issues in second language phonology.*

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics, 61,* 977–985.

Ferguson, S. H., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America, 112,* 259–271.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Baltimore: York Press.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49,* 585–612.

Hay, J. F., Sato, M., Coren, A. E., Moran, S. L., & Diehl, R. L. (2006). Enhanced contrast for vowels in utterance focus: A cross-language study. *The Journal of the Acoustical Society of America, 119,* 3022–3033.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America, 97,* 3099–3111.

Hirata, Y., & Tsukada, K. (2004). The effects of speaking rates and vowel length on formant movements in Japanese. In A. Agwuele, W. Warren, & S. H. Park (Eds.), *Proceedings of the 2003 Texas Linguistics Society Conference* (pp. 73–85). Somerville, MA: Cascadilla Proceedings Project.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America, 118,* 3267–3278.

Krause, J. C., & Braida, L. D. (2002). Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *The Journal of the Acoustical Society of America, 112,* 2165–2172.

Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America, 115,* 362–378.

Labov, W., Ash, S., & Boberg, C. (2006). *Atlas of North American English: Phonology and phonetics.* Berlin: Mouton de Gruyter.

Ladefoged, P. (1993). *A course in phonetics.* Fort Worth, TX: Harcourt Brace.

Ladefoged, P. (2001). *Vowels and consonants*. Oxford, England: Blackwell.

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics, 26,* 227–247.

Lively, S. E., Logan, J. S., & 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. *The Journal of the Acoustical Society of America, 94,* 1242–1255.

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

McClaskey, C. L., Pisoni, D. B., & Carrell, T. D. (1983). Transfer of training of a new linguistic contrast in voicing. *Perception & Psychophysics, 34,* 323–330.

Miller, J. D., Dalby, J. M., Watson, C. S., & Burleson, D. F. (2004). Training experienced hearing-aid users to identify syllable-initial consonants in quiet and noise. *The Journal of the Acoustical Society of America, 115,* 2387.

Nishi, K., & Kewley-Port, D. (2005). Second language vowel training using vowel subsets: Order of training and choice of contrasts. *The Journal of the Acoustical Society of America, 118,* 1900.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America, 24,* 175–184.

Pruitt, J. S., Jenkins, J. J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *The Journal of the Acoustical Society of America, 119,* 1684–1696.

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review, 82,* 225–260.

Shibatani, M. (1990). *The languages of Japan*. New York: Cambridge University Press.

Sperbeck, M. N., Strange, W., & Ito, K. (2005). Training L2 learners to perceive difficult American vowel contrasts. *The Journal of the Acoustical Society of America, 117,* 2400.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., & Nishi, K. (2001). Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *The Journal of the Acoustical Society of America, 109,* 1691–1704.

Strange, W., Bohn, O.-S., Nishi, K., & Trent, S. A. (2005). Contextual variation in the acoustic and perceptual similarity of North German and American English vowels. *The Journal of the Acoustical Society of America, 118,* 1751–1762.

Strange, W., Weber, A., Levy, E., Shafiro, V., Hisagi, M., & Nishi, K. (2007). Acoustic variability within and across German, French, and American English vowels: Phonetic context effect. *The Journal of the Acoustical Society of America, 122,* 1111–1129.

Studebaker, G. (1985). A "rationalized" arcsine transform. *Journal of Speech and Hearing Research, 28,* 455–462.

The MathWorks. (2004). MATLAB (Version 6.5) [Computer software]. Natick, MA: The MathWorks.

Received July 13, 2006

Accepted June 2, 2007

DOI: 10.1044/1092-4388(2007/103)

Contact author: Kanae Nishi, who is currently at Boys Town National Research Hospital, 555 North 30th Street, Omaha, NE 68131. E-mail: nishik@boystown.org.