Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Phonetics 32 (2004) 233–250
Journal of Phonetics
www.elsevier.com/locate/phonetics

Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/

Katsura Aoyama[a,*], James Emil Flege[b], Susan G. Guion[c], Reiko Akahane-Yamada[d], Tsuneo Yamada[e]

[a] *Department of Speech-Language & Hearing Sciences, Texas Tech University Health Sciences Center, 3601 4th St., Stop 6073, Lubbock, TX 79430 6073, USA*
[b] *Division of Speech and Hearing Sciences, School of Health Related Professions, University of Alabama at Birmingham, USA*
[c] *Department of Linguistics, University of Oregon, USA*
[d] *ATR Human Information Processing Research Labs, Japan*
[e] *National Institute of Multimedia Education, Japan*

Received 8 February 2002; received in revised form 30 April 2003; accepted 23 May 2003

Abstract

Previous research has demonstrated that English /r/ is perceptually more dissimilar from Japanese /r/ than English /l/ is for native Japanese (NJ) speakers. It has been proposed by the Speech Learning Model that the more distant an L2 sound (phonetic segment) is from the closest L1 speech sound, the more learnable the L2 sound will be (in: W. Strange (Ed.), Speech Perception and Linguistic Experience: Issues in Cross-language Research, York Press, Timonium, MD, 1995, p. 233). This hypothesis was evaluated in this study by investigating whether NJ speakers will have more success acquiring English /r/ than /l/. A longitudinal study examined the perception (Experiment 1) and production (Experiment 2) of English /l/, /r/, and/w/ by NJ adults and children who were living in the US at the time of testing. The results suggested that there was greater improvement for English /r/ than English /l/ among the NJ children. The NJ children's discrimination of /l/-/r/ and /r/-/w/ was significantly better at the second testing (T2) than 1 year earlier (T1). The NJ children also showed greater improvement from T1 to T2 in producing /r/ than /l/. The results are taken as support for a hypothesis of the Speech Learning Model (in: W. Strange (Ed.), Speech Perception and Linguistic Experience: Issues in Cross-language Research, York Press, Timonium, MD, 1995, p. 233) that degree of perceived phonetic dissimilarity influences L2 learners' success in acquiring L2 phonetic segments.

Keywords: Phonetic dissimilarity; /l/,/r/; Second language; Speech; Japanese; English; Production; Perception

*Corresponding author. Tel.: +1-806-743-5660x239; fax: +1-806-743-5670.
E-mail address: katsura.aoyama@ttuhsc.edu (K. Aoyama).

0095-4470/03/$ - see front matter
doi:10.1016/S0095-4470(03)00036-6

1. Introduction

Theories of cross-language and second-language (L2) speech perception (Best, 1995; Kuhl & Iverson, 1995) posit that the perceived relation between phonetic segments encountered in an L2 (or foreign language) play a key role in how those phonetic segments will be discriminated. Previous research has shown that the perceived phonetic relationship between sounds in L2 learners' first language (L1) and sounds in an L2 is closely related to how well the learner discriminates a particular pair of L2 contrastive phonetic segments (or "sounds" for short) (Best & Strange, 1992; Guion, Flege, Akahane-Yamada, & Pruitt, 2000; Best, McRoberts, & Goodell, 2001). When two L2 sounds are identified as being instances of two different L1 categories, learners can usually discriminate the two with relative ease, whereas discrimination may be difficult if two L2 sounds are both identified as being instances of a single L1 category. Differences in the discriminability of L2 speech sounds, in turn, may be related to how those speech sounds are produced.

Japanese speakers' difficulty in discriminating English /l/ and /r/ is well known (e.g., Goto, 1971; Strange & Dittmann, 1984; Miyawaki et al., 1975; Takagi & Mann, 1995). It is often attributed to the perceptual assimilation of English [l] and [ɹ] by a single Japanese sound, /r/, at least in syllable-initial position (Takagi, 1993; Komaki, Akahane-Yamada, & Choi, 1999; Best & Strange, 1992). However, the Japanese /r/ is phonetically an apico-alveolar tap [ɾ] (Price, 1981; Vance, 1987) that is quite different from American English [l] and [ɹ], which are lateral and central approximants, respectively (Ladefoged, 2001). In fact, Japanese [ɾ] is phonetically more similar to flapped /t/ and /d/ in American English (Price, 1981; Vance, 1987) than it is to either [l] or [ɹ].

Despite the articulatory difference between Japanese [ɾ] and the English liquids [l] and [ɹ], Japanese speakers seem to perceptually assimilate both English liquids to Japanese /r/. Best and Strange (1992) suggested that both English [l] and [ɹ] will be heard as poor exemplars of a single Japanese consonant, either /w/ or /r/. Takagi (1993) reported that syllable-initial English [l] and [ɹ] were usually identified as Japanese /r/ by Japanese adults. Similarly, the Japanese adults tested by Komaki et al. (1999) and Guion et al. (2000) identified English [lɑ] and [ɹɑ] as either Japanese /ra/ or /ura/ (i.e., [ɾa] preceded by a back unrounded vowel [ɯ]). The results all suggested a two-to-one cross-language mapping pattern.

Although English [l] and [ɹ] tokens have usually been found to be perceptually assimilated by Japanese /r/, the degree of perceived phonetic dissimilarity between English [l] and [ɹ] and Japanese [ɾ] may differ. Specifically, English [ɹ] may be more dissimilar phonetically from Japanese [ɾ] than English [l] is, despite the fact that the Japanese [ɾ] is often designated "R". In a study by Takagi (1993), Japanese speakers rated English [l] and [ɹ] tokens in terms of goodness of fit as an instance of Japanese [ɾ] using a scale ranging from 0 ("not like [ɾ] at all") to 7 ("perfect [ɾ]"). The Japanese speakers gave lower ratings to [ɹ] than [l] when tokens of these sounds occurred in syllable-initial, initial cluster and intervocalic positions. Komaki et al. (1999) also reported lower goodness-of-fit ratings for English [ɹ] than [l] with respect to Japanese [ɾ]. In addition, Japanese speakers identified synthesized [ɹ] tokens as Japanese [ɾ] less often than they identified [l] tokens as Japanese [ɾ] (71% vs. 90%) in a study by Iverson et al. (2001). The goodness ratings as Japanese [ɾ] were also lower for [ɹ] than for [l] tokens in the Iverson et al. study.

According to the contrastive analysis hypothesis (CAH) (e.g., Lado, 1957), cross-language differences result in learning difficulty. This proposal leads to the expectation that learners of an L2 will have more difficulty learning an L2 sound that has no equivalent in the L1 than in learning an L2 speech sound that resembles (but is not physically identical to) an L1 sound. Flege (1987) proposed that the CAH might hold true, but only in early stages of L2 learning. He proposed that for individuals who had become familiar with the sound system of an L2, large phonetic differences between an L2 speech sound the closest sound in the L1 inventory will promote phonetic learning. According to the Speech Learning Model (SLM) (Flege, 1995), the greater the perceived phonetic dissimilarity between an L2 speech sound and the closest L1 sound is, the more likely learners will be to discern the difference between the L1 and L2 sounds and show measurable progress in production and/or perception. Taken together with evidence (see above) that English [ɹ] is phonetically more dissimilar from Japanese [ɾ] than English [l], the SLM generated the prediction that Japanese learners of English will have more success in learning to produce and perceive English [ɹ] than English [l].

Previous research has supported the prediction that Japanese learners of English will exhibit more success in learning English [ɹ] than [l]. Sheldon and Strange (1982) found an asymmetry in Japanese adults' identification of [l] and [ɹ], with initial tokens of [ɹ] being identified correctly more often than [l] tokens were (96% vs. 82%). Similarly, in a study by Flege, Takagi, and Mann (1996), Japanese adults who had lived in the United States (US) for 21 years on average correctly identified [ɹ] more often than [l] in minimal pairs of equal familiarity (96% and 81%, respectively), as did Japanese adults who had lived in the US for just 2 years ([ɹ] 87%, [l] 53%). Finally, in a training study by Bradlow, Pisoni, Akahane-Yamada, and Tohkura (1997), Japanese speakers identified English [ɹ] more accurately than [l] in perception tasks, and the same Japanese participants' productions of English [ɹ] were identified as intended more often than their [l] productions were.

The purpose of this study was to evaluate the SLM hypothesis regarding the effect of perceived cross-language phonetic dissimilarity on the learning of L2 phonetic segments. The SLM predicts that if a difference exists in the perceived phonetic dissimilarity of two L2 speech sounds to the closest L1 speech sound, then the L2 speech sound which is more dissimilar will ultimately be produced and perceived more accurately. The learning of English [l] and [ɹ] by native Japanese (NJ) adults and children were examined in Experiment 1 (perception) and Experiment 2 (production). Based on the findings of previous studies that suggested English [ɹ] is phonetically more dissimilar from Japanese [ɾ] than English [l] is, the SLM predicts that the NJ speakers will show greater learning for [ɹ] than [l].

An important aspect of the current study is that NJ children as well as NJ adults participated. The SLM (e.g., Flege, 1995, 1999, 2002) makes the prediction that the earlier in life that L2 learning begins, the more likely L2 learners will be to establish new phonetic categories. This prediction is based on the hypothesis that as phonetic units making up the L1 sound system develop, they become more powerful attractors of L2 speech sounds, thereby inhibiting the development of new phonetic categories for L2 speech sounds. Most previous studies carried out to evaluate the SLM have examined adult participants only, not children actually in the process of acquiring an L2. A secondary purpose of this study was to investigate whether different learning patterns would be observed to exist between adult and child learners of English as an L2.

Table 1
Characteristics of the native English (NE) and Japanese (NJ) participants in Experiments 1 and 2

	Gender	Mean age at T1	Mean LOR		Mean PPVT scores	
			T1	T2	T1	T2
NE adults	Seven males	40.3	—	—	—	—
	Nine females	(4.7)				
NJ adults	Eight males	39.9	0.5	1.6	6.8	9.1
	Eight females	(3.8)	(0.2)	(0.3)	(2.9)	(3.1)
NE children	Ten males	10.6	—	—	—	—
	Six females	(2.1)				
NJ children	Nine males	9.9	0.4	1.6	2.8	6.0
	Seven females	(2.4)	(0.2)	(0.3)	(1.6)	(1.4)

LOR, length of residence in the United States in years; age, chronological age in years. Standard deviations in parentheses. PPVT, Peabody Picture Vocabulary Test.

2. Experiment 1

The hypothesis generated by SLM was tested here by examining the perception of English consonants by NJ adults and children. Age-matched native English-speaking adults and children participated as controls. A categorial discrimination test was employed to examine the following five contrasts: /l/-/r/,/r/-/w/,/b/-/s/,/s/-/θ/, and /b/-/v/. These contrasts were chosen because /l/-/r/, /r/-/w/,/s/-/θ/, and /b/-/v/ were considered 'difficult' for Japanese L2 speakers of English (Best & Strange, 1992; Guion et al., 2000; Yoshida & Hirasaka, 1983). The /b/-/s/ contrast was included as a control.

2.1. Method

2.1.1. Participants

Sixteen NJ adults and 16 NJ children participated (see Table 1). Most of the NJ participants were living in Houston and Dallas, Texas, when tested. The NJ participants' average length of residence (LOR) in the US was 0.5 years at the first time of testing (T1); their average LOR was 1.6 years at the second testing (T2). The interval between T1 and T2 was much the same for the NJ adults (1.1 years) and the NJ children (1.2 years). Sixteen adults and 16 children who spoke English as their native language also participated. In both the NJ and the native English (NE) groups, the adult participants were the parents of the child participants. The NE participants, who were tested in Birmingham, Alabama, were also tested twice, one year apart. All participants were tested in a quiet room at their homes, or at the University of Alabama at Birmingham.[1]

The NJ participants planned to return to their home country after several years, and so were not immigrants. All of the NJ adults had studied English for at least 6 years in Japan whereas only

[1] Eighty-three NE and NJ adults and children (approximately 20 participants in each group) participated at T1. However, only 16 each of Japanese adults and children were available for testing at T2 because of relocations. Sixteen each of NE adults and children were retained in the study in order to have an equal cell size in all groups

one NJ child had studied English before his arrival in the US (this child had studied English for 2 years). The NJ children began attending English-speaking schools upon arriving in the US.

The Peabody Picture Vocabulary Test[2] (PPVT) was administered for the NJ participants to estimate their knowledge of the English lexicon at T1 and T2. As shown in Table 1, the NJ adults obtained age-equivalent scores of 6.8 and 9.1 years at T1 and T2, respectively. The NJ children obtained age-equivalent scores of 2.8 and 6.0 years at T1 and T2. The PPVT scores were submitted to a (2) Age × (2) Time ANOVA. It yielded a significant main effect of time ($F(1,30)=96.54$, $p<0.01$) and age ($F(1,30)=20.34$, $p<0.01$), and a non-significant two-way interaction ($F(1,30)=2.69$, $p>0.1$). This finding suggested that both the NJ adults and children were actively learning English.

2.1.2. Stimuli

An adult male native speaker of American English produced tokens of [lɑ, ɹɑ, wɑ, bɑ, sɑ, θɑ, vɑ] in a carrier phrase, 'Then I saw /Cɑ/ there'. The /Cɑ/ tokens were then extracted from the carrier sentence. Five tokens for each consonant category were prepared. The recording was digitized at 22.05 kHz, and normalized to 50% of the dynamic range.

2.1.3. Procedure

A categorial discrimination test that has been used in previous studies of L2 speech learning (Flege, MacKay, & Meador, 1999; Guion et al., 2000) was used here because of its suitability as a test of category formation. Each consonant contrast of interest (/l/-/r/,/r/-/w/, /b/-/s/, /s/-/θ/, /b/-/v/) was tested by eight triadic change trials and eight triadic no-change trials with an inter-stimulus-interval of 0.5 s. The correct response to a change trial was a button ("1", "2", or "3") indicating the position of the odd item out, which occurred with equal frequency in all three possible serial positions. The correct response to no-change trials, which consisted of three physically different instances of a single category, was a fourth button marked "no". The change trials tested the participants' ability to distinguish consonants drawn from different phonetic categories. The no-change trials tested the ability to ignore audible but phonetically irrelevant within-category variation. Trials testing the five contrasts were presented in one randomized set, not in separate blocks.

The participants heard several stimuli over headphones to establish a comfortable volume level before the experiment began. The volume was fixed once selected. The participants were instructed, in English, that they were to decide if one consonant was different from the other two. If one consonant was different, they were instructed to click a button marked "1", "2", or "3", according to the serial position of the odd item out. They were instructed to click the fourth button ("no"), if they did not hear an odd item out. There was a practice session using /wɑ/ and /sɑ/ stimuli. The participants had to respond correctly to at least 9 out of the 10 practice items in order to advance to the actual experiment. When a participant did not reach this criterion, the 10 practice trials were repeated in a different order, until the criterion was met, up to four times. All participants reached the criterion.

[2] The Peabody Picture Vocabulary Test is an individually administered, norm-referenced, wide-range test. It is useful in testing pre-school children, screening for verbal ability, measuring English language proficiency, or evaluating non-readers and those with written-language problems (Dunn & Dunn, 1997).

2.1.4. Analysis

The dependent variable was A', which was calculated for each of the five contrasts for each participant. These scores were based on the proportion of "hits" (correctly selecting the odd item out in change trials) and "false alarms" (incorrectly saying that one member of a no-change trial was phonetically different from the other two members of the trial). The formula provided by Snodgrass, Levy-Berger, and Haydon (1985) was used. There were a maximum number of eight hits for each contrast. There were also a maximum eight false alarms for each contrast (four no-change trials for each of the two consonants). If the proportion of hits (H) equaled the proportion of false alarms (FA), then A' was set to 0.5. If H exceeded FA, then $A' = 0.5 + ((H-FA)*(1+H-FA))/((4*H)*(1-FA))$. If FA exceeded H, then $A' = 0.5-((FA-H)*(1+FA-H))/((4*FA)*(1-H))$. An A' score of 1.0 indicated perfect discrimination (i.e., correct responses to all eight change trials and all eight no-change trials), whereas a score of 0.5 or lower indicated insensitivity to a contrast.

2.2. Results

The two contrasts which included /l/ and /r/ (i.e., the /l/-/r/and /r/-/w/ contrasts) were of primary interest. Before turning our attention to those two contrasts, the results for the other three contrasts will be reported briefly.

The /b/-/s/ contrast served as a control. Given that the Japanese phonetic inventory includes both an /s/ and a /b/, it was likely that all of the NJ participants would generate two distinct phonetic codes on /b/-/s/ change trials. As expected, high /b/-/s/ scores (>0.90) were obtained for all four groups, including the two NJ groups. The scores for /b/-/s/ were submitted to a three-way ANOVA in which language (English vs. Japanese) and age (adult vs. child) served as between-subjects variables, and time (T1 vs. T2) served as a within-subjects variable. The three-way interaction, as well as all of the two-way interactions, were non-significant ($F(1,60) = 0.08$–2.55, $p > 0.1$). The main effect of age was non-significant ($F(1,60) = 0.03$, $p > 0.1$), as was the main effect of time ($F(1,60) = 0.39$, $p > 0.1$).

The main effect of language was significant ($F(1,60) = 8.23$, $p < 0.01$) indicating that the NJ participants' scores (*mean* 0.92) were lower than the NE participants' (*mean* 0.97). There are several possible explanations for this difference. First, it is likely that the task itself was more difficult for the NJ participants than for the NE participants because the instructions were administered entirely in English. Second, small phonetic differences in English and Japanese /s/ and /b/ might have contributed to errors. Finally, the NJ participants may have attended more to vowel duration than the NE participants did because of its importance in Japanese (see Strange, Akahane-Yamada, Kubo, Trent, & Nishi, 2001). The important point here is, however, that the NJ adults' and children's /b/-/s/ scores were very high both at T1 and T2, which indicated that they understood the instructions and were able to perform the experimental task well.

The NJ participants obtained lower scores on average for /s/-/θ/ (*means*: NJ adults T1 = 0.788, T2 = 0.855; NJ children T1 = 0.574, T2 = 0.656) and /b/-/v/ (*means*: NJ adults T1 = 0.806, T2 = 0.880; NJ children T1 = 0.845, T2 = 0.837) than they did for /b/-/s/. Neither of the three-way ANOVAs examining these contrasts yielded a significant effect of time for any group (/s/-/θ/ $F(1,15) = 0.1$–5.8, $p > 0.1$, /b/-/v/ $F(1,15) = 0.2$–4.1, $p > 0.1$), suggesting that the T2 scores did not differ from the T1 scores for any group.

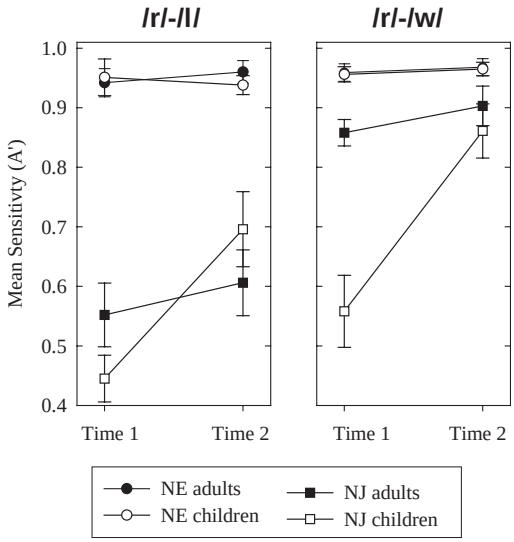

Fig. 1. Mean A' scores obtained in Experiment 1 for four groups for the /l/-/r/and /r/-/w/ contrasts. A' score of 1.0 indicates perfect discrimination, and a score of 0.5 or below indicates insensitivity to a contrast. The error bars represent standard errors.

The results for the two contrasts containing /r/ (i.e., /l/-/r/, /r/-/w/) are shown in Fig. 1. The scores for the NE adults and children were high for both contrasts. There is little evidence of change from T1 to T2 for either group. This is because the NE adults' and NE children's scores were near the ceiling at both T1 and T2. The amount of improvement from T1 to T2 observed for the NJ participants depended on contrast (/l/-/r/, /r/-/w/) and age. There was little difference between the two contrasts for the NE participants, but higher scores were obtained for /r/-/w/ than /l/-/r/ for the NJ participants. The NJ children's scores improved from T1 to T2, whereas the NJ adults' scores at T1 and T2 were similar.

The NJ participants' scores[3] for /l/-/r/ and /r/-/w/ were submitted to a three-way ANOVA in which age (adult vs. child) served as a between-subjects variable and contrast (/l/-/r/ vs. /r/-/w/) and time (T1 vs. T2) served as within-subjects variables. All three factors reached significance (age, $F(1,30)=4.5$, $p<0.05$; contrast, $F(1,30)=52.4$, $p<0.01$; time, $F(1,30)=19.4$, $p<0.01$). The age $\times$ contrast reached significance ($F(1,30)=7.1$, $p<0.05$), as did the age $\times$ time interaction ($F(1,30)=9.4$, $p<0.01$). However, the contrast $\times$ time interaction was non-significant ($F(1,30)=0.2$, $p>0.01$), as was the three-way interaction, $F(1,30)=0.4$, $p>0.01$).

The age $\times$ contrast interaction was explored through simple effects tests. The scores obtained for /l/-/r/ and /r/-/w/ at both times of testing were averaged. Both the NJ adults and the NJ children obtained higher scores for /r/-/w/ than for /l/-/r/ [adults: $F(1,15)=71.7$, $p<0.01$; children: $F(1,15)=7.9$, $p<0.05$]. The scores obtained for the NJ adults and children for /l/-/r/ did not differ significantly, ($F(1,30)=0.02$, $p>0.10$). However, the NJ adults obtained higher scores for /r/-/w/ than the NJ children did ($F(1,30)=14.8$, $p<0.01$).

[3]The NE participants' scores were not examined statistically because there was insufficient variation among their scores.

The age × time interaction was also explored through simple effects tests. The scores obtained for /l/-/r/ and /r/-/w/ at T1 were averaged, as were the scores obtained for /l/-/r/ and /r/-/w/ at T2. The NJ adults obtained significantly higher scores than the NJ children did at T1, ($F(1,30) = 19.2$, $p < 0.01$). However, scores obtained for the NJ adults and children did not differ significantly at T2 ($F(1,30) = 0.1$, $p > 0.10$). The scores obtained at T1 and T2 by the NJ adults did not differ significantly, ($F(1,15) = 1.0$, $p > 0.10$). However, the NJ children obtained significantly higher scores at T2 than at T1 ($F(1,15) = 26.5$, $p < 0.01$).

2.3. Discussion

These results showed that the NJ children but not the NJ adults discriminated /l/-/r/ and /r/-/w/ better at T2 than at T1. At T1, the NJ children's discrimination scores were lower than the NJ adults' for the /l/-/r/ and /r/-/w/ contrasts. However, the NJ children's scores improved from T1 to T2, while the NJ adults' scores did not. As a consequence, the adult-child differences among the NJ participants for the /l/-/r/ and /r/-/w/ contrasts disappeared at T2. As a consequence, the significant difference between the NJ adults and children that was evident at T1 for the /l/-/r/ and /r/-/w/ contrasts disappeared at T2. We infer that perceptual learning of [ɹ] took place because the NJ children showed a significant improvement on the discrimination of *both* /l/-/r/ and /r/-/w/.

It appears that the NJ adults had an initial advantage over the NJ children in discriminating /l/-/r/ and /r/-/w/. Similar findings were obtained for English-speaking adults and children who were learning Dutch as an L2 in the Netherlands. In a study by Snow and Hoefnagel-Höhle (1977, 1978), English-speaking adults who were learning Dutch had an initial advantage over the NE children in speaking, comprehension and discrimination. In the course of 1 year, however, the children's performance improved while the adults' leveled off.

The NJ adults examined in the present study may have had an initial advantage over the NJ children because they, but not the children, had studied English in school before arriving in the US. This is probably why the NJ adults' PPVT scores were higher than the children's at T1 (age equivalent scores of 6.8 years for the NJ adults, 2.8 years for the NJ children). The difference in PPVT scores between the NJ adults and children was smaller at T2 than at T1 (age equivalent norm of 9.1 years for the NJ adults, 6.0 years for the NJ children).

3. Experiment 2

The purpose of this experiment was to examine the production of English [l], [ɹ] and [w] by the NJ adults and children. The participants from Experiment 1 produced English words beginning with [l], [ɹ] and [w]. The data were collected twice (1 year apart) as in Experiment 1 in order to test for phonetic learning.

3.1. Method

3.1.1. Elicitation

Twenty-six frequently occurring English words were elicited from each participant. Each target word was elicited three times in the following manner. The participants wore a head-mounted

microphone (Shure model SM10A) connected to a DAT tape recorder (Sony model TCD-D8). First, the participants saw a picture on the screen of a laptop computer and heard the corresponding word via a loudspeaker. An equivalent word in Japanese was displayed in Japanese orthography along with the picture to reduce the NJ participants' uncertainty as to what word to say. After the 26 test words were elicited, the pictures were displayed a second and third time without an auditory model. (Only when the participant was not able to say a test word was the auditory model for the word played out.) The order of the 26 pictures was randomized differently for each participant and for each of the three elicitations of the test words.

The productions of the first and third productions were used in this experiment. The first productions of the 26 words will be referred to as 'cued' productions because an auditory model was provided. The participant's third productions will be referred to as 'non-cued' because these words were produced without an auditory model in most (>98%) instances.

3.1.2. Stimulus preparation

The participants' productions of *light*, *leaf*, *write*, *read*, *watch*, and *wing* were normalized for peak intensity (50% of the full scale) after being digitized at 22.05 kHz with 16-bit amplitude resolution. Two productions of each test word were examined for each participant, the first (cued) production and the third (non-cued) production. This yielded a total of 1536 tokens for analysis (4 groups × 16 participants × 6 words × 2 tokens × 2 times of testing).

The final consonants were deleted, and the initial consonants of interest ([l], [ɹ], [w]) and the following vowels were retained. In *light* and *write*, the second element of the diphthong /aɪ/was removed as well. Then the final 30 ms portion of the vowel was ramped off in amplitude. The purpose of the editing was to make the final consonant imperceptible so that judgments of the initial consonants were less likely to be affected by lexical identity.[4] The initial consonants were not edited. The edited stimuli typically sounded /lɑ/, /li/, /rɑ/, /ri/, /wɑ/ and /wɪ/.[5]

3.1.3. Listener judgments

Twelve native speakers of English (six males and six females, mean age = 26 years) auditorily evaluated the test words produced by the 64 NE and NJ speakers. None of the listeners was proficient in a language other than English. Seven listeners were from Alabama and five were from Texas.

The listeners were asked to identify the initial consonants in the CV stimuli. The 1536 stimuli described earlier were presented in four blocks. The adults' and children's productions were presented in separate counterbalanced blocks (2 blocks for adults, 2 blocks for children). Within adult and child blocks, the edited tokens of *light*, *write* and *watch* (all having a low vowel), and the

[4] It is uncertain whether the truncated syllables affected listener judgments in any way. /wɪ/is not a permissible string whereas all others are, and this fact might have influenced listener judgments. However, we do not believe that such an influence, if there was any, was substantial because all the initial consonants produced by the NE adults and children were correctly identified regardless of the following vowel.

[5] One token had an overlapping noise on the initial consonant that could not be edited, and was considered as missing data (one token of *light* produced by an NE child at T1). This *light* token was presented among the other stimuli, but responses to that stimulus were excluded when analysis was conducted.

leaf, *read* and *wing* tokens (all having a high front vowel) were presented in counterbalanced order. Each of these blocks contained 384 tokens (32 participants × 3 consonants × 2 repetitions × 2 times of testing).

As a preliminary, an NE speaker with phonetic training used an open-set classification procedure to identify all of the word-initial consonants. All consonants he heard were included among the response alternatives. The 12 listeners responded to each stimulus by clicking one of eight buttons shown on a computer screen: *r*, *l*, *w*, *d*, *dr*, *br*, *bl*, *ml*. These response alternatives were deemed all that were needed for the English-speaking listeners to accurately report what they heard.[6] Three keywords for each consonant (except for *ml*) were given above each key on the screen to help orient the listeners, and the experimenter demonstrated how [ml] might sound. The listeners had a chance to listen to each token again, but responses could not be changed once given. The tokens in each block were randomly presented to the listeners, and each listener received a unique randomized order. The inter-stimulus-interval was set to 1.2 s. Each block lasted approximately 20 min.

3.1.4. Analysis

The dependent variable calculated for each of the 1536 tokens was how many listeners heard each target consonant as intended (intelligibility score, maximum = 12). A total of 24 scores (3 words × 2 repetitions × 2 vowel contexts × 2 times of testing) were calculated for all but one NE child, for whom 23 scores were calculated due to missing data (*light* at T1). The scores obtained in the two vowel contexts (high vs. low vowel) were averaged because finding a difference in production accuracy as a function of the following vowel was not a focus of this experiment. The intelligibility scores for cued and non-cued productions were also averaged. This is because a preliminary ANOVA showed that the main effect of repetition (cued vs. non-cued words) was not significant ($F(1,60) = 0.29$, $p > 0.5$), and this factor did not interact significantly with any other factor ($p > 0.1$).

3.2. Results

Table 2 presents confusion matrices showing the percentage of times each response label was used to identify the NJ adults' and children's productions of [l], [ɹ] and [w]. The boldfaced values indicate the percentage of times that the production of each consonant was identified as intended. For the NJ adults, there was a small improvement from T1 to T2 for all three consonants: 3.9% for [l], 6.5% for [ɹ] and 0.5% for [w]. The NJ children showed more improvement than the adults did for all three consonants: 7.5% for [l], 35.1% for [ɹ], and 16.6% for [w]. The greatest improvement was for [ɹ].

The NE participants obtained high scores (adults 96.1% and 95.1% at T1 and T2, respectively; children 95.5% and 94.1%), indicating that the NE listeners understood the evaluation procedure. As expected, the NJ participants' scores were lower than the NE participants' scores were (NJ adults, 73.4% and 77.0% at T1 and T2, respectively, NJ children, 54.6% and 74.4%).

[6] For a few instances, the listeners nonetheless reported hearing sounds that were not among the alternatives (e.g., [gl]). They were instructed to choose the closest alternative possible among the ones available.

Table 2
Confusion matrices for the NJ speakers' productions in Experiment 2

	Listeners' judgements (%)							
	l	r	w	d	dr	br	bl	ml
NJ adults, Time 1								
[l]	**66.7**	20.8	3.0	6.0	0.4	0.3	2.9	0
[ɹ]	25.3	**59.5**	5.7	2.1	0.4	4.2	2.9	0
[w]	2.0	3.9	**93.9**	0	0	0.3	0	0
NJ adults, Time 2								
[l]	**70.6**	15.9	2.3	5.7	0.1	0.9	4.2	0.3
[ɹ]	19.4	**66.0**	6.9	2.3	0.4	2.5	2.0	0.5
[w]	1.8	2.9	**94.4**	0.5	0.1	0.1	0.1	0
NJ children, Time 1								
[l]	**59.4**	17.7	1.8	13.7	0.9	2.5	4.0	0
[ɹ]	42.8	**28.8**	11.6	10.5	0.9	2.6	2.5	0.3
[w]	16.3	5.7	**75.8**	0	0.4	0.4	1.3	0.1
NJ children, Time 2								
[l]	**66.9**	13.4	8.1	7.2	0	1.3	2.6	0.5
[ɹ]	13.8	**63.9**	13.3	5.1	0	3.0	0.8	0.1
[w]	3.8	2.9	92.4	0.4	0	0.1	0.3	0.1

Boldfaced percentages indicate the productions identified as intended by the listeners.

The intelligibility scores obtained for the NJ participants were examined in a three-way ANOVA.[7] Age (adult vs. child) served as a between-subjects variable and consonant ([l], [ɹ] and [w]) and time (T1 vs. T2) served as within-subjects variables. This analysis indicated that the intelligibility scores obtained by the NJ adults and children (*means* 75.2% vs. 64.5%) differed significantly ($F(1,30) = 6.26$, $p < 0.05$). The scores obtained at T1 and T2 by the NJ participants (*means* 64.0% vs. 75.7%) also differed significantly ($F(1,30) = 19.58$, $p < 0.01$), as did the scores obtained for the three consonants ([l] 65.9%, [ɹ] 54.6%, [w] 89.1%) ($F(2,60) = 17.88$, $p < 0.01$). In addition to these significant main effects, two of the three two-way interactions reached significance (consonant × time, $F(2,60) = 4.50$, $p < 0.05$; age × time, $F(1,30) = 9.29$, $p < 0.01$).

The three-way interaction (consonant × time × age) was marginally significant ($F(2,60) = 2.73$, $p = 0.07$). The implication of this finding, namely that equivalent changes from T1 to T2 were observed for the NJ adults and the NJ children, may be somewhat misleading. We decided to explore the marginally significant three-way interaction through simple effect tests in order to determine whether there was in fact a difference between the NJ adults and children. That is, we examined the simple effect of time (T1 vs. T2) for all six group × consonant combinations, the simple effect of consonant for each of the time × age combinations, and the simple effect of age for all six consonants × time combinations (see Fig. 2).

[7] As in Experiment 1, the NE participants' scores were not examined statistically because there was insufficient variation among their scores.

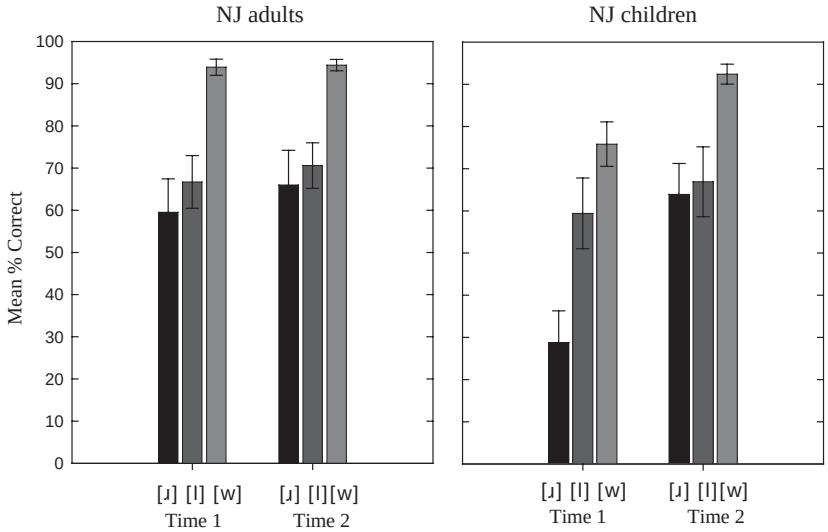

Fig. 2. Mean intelligibility scores by consonants for the NJ adults and children in Experiment 2. The error bars represent standard errors.

These tests revealed that the NJ children's scores improved significantly from T1 to T2 for [ɹ] ($F(1,15) = 18.88$, $p < 0.01$) and [w] ($F(1,15) = 11.94$, $p < 0.05$) but not [l] ($F(1,15) = 0.77$, $p > 0.1$). The NJ adults' scores did not differ significantly between T1 and T2 for any consonant ($F(1,15) = 0.07$–4.06, $p > 0.1$). In addition, the simple effect of consonant was significant for both the NJ adults and children at T1 and T2 ($F(2,30) = 7.03$–9.12, $p < 0.05$). Tukey's post-hoc tests indicated that the NJ children obtained significantly lower scores at T1 for [ɹ] than for either [l] or [w]. However, there was no significant difference between the NJ children's [l] and [ɹ] scores at T2, which were significantly lower than their scores for [w]. This indicated that the NJ children's productions of [l] were initially more intelligible than their productions of [ɹ], but their productions of [ɹ] improved from T1 to T2 so that there was no significant difference between [l] and [ɹ] at T2. The NJ adults' scores for [w] were higher than their scores for [l] and [ɹ], which did not differ at either T1 or T2. The simple effect of age was significant for [ɹ] ($F(1,30) = 7.93$, $p = 0.05$) and [w] ($F(1,30) = 10.42$, $p < 0.01$) at T1, whereas it was non-significant for [l] at both T1 and T2, and non-significant for all three consonants at T2 ($F(1,30) = 0.04$–0.52, $p > 0.1$).

In summary, the NJ adults obtained higher scores than the NJ children did for [ɹ] and [w] at T1. The NJ adults' and children's scores for [ɹ] and [w] did not differ significantly at T2, and their scores for [l] did not differ significantly at either T1 or T2. In other words, the intelligibility scores improved significantly from T1 to T2 only for [ɹ] and [w], and then only for the NJ children. Neither the NJ adults nor the NJ children showed a significant improvement for [l] from T1 to T2. The results also showed that the NJ adults' scores were higher than the NJ children's for [ɹ] and [w] at T1. The difference between the NJ adults and children was non-significant at T2, because the NJ children's but not the NJ adults' scores, improved from T1 to T2.

3.3. Discussion

The NJ children's production of English consonants improved over the course of 1 year in the US. More specifically, they showed improvement for [ɹ] and [w] but not for [l]. The NJ adults' production of [l], [ɹ] and [w] showed little change from T1 to T2. The NJ participants' [w]s were identified as intended more often than their [l]s and [ɹ]s. The NJ participants' [l]s were often misidentified as [ɹ], and their [ɹ]s were often misidentified as /l/ (see Table 2). For example, at T1, the NJ children's [ɹ]s were misidentified as /l/(42.8%), /w/(11.6%), or /d/(10.5%). Their [l]s were misidentified as /r/(17.7%) or /d/(13.7%).

The NJ children's [l]s were more intelligible than their [r]s at T1. This might seem to contradict the hypothesis that [ɹ] will be more learnable than [l] for NJ learners of English. However, the relatively high level of intelligibility for [l] at T1 may not actually have been the result of learning. Riney, Takada and Ota (2000) provided evidence that NJ adults substitute Japanese [ɾ] more often for English [l] than for [ɹ]. It has been reported that NE speakers tend to identify Japanese [ɾ] tokens as English /l/ (Sekiyama & Tohkura, 1993) or /d/ (Price, 1981). We speculate that the simple substitution of Japanese [ɾ] will lead to higher intelligibility for English /l/ than for English /r/ because of the relatively greater perceptual similarity between English [l] and Japanese [ɾ].

The difference in intelligibility among [l], [ɹ] and [w] at T1 makes it difficult to compare the NJ participants' degree of improvement for each consonant. One might argue that the NJ children's scores improved to a greater extent for [ɹ] than [l] because their [ɹ] scores were so low at T1. Bradlow et al. (1997) took the "room for improvement" factor into consideration when examining the scores obtained before and after /l/-/r/ training. These authors computed 'relative improvement' as post-test scores minus pretest scores divided by 100 minus pretest score (Bradlow et al., 1997, p. 2306). Following Bradlow et al., relative improvement was computed for each consonant. It was computed as the T2 score minus the T1 score divided by 100 minus the T1 score, i.e., (T2-T1)/(100-T1). For the NJ children, the relative improvement was 68.5% for [w], 49.3% for [ɹ] and 18.5% for [l]. The NJ adults' relative improvement was slightly greater for [ɹ] (16.0%) than [l] (11.7%). The relative improvement was 0.1% for [w] for the NJ adults, because their intelligibility scores for [w] were high at both T1 and T2. A consideration of the 'room for improvement' factor, therefore, indicated that there was more improvement by the children for [ɹ] and [w] than for [l], and slightly more improvement for [ɹ] than [l] by the adults.

4. General discussion

Flege (1987, 1995) suggested that L2 learners' relative degree of accuracy in producing L2 speech sounds will vary over time as a function of their perceived relation to sounds in the L1 inventory. An L2 sound that is similar, but not identical to an L1 speech sound may enjoy an advantage in early stages of L2 acquisition. This is because the simple substitution of the L1 sound for its counterpart in the L2 might result in a high degree of intelligibility. An L2 sound that is more dissimilar phonetically from the closest L1 sound, on the other hand, is expected to show a disadvantage in early stages of L2 learning. Several different L1 speech sounds might be used as substitutes for it; and the learner may struggle to find new articulatory patterns for producing it. According to the SLM (Flege, 1995, 1999, 2002), however, the initial disadvantage of the more

dissimilar L2 speech sound will ultimately prove to be an advantage. By hypothesis, a relatively high degree of perceived dissimilarity will eventually result in accurate segmental production and perception because it will promote the formation of a new phonetic category.

The purpose of this study was to test the SLM hypothesis regarding the effect of perceived cross-language phonetic dissimilarity on the learning of L2 speech sounds. The hypothesis tested here was that if two L2 consonants differed in perceived dissimilarity from the closest consonant in L1 inventory, the more dissimilar of the two L2 consonants would manifest the greater amount of learning. We tested this hypothesis by examining the learning of English [l] and [ɹ] by native speakers of Japanese.

The results provided support for the SLM hypothesis. In Experiment 1, the NJ children showed improvement in discriminating English [l] from [ɹ], and [ɹ] from [w], over the course of 1 year. The NJ children showed a substantial amount of improvement over the course of 1 year for both contrasts, leading to the inference that their perception of [ɹ] had improved.[8] In Experiment 2, the same NJ children showed more improvement in producing [ɹ] and [w] than in producing [l]. When 'the room for improvement' is considered, the relative improvement for [ɹ] was larger than [l] for both the NJ adults and children in production. From the results of Experiments 1 and 2, we conclude that the NJ children showed more learning for English [ɹ] than [l].

The results from this study seem to agree with the results of other studies that have also provided some evidence of greater learning for [ɹ] than [l] by NJ learners of English. It has been reported that NJ adults with various experience levels in English identify [ɹ] more accurately than [l] (Sheldon & Strange, 1982; Flege et al., 1996; Ingram & Park, 1998). Cochrane (1980) reported that both NJ adults and children who were learning English in the US produced [ɹ] more accurately than [l]. In a training study (Bradlow et al., 1997), it was found that the NJ adults' improvement was greater for [ɹ] than [l] in both perception and production. One might object that more improvement has been observed for [ɹ] than [l] because [ɹ] is easier to articulate. However, studies examining the acquisition of English by children learning English as their L1 show that [l] is typically mastered before [ɹ] (Snow, 1963). This implies that [ɹ] is a more difficult consonant to learn to produce than [l] is.

Flege et al. (1995) reported that the intelligibility scores obtained for Japanese adults who had lived in the US for an average of 2 years were higher for [l] than for [ɹ] (90% vs. 80% on average, respectively). The production data from Experiment 2 in the present study also suggested that NJ speakers' production of [l] initially received higher intelligibility scores than their production of [ɹ] did. Japanese speakers' production of [l] may appear to be more accurate than their production of [ɹ] at an early stage of learning, because English [l] is perceptually more similar to Japanese [ɾ] than English [ɹ] is (Sekiyama & Tohkura, 1993; Takagi, 1993), and because Japanese [ɾ] is substituted for English [l] more often than for [ɹ] (Riney et al., 2000). Flege et al. (1995) also found that both [l] and [ɹ] produced by the 'experienced' NJ adults (an average LOR of 21 years) were identified correctly more than 95% of the time. This suggests that NJ learners of English will eventually produce both [l] and [ɹ] with equal intelligibility. Thus, while the production of [l] may appear to be more accurate at an initial stage of L2 learning, more learning of [ɹ] seems to occur both in production and perception for NJ learners of English. According to the SLM, the

[8] The findings obtained here do not, of course, rule out the possibility that perceptual learning also took place for /w/ or /l/. The results of an identification experiment will be needed to supplement the results obtained here.

difference in the learning of [l] and [ɹ] by NJ speakers is due to differences in the degree of perceived phonetic dissimilarity between [l] and [ɹ] with respect to Japanese [ɾ].

English [w] has a near equivalent sound in Japanese. One might expect the learning of [w] to be affected in the same way as the learning of [l] if the SLM's hypothesis is correct. However, the NJ children's production scores for [w] improved from 75.8% to 92.4% from T1 to T2. This finding appeared to contradict the SLM's hypothesis that phonetic dissimilarity between L1 and L2 sounds, not similarity, facilitates L2 learning. The perceptual assimilation of English [w] by Japanese sound categories needs to be evaluated experimentally in order to address this issue. We speculate that a one-to-one mapping between an L1 and L2 consonant, as in the case of English and Japanese /w/, may be different from a two-to-one mapping, as in the case of English [l], [ɹ] and Japanese [ɾ]. That is, there appears to be no other English consonants that can be mapped on to Japanese /w/, and this may be the reason why the NJ children's scores on English [w] improved while their scores on [l] did not.

Flege et al. (1995) found that correctly identified [l] and [ɹ] tokens produced by NJ adults may still differ phonetically from those produced by NE adults. The NE listeners in the Flege et al. study were asked to identify [l] and [ɹ] and also to rate how confident they were in their identification response ("definitely L", "probably L" and so on). Even though experienced NJ speakers' productions of [l] and [ɹ] were correctly identified more than 95% of the time, their "% definite" scores were lower than NE adults' (Flege et al., 1995).

Only intelligibility scores were obtained in the present study. It is likely that the phonetic properties of [l], [ɹ] and [w] were different between the NE participants and the NJ participants, even for segments that were identified as intended. Further research will be needed to examine fine phonetic differences between the native and non-native productions of these segments.

Additional research will also be needed to clarify the role of the magnitude of perceived cross-language phonetic differences on L2 phonetic learning. First, it would be helpful to obtain perceptual assimilation data from the same participants whose L2 segmental production and perception is assessed. In particular, how Japanese children identify and rate English [l] and [ɹ] with respect to Japanese [ɾ] needs to be investigated in future research.

Second, it would be valuable to attempt to generalize these findings to additional phonetic contexts and to multiple talkers. The present study examined only singleton tokens of [l] and [ɹ] in syllable-initial position (and then only in the context of [ɑ]). Previous studies have shown that perceptual assimilation patterns of English [l] and [ɹ] to Japanese [ɾ] differ depending on vowel context (Komaki et al., 1999) and position within the syllable (Takagi, 1993; Komaki et al., 1999). Phonetic environment also influences how well Japanese speakers can perceive and produce English [l] and [ɹ] (e.g., Dickerson, 1974; Sheldon & Strange, 1982; Shimizu & Dantsuji, 1983; Bradlow et al., 1997; Ingram & Park, 1998). In order to generalize the findings of this study, it will be necessary to examine both perceptual assimilation patterns and perception of [l] and [ɹ] tokens produced in other phonetic contexts.

Finally, it will be important in future research to find ways to recruit adults and children who have received comparable L2 phonetic input at the start of the study interval. The NJ adults examined in this research had all studied English before their arrival in the US whereas most of the NJ children had not. After their arrival, it was likely that the NJ children received more native-speaker input than the NJ adults did because all the children went to school while some adults might not have had much opportunity to speak English. It may be difficult to find a context in

which adults and children receive comparable L2 input, but it is important to find such situations in order to help select among competing explanations for adult-child differences.

In conclusion, the results of this study indicated that more learning took place for English [ɹ] than for [l] among Japanese children. The results of this study were interpreted as supporting a hypothesis proposed by the Speech Learning Model (Flege, 1995), viz., that if two L2 sounds differ in perceived dissimilarity from the closest sound in the L1 inventory, the more dissimilar of the L2 sounds will manifest the greater amount of learning. Future research examining changes over time in perceived relations between L1 and L2 sounds will be needed to provide additional insight into the learning of L2 phonetic segments.

Acknowledgements

This study was supported by NIH grant (DC00257). We are grateful to Linda Legault for help with data collection and analyses, and to Kimiko Tsukada, Satomi Imai for their comments on a previous version of this article. We also thank the Associate Editor of *Journal of Phonetics* and the three reviewers for their valuable comments.

References

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.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, *109*, 775–794.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, *20*, 305–330.

Bradlow, A., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, *101*, 2299–2310.

Cochrane, R. M. (1980). The acquisition of /r/ and /l/ by Japanese children and adults learning English as a second language. *Journal of Multilingual and Multicultural Development*, *1*, 331–360.

Dickerson, L. J. H. (1974). *Internal and external patterning of phonological variability in the speech of Japanese learners of English: Toward a theory of second-language acquisition*. Ph.D. Dissertation, University of Illinois, Urbana-Champaign.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test* (3rd ed.) (PPVT-III). Circle Pines, MN: American Guidance Service.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*, 47–65.

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

Flege, J. E. (1999). Age of learning and second-language speech. In D. P. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 101–132). Hillsdale, NJ: Lawrence Erlbaum.

Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development: Papers in honor of Henning Wode* (pp. 217–244). Trier: Wissenschaftlicher Verlag.

Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, *106*, 2973–2987.

Flege, J. E., Takagi, N., & Mann, V. (1995). Japanese adults can learn to produce English /ɹ/ and /l/ accurately. *Language and Speech*, *38*, 25–55.

Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /l/. *Journal of the Acoustical Society of America*, *99*, 1161–1173.

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

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America*, *107*, 2711–2724.

Ingram, J. C., & Park, S.-G. (1998). Language, context, and speaker effects in the identification and discrimination of English /r/ and /l/ by Japanese and Korean listeners. *Journal of the Acoustical Society of America*, *103*, 1161–1174.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2001). A perceptual interference account of acquisition difficulties for non-native phonemes. *Speech, hearing and language: Work in progress*, Vol. 13 (pp. 106–118). London: University College London Department of Phonetics and Linguistics.

Komaki, R., Akahane-Yamada, R., & Choi, Y. (1999). *Effects of native language on the perception of American English /r/ and /l/: Cross-language comparison between Korean and Japanese.* Technical report of IEICE, SP 99-45, The institute of electronics, information and communication engineers (pp. 39–46).

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect". In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121–154). Timonium, MD: York Press.

Ladefoged, P. (2001). *A course in phonetics* (4th ed.). Boston: Heinle & Heinle.

Lado, R. (1957). *Linguistics across cultures*. Ann Arbor: University of Michigan Press.

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.

Price, P. (1981). *A cross-linguistic study of flaps in Japanese and in American English*. Ph.D. Dissertation, University of Pennsylvania.

Riney, T., Takada, M., & Ota, M. (2000). Segmentals and global foreign accent: The Japanese flap in ESL. *TESOL Quarterly*, *34*, 711–737.

Sekiyama, K., & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics*, *21*, 427–444.

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.

Shimizu, K., & Dantsuji, M. (1983). A study on the perception of /r/ and /l/ in natural and synthetic speech sounds. *Studia Phonologica*, *17*, 1–14.

Snodgrass, J. G., Levy-Berger, G., & Haydon, M. (1985). *Human experimental psychology*. New York: Oxford University Press.

Snow, K. (1963). A detailed analysis of articulation responses of 'normal' first grade children. *Journal of Speech and Hearing Research*, *6*, 277–290.

Snow, C. E., & Hoefnagel-Höhle, M. (1977). Age differences in the pronunciation of foreign sounds. *Language and Speech*, *20*, 357–365.

Snow, C. E., & Hoefnagel-Höhle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development*, *49*, 1114–1128.

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

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.

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

Takagi, N., & Mann, V. A. (1995). The limits of extended naturalistic exposure on the perceptual mastery of English /r/ and /l/ by adult Japanese learners of English. *Applied Psycholinguistics*, *16*, 379–405.

Vance, T. (1987). *An introduction to Japanese phonology*. New York, NY: State University of New York Press.

Yoshida, K., & Hirasaka, F. (1983). The lexicon in speech perception. *Sophia Linguistica*, *11*, 105–116.