Journal of Phonetics (1992) **20**, 305–330

Effects of phonological and phonetic factors on cross-language perception of approximants

Catherine T. Best

Haskins Laboratories, New Haven CT 06511, U.S.A. and Wesleyan University, Middletown, CT 06457, U.S.A.

Winifred Strange

University of South Florida, Tampa, FL 33620, U.S.A

Received 16th November 1990, and in revised form 4th February 1992

Past research suggests that the degree of difficulty adults have with discriminating non-native segmental contrasts varies considerably across contrasts and languages. According to a recent proposal, this variation may be explained by differences in how the non-native phones are perceptually assimilated into native phoneme categories [Best, McRoberts & Sithole (1988) *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 345–360]. The present study examined that proposal by testing identification and discrimination of three synthetic series of American English approximant contrasts, presented to American English-speaking subjects and native Japanese-speaking learners of English. The English approximants differ with respect to their phonemic status in Japanese, as well as in the phonetic details of the most similar Japanese phonemes. The perceptual assimilation hypotheses were strongly upheld in cross-language comparisons. Moreover, on the assumption that perceptual assimilation may be modified by learning the second language (L2), we also evaluated differences between subgroups of the Japanese subjects who had two different levels of English conversation experience. Those with intensive English conversation experience showed identification and discrimination patterns that were more similar (but not identical) to the Americans' performance than did those who had had little English experience.

1. Introduction

Language-specific experience influences the perception of phoneme contrasts. Adults are often hampered in their identification and/or discrimination of phones that are not employed contrastively in the phonological system of their language. For example, monolingual Japanese and Korean speakers have difficulty distinguishing the American English liquids /r/ and /l/, which do not occur contrastively in their native languages (Gillette, 1980; Goto, 1971; Miyawaki, Strange, Verbrugge, Liberman, Jenkins & Fujimura, 1975; Sheldon & Strange, 1982). Analogously, English speakers have difficulty with some nonnative contrasts such as the Czech retroflex *vs.* palatal fricatives (Trehub, 1976), Thai voiced *vs.* voiceless unaspirated stops (Lisker & Abramson, 1970), Hindi dental *vs.* retroflex stops, and

0095-4470/92/030305 + 26 $03.00/0

Salish velar *vs.* uvular ejectives (Tees & Werker, 1984; Werker & Tees, 1984; Polka, 1991). This perceptual difficulty, however, appears to be neither universal nor immutable. Some non-native contrasts are relatively easy to discriminate even without prior exposure or training (e.g., Best, McRoberts & Sithole, 1988; Best, 1992). Perceptual difficulties with particular contrasts also vary depending on syllable position and phonetic context (e.g., Mochizuki, 1981). Other contrasts are distinguishable when listening conditions minimize memory demands or phonemic categorization (Carney, Widin & Viemeister, 1977; Werker & Logan, 1985). Discrimination of non-native contrasts that are initially difficult for adults can sometimes be improved rapidly through laboratory training (e.g., Pisoni, Aslin, Perey & Hennessy, 1982), while others are resistant to change (Strange & Dittmann, 1984). Perception of non-native contrasts improves in the course of learning to speak a second language (L2), even in adulthood (e.g., MacKain, Best & Strange, 1981), although improvement is often more marked if exposure to L2 occurs before puberty (Tees & Werker, 1984; Yamada & Tohkura, 1991; see Flege, 1988). Furthermore, some individuals appear to be more sensitive than others to non-native distinctions even without experience or training (e.g., subject M.K. in MacKain *et al.* 1981; see also Polka, 1991; Pruitt, Strange, Polka & Aguilar, 1990).

The fact that native language experience constrains perception of non-native contrasts, but that further experience with non-native sounds may nonetheless alter those perceptual constraints even in adults, raises questions about the nature of the native-language influence. Specifically, what properties do listeners perceive in non-native sounds, and how might those properties relate to the perceived properties of native phonemes?

Recently, it has been proposed that mature listeners perceptually assimilate most non-native phones to native categories (Best, 1992; Best *et al.*, 1988; cf. Flege, 1990). That is, the non-native phones are perceived in terms of their similarities (and dissimilarities) to native phonemes. According to this model, mature language users assimilate non-native speech sounds to native categories on the basis of their perceived gestural (articulatory-phonetic) similarities to native phones (Best, 1992). The gestural similarities and dissimilarities referred to are based on the model of gestural phonology proposed by Browman & Goldstein (e.g., 1986, 1989; Goldstein & Browman, 1986), i.e., they refer to temporal and spatial properties (i.e., degree and location of constrictions) of the dynamic movements of vocal tract articulators such as lips, jaw, tongue body, glottis, etc.

Four perceptual assimilation patterns are possible: (1) the two members of the non-native contrast may be assimilated into *two categories* in the native phonology; (2) both non-native phones may be assimilated equally well (or poorly) into a *single category*; (3) both may be assimilated into a single category, but unequally, thus showing a *category goodness difference* in their fit to the native phoneme; or (4) The non-native phones may differ so much from the phonetic properties of native phonemes that they are *non-assimilable.* Note that the assimilation pattern depends on the listener's *perception* of similarities; listeners may differ from one another, even within the same native language, with respect to which phonetic properties of a non-native phone they may detect or attend to in perception. (Although it might be argued that non-native phones are assimilated on the basis of *acoustic*-phonetic similarities rather than, or in addition to, gestural similarities, the distinction is difficult to make because articulatory- and acoustic-phonetic properties are confounded in the signal).

Best and colleagues (Best *et al.*, 1988; Best, 1992) predicted that phones that are assimilated equally to a single category should prove most difficult to discriminate. Discrimination of phones assimilated to two different native categories should be quite good, while contrasts that are non-assimilable, or those that show a category goodness difference in assimilation, should result in intermediate and variable levels of discrimination difficulty. The level of discrimination of non-native phones that differ in category goodness should depend on the degree of perceived *phonetic* similarity between the native phoneme category and each of the non-native phone categories. Non-assimilable contrasts are perceived as non-speech sounds rather than as phonological segments; for them, discrimination difficulty should be a function of *acoustic* similarity.

Thus, the issue of native-language (L1) influence on perception of non-native speech contrasts focuses on the relation between phonetic details and phonemic categories. In turn, any readjustment in perception as a result of further experience with non-native phones would seem to involve an adjustment in the perceived phonetic details of the second language (L2) phoneme categories (cf. Flege, 1990; Flege & Bohn, 1989). That is, non-native phones may be assimilated to native phonemes to the strongest degree by listeners who have had little or no L2 experience. However, increased L2 experience may foster improved recognition of the *discrepancies* between the L1 and L2 phones. This could lead to a decline in degree of assimilation of L2 phones to L1 categories, and perhaps ultimately to the emergence of a separate L2 phoneme category due to improved recognition of phonetic properties within the L2 phonological system. We pursued these issues in the present study by examining the perception of three English approximant contrasts by American English listeners and by Japanese listeners at two levels of English experience.

Contrasts between approximant consonants (/w – j/, /w – r/ and /r – l/) in syllable-initial position offer a rich context for studying the perceptual influence of both phonetic and phonological differences between American English and Japanese. The contrasts differ across these languages in their phonological status; /r – l/ is a phonemic contrast in English but not in Japanese. The remaining two contrasts can be said to represent abstract phonological oppositions in both languages. However, /w – j/ and /w – r/ differ in terms of the similarities between American and Japanese phonetic realizations of the phonemic categories.

Realizations of /j/ are quite similar in the two languages, differing only slightly in phonetic and phonotactic details. Both are glide consonants with a palatal place of articulation and spread or neutral lip posture. However, Japanese phonotactic constraints disallow the occurrence of /j/ before the high front vowels /i/ and /e/, whereas no such restrictions occur in English. Also, the starting tongue posture has been described as somewhat lower and further back for Japanese /j/ (Vance, 1987) than for English /j/ preceding /a/ (the context used in this study), which should, if true, result in slightly higher F_1 and lower F_2 and F_3 onset frequencies for Japanese /j/.

The phonetic realization of /w/ differs more obviously between languages. In English, /w/ is realized with lip-rounding or protrusion ([w]), similar to the back rounded English vowel /u/, whereas in Japanese, /w/ is produced with spread lips [(ɯ̟)], similar to the back unrounded Japanese vowel [ɯ] (Bloch, 1950; Vance, 1987). Because lip rounding/protrusion lowers the frequency of all formants (especially upper formants), F_2 and F_3 onset frequencies should be higher (hence

more similar to English /j/) in Japanese than English (see Kasuya, Takeuchi, Sato & Kido, 1982; Lisker, 1957; O'Connor, Gerstman, Liberman, Delattre & Cooper, 1957).

The cross-language discrepancy in the phonetic realization of /r/ is even greater, involving a difference in both manner of articulation and tongue posture. Whereas American English /r/ is a retroflex or palato-alveolar central approximant ([ɻ] or [ɹ], respectively), Japanese /r/ is usually an alveolar tap [ɾ] rather than an approximant. (Bloch, 1950; Price, 1981; Vance, 1987). In addition, while English /l/ is an alveolar lateral approximant, Japanese does not employ a distinct /l/ phoneme. Japanese /r/ is, in fact, variably pronounced, and is occasionally realized in some positions by some speakers as an approximant [ɻ] or [ɹ], as a retroflex [ɖ], as an alveolar trill [r], or even as a lateral alveolar tap [l]. Thus, the lateral alveolar is a rare allophone of /r/ in Japanese and is apparently not even then an approximant; rhotic approximants may occur but are also quite rare (Bloch, 1950; Miyawaki, 1973; Vance, 1987).

According to the perceptual assimilation model (Best *et al.*, 1988; Best, 1992), Japanese listeners would be expected to assimilate the English /w – j/ contrast as a two category contrast *vis-à-vis* their native phonology. However, the phonetic boundary between categories may be shifted toward /j/ (that is, Japanese may hear more /w/s), since the Japanese /w/ is unrounded and is more similar to English /j/ acoustically and articulatorily than is the American English /w/. Nonetheless, categorization and discrimination should be quite good. English /w-r/ might be expected to be assimilated to a single Japanese phoneme category, but as a contrast involving a category goodness difference. That is, since English /r/ is an approximant, not a tap as in Japanese, it seems likely to be assimilated as a "poor" exemplar of the Japanese approximant /w/, whereas English /w/ would be assimilated as a "better" exemplar of Japanese /w/. The possibility that [ɹ] would assimilate to Japanese /w/ is supported by evidence from Mochizuki (1981) and Yamada and Tohkura (1991). The alternative possibility, though less likely, is that English /r/ might be assimilated as a very poor exemplar of the Japanese tapped /r/, which would lead to two category assimilation for /w-r/. In either case, Japanese discrimination of /w-r/ should be good. Finally, English /r-l/ should result in single category assimilation by Japanese, in which both phones are equivalently poor exemplars either of their approximant /w/ or (less likely) of their tapped /r/. Japanese categorization and discrimination are known to be rather poor for syllable-initial /r/ and /l/, particularly for those who have had little conversational English experience (Miyawaki *et al.*, 1975; Mochizuki, 1981).

Best *et al.* (1988; 1992) discussed assimilation of non-native speech contrasts only in terms of their relative levels of discriminability. In the present study, the concept of perceptual assimilation was extended to predict cross-language differences in phonetic category boundaries along synthetic approximant series that interpolated on multiple, phonetically relevant acoustic parameters. Specifically, in identification tests of /w-j/ and /w-r/ series, the Japanese listeners were expected to label more of the acoustically intermediate stimuli as /w/ than American listeners. For /w-j/, which are distinguished primarily by F_2 and F_3 onsets and transitions, stimuli with higher F_2 and F_3 values are more similar to Japanese [ɯ̞] than to American [w]. Thus, the Japanese /w-j/ boundary should be shifted toward /j/, relative to the American boundary. However, the steepness of the category boundary should be

equivalent in the two language groups because the contrast reflects a phonological opposition for both.

In the case of /w-r/, Japanese listeners might be expected to label more intermediate stimuli as /w/ rather than as /r/, as compared to American listeners, because the slow transitions of these approximants are more similar to the Japanese /w/ than to their tapped /r/ (see also Mochizuki, 1981). Yet because neither the English /w/ nor /r/ are ideal exemplars of Japanese phoneme categories, and because /w-r/ was expected to be assimilated as a category goodness difference within the Japanese /w/ category, their identification function was expected to be less steep in the region of the category boundary than that of American listeners.

No clear predictions can be made about the *location* of the /r-l/ boundary for Japanese. However, the predicted single category assimilation pattern is consistent with previous findings that the labeling function is less clearly defined for Japanese than for American listeners, resulting in a shallower slope at the category boundary (e.g., MacKain *et al.*, 1981; Miyawaki *et al.*, 1975).

If increased L2 experience serves to shift adults' perception of the phonetic details of non-native phonemes toward improved recognition of the discrepancies between L2 phones and the L1 categories to which they were initially assimilated (cf. Flege & Bohn, 1989; Flege, 1990), additional predictions can be made about relative performance on the three contrasts by Japanese subjects with more or less spoken English experience. According to perceptual assimilation predictions (Best *et al.*, 1988; Best, 1992), Japanese listeners with little English experience should discriminate the /w-j/ contrast best, as a two category contrast, with a peak in discrimination functions at their category boundary (i.e., shifted toward the /j/ end of the series). They should show lower discrimination levels and a lower, broader boundary-related peak (also shifted toward /r/) in discrimination of the English /w-r/ contrast, which shows a category goodness difference with respect to Japanese /w/. Their discrimination should be poorer still on the English /r-l/ contrast, a single category assimilation type. Thus, discrimination performance by inexperienced Japanese listeners should be equivalent to that of American listeners on the /w-j/ contrast, somewhat lower on the /w-r/ contrast, and perhaps even lower on the /r-l/ contrast. In comparing identification performance of Americans and the two Japanese subgroups, we expected that category boundary steepness for /w-j/ would be equivalent across all three groups, but less steep for the inexperienced Japanese than the other two groups on the /w-r/ and /r-l/ series. Japanese with more extensive English conversational training were expected to discriminate and identify all three contrasts in a pattern more similar to that of American adults than their peers who had had minimal English experience, i.e., the position and steepness of their category boundaries should have become shifted toward the values found in Americans. However, according to earlier work showing residual differences from Americans on syllable-initial /r-l/ (Mackain *et al.*, 1981), even the experienced Japanese listeners were expected to differ somewhat from the Americans on the /w-r/ and /r-l/ series in both boundary position and steepness, as well as in discrimination levels.

2. Experiment 1

2.1. *Method*

The aim of this study was to compare identification and discrimination of synthetic /r-l/, /w-r/, and /w-j/ series by American and Japanese listeners. A previous report

had examined perception of an /r-l/ series by these two language groups (MacKain *et al.*, 1981). The stimuli and methods for the /r-l/ tasks, as well as the results for a larger group of Japanese subjects on that contrast, were presented in the earlier publication. For the present paper, we reanalyzed a subset of those earlier-reported data for comparison with responses of the same listeners on the other two approximant series.

2.1.1. *Subjects*

Nine of the 10 original American participants in the MacKain *et al.* study returned within the subsequent two weeks for two additional test sessions on the /w-r/ and /w-j/ contrasts. All were college undergraduates (four males, five females) recruited through notices posted at Yale University.

Nine of the 13 Japanese who participated in the original study returned within two weeks for tests on the other two approximant contrasts. Four Japanese (two males, two females) had had intensive English conversational instruction with native American English speakers (8–10 hours/week) and had been in residence in the U.S.A. for 18 to 48 months at the time of testing (Ss7–10 in MacKain, *et al.*, 1981). These subjects are hereafter referred to as the Experienced Japanese. Five others (four male, one female) had had little or no English conversational instruction (0–3 hours/week) and had resided in the U.S.A. less than seven months (Ss1–4 and S13 in MacKain, *et al.*, 1981). These are hereafter referred to as the Inexperienced Japanese. Note that S13 was subject M.K., an anomalous listener who showed remarkably good /r-l/ perception even though he had been in the U.S.A. only briefly and had had little conversational experience with English. He was discussed separately in MacKain *et al.*, but was incorporated into the Inexperienced group for the present study because of the small number of subjects in each subgroup.

All subjects were paid. All reported good hearing in both ears and could read written English.

2.1.2. *Simulus materials*

The /r-l/ series was a /rak/-/lak/ continuum, and is described in detail in MacKain *et al.* (1981). Two additional series, /wak/-/jak/ and /wak/-/rak/, were generated in analogous manner on the OVE-IIIc cascade formant synthesizer at Haskins Laboratories. Synthesis parameters for series endpoints, /jak/, /wak/, /rak/ (and /lak/), were derived from an analysis of real speech tokens produced by an adult male speaker of American English. These endpoint synthetic stimuli were equated for overall duration (330 ms including the silence and burst of a natural /k/), amplitude and intonation contour (rising-falling), and spectral pattern of the final 105 ms of the 210 ms vocalic portion of the syllable. The initial 105 ms of the four stimuli differed in frequency of onset and the subsequent pattern of transitions of the first three oral formants (F_1, F_2, F_3, respectively). Table I gives the onset frequencies of these formants for the four endpoint stimuli, and Fig. 1 provides a schematic diagram of the formant patterns for the endpoint stimuli of each continuum.

The 10-step /wak/-/jak/ series was generated by interpolating on the F_2 and F_3 onset frequencies in approximately equal steps of 162 Hz and 57 Hz, respectively, from the /wak/ pattern (item 1) to the /jak/ pattern (item 10). The initial

TABLE I. Nominal stimulus parameters for endpoint stimuli

Stimuli	Formant onset frequencies (Hz)		
	F_1	F_2	F_3
/jak/	275	2105	2809
		*[a]	*
/wak	275	644	2295
	*	*	*
/rak/	349	1067	1477
		*	*
/lak/	349	1207	2594

[a] Asterisk indicates that the parameters are interpolated to produce series between endpoints.

steady-state portion was 28 ms for F_2. F_3 was steady-state for 21 ms, followed by a linear transition of 49 ms to a common frequency (2379 Hz). As can be seen in Fig. 1, this produced a "dip" in F_3 for stimuli toward the /jak/ end of the series, which is characteristic of /j/ in natural utterances.

The 10-step /wak/-/rak/ series was generated by interpolating between /wak/ (item 1) and /rak/ (item 10) on F_1, F_2 and F_3 onset frequency (and subsequent transitions) in approximately equal steps of 8 Hz, 47 Hz, and 91 Hz, respectively. An inflection point 28 ms after onset of F_2 and F_3, and 21 ms after onset for F_1, produced an initial quasi-steady-state pattern (see Fig. 1).

For comparison, the endpoints of the /rak/-/lak/ series are included in Table I and in Fig. 1. In this series, onsets and transitions of F_2 and F_3 were varied, as well as the temporal pattern of the F_1 transition (see MacKain *et al.*, 1981, for a detailed description).

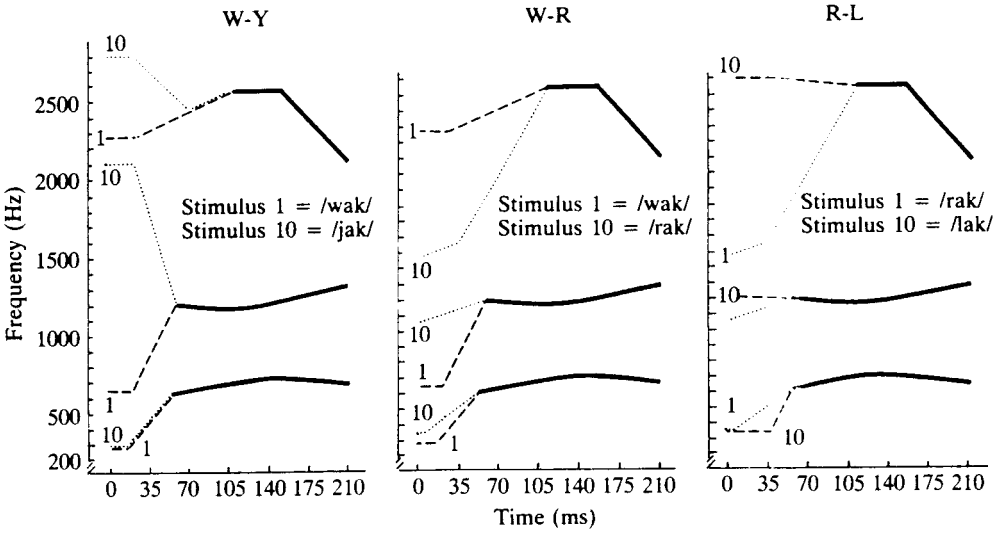

Figure 1. Schematic diagram of the center frequencies of F_1, F_2 and F_3 in the endpoint stimuli for the three stimulus series.

2.1.3. *Procedure*

The tests for the /rak/-/lak/ series are described in MacKain *et al.* (1981). The tests for the other two series were similar in format, except that the oddity discrimination test used in the previous study was not employed; only the AXB discrimination task was used for the present report. All subjects completed two sessions consisting of two tests each, with a 15 min break between the first and second test of the session. In one session subjects completed a two-choice forced choice identification test followed by an AXB discrimination test of the /w-j/ series. The other session included identification and AXB discrimination tests of the /w-r/ series. Testing was conducted in a sound-attenuated chamber with two to four subjects at a time (all from a single language group during a given test session). Subjects listened over headphones (Telephonics TDH-39) to stimuli presented via a Crown reel-to-reel tape deck at a comfortable loudness level (approximately 75 db SPL).

Each identification test included 20 repetitions of each of the 10 stimuli in the series being tested, presented singly and randomized within each block of 10 trials. Intertrial intervals (ITIs) were 2.5 s; interblock intervals (IBIs) were 4 s. For each trial, subjects were asked to write one of two letters to indicate the initial consonant of the syllables they heard; that is they wrote "W" or "Y" during the /w-j/ identification tests, and "W" or "R" during the /w-r/ identification tests.

The AXB discrimination procedure was chosen bause of its relatively low memory demands and low sensitivity to observer bias, by comparison to other standard discrimination procedures such as oddity, 2IAX and 4IAX (e.g., Best, Morrongiello & Robson, 1981; MacKain *et al.*, 1981; cf. Pollack & Pisoni, 1971). Each AXB discrimination test contained 10 repetitions of each of the two AXB orders for the seven possible pairings of stimuli that differed by three steps along the continuum being tested (1–4, 2–5, 3–6, 4–7, 5–8, 6–9 and 7–10). Trials occurred in blocks of 14 (two orders × seven AXB pairings), and were randomized within blocks. Within-trial interstimulus intervals (ISIs) were 1 s, ITIs were 3 s and IBIs were 6 s. For each trial, the subject circled the number "1" or the number "3" to indicate whether the second item of the trial (X) matched the first (A) or the third (B) item of that trial.

2.2. *Results*

The results of identification tests are reported first, followed by the results of discrimination tests. Differences between the American group and the Japanese group as a whole were statistically analyzed. Performance by Experienced and performance by Inexperienced Japanese subgroups were compared with that of the American group in separate analyses. For all analyses, data on the perception of /r-l/ by the nine Americans and nine Japanese, which were a subset of the data reported previously in MacKain *et al.* (1981), were included for comparison with results on the /w-r/ and /w-j/ series.

2.2.1. *Identification tests*

Figure 2 presents the pooled identification functions for the American and Japanese groups on the /w-j/, the /w-r/ and the /r-l/ continua. These functions represent the raw identification data, averaged over nine subjects in each group. As the figure

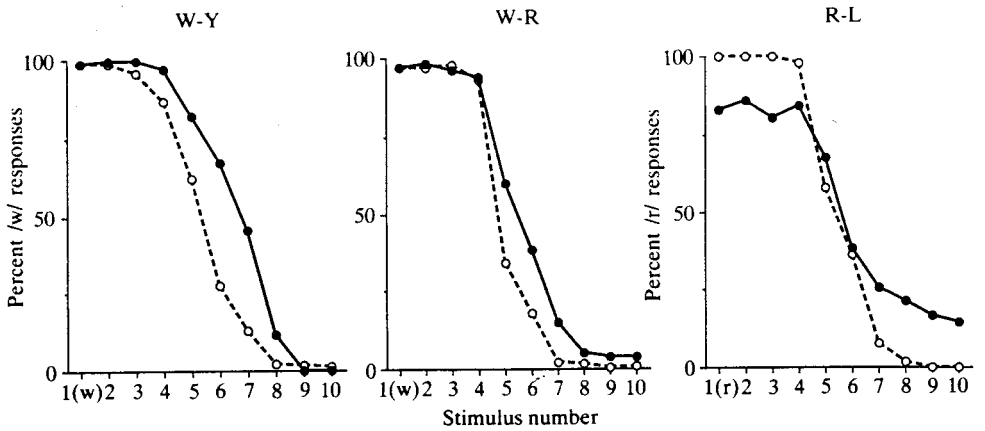

Figure 2. Average identification functions for the American (O- - -O) and Japanese (●——●) listener groups on the three series.

shows, the American listeners labeled /w-j/ and /w-r/ categorically, with abrupt crossovers at category boundaries and highly consistent labeling of within-category stimuli. Performance was commensurate with their identification of the /r-l/ series. The Japanese as a group also labeled /w-j/ and /w-r/ categorically. This contrasts with their identification performance on the /r-l/ series, which showed less consistency in labeling within-category stimuli. As previously reported, performance by the Japanese was markedly different from that of the American listeners on the /r-l/ series.

In order to make between-group comparisons on the location and steepness of category boundaries for the three series, best fit ogives of individual subject's identification functions were determined through narrow-range PROBIT analyses, using the labeling probabilities on the three stimuli closest to the 50% crossover. This statistical procedure fits a cumulative normal curve to the raw data, thus smoothing the function. Category boundaries were defined as the 50% intercept of the ogives. The slopes of these ogives (1/s.d.) indicate the peak rate of change in category labeling at the crossover, and were used as a reflection of the steepness of the category boundaries, i.e., larger slope values indicate steeper functions.

The ogives for the Americans and the two Japanese subgroups are displayed in Fig. 3. χ^2 values were significant, indicating a significant deviation between the raw data and the fitted ogives, for only six out of the 54 PROBIT analyses (two groups × nine subjects × 3 series): three Americans on /w-r/, one American on /r-l/, and two Experienced Japanese on /w-j/. In all cases, the significant χ^2 resulted from extremely sharp category boundaries that were not well-fitted to three data points, and would fit better for two points. There were only two cases of grossly non-monotonic raw identification functions for two Inexperienced Japanese on /r-l/. In neither case was the PROBIT χ^2 significant, i.e., the ogives provided a good fit to the raw data.

Boundary location analyses. The boundary locations for American and Japanese groups (expressed in terms of stimulus number) on each series are given in Table II. These data indicate that, on average, the boundaries for the Japanese on all three

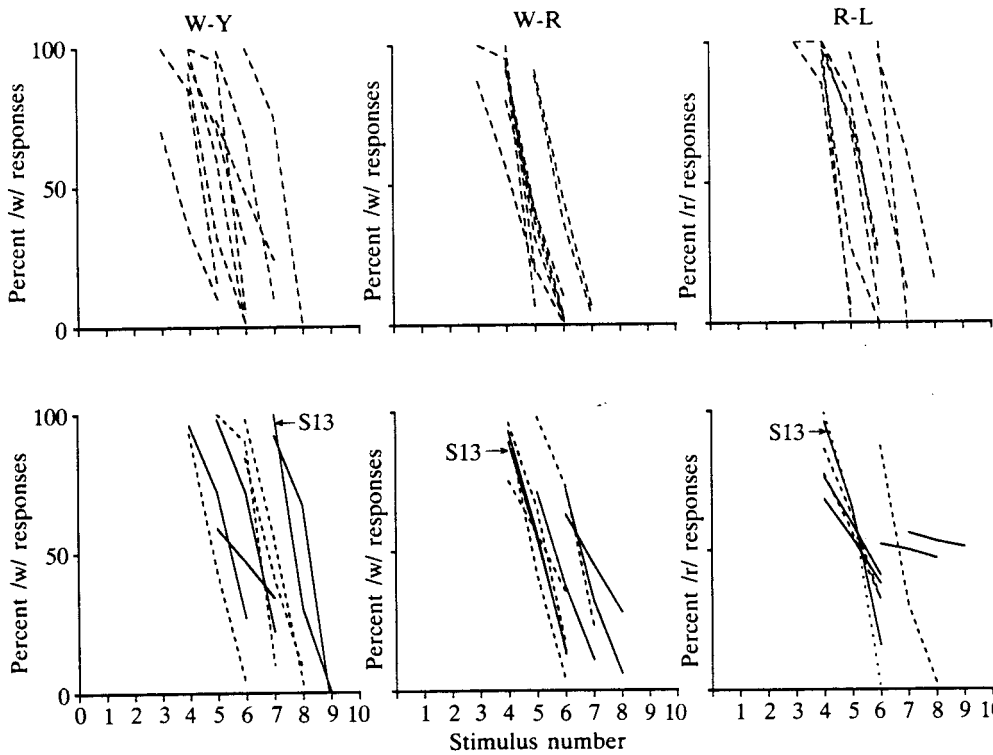

Figure 3. Narrow-range fitted ogive functions for individual subjects in the American (top) and Japanese (bottom: – – –, experienced; ——, inexperienced) groups. The S13 lines indicated in the Japanese plots refer to the data from subject M.K., discussed in MacKain *et al.* (1981) as being inexperienced with American English conversation yet similar to Americans in categorization of /r/ and /l/.

series fell to the right of the American boundaries. That is, the mean boundary values show that the Japanese labeled more stimuli as /w/ on the /w-j/ and /w-r/ series, and more stimuli as /r/ on the /r-l/ series. Note also that the variability of boundary locations appears to be greater on the /w-j/ series than on the /w-r/ series for both Japanese and American subjects, as reflected in the standard deviations (SDs).

To test the reliability of these boundary differences, a Groups (American *vs.* Japanese) × Series (/w-j/, /w-r/, /r-l/) analysis of variance (ANOVA) of the 50% intercept values of best fit ogives for individual subjects was conducted. The main effect of Groups was significant, $F(1, 16) = 10.82$ $p < 0.005$, indicating that the Japanese boundaries were indeed shifted significantly rightward in comparison to the American boundaries. Neither the Series main effect nor the Groups × Series interaction approached significance ($p = 0.17$ and 0.64, respectively), suggesting that the rightward shift of the Japanese boundary occurred in all three series, and to approximately the same degree in each. However, *a priori* predictions about possible cross-language differences on the boundaries for each series warranted an analysis of simple effects, which indicated that the language difference was significant for /w-j/ $F(1, 48) = 6.44$, $p < 0.02$, but was marginal for /w-r/ ($p = 0.10$)

TABLE II. Boundary locations for American English and Japanese listeners, including Japanese subgroups. Numerical values represent stimulus numbers along each of the test series

	/w-j/ Mean (SD)	/w-r/ Mean (SD)	/r-l/ Mean (SD)
Americans	5.36 (1.05)	4.93 (0.57)	5.53 (0.96)
Japanese: Pooled	6.55 (1.07)	5.72 (0.70)	6.08 (1.40)
Experienced	6.32 (0.98)	5.59 (0.69)	5.60 (0.74)
Inexperienced	6.73 (1.22)	5.82 (0.77)	6.47 (1.76)

and non-significant for /r-l/ ($p = 0.24$). That is, the boundary shift between language groups was reliable only for /w-j/.

To assess the statistical reliability of the differences between Experienced and Inexperienced subgroups in comparison with American listeners, an English Experience (American *vs.* Experienced Japanese *vs.* Inexperienced Japanese) × Series ANOVA was computed. (Because group sizes were small and unequal, these statistical results should be interpreted cautiously, although these factors decrease rather than increase the likelihood of attaining statistical significance). The main effect of English Experience was significant, $F(2, 15) = 6.75$, $p < 0.01$, while the main effect of Series and the English Experience × Series interaction were non-significant. Planned linear contrasts among the three groups, based on *a priori* predictions, yielded reliable evidence that the boundary for the Experienced Japanese subjects was intermediate between that of the Americans and that of the Inexperienced Japanese, $F(1, 15) = 13.12$, $p < 0.003$. Table II summarizes these differences in boundary locations for the Experienced and Inexperienced Japanese subjects.

Slope analyses. Table III presents the data on steepness of category boundaries for American and Japanese groups (expressed as the mean slope of their ogives). The Japanese showed a pattern across the three series that was strikingly different from that of the Americans. The slope for /w-j/ was steepest and most similar to the Americans', while those for /w-r/ and /r-l/ were less steep than the Americans'. This was as predicted on the reasoning that /w-j/ would constitute a two category distinction for the Japanese, while /w-r/ would show a category goodness difference within a Japanese category, and both /r/ and /l/ would show a poor fit to one Japanese category.

TABLE III. Slope values for American and Japanese listeners, including Japanese subgroups. Numerical values represent the peak rate of percentage change in category responses per step along each stimulus series

	/w-j/ Mean (SD)	/w-r/ Mean (SD)	/r-l/ Mean (SD)
Americans	2.19 (1.29)	1.99 (1.05)	2.65 (1.67)
Japanese: Pooled	2.04 (1.29)	1.09 (0.41)	1.04 (1.23)
Experienced	1.84 (0.57)	1.24 (0.49)	1.75 (1.55)
Inexperienced	2.20 (1.74)	0.97 (0.33)	0.48 (0.55)

The statistical reliability of these differences was assessed in a Groups (American *vs*. Japanese) × Series ANOVA of slope values. The main effect of Groups was significant, $F(1, 16) = 5.47$, $p < 0.04$, indicating that, overall, the American boundaries were significantly more abrupt than the Japanese boundaries. Neither the Series main effect nor the Groups × Series interaction was significant. However, *a priori* predictions about cross-language differences warranted simple effects tests, which indicated that the American slopes were steeper than the Japanese slopes on /w-r/, $F(1, 16) = 5.77$, $p < 0.03$, and /r-l/, $F(1, 16) = 11.58$, $p < 0.04$, but not on /w-j/ ($p = 0.80$).

Again, the Japanese data for Experienced and Inexperienced subjects were analysed in an English Experience × Series ANOVA which included comparisons to the American group. Although the main effect of English Experience was only marginally significant, $F(2, 15) = 2.91$, $p < 0.09$, planned linear contrasts were warranted by *a priori* predictions (American > Experienced Japanese > Inexperienced Japanese). These tests revealed the predicted direction of group differences was significant for /r-l/, $F(1, 15) = 7.36$, $p < 0.02$, and /w-r/, $F(1, 15) = 5.03$, $p < 0.05$, but not for /w-j/ ($p = 0.99$), all as expected. No other effects were significant.

To summarize, the Japanese /w-j/ boundary was shifted toward /j/ relative to the American boundary. Both Experienced and Inexperienced Japanese labeled more intermediate stimuli as /w/ than the Americans, as predicted from cross-language differences in the phonetic details of /w/. Also as predicted, the steepness of the category boundary slope on this series did not differ between language groups, indicating that the division between /w/ and /j/ categories was equally sharp for all groups of listeners. These findings suggest that the American /w-j/ distinction was assimilated as a two category contrast by the Japanese listeners, with /w/-like and acoustically intermediate stimuli assimilating to the phonetically different Japanese /w/, and /j/-like stimuli assimilating to the phonetically similar Japanese /j/ phoneme category. This characterization is somewhat qualified, however, by the discrimination results on /w-j/ (see below).

The identification results for the /w-r/ and /r-l/ series were different from those for /w-j/. As previously reported, the Japanese listeners showed significantly shallower category boundary slopes on /r/-/l/, but failed to show a significant difference in boundary location, relative to Americans. On /w-r/, the Japanese again showed a shallower boundary slope than Americans, and their boundary location differed marginally from Americans' ($p = 0.10$) in the predicted direction (i.e., they identified more stimuli as /w/). The /w-r/ and /r-l/findings are consistent with the reasoning that American English /w-r/ should constitute a category-goodness difference within the Japanese /w/ category, and that English /r-l/ should represent rather poor examples of a single phoneme category in Japanese (either their glide /w/ or, less likely, their tapped /r/).

As for the effect of experience with L2, the patterns of identification performance differed as expected between the two levels of English conversation experience of the Japanese subjects. On all counts, the data of the Experienced Japanese subjects were more similar (but not identical) to the American results than were those of the Inexperienced Japanese. More intensive English conversation experience was associated with a more American-like boundary location on the English /w-j/ contrast and with steeper category boundaries for the English /w-r/ and /r-l/ contrasts.

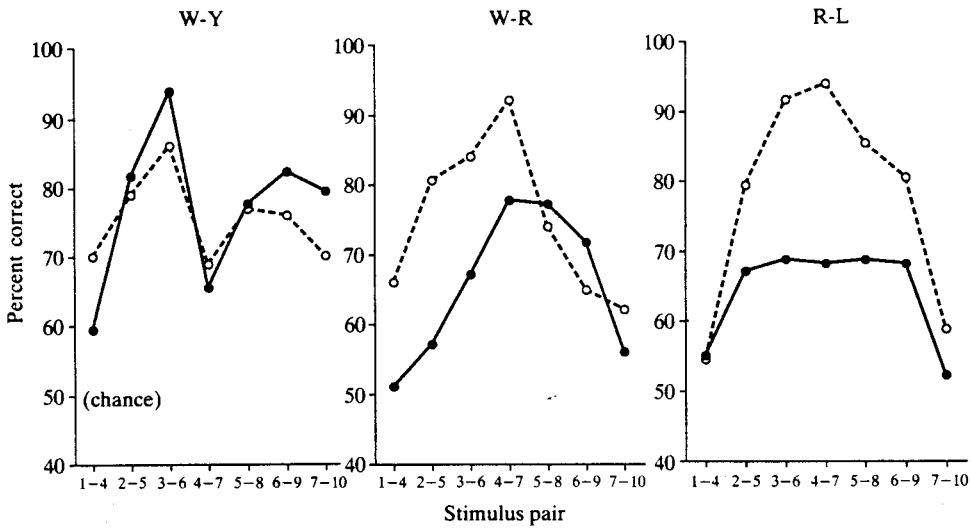

Figure 4. Average discrimination functions for the American (O– – –O) and Japanese (●——●) groups on the three series.

2.2.2. *Discrimination tests*

Discrimination test results were also examined for evidence of native language differences and influences of L2 English experience. Percent correct responses for each of the AXB comparison pairs on each stimulus series were computed for the American and Japanese groups. Pooled discrimination functions for the Japanese and American groups are displayed in Fig. 4, and mean performance levels (overall percent correct) are presented in Table IV. The relationship between American and Japanese discrimination function varied considerably across the three series.

The data were entered into a Groups × Series × Comparison Pairs (1–4, 2–5, 3–6, 4–7, 5–8, 6–9, 7–10) ANOVA. A significant Groups main effect, $F(1, 16) = 8.55$, $p < 0.01$, indicated that Japanese were less accurate overall in discrimination than were Americans. The significant main effect for Comparison Pairs, $F(6, 96) = 30.87$, $p < 0.001$, indicated that overall there were peaks and troughs in discrimination performance across the three series. The latter effect was qualified, as expected, by a Comparison Pairs × Groups interaction, $F(6, 96) = 3.39$, $p < 0.005$, indicating that, in general, the Japanese showed smaller discrimination peaks than the American listeners. The significant Series effect, $F(2, 32) = 3.64$, $p < 0.04$, revealed

TABLE IV. Mean correct performance levels for American and Japanese listeners on the AXB discrimination task, including Japanese subgroups

	/w-j/ Mean (SD)	/w-r/ Mean (SD)	/r-l/ Mean (SD)
Americans	74.52 (11.94)	74.68 (17.46)	77.78 (19.32)
Japanese: Pooled	77.14 (12.97)	65.48 (15.86)	64.13 ((14.99)
Experienced	78.04 (11.57)	66.43 (18.00)	67.50 (15.55)
Inexperienced	76.43 (14.12)	64.71 (14.14)	61.43((14.17)

that discrimination performance was somewhat higher overall for /w-j/ than for the other two series. However, Series interacted with Group, $F(2, 32) = 6.68$, $p < 0.004$; as expected, cross-series mean performance differed between language groups. Simple effects tests of this interaction revealed that mean performance differed among series for the Japanese, $F(2, 16) = 12.77$, $p < 0.0005$, being substantially better for /w-j/ (77% correct) than for /w-r/ (65%) or /r-l/ (64%). Planned comparisons provided support for the order of performance that had been predicted on the basis of expected phonemic assimilation patterns (/w-j/ > /w-r/ ≥ /r-l/), $F(1, 16) = 25.313$, $p < 0.0001$. However, a test of simple effects showed that the Americans' mean discrimination did not differ significantly across series, $p = 0.58$.

Comparison Pairs and Series also interacted significantly, $F(12, 192) = 6.48$, $p < 0.001$, indicating differences in the cross-series patterns of discrimination peaks for both groups, which were further qualified by a significant Groups × Comparison Pairs × Series interaction, $F(12, 192) = 3.04$, $p < 0.002$. To interpret these interactions, separate ANOVAs for Groups × Comparison Pairs were computed for each stimulus series. As predicted, analysis of the /w-j/ series yielded no significant difference between groups in overall discrimination accuracy. A significant main effect of Comparison Pairs, $F(6, 96) = 21.14$, $p < 0.001$, revealed that both groups showed two peaks of relatively accurate discrimination. The occurrence of a double peak suggests that both Japanese and American listeners differentiated three rather than two categories along this synthetic continuum, although they could not indicate this in the two-category forced-choice identification test. (This possibility is considered further below and in Experiment 2.) The significant Groups × Comparison Pairs interaction, $F(6, 96) = 3.46$, $p < 0.01$, was due to the fact that Japanese and American listeners performed differently on both within-category extremes of the series (Pairs 1–4 and 7–10). As indicated in Fig. 4, Japanese subjects discriminated Pair 7–10 (within-category for /j/) more accurately, while Americans discriminated Pair 1–4 (within-category for /w/) more accurately. This asymmetry in discrimination of the endpoint within-category comparison pairs is compatible with the fact that the Japanese category boundary was shifted significantly more toward /j/ than was the American boundary. That is, both stimuli 10 and 7 fell within the /j/ category for Americans (99% and 87% of identification responses, respectively), but for the Japanese stimulus 7 was quite near the /w-j/ boundary (59% identification as /j/) while stimulus 10 was a clear /j/ (100%), which resulted in better discrimination by the latter language group. Conversely, at the other end of the series, the Japanese and Americans agreed that stimulus 1 was a clear /w/ (97 and 98%, respectively), but whereas the Japanese also identified stimulus item 4 as /w/ 98% of the time, the Americans gave only 87% /w/ identifications. Thus the Japanese discriminated comparison pair 1–4 near chance, while the Americans discriminated that pair more readily. In fact, Americans showed the same level of performance as on pair 7–10, which had received quite similar identification scores. No other /w-j/ discrimination pairs differed between language groups.

The pattern of discrimination was quite different on the /w-r/ series. A significant Comparison Pairs effect, $F(6, 96) = 9.70$, $p < 0.001$, reflected a single peak in discrimination performance, with troughs on either side. A significant Groups effect, $F(1, 16) = 8.64$, $p < 0.01$, indicated that discrimination was less accurate overall for Japanese than for American listeners. This was due to their poorer performance on

pairs at the /w/ end of the continuum (1–4, 2–5) and on cross-category pairs (3–6, 4–7), as indicated by a significant Groups × Comparison Pairs interaction, $F(6, 96) = 3.23$, $p = 0.01$, and simple effects tests of individual pairs. Thus, while both groups showed a single discrimination peak, the Japanese peak was shifted slightly toward the /r/ end of the continuum, and was broader and lower than the American peak. Both of these effects are consistent with cross-language phonemic and phonetic differences, as discussed in the Introduction. The identification test had provided marginal evidence that the Japanese /w-r/ boundary was shifted toward the /r/ end of the continuum, relative to the American's boundary, a pattern now corroborated by the small rightward shift of the peak in the Japanese discrimination function. This shift, although slight, is compatible with the greater cross-language phonetic similarities for /w/ than for /r/. As was argued earlier, the lack of rounding in the Japanese /w/ should lead Japanese listeners to identify more /w/s in the /w-r/ (as well as the /w-j/) series. Correspondingly, the poor fit of English /r/ to either the Japanese /w/ or the Japanese /r/ categories should converge on perception of fewer /r/s by the Japanese on the /w-r/ series. English /w-r/was expected to be assimilated as a category goodness difference within Japanese /w/, English /r/ being heard as a poor Japanese /w/. The lower, broader peak in Japanese discrimination, relative to the American /w-r/ peak and to the Japanese /w-j/ peak(s), is compatible with this hypothesis.

Finally, as previously reported for larger groups (MacKain *et al.*, 1981), results on /r-l/ indicated significant differences between Groups, $F(1, 16) = 10.14$, $p < 0.006$, and between Comparison Pairs, $F(6, 96) = 17.74$, $p < 0.001$, as well as a significant Groups × Comparison Pairs interaction, $F(6, 96) = 2.90$, $p < 0.02$. Japanese subjects discriminated cross-category pairs (3–6, 4–7, 5–8) much more poorly than Americans. This was expected, and is compatible with the hypothesis that Japanese listeners assimilate English /r-l/ as poor exemplars of a single category in their own language. Note also the difference in Japanese performance on /w-r/ *vs*. /r-l/ in Fig. 4. Their minimal "peak" in discrimination of the cross-category /r-l/ pairs is clearly lower and broader than their peak in discrimination of /w-r/. This relation is compatible with the hypothesis that /r/ and /l/ are assimilated to a single native category, whereas the /w-r/ contrast constitutes a category goodness difference for Japanese.

Differences in discrimination performance by Experienced and Inexperienced Japanese subgroups were also considered. Overall accuracy across English Experience and Series is shown in Table IV and Fig. 5. Both Japanese subgroups performed relatively well on the /w-j/ series; overall levels were similar to that of the Americans'. For /w-r/ the Japanese subgroups showed similar performance levels (but note the difference in the position of their performance peaks, Fig. 5), although their performance was lower than Americans. Inexperienced Japanese showed lower /r-l/ performance than Experienced Japanese, but again both groups performed less well than Americans.

An English Experience (Americans, Experienced Japanese, Inexperienced Japanese) × Series × Comparison Pairs ANOVA revealed significant effects of English Experience, $F(2, 15) = 4.70$, $p < 0.03$, Series, $F(2, 30) = 6.16$, $p < 0.01$, and Comparison Pairs, $F(6, 90) = 25.40$, $p < 0.01$, as well as significant two-way and three-way interactions [Series × English Experience, $F(4, 30) = 3.34$, $p < 0.03$; Comparison Pairs × English Experience, $F(12, 90) = 1.93$, $p < 0.05$; Series ×

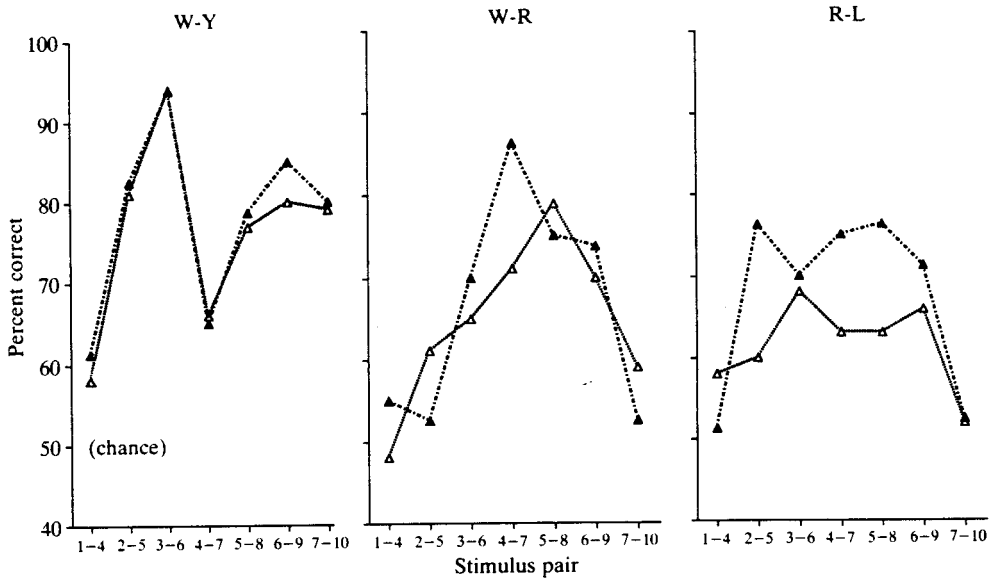

Figure 5. Average discrimination functions for the Experienced (▲–·–·▲) and Inexperienced (△·····△) Japanese subgroups on the three series.

Comparison Pair, $F(12, 180) = 6.24$, $p < 0.001$; Series × Comparison Pair × English Experience, $F(24, 180) = 2.04$, $p < 0.01$]. Analyses of simple effects for Series within Japanese subgroups showed no significant differences in overall accuracy across series for the Experienced Japanese ($p = 0.10$), although peaks and troughs were positioned differently across series, as indicated by their significant Series × Comparison Pairs interaction, $F(12, 36) = 4.84$, $p < 0.01$. In contrast, a significant Series effect for the Inexperienced Japanese indicated more accurate dicrimination of /w-j/ pairs than of /w-r/ or of /r-l/ pairs, $F(2, 8) = 9.31$, $p < 0.01$. A planned linear contrast on the predicted performance pattern (/w-j/ > /w-r/ > /r-l/) was also significant for the latter subgroup, $F(1, 2) = 16.85$, $p < 0.01$.

Experienced and Inexperienced subjects performed almost identically on the /w-j/ series; both groups displayed double peaked functions, which suggest that all the Japanese subjects could differentiate acoustically intermediate stimuli from both /w/ and /j/ phonetic endpoints. An English Experience × Comparison Pairs simple effect ANOVA for /w-j/ revealed no significant effect of English Experience ($p = 0.66$) and a marginally significant English Experience × Comparison Pairs interaction ($p = 0.08$). The latter suggests a tendency for the discrimination peaks to be higher, and for the peak between /j/ and the intermediate stimuli to be shifted toward /j/, in both Japanese subgroups relative to the Americans.

There were obvious differences in the pattern of discrimination for Experienced and Inexperienced subgroups on /w-r/ and /r-l/. Separate English Experience × Comparison Pairs analyses revealed significant overall group differences in discrimination of /w-r/, $F(2, 15) = 4.16$, $p < 0.04$) and of /r-l/, $F(2, 15) = 5.56$, $p < 0.02$. Planned linear contrasts indicated that the expected ordering of performance (American > Experienced Japanese > Inexperienced Japanese) was significantly upheld for both series [$F(1, 2) = 6.85$, $p < 0.02$ and $F(1, 2) = 10.38$,

$p < 0.01$, respectively]. Performance by the two Japanese subgroups on /w-r/ suggested an effect of experience on the location of the phonetic boundary. This was corroborated by a significant English Experience × Comparison Pairs interaction, $F(6, 90) = 2.08$, $p < 0.04$. While discrimination for Experienced Japanese was most accurate for comparison pair 4–7 (as it was for Americans), the Inexperienced Japanese performed best on pair 5–8. For /r-l/, English experience instead affected the height of the discrimination peak across the category boundary. Consistent with the larger dataset reported in the MacKain *et al.* (1981), Experienced Japese showed better discrimination than Inexperienced Japanese on cross-category pairs (4–7, 5–8).

2.3. *Discussion*

Both the identification and the discrimination results are consistent with predictions based on the perceptual assimilation model (Best, 1992; Best *et al.*, 1988). That is, American English /w/ *vs*. /r/ appears to be perceived as a category goodness difference within one Japanese phoneme category (/w/), and /r/ and /l/ are perceived as poor examples of a single category. The identification results and the mean discrimination performance levels on /w-j/ are compatible with the hypothesis that the phones are assimilated to two different Japanese categories (but see the qualifications discussed below). Analyses of the two Japanese subgroups further corroborated predictions. Specifically, Experienced Japanese performed more like Americans than did the Inexperienced Japanese on all series and measures except for discrimination of /w-j/. On that series, there were no cross-language differences (as expected) except for the within-category comparison pairs at the endpoints of the series; this pattern is compatible with language differences in the phonetic properties of /w/.

There was a surprise, however, in the discrimination results for the /w-j/ series. The double peak in discrimination by the Americans and both Japanese subgroups suggested that all listeners may have perceived three rather than two categories along the series, with some category intermediate between /w/ and /j/ perceived in the central portion of the series. This suggests the possibility that the /w-j/ series actually constitutes a combination of a two category distinction for Japanese (/w-j/), along with a category goodness difference within one of those categories. Comparison between the Japanese identification function and their discrimination performance indicates that most of the intermediate category tokens (5–7) were labeled as ambiguous /w/s. These items were apparently difficult to discriminate from one another but easy to discriminate from "good" /w/'s (i.e., items 1–3, consistently labeled as /w/), suggesting a goodness-of-fit distinction within the Japanese /w/ category. Indeed, when the experimenters listened to this synthetic series, several items near the center of the series were perceived as /l/-like. Consistent with this perception, the F_1, F_2, and F_3 onset frequencies and transition patterns in the central stimuli of the /w-j/ series were quite similar to those of the stimuli in the /r-l/ series that were identified by Americans as /l/. The suggestion that the /w-j/ series actually contained three identifiable categories, /w-l-j/, was examined further with a naive group of Americans in Experiment 2.

3. Experiment 2

3.1. *Method*

3.1.1. *Subjects*

As the original subjects were no longer available for testing, nine new native English-speaking American subjects (three males, six females) participated in the study. Seven were graduate students; the other two were faculty members. All reported normal hearing in both ears. Two additional subjects were eliminated from the final sample after testing, when they indicated that they had been diagnosed as learning disabled in childhood. Both had phonemic categorization difficulties, having failed to consistently categorize and discriminate synthetic /ra/-/la/ in a separate but concurrently-run study.

3.1.2. *Stimuli and procedures*

The /w-j/ series from Experiment 1 was again employed. The procedure and testing conditions were identical to those of Experiment 1, except that the forced-choice identification test included three response alternatives ("W", "L", "Y") rather than two.

3.2. *Results*

3.2.1. *Identification test*

As illustrated in the left side of Fig. 6, subjects consistently divided the continuum into three sharply defined categories. Table V lists the means and standard deviations of the boundary locations and slope values for both boundaries, computed from PROBIT analyses as in Experiment 1. Three of the 18 fitted ogives deviated significantly from the raw data, according to χ^2 analyses, two on the /l-j/ boundary and a third on the /w-l/ boundary. In all cases, the ogive was the best fit obtainable, and the significant χ^2s were due to extremely steep category boundary slopes.

The location of /w-l/ and /l-j/ boundaries obtained in the three-choice identification task was compared with the /w-j/ boundaries obtained in the two-choice task of Experiment 1. A Groups (Americans-Exp. 2 *vs.* Americans-Exp. 1 *vs.* Japanese-Exp. 1) × Comparison Pairs ANOVA comparing the /w-l/ boundary with the /w-j/ boundaries yielded a significant main effect of Groups, $F(2, 24) = 25.04$, $p < 0.001$. Sheffe's tests showed that the /w-l/ boundary differed from both the American and Japanese /w-j/ boundaries in Experiment 1 ($p < 0.01$). In a separate ANOVA comparing the /l-j/ boundary with /w-j/ boundaries from Experiment 1, there was again a significant main effect of Groups, $F(2, 24) = 7.86$, $p = 0.001$. Scheffe's tests indicated that the /l-j/ boundary again differed from the Americans-Exp. 1 /w-j/ boundary ($p < 0.01$). However, it did not differ from the Japanese /w-j/ boundary ($p = 0.35$). Thus, while the Experiment 1 discrimination results suggest that the Japanese had actually perceived three categories along the /w-j/ series, as do Americans, the latter result suggests that the Japanese assimilated the intermediate tokens to their /w/ category but as perceptibly poorer exemplars of that category.

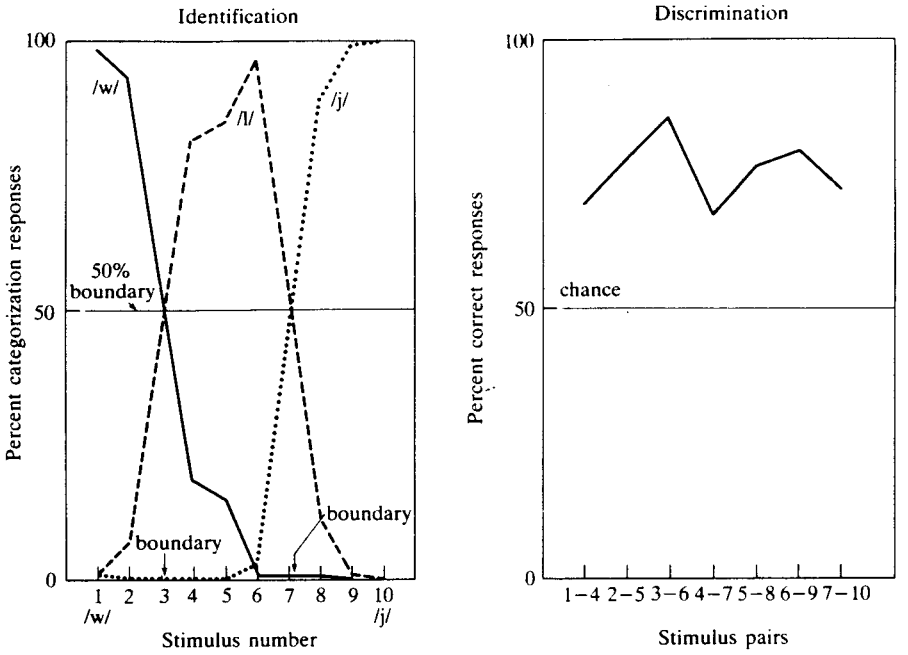

Figure 6. Identification and discrimination functions for the three-category tests on the /w-j/ series with Americans in Experiment 2.

Neither the /w-l/ nor the /l-j/ slope values differed from those found for either group in Experiment 1.

3.2.2. *Discrimination test*

As can be seen in the right side of Fig. 6, the discrimination function again showed two peaks of relatively accurate performance, which coincided with the two category boundaries revealed in the three-choice identification task. For comparison with Experiment 1, a Groups (Japanese-Exp. 1, Americans-Exp. 1, Americans-Exp. 2) × Comparison Pairs ANOVA was conducted. The Groups main effect was non-significant ($p = 0.66$), indicating no systematic differences among groups in overall discrimination performance. The significant Comparison Pairs effect, $F(6, 144) = 29.68$, $p < 0.001$, revealed that there were two reliable peaks in discrimination. Finally, the Groups × Comparison Pairs interaction was significant, $F(12, 144) = 2.48$, $p < 0.01$, due primarily to differences among the groups in

TABLE V. Category boundary locations and slope values for American's three-choice identification of the /w-j/ series (Experiment 2)

	/w-l/ Mean (SD)	/l-j/ Mean (SD)
Boundary location	3.26 (0.84)	7.20 (0.59)
Boundary slope	2.53 (1.39)	3.17 (1.17)

discrimination of the within-category Pairs (1–4, 3–6, 7–10). However, the locations of discrimination peaks did not differ among the three subject groups.

3.3. *Discussion*

The results of Experiment 2 confirm that the intermediate category suggested by the double peak in the Experiment 1 discrimination functions was identified by Americans as /l/. As suggested earlier, this categorization is interpretable on the basis of the similarity between the acoustic properties of /l/ and those of the intermediate tokens in the /w-j/ series (see Fig. 1). For intermediate tokens, F_1 had a steady-state onset, followed by a moderately steep transition, like /l/ but unlike /r/ in the /r-l/ series. They had F_2 onsets around 1200–1400 Hz, with a shallow falling transition, again like /l/ in the /r-l/ series. Moreover, their F_3 transitions were nearly flat or slightly falling, like that of /l/ in the /r-l/ series, except for a slight dip in frequency just before reaching the vowel steady-state. In particular, the F_3 onset frequency of these stimuli was not close to the frequency of F_2, which is needed for good /r/ perception. Given that Japanese does not employ an /l/ phoneme, this intermediate category may have been discriminated from both /w/ and /j/ as a category goodness distinction, most likely within the Japanese /w/ category.

4. General discussion

The results of Experiment 1 revealed language-specific influences in the perception of English approximant contrasts by adult native speakers of American English and Japanese. Identification and discrimination performance were consistent with cross-language differences in both the phonemic status and the phonetic details of the three contrasts. Both language groups showed sharp category boundaries and high discrimination peaks on the /w-j/ series, which represents a phonemic contrast in both languages. However, there were group differences in the location of the /w-j/ category boundary. The Japanese identified more items as /w/, consistent with cross-language phonetic differences in degree of lip-rounding during production of /w/. On the /w-r/ series, the Japanese showed a more gradual crossover in identification functions and less accurate between-category discrimination than the Americans. In addition, a marginal shift in boundary location and discrimination peak suggested that Japanese categorized more intermediate tokens as /w/ than Americans did. This pattern is also consistent with cross-language differences in the phonetic realization of the /w-r/ contrast. Thus, while in abstract phonological terms /w/ *vs.* /r/ is a distinctive contrast in Japanese, the phonetic differences across languages led to distinctly different patterns of perception of the synthetic /w-r/ stimuli. As for /r-l/, the Inexperienced Japanese showed much less consistent identification functions and markedly poorer discrimination than the Americans. However, there was no significant shift in boundary location relative to Americans, in keeping with earlier reports (MacKain *et al.*, 1981; Miyawaki *et al.*, 1975). This group difference is compatible with the fact that /r-l/ is a phonemic distinction only in English, and that neither segment is phonetically similar to the Japanese /r/.

This pattern of cross-language differences supports predictions based on the perceptual assimilation model proposed by Best and colleagues (Best, 1992; Best *et*

al., 1988) to explain variations in the difficulty of discriminating non-native segmental contrasts. Specifically, Japanese listeners were expected to assimilate the English /w-j/ contrast as a two category contrast. The pattern of Japanese listeners' sharp category boundary and high discrimination performance on the /w-j/ series was consistent with this prediction. English /w-r/ was expected to be assimilated to Japanese as a contrast involving a category goodness difference, with /r/ most likely being assimilated as a "poor" exemplar of Japanese /w/. Japanese listeners' more gradually sloping identification function and lower discrimination peak for the /w-r/ series were compatible with this prediction. Finally, English /r-l/ was expected to be assimilated to a single category by Japanese, with both phones representing poor exemplars of either the Japanese /w/ or, less likely, of their tapped /r/. Once again, the more poorly defined category boundary and lower discrimination performance of the Japanese listeners were consistent with this prediction.

The present study extended the model of perceptual assimilation from simple predictions about discriminability of non-native segmental contrasts to two measures of how non-native segments are actually categorized by listeners. The location of the category boundary differed between the two groups, consistent with the articulatory-phonetic (and acoustic-phonetic) differences between the American English and Japanese /w-j/ contrast. Specifically, the Japanese perceived more tokens as /w/ than did the Americans, in keeping with observations that Japanese /w/ is more similar to /j/ acoustically and articulatorily than is English /w/. The stimulus items in the /w-j/ series that were identified as /w/ by Japanese but as /j/ by Americans in Experiment 1 were just those items perceived as /l/-like by Americans when they were given three-way choice (/w-l-j/) in Experiment 2. Language-specific differences in the phonetic details of the phoneme contrast "shared" by the two languages resulted in a divergence between language groups in the location but not the steepness of the /w-j/ category boundaries across Experiments 1 and 2, which supports the notion that the Japanese listeners assimilated the non-native segments to the familiar categories of their native phonological system. This language-specific boundary shift extends Lisker & Abramson's (1970) classic findings on cross-language differences in the voice-onset-time boundary for stop consonants to a place-of-articulation distinction for approximants. Moreover, the cross-language differences in identification and discrimination of /w-r/ (and /r-l/) are quite consistent with differences in the phonemic status and phonetic details of those contrasts with respect to the two languages.

The results of this study are also relevant to Flege's account of cross-language differences in speech perception. According to his Speech Learning Model (Flege, 1988, 1990) adult learners perceive phones of the L2 on the basis of their "phonetic similarity" to native language (L1) categories. Highly dissimilar phones (referred to as New phones) are initially difficult to categorize perceptually, but with L2 experience, learners form distinct L2 phonetic representations of these categories, which leads to improvement in both their perception and production. Phones which are identical to or highly similar to native phones (Identical phones) are easily perceived even by beginning L2 learners, because they "fit" L1 categories. Phones which are similar to but not identical with L1 categories ("Similar" phones) are the most problematic for L2 learners. They continue to classify Similar phones according to L1 categories even after considerable experience, which leads to continued "accented" production and difficulties perceiving that the L2 phones

differ from those of L1. Thus, Flege's model assumes that L2 phones are equated with L1 phonemes in a dichotomous, all-or-none fashion; i.e., they are either fully equated with an L1 phone or fail to be equated to an L2 phone. By comparison, the perceptual assimilation model (Best, 1992) instead assumes that listeners can perceive variations in the goodness of fit of an L2 phone to an L1 phoneme category. The latter assumption is compatible with findings that listeners are sensitive to the category goodness of stimulus variations within a given native category (e.g., Grieser & Kuhl, 1989; Miller & Volaitis, 1989). Also note that Flege's model was developed to address perceived similarities between individual L2 phones and individual L1 phoneme categories, whereas the perceptual assimilation model was developed to address the perception of L2 *contrasts*.

If we extend the Flege model to perception of non-native contrasts between phones, the results of Experiment 1 are partially consistent with that model. According to Flege's classification scheme, English /j/ is Identical, /w/ is Similar, and /r/ and /l/ are New phones for Japanese learners of English. Both inexperienced and experienced (re. spoken English) Japanese would thus classify stimuli of the /w/-/j/ contrast according to two Japanese categories, resulting in good identification and discrimination. His model would also predict a shift in the category boundary (relative to Americans), reflecting differences between the Japanese and English /w/. The results of experiment 1 are consistent with both expectations. For the /r-l/ series, inexperienced Japanese would be expected to have considerable difficulty, but experienced Japanese would show improved perception, reflecting the establishment of new phonetic categories. This was indeed the case in Experiment 1. In addition, the fact that the category boundary for experienced Japanese was not different from the Americans' supports the prediction that they had established new L2 categories. However, predictions for the /w-r/ series are somewhat more difficult to generate from Flege's model. The model should predict good identification and discrimination of these stimuli by experienced Japanese, who should have formed a New L2 category for /r/ to contrast with the Similar category of /w/. Their performance levels should therefore equal those of the Americans. However, it is less clear how inexperienced Japanese should perform with /w-r/. Although they would be predicted to identify /w/ well, and /r/ poorly, their discrimination performance is more difficult to predict. Should their performance be poor because they have difficulty with the /r/ that has not yet been established as a New L2 category, or should their performance be moderately good because they perceive /w/ as Similar and recognize that /r/ is different from /w/? In either case, we might expect, nonetheless, that discrimination performance would be lower for inexperienced Japanese than for Americans or for Japanese who are more experienced with spoken English. The shift in discrimination peak for the experienced Japanese toward the location of the American boundary in Experiment 1 suggests that those subjects may indeed have established a New /r/ category, which contrasts with the Similar /w/ category. Note, however, that the overall level of discrimination performance did not differ significantly among inexperienced Japanese, experienced Japanese, and Americans, as would be predicted from Flege's model.

Flege's model might also appear to address the existence of the intermediate category in the /w-j/ series, even for Japanese listeners, i.e., they may have begun to form a new /l/ category as a result of English experience. However, two

observations are at odds with this possibility. First, there was no difference on that contrast between the Inexperienced Japanese, who had had very little experience with spoken American English at the time of testing, and the Experienced Japanese. Both groups provided equally strong evidence of perceiving the intermediate category in the /w-j/ series; the intermediate category in the double-peaked discrimination functions was no less clear for the Inexperienced Japanese than for the Experienced Japanese, or in fact for the Americans. Second, if even the Inexperienced Japanese were truly developing a new phonetic category on the basis of their limited English exposure, then we would expect this /l/ category to emerge in their responses to the /r-l/ series as well. Such was not the case.

Flege's notion that L2 experience may lead to the formation of new phonetic categories is not incompatible with Best's perceptual assimilation model. The assumption that experience with spoken L2 may lead to a reorganization of perceptual assimilation of non-native phones, in fact, motivated the comparison between the Japanese subgroups differing in English conversation training and experience. The assimilation model assumes that listeners are sensitive to degrees of similarity and dissimilarity between the non-native and native phones. This is most obvious when there are category goodness differences in assimilation, or when the non-native phones are non-assimilable. Indeed, adult L2 learners should be expected to form new phonetic categories most readily for L2 phones perceived as discrepant exemplars of a native category, i.e., for the non-prototypical member of a contrast that is assimilated as a category goodness difference from a native phoneme. If no discrepancies are perceived between the L2 and L1 phone—that is, for the L2 phone that is perceived as a good exemplar of the native phoneme—it should be quite difficult for the L2 learner to form a new category. Conversely, if the L2 phone is so *dis*similar from L1 phonemes that it cannot readily be related to any L1 category, we may expect the L2 learner to have some difficulty forming a new phonetic category, because a clear *contrast* between a specific familiar phoneme and an unfamiliar phone may be particularly informative to the learner.

The one unexpected finding—that listeners from both language groups apparently discriminated a third, intermediate phonetic category between the two endpoint categories of the /w-j/ series—is consistent with the above suggestion. Experiment 2 with a new group of American listeners verified that this third category was highly identifiable as /l/ (although it remains to be determined whether Japanese listeners at either level of English experience would reliably label those items as "L"). Although the Japanese language does not employ an /l/ phoneme, even the Inexperienced Japanese clearly distinguished a third phonetic category for the /w/ and /j/, according to the two marked peaks in their /w-j/ discrimination function, which was virtually identical to the discrimination functions of the two groups of Americans. This observation, together with the Inexperienced Japanese listeners' better discrimination performance on /w-r/ than on /r-l/, suggests the possibility that adults' recognition of the phonetic properties of a non-native segment might be aided by direct comparison between exemplars of that segment presented in context with exemplars of the most similar (in articulatory-phonetic or acoustic-phonetic terms) native phoneme. That is, perceptual learning about the novel L2 segment may benefit from contextual comparisons which exemplify differences between the native phoneme and the non-native phone that is perceived as a poorer exemplar of that familiar category. In the present context, Japanese listeners' recognition of a

third category in the /w-j/ series, which was identified as /l/ by the Americans in Experiment 2, apparently benefited from its contrast to the flanking categories of Japanese /w/ and /j/, i.e., the intermediate, non-native category constituted a noticeably poor fit to one or both of the familiar Japanese categories. While this observation is consistent with Flege's (1988, 1990) claim about the importance of similarity versus "newness" of non-native phones to the degree of perceptual adjustments to L2 learning, it is also compatible with the perceptual assimilation hypothesis that category goodness differences are relatively discriminable as a difference between the native category "ideal" and less-good exemplars. Further research is obviously needed to determine whether presenting a non-native phone in juxtaposition to the most similar native phoneme contrast may actually improve perception of the new category.

In either event, the data presented here are generally consistent with the suggestion that language-specific attunement of phonetic perception may remain somewhat malleable even in adulthood (see also Flege, 1988; MacKain *et al.* 1981; Pisoni *et al.*, 1982; Strange & Dittmann, 1984; Tees & Werker, 1984; Werker & Tees, 1984). The subgroup of Japanese listeners who had had more intensive conservation experience with American English speakers showed greater similarities to the Americans than did the Inexperienced Japanese in their performance on all three stimulus series. Thus, English conversation experience may have shifted those Japanese listeners' categorization and discrimination toward the phonemic and phonetic properties of the approximant contrasts employed in American English. Note, however, that the performance of the Experienced Japanese was not identical to the Americans', instead falling intermediate between the latter group and the Inexperienced Japanese (see also Yamada & Tohkura, 1991).

Further research is needed to determine which factors may influence adults' perceptual adjustments to the phonemic and phonetic properties of L2 segmental contrasts, and to what extent there may be limitations on such L2 influences in adulthood. It is important to recognize that we had no control over, or access to, the factors that led to the group differences in English conversation experience. For example, in our Japanese subgroups, level of English conversation experience may have been affected by individual differences in phonetic ability (recall the categorical /r-l/ performance of the Inexperienced Japanese subject M.K.: MacKain *et al.*, 1981), by differences in the necessity of speaking English, by differences in motivation to use English "like a native", and/or by differences in the nature of exposure to English (e.g., traditional classroom *vs*. immersion program), in addition to duration and intensity of exposure to spoken English. Another factor that appears to have strong impact on an adult's ability to perceive a given non-native contrast is whether the individual had any substantive exposure during early childhood to languages using that contrasts (e.g., Flege, 1988; Tees & Werker, 1984).

Although we cannot verify that the Japanese subgroup difference we found was due to differences in L2 experience in adulthood, rather than to earlier-occurring factors, several observations suggest the likelihood that the relevant experience with spoken L2 was limited to adulthood. Three of the Experienced Japanese had come to live in the U.S. as adults, the fourth at 19 years, all past the presumed "critical period" for language-learning which ends at puberty. All had begun intensive English conversation training either after their arrival in the U.S. or less than a year

before they left Japan. Moreover, while most Japanese are formally taught English in school beginning at age 12 years or earlier, the instructors are typically native Japanese rather than English speakers, and the emphasis is on reading/writing and not on speaking/hearing (Mochizuki, 1981; Yamada & Tohkura, 1991). Nonetheless, further research is needed to clarify the contribution of various factors to subgroup differences in perception of L2 contrasts, including studies of longitudinal changes within a given group of listeners.

This work was supported by NIH grants NS-05877 and DC-00403 to the first author and grant HD-01994 to Haskins Laboratories, as well as DC00323 to Winifred Strange, which supported her during the preparation of this manuscript, and NIMH grant MH-21153 to James Jenkins, which supported the second author while she was a visiting scholar at Haskins Laboratories. We gratefully acknowledge Leon Seraphim for his help in establishing contact with our Japanese subjects; Arlene Antilla, Susan Koski, and Marshall Gladstone for their help with data collection; and James Jenkins for help with final data analyses. We also thank Hajime Hirose and Arthur Abramson for helpful discussions about the phonetic properties of Japanese approximants, as well as James Flege, Virginia Mann, and an anonymous reviewer for comments on an earlier draft of this manuscript.

Reprint requests should be directed to the first author at Haskins Laboratories, 270 Crown Street, New Haven CT 06511.

References

Best, C. T. (1992) The emergence of language-specific phonemic influences in infant speech perception. In *The transition from speech sounds to spoken words: the development of speech perception* (H. C. Nusbaum & J. Goodman, editors). Cambridge, MA: MIT Press. (In press.)

Best, C. T., McRoberts, G. W. & Sithole, N. N. (1988) The phonological basis of perceptual loss for non-native contrasts: maintenance of discrimination among Zulu clicks by English-speaking adults and infants, *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 345–360.

Best, C. T., Morrongiello, B. & Robson, R. (1981) Perceptual equivalence of acoustic cues in speech and nonspeech perception, *Perception and Psychophysics*, **29**, 191–211.

Bloch, B. (1950) Studies in colloquial Japanese IV: Phonemics, *Language*, **26**, 86–125.

Browman, C. P. & Goldstein, L. (1986) Towards an articulatory phonology, *Phonology Yearbook*, **3**, 219–252.

Browman, C. P. & Goldstein, L. (1989) Articulatory gestures as phonological units, *Phonology*, **6**, 2.

Carney, A. E., Widin, G. P. & Viemeister, N. F. (1977) Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, **62**, 961–970.

Flege, J. E. (1988) The production and perception of speech sounds in a foreign language. In *Human communication and its disorders: a review 1988* (H. Winitz, editor). Norwood, NJ: Ablex.

Flege, J. E. (1990) Perception and production: the relevance of phonetic input to L2 phonological learning. In *Crosscurrents in second language acquisition and linguistic theories* (C. Ferguson & T. Huebner, editors). Philadelphia: John Benjamins.

Flege, J. E. & Bohn, O.-S. (1989) The perception of English vowels by native Spanish speakers, *Journal of the Acoustical Society of America*, **85** (Suppl.), S85(A).

Gillette, S. (1980) Contextual variation in the perception of L and R by Japanese and Korean speakers. *Minnesota Papers in Linguistics and the Philosophy of Language*, **6**, 59–72.

Goldstein, L. & Browman, C. P. (1986) Representation of voicing contrasts using articulatory gestures. *Journal of Phonetics*, **14**, 339–342.

Goto, H. (1971) Auditory perception by normal Japanese adults of the sounds "L" and "R", *Neuropsychologia*, **9**, 317–323.

Grieser, D. L. & Kuhl, P. K. (1989). The categorization of speech by infants: support for speech sound prototypes, *Developmental Psychology*, **25**, 577–588.

Kasuya, H., Takeuchi, S., Sato & Kido, K. (1982) Articulatory parameters for the perception of bilabials, *Phonetica*, **39**, 61–72.

Lisker, L. (1957) Minimal cues for separating /w, j, r, l/ in intervocalic position, *Word*, **13**,256–267.

Lisker, L. & Abramson, A. S. (1970) The voicing dimension: some experiments on comparative phonetics. In *Proceedings of the 6th international congress of phonetic sciences*. Prague: Academia.
MacKain, K. S., Best, C. T. & Strange, W. (1981) Categorical perception of English /r/ and /l/ by Japanese bilinguals, *Applied Psycholinguistics*, **2**, 369–390.
Miller, J. L. & Volaitis, L. E. (1989) Effect of speaking rate on the perceptual structure of a phonetic category, *Perception and Psychophysics*, **46**, 505–512.
Miyawaki, K. (1973) A study of lingual articulation by use of dynamic palatography. Unpublished masters thesis, Department of Linguistics, University of Tokyo.
Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J. & Fujimura, O. (1975) An effect of linguistic experience: the discrimination of [r] and [l] by native speakers of Japanese and English, *Perception and Psychophysics*, **18**, 331–340.
Mochizuki, M. (1981) The identification of /r/ and /l/ in natural and synthesized speech, *Journal of Phonetics*, **9**, 283–303.
O'Connor, J. D., Gerstman, L. J., Liberman, A. M., Delattre, P. C. & Cooper, F. S. (1957) Acoustic cues for the perception of initial /w, j, r, l/ in English, *Word*, **13**, 24–43.
Pisoni, D. B., Aslin, R. N., Perey, A. J. & Hennessy, B. L. (1982) Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants, *Journal of Experimental Psychology: Human Perception and Performance*, **8**, 297–314.
Polka, L. (1991) Cross-language speech perception in adults: phonemic, phonetic, and acoustic contributions, *Journal of the Acoustical Society of America*, **89**, 2961–2977.
Pollack, I. & Pisoni, D. B. (1971) On the comparison between identification and discrimination tests in speech perception, *Psychonomic Science*, **24**, 299–300.
Price, P. J. (1981) A cross-linguistic study of flaps in Japanese and in American English. Unpublished Ph.D. thesis, University of Pennsylvania.
Pruitt, J. S., Strange, W., Polka, L. & Aguilar, M. C. (1990) Effects of category knowledge and syllable truncation during auditory training on American's discrimination of Hindi retroflex-dental contrasts, *Journal of the Acoustical Society of America*, 87 (Suppl.), S72(A).
Sheldon, A. & Strange, W. (1982) The acquisition of /r/ and /l/ by Japanese learners of English: evidence that speech production can precede speech perception, *Applied Psycholinguistics*, **3**, 243–261.
Strange, W. & Dittmann, S. (1984) Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English, *Perception and Psychophysics*, **36**, 131–145.
Tees, R. C. & Werker, J. F. (1984) Perceptual flexibility: maintenance or recovery of the ability to discriminate non-native speech sounds, *Canadian Journal of Psychology*, **38**, 579–590.
Trehub, S. E. (1976) The discrimination of foreign speech contrasts by adults and infants, *Child Development*, **47**, 466–472.
Vance, T. J. (1987) *An introduction to Japanese phonology*. Albany, NY: State University of New York Press.
Werker, J. & Logan, J. (1985) Cross-language evidence for three factors in speech perception, *Perception and Psychophysics*, **37**, 35–44.
Werker, J. F. & Tees, R. C. (1984) Phonemic and phonetic factors in adult cross-language speech perception, *Journal of the Acoustical Society of America*, **75**, 1866–1878.
Yamada, R. A. & Tohkura, Y. (1991) Age effects on acquisition of non-native phonemes: Perception of English /r/ and /l/ for native speakers of Japanese. In *Proceedings of the 12th international congress of phonetic sciences*, vol. 4, pp. 450–453.