Effects of stimulus order in categorial AX discrimination training on improvements of the ability to perceive the English/b/-/v/contrast among Japanese learners of English.

Teruaki TSUSHIMA

1. Introduction

In normal listening conditions, native listeners have little difficulty categorizing speech sounds, despite a great deal of variability that exists in speech signals (cf. Perkell, Klatt, & Stevens, 1986). This gives an impression that perception of speech sounds is largely categorical with stable phonetic boundaries between phonemes. However, early studies provided ample evidence that the phonetic boundaries are susceptible to a number of contextual influences, which include range effects (Brady & Darwin, 1978), frequency effects (Rosen, 1979) and stimulus order effects (Cowan & Morse, 1986). For example, it has been shown that, in an experimental procedure where listeners determine whether a pair of successively presented sounds belongs to the same phonetic category or different categories, they respond "same" significantly more often in one presentation order than in the reversed order. Evidence accumulates in the literature that such directional asymmetry exists in discrimination of within-category native vowels, between-category native vowels and between-category nonnative vowels (see Poka & Bohn, 2003; in press, for a review). The present study attempted to provide further evidence that would add to our understanding of the directional asymmetry by focusing on discrimination of the English /b/-/v/ contrast by Japanese-speaking adults during AX categorial discrimination training.

Early evidence of the stimulus order effects came from within-category and between-category discrimination of native vowels (Cowan & Morse, 1986; Repp & Crowder, 1990). The results of their studies generally found that discriminability of a contrast (e.g., /i/-/ɪ/) in an AX discrimination procedure is significantly lower when the first vowel is more peripheral (e.g., /i/) in a vowel space. More recent research has provided corroborating evidence on between-category vowel contrasts that are not phonemic in the native language. Polka & Bohn (in press), for example, tested Danish adults on discrimination of the Southern British English /ɒ/-/ʌ/. In Danish, there is no vowel contrast in the area of the vowel space that corresponds to the English /ɒ/-/ʌ/. The results showed that the participants had significantly greater difficulty discriminating the contrast when a more peripheral vowel /ɒ/ was followed by a more central vowel /ʌ/ than in the reversed order. Polka & Bohn further tested English-speaking adults on discrimination of the German contrast /u/-/y/ and /ʊ/-/ʏ/, using a go/no go task. The results found that the participants showed greater difficulty discriminating both contrasts when more peripheral vowels (/u/ and /ʊ/) were the background than otherwise. In addition, they tested Cantonese adults on the same German contrasts (Polka, Sundara, Zhaoo, Pang, & Ciocca, in preparation). In Cantonese, the /u/-/y/ contrast is phonemic while the /ʊ/-/ʏ/ contrast is not. It was found that significant order effects were found only for the nonnative /ʊ/-/ʏ/. Finally, research on infant perception has provided convincing evidence that infants during the first year of life show the same pattern of asymmetry, including the English /ɒ/-/ʌ/ and the German /u/-/y/ contrast and the German /ʊ/-/ʏ/ contrast (Polka & Bohn, 1996, 2003, in press; Polka & Werker, 1994). Based on these findings, Polka & Bohn proposed a framework called "The Natural Referent Vowel Framework" that vowels with extreme articulatory-acoustic properties (peripheral vowels) serve as perceptual anchors in discrimination. It is proposed that the organization of perceptual space primarily defined by these vowels underlie the observed asymmetry. It is hypothesized that the directional asymmetry in discrimination of vowels present during infancy may be maintained, enhanced or reduced through linguistic experience depending on a particular phonological system of the native language.

The author and his colleagues hypothesized that significant stimulus order effects would be obtained in discrimination of certain nonnative consonants as well. First, although perception of native consonants is mostly categorical (Liberman, Harris, Hoffman, & Griffith, 1957; Van Hessen & Schouten, 1992), that of nonnative consonants is hardly so and easily influenced by contextual factors (e.g., Miyawaki, Strange, Verbrugge, Liberman, & Jenkins, 1975; Yamada & Tohkura, 1992). Second, when both of the nonnative phonemes of a contrast are perceptually assimilated to one native phoneme, it is often the case that one phoneme is more directly assimilated than the other due to differences in the perceptual distance to the native phoneme (e.g., Best, 1995). It was hypothesized that, analogous to the referent vowels, a phoneme which is perceptually closer to the native phoneme might act as a perceptual anchor in a discrimination task, which would give rise to significant order effects in nonnative consonant discrimination.

As the first investigation, Japanese-speaking adults were tested on discrimination of the

English contrast /b/-/v/ (Tsushima, Shiraki, Yoshida, & Sasaki, 2003; Tsushima, Yoshida, Shiraki, & Sasaki, 2003). Japanese has a phoneme /b/, which is phonetically similar to the English /b/, although there are some differences in phonetic details (e.g., VOT). On the other hand, Japanese does not have a phoneme /v/. Both phonemes are normally assimilated to the Japanese /b/, but the English /b/ is more directly assimilated to the Japanese /b/ than the English /v/. The participants (N = 72) were divided into six groups across three ISI (inter-stimulus-interval) conditions (100 ms, 300 ms and 1500 ms) and two talker conditions (single- and multiple-talker). Under the multiple-talker condition, stimuli consisted of two tokens of /ba/ and of /va/ produced by two native speakers of English, while one token from a single speaker was used in the single-talker condition. The results supported the hypothesis, showing that discriminability was significantly lower when /b/ was presented as the first stimulus than in the reversed order in both talker conditions at the ISI of 100 ms and 1500 ms. The finding that significant order effects were obtained under the long-ISI supported a hypothesis that the effects were based on phonetic processing rather than general auditory processing.

In order to replicate the finding regarding the direction of the order effects and further investigate the conditions under which the effects occurred, the next study (Tsushima, Shiraki, Yoshida, & Sasaki, 2005; Tsushima, Yoshida, Shiraki, & Sasaki, 2005) investigated the order effects in different experimental procedures. In Experiment 1, a procedure similar to go/no go procedure was employed, where the first set of stimuli was repeatedly presented seven or fifteen times before the second set of stimuli was presented with or without a change in the phoneme category. Significant order effects with the expected direction were obtained in both repetition conditions, showing that the effects were not limited to the AX discrimination procedure. In Experiment 2, an irrelevant vowel /ʌ/ was inserted between the pair of stimuli, in order to prevent listeners from utilizing a rehearsal strategy. Significant order effects with the same direction were found in the multiple-talker condition but not in the single-talker condition. In Experiment 3, the participants were required to identify a pair of stimuli in an AX procedure (i.e., AX identification procedure) across the three ISI conditions (100 ms, 300 ms and 2000 ms) in the multiple-talker condition. Significant order effects with the same direction were obtained in all the ISI conditions.

The overall results not only replicated the previous findings in terms of the direction of the effects, but also provided further support for the hypothesis that some form of categorized memory was the basis of the order effects. Another important finding in Experiment 3 (i.e., AX identification procedure) was that a relatively large proportion of incorrect responses in the order of /b/ to /v/ were accounted for by the /b/-/b/ responses. This suggested that the order effects might be accounted for by an assimilation process where the less native-like sound presented as the second stimulus was perceptually assimilated to the more native-like sound presented as the first stimulus, which served as a perceptual reference point at the time of stimulus comparison.

The overall findings supported the hypothesis that the order effects were due to the differences in perceptual distance of the nonnative phonemes to the corresponding native phoneme. However, it was also be the case that /b/ served as the perceptual anchor as it is phonetically more basic than /v/. In fact, the voiced bilabial plosive /b/ is one of the most common phonemes in the world languages while the voiced labio-dental fricative /v/ is one of the least common (cf. Ladefoged & Maddieson, 1996). In addition, /b/ is one of the phonemes acquired relatively early in L1 development, while /v/ is one of the late-acquired phonemes (cf. Jacobson, 1968). It might be the case that the order effects were largely due to the language-universal characteristics of these phonemes rather than the language-specific phonetic processing of nonnative speech sounds.

To sort out these hypotheses, the next series of studies focused on the American English /r/-/l/ discrimination by Japanese-speaking adults (Tsushima, 2007a, 2007b). Neither of the lateral liquid /l/ and the central approximant /r/ was among the common phonemes in the world languages (cf. Ladefoged & Maddieson, 1996). Japanese does not have a phonemic distinction of these phonemes. Although it has a phoneme /r/, the phoneme is normally realized as a dental flap, [ɾ] (Vance, 1987). Both /l/ and /r/ are perceptually assimilated to the Japanese /r/, but there is evidence that the English /l/ is perceptually closer to the Japanese /r/ than the English /r/ (Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004; Takagi, 1993). It was hypothesized that discriminability of the contrast was significantly lower when /l/ was followed by /r/ than otherwise. In Tsushima (2007b), Japanese-speaking adults were tested on discrimination of the contrast in an AX discrimination procedure with an ISI of 300 ms and 1500 ms. Speech stimuli were natural tokens (N = 4 for each category) produced by two native speakers of American English. The results found a significant effect of stimulus order with the expected direction at the ISI of 1500 ms, but not at that of 300 ms. In Tsushima (2007a), the same participants were tested on discrimination of the contrast under both ISI conditions, using ten-step synthetic stimuli of /ra/ to /la/. The paired stimuli were 1) a typical (end-point) /l/ with an ambiguous /r/ (3 steps away from the end-point), 2) an ambiguous /r/ and a typical /l/, and 3) a typical /l/ with a typical /r/. The results found no significant order effects even in the pair of the end-point stimuli in either ISI condition, although discriminability (averaged over the order conditions) was significantly lower in the pair of the typical /l/ with the ambiguous /r/ than in that of the typical /r/ with the ambiguous /l/. The overall results suggested that the effects were due to language-specific speech processing rather than language-universal or psychophysical auditory processing.

The next step in the investigation was to examine whether significant order effects were obtained during discrimination training. Learning to perceive a nonnative speech contrast involves learning to attend to some critical acoustic cues not functional in the native language, but at the same time, learning to ignore those cues that are irrelevant in distinguishing the contrast. In discrimination training on nonnative contrasts, listeners presumably attempt to sort out the critical acoustic cues by changing their attentional patterns based on feedback. It was of interest to examine whether significant order effects would occur when listeners were engaged in such perceptual learning.

In Tsusima et al. (2003) described above, the Japanese-speaking adults received six sessions of AX discrimination training on the English /b/-/v/ contrast in the same experimental conditions as in the pretest (i.e., across three ISI conditions (100 m, 300 ms and 2000 m) and two talker conditions (a single or multiple-talker)). The results showed that, at the ISI of 300 ms and 2000 ms, significant order effects were not obtained in either talker condition at any training sessions, although the tendency with the expected direction of the effects appeared at a few beginning sessions. At the ISI of 100 ms, significant order effects continued to be obtained throughout the sessions. In the single-talker condition, the direction of effects remained unchanged, while in the multiple-talker condition, the direction became reversed in the middle of the sessions. The results showed that, except in the very short ISI condition, stimulus order effects did not occur during discrimination training

The next study examined whether the order of stimuli had significant effects on learning to perceive the American English /r/-/l/ and /gl/-/gr/ contrast during a fixed categorial AX discrimination procedure. In the procedure, /l/ was always presented as the first stimulus (i.e., /l/ to /l/ or /l/ to /r/) in one group (/l/-first group; N = 16), while in the other group (/r/-first group; N = 15), /r/ was always the first stimulus (i.e., /r/ to /r/ or /r/ to /l/). It was hypothesized that discrimination learning was negatively influenced in the /l/-first condition as the effects of the perceptual anchor would be pronounced when the phonemic category of the first stimulus was fixed. They received twelve sessions of discrimination training. The ISI increased from 500 ms to 1500 ms, while the single-talker and multiple-talker conditions were intermixed during training. In order to encourage categorization learning, lexical items (14 items each with /r/-/l/ and /gr/-/gl/; e.g., right-light, grass-glass) were used as training stimuli. They were tested on discrimination and identification of the stimuli in the pretest and posttest. They were also tested on identification of a 10-step, synthetic /ra/-/la/ continuum to examine how the identification functions changecl before and after the training.

The results found that the /l/-first group significantly improved the ability to discriminate the /r/-/l/ contrast, to identify /l/ in the /r/-/l/ contrast, and showed a significantly better-defined category boundary of the /r/-/l/ contrast in the posttest than in the pretest. None of the improvements, however, was significant in the /r/-first group. The overall results clearly refuted the hypothesis that discrimination learning was negatively affected when the more native-like stimulus was presented as the first stimulus. One possible account of the failure to obtain the significant order with the expected direction, however, was that discrimination training on the difficult-to-learn contrasts (i.e., /r/-/l/ and /gr/-/gl/) using lexical items might have been too demanding for the participants. In fact, discrimination performance in the "different" pair (e.g., /l/ to /r/ or /r/ to /l/) in the post was only slightly above the chance level.

The present study was designed to extend the previous studies on possible effects of the stimulus order in nonnative discrimination training. It focused on the English /b/-/v/ contrast, which is known to be easier to perceive than the /r/-/l/ contrast by Japanese listeners (Brown, 1994). In addition, nonsense syllables, rather than lexical items, were used as training stimuli. Furthermore, the design of training was modified such that participants went through four stages of training which were incrementally more difficult to pass. The change in the design was expected to encourage participants to attend to the discrimination tasks through the end of the training. As the preliminary study (Tsushima, 2010) which used stimuli from two speakers turned out to be too easy for the participants, stimuli from four speakers were used to increase stimulus variability. The purpose of the present study was to examine whether and how the stimulus order in the fixed categorial AX discrimination training influenced the ability to learn to discriminate the English /b/-/v/ contrast during training. It specifically examined whether perceptual learning would be negatively affected if the first stimulus in the pair was always /b/, which may serve as a perceptual anchor for discrimination of the contrast.

2. Method

2–1. Participants

Participants were 27 monolingual Japanese students at Tokyo Keizai University, in Japan. They were at a basic or pre-intermediate level in terms of their English ability. None of them reported any history of hearing or neurological disorders. Based on the results of the pretest, they were assigned to the /b/-first group ($N=14$) or the /v/-first group ($N=13$) such that the identification ability did not significantly differ between the two groups. The participants were not informed of which group they belonged to. Nor were they informed that the first stimulus of the pair in the AX discrimination training was always fixed to one category.

2–2. Speech Stimuli

Speech stimuli were natural tokens produced by six native speakers of English (three females and three males). The speakers recorded multiple tokens of /ba/ and /va/, using a frame sentence, “I will say X”. The speech was digitized at a sampling rate of 44,000 Hz with 16 bits of resolution. For the training stimuli, four tokens for each category were selected from each of the four speakers (two females, two males), totaling 32 stimuli. For the stimuli used only in the pretest/posttest (i.e., generalization stimuli), two tokens for each category were selected from each of the two speakers (one female, one male), totaling 8 stimuli. Selection of the stimuli was made such that non-critical acoustic cues to the contrast did not systematically differ between the two categories (e.g., VOT, F0 and duration). The stimuli were further edited by truncating a prevoiced portion and an offset portion of a vowel to make adjustments in terms of the duration from the onset of voicing to the release of closure or frication and the overall duration of the stimuli, using sound editing software (Acoustic Core 8: Arcadia). Table 1 shows descriptive statistics of acoustic properties among training and generalization stimuli. The properties included 1) overall duration, 2) duration before a release, 3) the first formant at the beginning of a vowel portion (F1), 4) the second formant at the beginning of the vowel portion (F2), 5) a fundamental frequency (F0) at the beginning of a vowel and 6) F0 at the offset of a vowel. It is clearly shown that none of the acoustic properties was not a reliable cue to distinguishing the contrast. The available evidence indicates that the English labio-dental fricatives are characterized by the following acoustic properties (Maniwa, Jongman, & Wade, 2009). First, a spectrum of frication is relatively flat without dominant peaks in any particular frequency regions. Second, noise duration is relatively short. Third, the F2 onset frequency is relatively low, as compared to alveolar and palato-alveolar fricatives. Due to the first two properties, the voiced frication portion of /v/ tends to be short and perceptually non-salient, which probably makes it difficult for nonnative listeners to perceive it. The third property (i.e., relatively low F2 onset frequency) is shared with /b/. Although there is some evidence that the F2 onset frequency tends to be higher for /v/ than /b/ (Olive, Greenwood, & Coleman, 1993), the tendency was not observed among the training stimuli, but only among the generalization stimuli. Thus, the most reliable cue to distinguishing the contrast includes presence (or absence) of frication noise of /v/, which spreads without predominant peaks in the mid-fre-

Table 1. Acoustic properties of training and generalization stimuli (DUt=total duration; DUr=duration from the beginning of voicing to the point where the bilabial closure for /b/ or the labio-dental constriction for /v/ is released; F0s=F0 at the start of a vowel; F0e=F0 at the end of a vowel). Durational values are shown in ms, and those of formants are in Hz.

Training		DUt	DUr	F1	F2	F0s	F0e	N
/b/	M	408	42	603	1231	147	110	16
	SD	92.2	18.9	42.4	92.4	19.9	24.9	
	Max	637	78	668	1434	178	147	
	Min	295	9	524	1118	122	85	
/v/	M	414	48	610	1255	149	112	16
	SD	51.4	12.4	70.0	33.6	46.8	43.1	
	Max	543	71	789	1361	174	146	
	Min	299	38	503	1137	127	85	
Generalization	M	341	43	543	1268	170	136	4
/b/	SD	78.8	10.0	77.5	73.1	17.8	25.1	
	Max	415	58	632	1297	213	179	
	Min	299	32	478	1233	129	97	
/v/	M	345	43	623	1356	169	133	4
	SD	31.0	5.3	64.6		48.8	33.2	
	Max	391	51	681	1487	217	163	
	Min	328	40	543	1256	125	99	

quency range (Jongman, Wayland, & Wong, 2000) and presence (or absence) of a burst produced by a bilabial plosive release of closure in production of /b/ (cf. Lisker & Abramson, 1964). The primary task for the participants was to learn to selectively attend to these critical cues in the speech stimuli through the discrimination training. An adult native speaker of English was tested on identification of the stimuli using the same format as the pretest below. As expected, the identification accuracy was over 95%.

2–3. Experimental Procedure

2–3–1. Pretest and Posttest

A two-alternative, forced-choice identification test was employed to test the participants' ability to identify the speech stimuli. In this procedure, participants were asked to indicate whether a presented stimulus belonged to either /b/ or /v/. The test consisted of one practice block of 8 trials, followed by 4 blocks of 16 trials. Except for the practice block, stimuli used for the training were presented in half of the trials, while generalization stimuli were presented in the other half in one block. All the training stimuli (N = 32) were presented once in four blocks, while all the generalization stimuli (N = 8) were presented once in every block. In the practice block, stimuli consisted of randomly selected training stimuli (N = 4 in each category) and all the generalization stimuli (N = 8).

The test program began with a message on a monitor that participants should press a button "b" on the keyboard when they heard /b/, or a button "v" when they heard /v/. When they pressed the space bar, a visual stimulus that indicated a block number (i.e., "Block 1") appeared for 2000 ms. A speech stimulus was presented 1000 ms after the offset of the visual stimulus. When participants responded within 2000 ms after the offset of each stimulus, a message which confirmed the button press (i.e., "OK!! Good!!") appeared on the monitor for 500 ms, regardless of whether the response was correct or not. If the participants did not respond within the time limit or pressed a wrong button, a warning message appeared on the monitor for 500 ms. The next stimulus presentation began 1000 ms after the offset of the message.

2–3–2. Training

A fixed categorial AX discrimination training procedure was used for discrimination training. In this procedure, participants were asked to indicate whether a pair of the stimuli belonged to the same category or different categories by pressing "s" for the "same" response and "d" for the "different" response on the keyboard. The training program consisted of a practice block, followed by four stages of discrimination training (i.e., First, Bronze, Silver and Gold Stage). Each stage consisted of six blocks, each of which contained 12 trials. Participants could pass a certain stage if the percent correct was greater than a predetermined criterion in two consecutive blocks. The criterion for First, Bronze, Silver and Gold Stage was 60%, 70%, 80% and 90%, respectively.

In one block, stimuli from two (out of four) speakers (e.g., Speaker A & Speaker B) were presented. In each stage, six possible combinations in terms of the pair of speakers were randomly ordered, except that the pair in the last block was not the same as that of the first block in the following stage. In half of the trials in one block, two different tokens from the same category were paired (i.e., "same" trials), while in the other half, those from different categories were paired (i.e., "different" trials). In each group, the order of stimuli was "AA" (i.e., "/b/ to /b/" or "/v/ to /v/") in 6 trials, and was "AX" (i.e., "/b/ to /v/" or "/v/ to /b/") in the other 6 trials in one block. The order of presentation with respect to the stimulus order was randomized within a block. The pair of stimuli always consisted of tokens from different speakers (as opposed to the same speaker).

The ISI between a pair of stimuli was set at 1000 ms in order to encourage the participants to utilize categorized information for a stimulus comparison. After the second stimulus was presented, the participants had 2000 ms to respond. When the response was correct, a message (i.e., "Correct") appeared on the monitor as a positive feedback for a period of 500 ms. A message (i.e., "Not Correct") appeared as a negative feedback when the response was incorrect or when no response was made within the time limit. The next trial began 2000 ms after the offset of the visual feedback.

The training session started with oral instruction with regard to how the participants should complete the training by passing all the four stages, as well as what they were asked to do in the discrimination task. Then, they were asked to launch the program, type in necessary information (e.g., the participant's name) and initiate the training program. At the beginning of the program, a message appeared on a monitor saying that participants should press a button, /d/, when a pair of the sounds belonged to different categories (e.g., /ba/-/va/ or /va/-/ba/), and a button, /s/, when they belonged to the same category (e.g., /ba/-/ba/ or /va/-/va/). The message also explained about the positive and negative feedback.

When they reached a criterion in one block at a certain stage, a message appeared on the monitor saying that they would go to the next stage if they surpassed the criterion in the following block. When they passed a certain stage, a message congratulated them for having done so, and instructed them to go to the next stage. The training was terminated 15 minutes after the initiation of the program.

The test and training programs were created by the author using a stimulus-presentation software (Superlab: Cedrus Corporation), which controlled the presentation of the audio and visual stimuli, and recorded the participants' responses and reaction times. On the day of the test or training, each participant sat at a computer in a quiet room, wearing a high-quality headset. The audio stimuli were presented at a comfortable level of approximately 65 dB. The training was conducted four or seven days after the pretest. The posttest was conducted on the same day as the training following a few minutes' break in order to assess the immediate effects of the training on identification of the training and generalization stimuli.

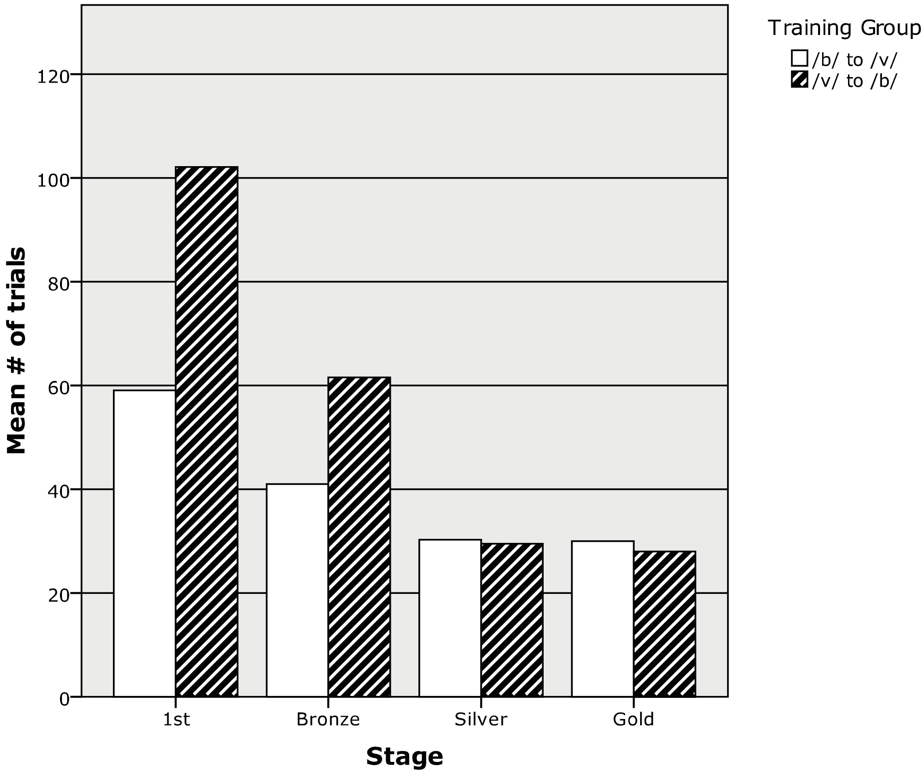

Figure 1. Mean number of trials attempted in each stage as a function of group.

3. Results

3–1. Training

3–1–1. The Number of Participants Who Passed the Stages

Table 2 shows the number of participants who reached each of the four stages in the two groups. All the participants who reached the Gold stage successfully passed the stage and finished the program within the time limit. It is shown that a half of all the participants (N = 7) were not able to pass the First stage in the /v/-first group, while only one participant was not able to so do in the /b/-first group. A little less than half of all the participants (N = 6) were able to pass the Gold stage in the /b/-first group, while only three participants were able to do so in the /v/-first group. A Chi-square test using the number of participants who remained in the First Stage and who passed the Gold stage showed significant differences in the distribution of the participants between the two groups (p = .024). It was indicated that the /v/-first

Table 2. The number of participants as a function of stage at which they ended the training program. All the participants who reached the Gold Stage passed the stage and finished the program within the time limit.

	Stage				
Group	First	Bronze	Silver	Gold	Total
/b/ to /v/	1	4	2	6	13
/v/ to /b/	7	3	1	3	14
Sum	8	7	3	9	27

group had significantly greater difficulty passing the stages than the /b/-first group in the training program.

3–1–2. The Number of Trials Attempted in Each Stage

The next analyses compared the number of trials attempted at each stage in the two groups. As is shown in Figure 1, the mean number of trials was much greater in the /v/-first group in the First stage and Silver stage. A statistical analysis using a non-parametric test[1] (i.e., Mann-Whitney test) showed, however, that the difference was only marginally significant at the First stage, $Z = 1.93$, $p = .054$, but not significant at the Bronze stage, $Z = 1.42$, p = 1.55. This was due to fairly large individual differences in the number of trials needed to pass the First stage in each group. In the /b/-first group, it took three participants more than six blocks (i.e., equal or more than 84 trials) to pass the First stage, while it took six participants just two blocks to do so. In the /v/-first group, seven participants could not pass the First stage, while four could pass it within four blocks (i.e., equal or less than 48 trials).

3–1–3. Percents Correct as a Function of Stage

Figure 2 shows mean percents correct at each stage in the two groups. The percents correct increased as the criterion to pass each stage increased from 60% in the First stage to 90% in the Gold stage, as expected. Compatible with the data presented above, the percent correct in the /b/-first group was higher than that of the /v/-first group in the First stage. A statistical analysis using a t-test, however, revealed no statistical difference between the two groups, $t(25) = 1.61$, $p = .121$, again due to large individual differences in each group.

In sum, the present data on the overall performance in the training strongly indicate that the participants in the /v/-first group experienced greater difficulty learning to discriminate the contrast, while there was fairly large individual variability in each group.

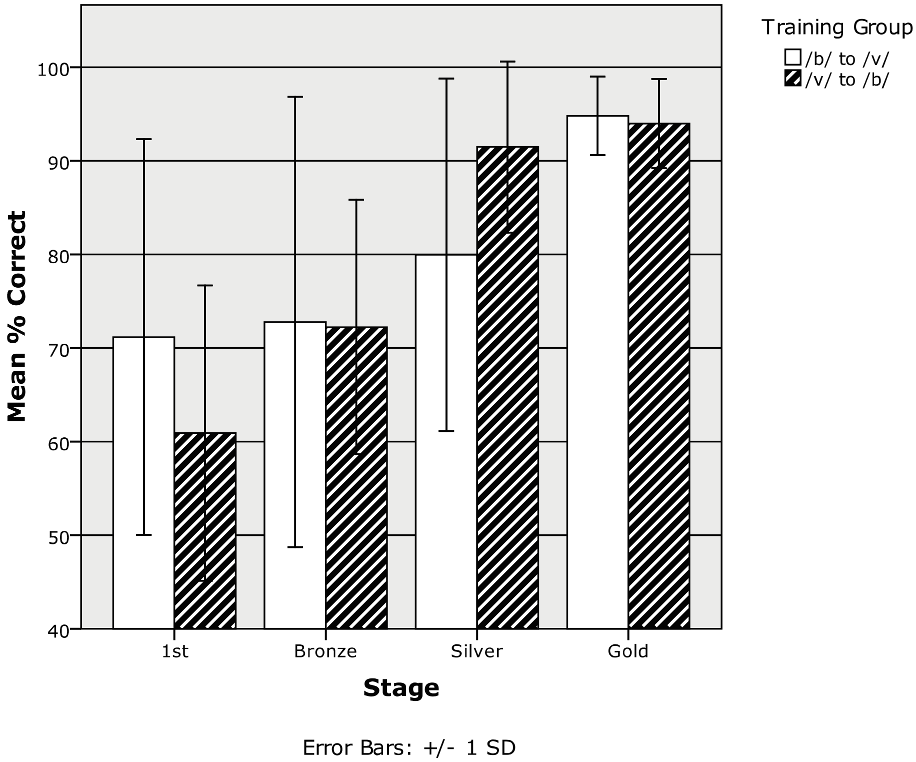

Figure 2. Mean percent correct as a function of group and stage.

3–1–4. Percents Correct as a Function of Stage and Stimulus Order

Figure 3 shows the mean percents correct as a function of stage and stimulus order in the /b/-first group. First, it is shown that the percent correct was substantially higher in the /b/ to /b/ than in the /b/ to /v/ condition in the Bronze and Silver stage. The percent correct in the /b/ to /b/ condition steadily improved to over 90% in the Gold stage, while that of the /b/ to /v/ condition gradually increased from the First to the Silver stage. A Wilcoxon singed-ranks test showed that the difference in the percent correct between the /b/ to /b/ and /b/ to /v/ condition in the Bronze and Silver stage did not reach significance ($p = .138$ and $p = .175$, respectively). Figure 4 shows the mean percents correct as a function of stage and stimulus order in the /v/ -first group. As is shown in the figure, the difference between the two stimulus order conditions was not substantial across the stages.

Finally, the percent correct in the /b/ to /v/ and the /v/ to /b/ condition was compared across the two groups. A comparison of Figure 3 and Figure 4 showed no indication that the percent correct in the /b/ to /v/ condition was significantly lower than that of the /v/ to /b/ condition in the First stage. On the contrary, that of the /v/ to /b/ condition was substantially lower

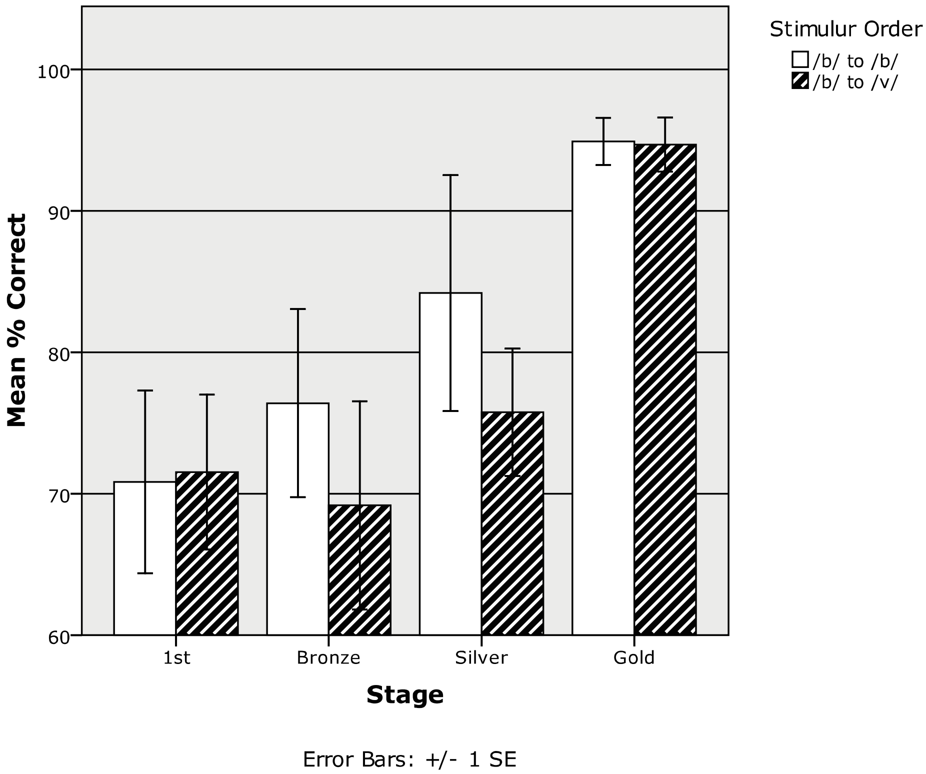

Figure 3. Mean percent correct as a function of group, stage and stimulus order in the /b/-first group.

than that of the /b/ to /v/ condition in the First stage. A Mann-Whitney test found that the difference was marginally significant, $Z = 1.80$, $p = .076$.

In sum, the analyses of the percents correct as a function of stage and stimulus order did not find any significant differences in the level of performance between the /b/ to /b/ and /b/ to /v/condition in the /b/-first group, or between the /v/ to /b/ and the /v/ to /v/condition in the /v/-first group. There were no significant differences between the /b/ to /v/ and the /v/ to /b/ condition, and between the /b/ to /b/ and /v/ to /v/condition, either.

3–1–5. Reaction Times as a Function of Stimulus Order

The next analyses examined how reaction times changed across the stages in the two groups. Following a conventional analysis procedure, mean reaction times were calculated on the trials where the responses were correct. Figure 5 and Figure 6 show the mean reaction times as a function of stage and stimulus order in the /b/-first and /v/-first group, respectively. First, reaction times showed a substantial decrease from the First stage to the Gold stage, indi-

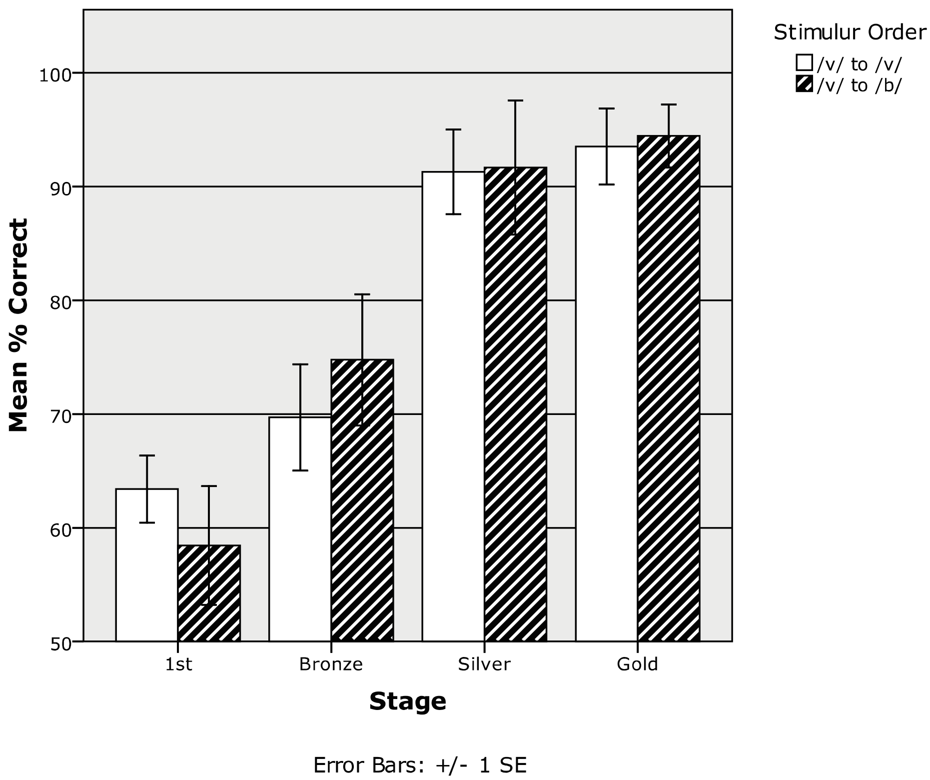

Figure 4. Mean percent correct as a function of group, stage and stimulus order in the /v/ -first group.

cating that the participants discriminated the contrast more efficiently through the training. Second, in the /b/-first group, the reaction time in the /b/ to /b/ condition was substantially longer than that of the /b/ to /v/ condition in the First stage. A t-test revealed that the difference was significant, $t(12)=2.67$, $p=.021$, indicating that it took the participants longer to reach a decision when the second stimulus in a pair was /b/ than /v/. Although the difference did not reach significance (Wilcoxon singed-ranks tests, $p>.05$), the same tendency was observed in the Silver and Gold stage. Third, none of the difference was significant between the /v/ to /v/ and the /v/ to /b/ condition in any stage in the /v/-first group (Wilcoxon singed-ranks tests, $p>.05$). In sum, the analyses provided an unexpected finding that, in the /b/-first group, reaction time was longer when /b/ was followed by /b/ than /v/ in the First stage.

3–2. Pretest and Posttest

Figure 7 shows the mean percents correct in the identification pretest and posttest in the

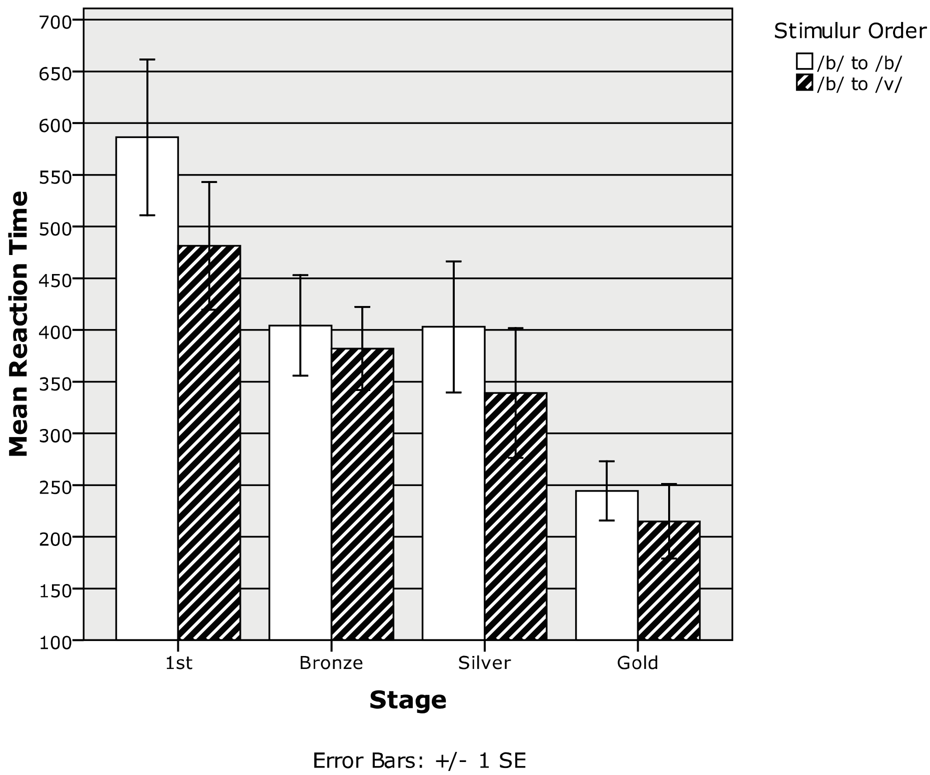

Figure 5. Mean reaction time as a function of stage and stimulus order in the /b/-first group.

two groups. Table 3 shows mean percents correct as a function of stimulus type (generalization and training stimuli) and phoneme (/b/ and /v/). A repeated-measures ANOVA was conducted using test (pretest, posttest) and phoneme (/b/, /v/) as within-subject factors and group (/b/-first, /v/-first group) as a between-subject factor. The main effect of time was highly significant, $F(1, 52) = 53.0$, $p = .000$, $\eta^2 = .680$, indicating that both groups significantly improved the ability to identify /b/ and /v/ through the training. Although three-way interaction of test, phoneme and group was not significant, $F(1, 52) = 1.75$, $p = .198$, $\eta^2 = .065$, simple-effects analyses were conducted as the data showed some complex relations among some factors (see Table 3). First, the effects of group were not significant in either phoneme in the pretest, indicating that the identification ability in both groups did not significantly differ before the training. Second, although the effect of phoneme was not significant, there was a tendency that the percent correct for /b/ was higher than that of /v/ in the pretest, as expected. Third, the effect of group was significant for the phoneme /b/ in the posttest, $F(1, 25) = 6.73$, $p = .015$, $\eta^2 = .213$, indicating that the /b/-first group performed significantly better in identification of /b/ than the /v/-first group in the test. Forth, the effect of phoneme was significant in the post-

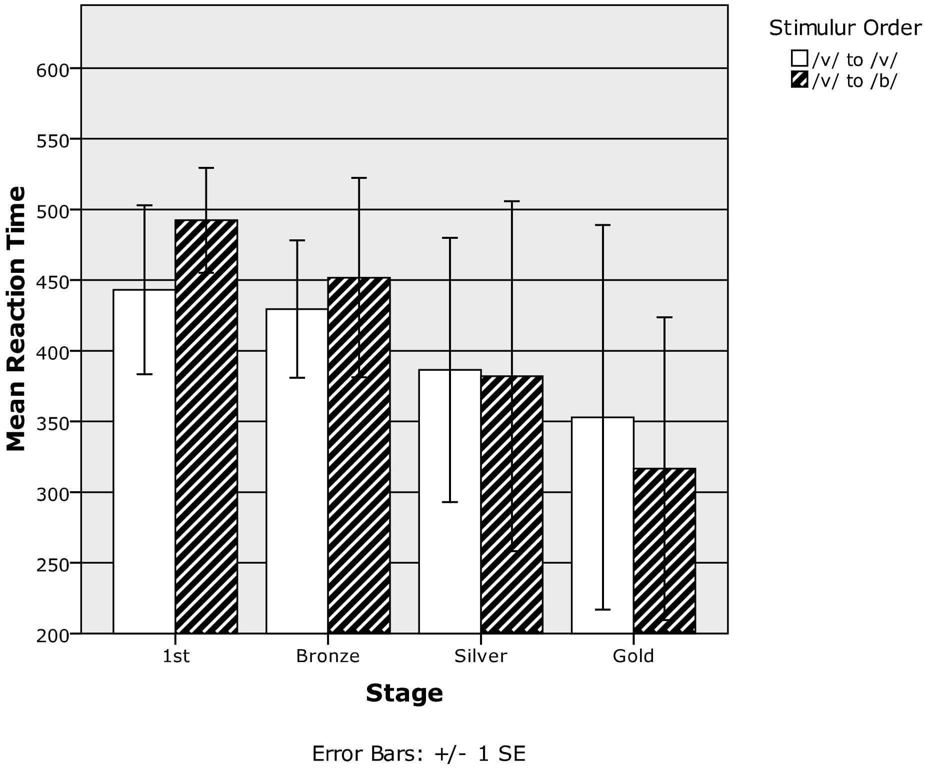

Figure 6. Mean reaction time as a function of stage and stimulus order in the /v/-first group.

test in the /b/-first group, $F(1, 25) = 4.56$, $p = .043$, $\eta^2 = .154$, indicating that the group performed significantly better in identification of /b/ than /v/. As is shown in Table 3, their performance in identification of /v/ was especially poor for the generalization stimuli (63.9%) as compared to the training stimuli (79.8%).

3–3. Relation between discrimination and identification

The following analyses examined whether and how the participants' performance in the discrimination training was related to that of the identification test in the posttest. Figure 8 shows the mean percents correct in the identification test at the posttest and the stage to which participants reached during training in the two groups. As was mentioned above, all the participants who reached the Gold stage finished the program within the time limit (see the number of the participants who reached each stage in Table 2). As expected, the percents correct were substantially lower for the participants who couldn't finish the program than those who could. The mean percent correct for the participants in the /v/-first group (N = 7) was especially low. Examination of the individual data showed the maximum percent correct in this group was 71.2

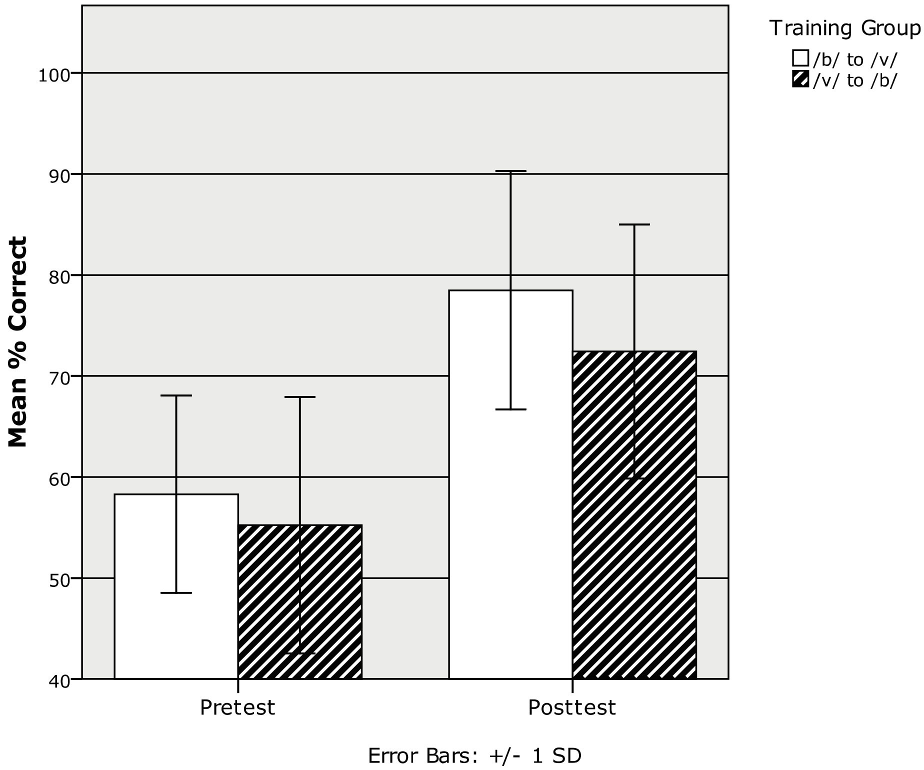

Figure 7. Mean percent correct as a function of group and test.

%. Correlation analyses on the mean percent correct averaged over all the trials in the training and that of the identification test in the posttest showed that the correlation was positive and significant in the /b/-first group, $r = .71$, $p = .007$, as well as in the /v/-first group, $r = .75$, $p = .002$. Overall, the results indicated that the ability to learn to discriminate the /b/-/v/ contrast during the training was significantly related to the ability to identify the contrast after the training.

3–4. Summary of the results

The present study provided the following major findings. The analyses on the participants' performance during the training showed that the participants in the /v/-first group had significantly greater difficulty with the training than the /b/-first group, as shown by the comparison in the number of participants who could not pass the First stage and could pass the Gold stage. Although the same tendency was observed in the number of trials attempted or the percents correct across the stages, the group difference did not reach statistical significance due to large individual variability. The analyses on the results of the identification pretest and posttest

Table 3. Mean % correct (standard deviation in parentheses) as a function of group, test, stimulus type and phoneme (S Type=stimulus type; Gen=generalization stimuli; Tr=training stimuli; Sub (ph)=Subtotal of phonemes averaged over stimulus types; Sub (st)=subtotal of stimulus types averaged over phonemes.

		Group			
		/b/ to /v/ (N=13)		/v/ to /b/ (N=14)	
		Test		Test	
S Type	Phoneme	Pretest	Posttest	Pretest	Posttest
Gen	/b/	61.1 (16.4)	85.6 (18.1)	56.3 (19.4)	75.4 (15.6)
	/v/	53.8 (14.3)	63.9 (22.1)	47.8 (22.0)	71.4 (22.0)
Tr	/b/	62.5 (11.1)	84.6 (14.1)	59.4 (22.2)	65.6 (16.2)
	/v/	55.8 (15.8)	79.8 (10.3)	57.6 (22.8)	77.2 (23.6)
Sub (ph)	/b/	61.8 (11.4)	85.1 (15.2)	57.8 (19.6)	70.5 (14.0)
	/v/	54.8 (13.1)	71.9 (14.8)	52.8 (20.5)	74.3 (21.1)
Sub (st)	Gen	57.5 (12.4)	74.8 (15.5)	52.0 (14.0)	73.4 (12.7)
	Tr	59.1 (10.3)	82.2 (9.7)	58.5 (14.2)	71.4 (15.2)
Total		58.2 (9.8)	78.5 (11.8)	55.2 (12.7)	72.4 (12.6)

found that both groups significantly improved their ability to identify /b/ and /v/ from the pretest to the posttest. Averaged over the phonemes (/b/ and /v/) and stimulus types (training and generalization), significant differences were not found between the two groups either in terms of the degree of improvement or the level of achievement. However, the /b/-first group performed significantly better in identification of /b/ than the /v/-first group in the posttest. In addition, the /b/-first group showed better performance in identification of /b/ than that of /v/. Finally, the participants' discrimination performance during the training was significantly correlated with that of the identification test in the posttest in both groups.

4. Discussion

The purpose of the present study was to examine whether and how the stimulus order in the fixed categorial AX discrimination training on the English /b/-/v/ contrast influenced perceptual learning among Japanese-speaking adults. In particular, it was hypothesized that the perceptual learning was adversely influenced under the condition where the first stimulus in the discrimination pair was fixed to /b/. The results clearly refuted the hypothesis, showing few adverse effects of the /b/-first condition on discrimination learning. The present results were compatible with Tsushima et al. (2003), which showed that significant order effects in discrimination of the /b/-/v/ contrast mostly disappeared at the beginning of the AX discrimination training in the long-ISI condition (i.e., 1500 ms). The present results were also consonant with Tsushima (2007b), which examined the order effects in discrimination of the American English /r/-/l/ contrast by Japanese-speaking adults using the AX discrimination training procedure with the first stimulus fixed either to /r/ or /l/. As mentioned in Introduction, the English /l/ is perceptually more similar than the English /r/ to the Japanese /r/ to which both phonemes are perceptually assimilated (Aoyama, et al., 2004). The results found few negative effects of the /l/-first condition over the /r/-first condition. There is also some corroborating evidence from vowel discrimination research (Polka, et al., in preparation). In the study which tested Cantonese-speaking adults on discrimination of the German contrast /u/-/y/ and /ʊ/-/ʏ/ (see Introduction), one group of participants received phonetic training (i.e., transcription of vowels) while the other group did not. It was found that, while the untrained group showed asymmetries in discrimination of the nonnative /ʊ/-/ʏ/ contrast, the trained group showed no asymmetries in either contrast. These results indicated that significant order effects with the same direction as found in an initial exposure to the speech stimuli would not be obtained during discrimination training.

As reviewed above, the pattern of results that emerged from the previous studies was the following. First, significant order effects in discrimination of nonnative consonant contrasts were obtained when the ISI was relatively long (i.e., 1500 ms or longer). Second, the effects were more likely to be obtained when stimulus variability was relatively high (i.e., the multiple-talker condition with natural stimuli) than otherwise (i.e., the single-talker condition or synthetic stimuli). Third, the effects were obtained when listeners were required to label the stimuli for a stimulus comparison (e.g., vowel insertion condition, AX identification). Under these conditions, listeners are presumably required to categorize the first stimulus by giving it some kind of a label and store it in memory for a comparison with the next stimulus. It is presumed that, in an initial exposure to the speech stimuli without feedback, perception of nonnative sounds is heavily influenced by the native-language attentional patterns, which may perceptually assimilate the nonnative phoneme to the perceptually similar native phoneme. It might be the case that perception and storage of a relatively native-like phoneme (i.e., /b/ or /l/) activates the native-language attentional patterns, which perceptually assimilates the following less native-like phoneme (i.e., /v/ or /r/) at the time of discrimination. During the training, on the other hand, listeners are encouraged to suppress the native-language attentional patterns, by weighing the available acoustic properties to sort out the critical acoustic cues to the trained contrast.

The present results did not only show the hypothesized adverse effects of the /b/-first condition, but also showed that the /b/-first group performed significantly better in the discrimination training. It was found that only one participant (out of 12) in the /b/-first group was not able to pass the First stage, while a half of the /v/-first group (7 out of 14) were not able to so do, and that twice as many participants in the /b/-first group passed the Gold stage as in the /v/-first group. One critical feature of the training procedure in the present study was that the first stimulus was fixed to one category of the contrast. Although it was designed to encourage listeners to utilize the first stimulus as a perceptual anchor in discrimination, it appears to have facilitated discrimination learning. The constant presentation of the /b/ stimuli as the first stimulus might have helped the participants to learn to focus attention on the critical acoustic cues (e.g., a burst) to the /b/ category, and correctly identify the first stimulus as /b/ at the relatively early stage of the training stages. It might be the case that the learning process was facilitated by familiarity with some critical acoustic cues (e.g., voiced bilabial stop) that are also used in the native language.

The participants in the /v/-first group, on the other hand, appear to have been unable to take advantage of the frequent presentation of /v/ as the first stimulus. It might be the case that they were not able to focus on the critical acoustic cues to the /v/ category (e.g., voiced frication) due to unfamiliarity with the acoustic property. It might have been difficult to discriminate the contrast as the perceptual representation of the first stimulus remained unstable. This finding was compatible with those of Tsushima (2007b) that the /l/-first group achieved significantly better discrimination ability than the /r/-first group at the end of the discrimination training. The overall findings support the conclusion that, in the fixed AX categorial discrimination training, discrimination learning is better facilitated when the first stimulus is perceptually similar to the native phoneme (e.g., /b/ or /l/) than when the order is reversed.

Although the /b/-first group surpassed the /v/-first group during the training, it showed some asymmetry in their identification of /b/ and /v/ in the posttest. For example, identification of /b/ was very high (85.1%), regardless of whether the stimuli were training stimuli (84.6%) or generalization stimuli (85.9%). It was shown, however, that identification of /v/ was significantly poorer (71.9%) than that of /b/ (85.1%), and that the difference was largely due to the relatively poor performance on generalization stimuli (63.9%) than that of training stimuli (79.8%). The results suggested that the participants were not able to develop a perceptual representation of the /v/ category that is robust enough to correctly perceive the new stimuli. It was found, on the other hand, that the identification accuracy in the posttest did not significantly differ between /b/ and /v/ in the /v/-first group. These findings were also consonant with those of Tsushima (2007b), which showed that the /l/- first group showed significant improvement in identification of /l/, but little improvement for /r/ between the pretest and posttest, while the /r/-first group improved their identification of both phonemes equally well. The overall results suggest that, although discrimination learning may appear to be facilitated when the first stimulus is fixed to the phoneme perceptually closer to the corresponding native phoneme, it does not necessarily result in improvement of the ability to perceive the perceptually more deviant phoneme of the contrast.

As described in Introduction, a growing body of evidence accumulates on the directional asymmetries in discrimination of vowels (Polka & Bohn, in preparation). Although the data on the directional asymmetries in discrimination of consonants are still sparse, the available data suggest that some parallels exist between them. First, developmental data indicate that vowels with extreme articulatory-acoustic properties (i.e., peripheral vowels), which define the basic structure of the vowel space, act as a perceptual anchor for vowel discrimination (Polka & Bohn, in preparation). For consonants, those which are acquired early in development and thus are phonetically more basic, appear to act as a perceptual anchor (Altvater-Mackensen & Fikkert, 2010), although some conflicting data exist (Kuhl, et al., 2006). Second, the directional asymmetries have been observed among adult listeners under the long-ISI conditions with relatively large stimulus variability, where phonetic (as opposed to auditory) processing is required for vowels as well as consonants. It appears that, for both vowels and consonants, the asymmetries occur when one sound is more directly categorized into one of the existing perceptual categories than the other, whether in L1 or L2 speech discrimination.

The present study is limited in a number of ways. First, the length of training was very short (i.e., 15 minutes) as compared with the previous studies. Although the length was determined by pilot data that showed it had not taken a long time for the participants to pass all the stages, it would be interesting to see whether and how the /v/-first group would improve their discrimination ability if they continue to be trained. Second, the identification test was conducted after a few minutes' break from the discrimination training, so that the present results do not necessarily represent long-term effects of the training on perceptual representations of nonnative phonemes. This measure was taken because it was of interest to examine the immediate effects of training on identification of training and generalized stimuli. Finally, as the size of the data sample was relatively small, the results might have been influenced by some individual factors to some degree. Despite these limitations, it was encouraging to find, as described above, that many aspects of the present results were compatible with those of the previous study (Tsushima, 2007b), which had longer discrimination training (i.e., 12 training sessions separated over four days), a longer break between the training and the posttest (i.e., a week), and a slightly larger sample size (N = 16 and N = 15 in each group).

5. Conclusion

The present study attempted to examine the effects of stimulus order in the fixed AX categorial discrimination procedure on learning to perceive the English /b/-/v/ contrast by Japanese-speaking adults. The study specifically tested the hypothesis that perceptual learning would be adversely affected when the first stimuli were fixed to the category which was perceptually closer to the native phoneme (i.e., /b/) than the other (i.e., /v/). The present results refuted that hypothesis, showing that discrimination learning was actually facilitated when the first stimulus was /b/ rather than /v/. Examination of the performance in the identification pretest and posttest showed that the /b/-first group performed significantly better than the /v/-first group in identification of /b/ stimuli, while that of /v/ stimuli did not show any significant differences between the two groups. In conclusion, the significant stimulus order effects with the same direction as has been found in an initial exposure to the speech stimuli would not be obtained in the discrimination training where listeners attempt to suppress the native-language attentional patterns to perceive often unfamiliar critical acoustic cues that differentiate the nonnative contrast.

Acknowledgements

This research was generously supported by a research grant from Tokyo Keizai University in the academic year of 2009.

References

Altvater-Mackensen, N., & Fikkert, P. (2010). The acquisition of the stop-fricative contrast in perception and production. Lingua, 120, 1898–1909.

Aoyama, K., Flege, J. E., Guion, S., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/. Journal of Phonetics, 32, 233–250.

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). Timonium, MD: York Press.

Brady, S. A., & Darwin, C. J. (1978). Range effect in the perception of voicing. Journal of the Acoustical Society of America, 63, 1556–1558.

Brown, C. (1994). The role of the L1 grammar in the L2 acquisition of segmental structure. McGill Working Papers in Linguistics, 9, 180–210.

Cowan, N., & Morse, P. A. (1986). The use of auditory and phonetic memory in vowel discrimination. Journal of the Acoustical Society of America, 79(2), 500–507.

Jacobson, R. (1968). Child Language, Aphasia, and Phonological Universals. The Hague: Mouton.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. Journal of the Acoustical Society of America, 108 (3 Pt 1), 1252–1263.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. Developmental Science, 9(2), F13–F21.

Ladefoged, P., & Maddieson, I. (1996). The sounds of the world's languages: Blackwell Publishers.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology, 54, 358–368.

Lisker, L., & Abramson, A. (1964) . A cross-language study of voicing in initial stops: Acoustical measurements. Word, 20, 384–422.

Maniwa, K., Jongman, A., & Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. Journal of the Acoustical Society of America, 125(6), 3962–3973.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., & Jenkins, J. J. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. Perception and Psychophysics, 18(5), 331–340.

Olive, J. P., Greenwood, A., & Coleman, J. (1993). Acoustics of American English speech: A dynamic approach. New York, Berlin, Heidelberg, London, Paris, Tokyo, Hong Kong , Barcelona, Budapest: Springer-Verlag.

Perkell, J. S., Klatt, D. H., & Stevens, K. N. (1986). Invariance and variability in speech processes. Hillsdale, N.J.: L. Erlbaum Associates.

Polka, L., & Bohn, O.-S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. Journal of the Acoustical Society of America, 100(1), 577–592.

Polka, L., & Bohn, O. -S. (2003). Asymmetries in vowel perception. Speech Communication, 41, 221–231.

Polka, L., Sundara, M., Zhaoo, V., Pang, M., & Ciocca, V. (in preparation). Directional assymmetries in adult vowel perception. The Journal of the Acoustical Society of America.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. Journal of Experimental Psychology: Human Perception and Performance, 20(2), 421–435.

Repp, B. H., & Crowder, R. G. (1990). Stimulus order effects in vowel discrimination. Journal of the Acoustical Society of America, 88(5), 2080–2090.

Rosen, S. M. (1979). Range and frequency effects in consonant categorization. Journal of Phonetics, 7, 393–402.

Takagi, N. (1993). Perception of American English /r/ and /l/ by adult Japanese learners of English: A unified view. Unpublished Dissertation, University of California, Irvine, Irvine, CA.

Tsushima, T. (2007a). Asymmetries in perception of an American English /r-l/ by adult Japanese learners of English. Journal of the Japan Society for Speech Sciences, 8, 45–62.

Tsushima, T. (2007b). Stimulus order effects in discrimination of the American English contrast /r-l/ by adult Japanese learners of English. Journal of the University of Marketing and Distribution Sciences, 19(3), 15–27.

Tsushima, T. (2010). A preliminary study on stimulus order effects in discrimination of the English /b/-/v/ contrast during categorial AX discrimination training. The Journal of Communication Studies, 31, 153–170.

Tsushima, T., Shiraki, S., Yoshida, K., & Sasaki, M. (2003). On stimulus order effects in discrimination of nonnative consonant contrasts. Acoustical Science and Technology, 24(6), 410–412.

Tsushima, T., Shiraki, S., Yoshida, K., & Sasaki, M. (2005). Stimulus order effects in discrimination of a nonnative consonant contrast, English /b-v/, by Japanese listeners in the AX discrimination procedure. Paper presented at the First Acoustical Society of America Workshop on L2 Speech Learning, Vancouver, Canada.

Tsushima, T., Yoshida, K., Shiraki, S., & Sasaki, M. (2003). How do stimulus variability and ISI influence stimulus order effects during discrimination training of nonnative consonant contrasts? Journal of the Marketing and Distribution Sciences, 15(3), 73–89.

Tsushima, T., Yoshida, K., Shiraki, S., & Sasaki, M. (2005). Investigation into the mechanisms of the stimulus order effects in discrimination of nonnative speech contrasts. Journal of the Marketing and Distribution Sciences, 17(3), 91–101.

Van Hessen, A. J., & Schouten, M. E. H. (1992). Phoneme perception. II: A model of stop consonant discrimination. Journal of the Acoustical Society of America, 92(4), 1856–1868.

Vance, T. J. (1987). Introduction to Japanese Phonology. Albany, NY: State University of New York Press.

Yamada, R. A., & Tohkura, Y. (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. Perception and Psychophysics, 52(4), 376–392.

注

1) A Shapiro-Wilks test was used to test whether the sample distribution was significantly deviant from normality. Non-parametric tests were used when the assumption of normality was judged to be violated in the following analyses.