An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants

Susan G. Guion[a)]
The University of Oregon, Department of Linguistics, Eugene, Oregon 97403-1290

James E. Flege
The University of Alabama at Birmingham, Department of Rehabilitation Sciences, VH 503, Birmingham, Alabama 35294

Reiko Akahane-Yamada and Jesica C. Pruitt
ATR Human Information Processing Research Labs, Soraku-gun, Kyoto, 619-0288 Japan

(Received 18 June 1999; accepted for publication 5 January 2000)

This study reports the results of two experiments with native speakers of Japanese. In experiment 1, near-monolingual Japanese listeners participated in a cross-language mapping experiment in which they identified English and Japanese consonants in terms of a Japanese category, then rated the identifications for goodness-of-fit to that Japanese category. Experiment 2 used the same set of stimuli in a categorial discrimination test. Three groups of Japanese speakers varying in English-language experience, and one group of native English speakers participated. Contrast pairs composed of two English consonants, two Japanese consonants, and one English and one Japanese consonant were tested. The results indicated that the perceived phonetic distance of second language (L2) consonants from the closest first language (L1) consonant predicted the discrimination of L2 sounds. In addition, this study investigated the role of experience in learning sounds in a second language. Some of the consonant contrasts tested showed evidence of learning (i.e., significantly higher scores for the experienced than the relatively inexperienced Japanese groups). The perceived phonetic distance of L1 and L2 sounds was found to predict learning effects in discrimination of L1 and L2 sounds in some cases. The results are discussed in terms of models of cross-language speech perception and L2 phonetic learning.
[S0001-4966(00)02404-8]

PACS numbers: 43.71.Hw, 43.71.Es, 43.71.An [JMH]

INTRODUCTION

A great deal of recent research has examined the perception of vowels and consonants (or ''sounds,'' for short) in a second language. In many but not all instances, adults who learn a second language (L2) perceive L2 sounds differently than monolingual native speakers of the target L2 do. The general aim of this study was to provide insight into the perception of English consonants by native speakers of Japanese differing in English-language experience. The more specific aim of the study was to address questions relating to a model of cross-language perception, the perceptual assimilation model (PAM) developed by Best and her colleagues (e.g., Best, 1995), and a model of L2 speech acquisition, the speech learning model (SLM) developed by Flege and his colleagues (e.g., Flege, 1995).

A. Previous research

Many studies have examined Japanese learners' acquisition of English /ɹ/ and /l/. Adult Japanese learners have great difficulty in distinguishing /ɹ/ from /l/ (Goto, 1971; Miyawaki *et al.*, 1975; MacKain, Best, and Strange, 1981; Mochizuki, 1981). This is because Japanese does not have an /ɹ/–/l/ contrast, and neither English liquid is similar phonetically to any Japanese consonant. At a more abstract phonological level, the single liquid consonant found in Japanese might be considered similar to both English /ɹ/ and /l/. However, reports on the Japanese liquid (often referred to as an ''r'') suggest that it is usually produced as an apico-alveolar tap /ɾ/ (Miyawaki *et al.*, 1975; Vance, 1987). In some instances it is produced with a lateral or retroflex articulation, although these variants do not occur in any predictable phonological context.[1]

The perceptual relationship between English /ɹ/ and /l/ and the Japanese /ɾ/ is uncertain. However, it appears that English /l/ is perceived as being phonetically more similar to Japanese /ɾ/ than English /ɹ/ is. In a cross-language mapping study, Takagi (1993) found that inexperienced Japanese listeners identified English word-initial /ɹ/ and /l/ tokens as instances of Japanese /ɾ/. However, in a rating experiment, the English /l/ tokens were judged to be more similar to Japanese /ɾ/ than the English /ɹ/ tokens were. Sekiyama and Tohkura (1993) found that Japanese listeners identified word-initial tokens of English /ɹ/ most often as English /ɹ/ but also as Japanese /ɾ/, /ɰ/ (an unrounded velar approximant), and /g/. Conversely, English listeners identified syllable-initial tokens of Japanese /ɾ/ most often as /l/. Best and Strange (1992) suggested that both English /ɹ/ and /l/ may be identi-

[a)]Electronic mail: guion@oregon.uoregon.edu

fied as a poor exemplar of Japanese /ɰ/ because English liquids are approximants, and because the closest Japanese consonant, in articulatory terms, is /ɰ/. Support for this suggestion was obtained in studies by Yamada and Tohkura (1992) and Mochizuki (1981). Japanese listeners identified the middle stimuli in synthetic /ɹ/-to-/l/ continua as ''w.'' However, the synthetic stimuli did not contain all the information found in natural stimuli. Therefore, the ''w'' identifications do not guarantee that naturally occurring English /ɹ/ and /l/ tokens will be heard as Japanese /ɰ/.

The difficulty adult Japanese learners of English have in identifying English /ɹ/ and /l/ is well-known. However, there is evidence that Japanese learners' discrimination of /ɹ/ and /l/ may improve as they gain experience in English. MacKain *et al.* (1981) found that a group of Japanese participants with little exposure to native-produced English showed near-chance performance on /ɹ/ and /l/ identification and discrimination, whereas Japanese participants with intensive English conversational training showed categorical perception of /ɹ/ and /l/. Flege, Takagi, and Mann (1996) found that Japanese adults who had been in the United States for an average of 21 years identified /ɹ/ and /l/ tokens more accurately than a group of participants who had been in the United States for just 2 years. Best and Strange (1992) also found an effect of English-language experience on the perception of the English contrasts /ɹ/–/l/ and /ɹ–w/. These authors found a significant difference in identification and discrimination tasks between an inexperienced Japanese group who had been in the United States less than 7 months and had no conversational training and an experienced Japanese group who had been in the United States for 18–48 months and had English conversation training. The experienced group performed more like a native English group in labeling and discriminating English contrasts than the inexperienced group of Japanese participants.

Other studies have investigated the effect of laboratory training on /ɹ/–/l/ identification and discrimination (Logan, Lively, and Pisoni, 1991; Lively, Logan, and Pisoni, 1993; Lively *et al.*, 1994; Bradlow *et al.*, 1997). These studies have shown that perceptual training using highly varied stimuli from multiple talkers in multiple phonetic environments, but lasting only a few weeks, can yield a small but significant improvement in /ɹ/–/l/ identification.

Although most studies investigating Japanese learners of English have focused on /ɹ/ and /l/, a few studies have examined other consonants as well. Japanese learners of English are reported to produce English /θ/ as /s/ (Lado, 1957; Ritchie, 1968). There is also evidence that Japanese listeners often misidentify voiceless English fricatives. Lambacher *et al.* (1997) found that Japanese listeners had the most difficulty distinguishing between /θ/ and /s/. When presented with a syllable containing /θ/, 28% of the participants chose /s/, and when presented with a syllable containing /s/, nearly 25% chose /θ/. The number of /f/-for-/θ/ responses was also quite high (13%). Yoshida and Hirasaka (1983) investigated the identification of minimal-pair contrasts between English /b/–/v/, /ɹ/–/l/, and /s/–/θ/. Some pairs consisted of real words and some consisted of nonwords. The results obtained from 96 Japanese listeners indicated that, overall, the rate of errors decreased from /ɹ/–/l/ (27%) to /b/–/v/ (23%) to /s/–/θ/ (16%). An effect of lexical status was also observed in that more errors were found for consonants in nonwords than real words (see also Flege, Takagi, and Mann, 1996). As mentioned above, the English /w/–/r/ contrast has also been investigated. Best and Strange (1992) found that Japanese listeners correctly labeled and discriminated English /w/–/r/ at higher rates than /ɹ/–/l/. Listeners with more English-language experience responded to both contrasts more like native English speakers than those with less experience. Recently, Pruitt *et al.* (1999) used a training procedure similar to that used by Lively *et al.* (1994) for /ɹ/–/l/ perception to train other English consonants. They found a significant increase in pretest to post-test accuracy scores for the /b/–/v/, /s/–/θ/, and /z/–/ð/ contrasts.

The results obtained in the studies just cited suggest that certain English consonant contrasts are more readily learned than others (either during the process of naturalistic acquisition or laboratory training). This raises the issue of why this might be. English /ɹ/ and /l/ appear to be especially difficult because these liquids do not map well onto any Japanese category(s) (although English /l/ may be phonetically more similar to the Japanese /ɾ/ than English /ɹ/ is). The likelihood of hearing two English sounds in terms of one Japanese sound may be the cause of the difficulty Japanese learners of English have perceiving and producing English liquids. Contrasts involving other English consonants (e.g., /θ/, /w/, and /v/) may also pose a perceptual challenge for Japanese adult learners of English. Taken together, the literature reviewed here leads us to formulate the following questions: Why are certain English consonants more difficult for Japanese learners of English than others? Will experience with English affect the discrimination of various English consonants by Japanese adult learners differentially, and if so, why?

B. Theoretical models

Two existing models relate to the differential learnability of second language (L2) consonants. The speech learning model (SLM) developed by Flege (1995) and the perceptual assimilation model (PAM) developed by Best and colleagues (Best, McRoberts, and Sithole, 1988; Best, 1993, 1995) both model the degree of success listeners will have in perceiving non-native sounds. These models posit that success will depend on the perceived relationship between phonetic elements found in the first language (L1) and the L2 systems. The models make predictions about performance in non-native segmental perception based on the perceived phonetic distance between L1 and L2 sounds.

PAM starts with the observation that certain pairs of sounds from an unknown foreign language are easier to discriminate than other pairs are. In fact, certain foreign contrasts are easy to discriminate, even for listeners who have never heard them before. Other contrasts, on the other hand, are quite difficult to discriminate. PAM proposes that sounds in a foreign language are perceived according to their similarities to, or discrepancies from, native-language sounds that are closest articulatorily. PAM proposes that listeners will detect similarities and dissimilarities to native sounds based on perceived articulatory properties (e.g., constriction locations, active articulators, constriction degree, and phasing). The perceived distance between the unknown foreign sounds and the closest L1 sound (if any) leads to differences in discriminability.

The SLM differs from PAM in that it focuses on L2 *learning*. The purpose of the SLM is to account for changes across the life span in L2 speech learning. Hypotheses of the SLM can generate predictions concerning the accuracy with which highly experienced learners will produce and perceive L2 sounds. The SLM hypothesizes that basic speech learning mechanisms, including the ability to establish long-term memory representations for speech sounds (''phonetic categories''), remain intact across the life span. It also hypothesizes that L2 learners can establish new L2 phonetic categories if they detect phonetic differences between an L2 sound and the nearest L1 sound. The SLM predicts that the greater the perceived phonetic distance between an L2 sound and the closest L1 sound is, the more likely it is that phonetic differences between the sounds will be detected and a phonetic category eventually established. The acquisition of phonetic categories is thought to make L2 segmental perception more native-like because it enables the learner to base perception on L2 phonetic input without interference from prior learning.

The extent to which foreign (or L2) sounds resemble sounds in the naive listener's (or L2 learner's) L1 phonetic inventory plays a crucial role in both PAM and the SLM. According to PAM the degree of perceived phonetic distance is based on the perceived resemblance of articulatory gestures used to produce the foreign phones being discriminated and those used to produce the closest L1 sound. The SLM does not take a specific position regarding how cross-language phonetic distance is gauged by speakers of an L2 (see, e.g., Flege, 1995, p. 264). However, both models agree that perceived phonetic distance must be assessed empirically through cross-language mapping experiments. For predictions to be generated by either PAM or the SLM, cross-language phonetic distance data are needed. Degree of phonetic distance has been examined using an identification and rating methodology (see, e.g., Schmidt, 1996; Strange *et al.*, 1998). The foreign (or L2) sounds are first classified as instances of a phonetic category(s) in the listener's L1, then rated for goodness-of-fit to the L1 category.

Unfortunately, very few studies thus far have provided the needed phonetic distance data *and also* obtained the relevant discrimination data. Two such studies examined vowels in an unknown foreign language; one other examined consonants. Best, Faber, and Levitt (1996) showed that cross-language mapping can predict vowel discrimination for unknown foreign languages. Native speakers of English used English keywords to classify vowels drawn from three unknown foreign languages. The same participants also participated in a categorial AXB discrimination experiment using the same foreign vowel stimuli. The study showed that the identification data predicted accuracy in discriminating the foreign vowel contrasts. Polka (1995) found that monolingual English listeners differed in their accuracy in discriminating German vowels. Discrimination accuracy was related to the observed cross-language mapping pattern. Two vowels that were both considered to be good instances of the same L1 category received lower discrimination scores than two vowels identified in terms of a single L1 category but differing in goodness ratings did. In the case of consonants, Best (1990) obtained cross-language identification data for ejectives examined earlier in a discrimination study. Again, discrimination accuracy was related to the observed cross-language mapping pattern.

The results from the studies just cited provide important insight into how naive listeners perceive the vowels and consonants of an unknown foreign language. One important question is whether the relationship between cross-language mapping patterns and discrimination in an unknown foreign language will also apply to individuals who are learning a second language. As far as we know, no previous study has examined both perceived cross-language phonetic distance and the discrimination of a wide range of L2 consonants. Therefore, the present study addressed two questions pertaining to the models discussed earlier. The first question was whether the PAM framework can be extended to naturalistic L2 acquisition. The second was whether the SLM framework can be extended to the acquisition of an L2 by relatively inexperienced L2 learners.

C. Present study

The primary aim of the present study was to examine the relation between the perceived phonetic distance of L2 and L1 consonants and discrimination of those consonants. In order to address questions pertaining to PAM and the SLM (see above), cross-language mapping data were needed. Thus, in experiment 1, a group of Japanese listeners who had little experience with spoken English identified English and Japanese consonants in terms of Japanese categories, then rated the consonants for goodness-of-fit to that Japanese category. In experiment 2, a categorial discrimination experiment was carried out. Given that the SLM predicts learnability, three groups of Japanese learners of English who differed in English experience were recruited, as well as a native English control group. The stimuli examined in the cross-language mapping experiment were used again in the discrimination experiment, which tested English–English, English–Japanese, and Japanese–Japanese consonant contrasts.

The organization of the paper is as follows: The results from the cross-language mapping experiment will be presented first, then the results from the categorial discrimination test. Finally, the research questions pertaining to PAM and the SLM (see above) will be addressed.

I. EXPERIMENT 1

The purpose of this experiment was to assess the perceived relation between English and Japanese consonants. Native speakers of Japanese identified English and Japanese consonant stimuli in terms of Japanese consonant categories, then rated the same stimuli for goodness-of-fit to the Japanese category. As an experimental control, they also identified and rated Japanese consonants.

A. Method

1. Speech materials

The speech materials were produced by eight male native speakers each of English and Japanese. The Japanese speakers (mean age = 37 years) had been living in the United States for an average of 3.3 years, and were from a variety of cities in Japan (Tokyo, Fukuoka, Oita, Kobe, Chiba). The native English speakers (mean age = 33 years) were from several places in the United States (Alabama, Ohio, South Carolina, Illinois, Virginia, Wisconsin, Georgia, Washington). These speakers produced English or Japanese consonants followed by /a/ in a carrier phrase (''Then I saw ____ there,'' for English; ''Korewa ____ desu,'' meaning ''This is ____'' for Japanese). The speakers produced each phrase at a relatively slow speech rate, then at a faster rate.

The following method was used to obtain two different speech rates: First, the speakers listened to a tape that modeled the task. They heard isolated consonant + /a/ combinations (/Ca/) then another voice saying, ''Then I saw /Ca/ there'' (or ''Korewa/Ca/desu''). Three such examples were given. The rate modeled was fairly slow and careful. After practicing, the speakers were given a written list of /Ca/ syllables to repeat in the carrier phrase following an auditory model. After the slow tokens were recorded, the faster tokens were collected. The speakers were asked to repeat the task, but this time they were asked to speak more rapidly. (They heard some examples of the task modeled at a faster rate, practiced, and then repeated the task.) The talkers were also encouraged to speak more rapidly by a smaller interstimulus interval (ISI) at the faster rate (1.6 vs 2.2 s in the earlier block). It was assumed that the increase in the speaking rate that was modeled would lead to a somewhat less careful speech style. Productions of most English and Japanese consonants (and vowels) were elicited in this way. However, only a subset of the consonant stimuli (English /b v w θ t s ɹ l/, Japanese /b ɰ t d s ɾ h/) was used in the present study.

As expected, the duration of the stimuli varied as a function of the modeled speaking rate. The /a/ in the English tokens averaged 184 ms in the faster condition vs 237 ms in the slower condition. The /a/ in the Japanese stimuli averaged 86 and 129 ms in duration, respectively. The Japanese /a/ was consistently shorter than the English /a/ (*mean* = 108 vs 210 ms). To prevent vowel length from being used as a cue in consonant discrimination, all vowels were first normalized to 50% of peak intensity, then truncated to the same duration (50 ms for the faster rate tokens, and 75 ms for the slower rate tokens). To minimize a ''clipped'' percept at the vowel ending, the intensity of the last 20 ms of the vowel ending was ramped off from 100% to 0%.

The following procedure was used to ensure that only good examples of the English and Japanese consonants were used as stimuli. Four native English speakers (two female, two male) from urban areas in the southern United States and four native Japanese speakers who had been in the United States for less than a year (two female, two male), all from the Tokyo area, judged consonants from their native language. The stimuli were presented simultaneously with a text stimulus (a key word for English, a *Katakana* transcription for Japanese). The listeners rated the stimuli for goodness on a scale ranging from ''bad example'' (1) to ''very good example'' (5). The best five (out of the eight) tokens, as judged by native speakers, were selected for use as stimuli. The ratings for both the Japanese and English stimuli were high, with means around the 4.0 level.

2. Participants

Nine native Japanese speakers living in Japan participated. All were college students (*mean* age = 20.1 years) who had never lived outside of Japan. It is difficult to find normal young adults in Japan who have not been exposed to English. However, the participants selected for the present study had the minimum possible exposure to English. They began to study English at about the age of 12 ($M = 11.7$ years) at school. Most of their exposure to English had taken place in the classroom. No participant reported a history of hearing or speech disorders, and all passed a pure-tone hearing screening at octave frequencies from 250 to 4000 Hz at 15 dB SPL in both ears. The participants were recruited in the same location (Kyoto) as the groups tested in the discrimination experiment (experiment 2, below).

3. Procedure

The English and Japanese speech stimuli (/Ca/ tokens) were presented to the participants for two kinds of auditory evaluation. They were first asked to identify each token as an instance of some Japanese consonant category. Then, immediately after, they were asked to rate the token for goodness-of-fit to the (just-selected) Japanese category. The opportunity to use all Japanese consonants might have overwhelmed the participants. Therefore, a set of likely consonant response categories was determined based on the results of a pilot experiment. The pilot used free transcription of the stimuli to determine which Japanese orthographic symbols should be presented to the participants as possible response alternatives. The choices of Japanese *Katakana* orthography (IPA representations given) were as follows:

ワ /ɰa/, ラ /ɾa/, ウラ /ɯɾa/, リャ /ɾʲa/, ヤ /ja/, ハ /ha/, パ /pa/, バ /ba/, ヴァ /va/[2], ファ /ɸa/, タ /ta/, ダ /da/, サ /sa/, ザ /za/, シャ /ʃa/, ツァ /tsa/, ア /a/.

TABLE I. Mean percent identification and goodness rating (in parentheses) of Japanese consonant stimuli in terms if Japanese categories. Boldfaced values indicate the modal identification response. The goodness ratings are based on a scale that ranged from ‘‘bad example’’ (1) to ‘‘very good example’’ (7).

Consonant	Percent Identification and Rating																
	ワ	ラ	ウラ	リャ	ヤ	ハ	パ	バ	ヴァ	ファ	タ	ダ	サ	ザ	シャ	ツァ	ア
Stimuli	ɰa	ɾa	ɯɾa	ɾʲa	ja	ha	pa	ba	va	ɸa	ta	da	sa	za	ʃa	tsa	a
/b/							8 (1.7)	**81** **(5.5)**	11 (4.8)								
/ɰ/	**62** **(4.2)**	28 (4.0)	4 (1.8)			1 (2.0)		1 (1.0)	4 (2.8)								
/t/											**99** **(5.9)**	1 (1.0)					
/d/												**88** **(5.1)**		12 (3.7)			
/s/													**85** **(5.0)**		1 (1.0)	14 (3.8)	
/ɾ/		**99** **(4.8)**		1 (3.0)													
/h/						**100** **(5.5)**											

After identifying each stimulus using one of these labels, the participants rated the goodness-of-fit to the selected Japanese consonant category using a scale ranging from bad example (1) to very good example (7).

Each of the nine participants responded to 150 trials, 10 tokens (5 fast and 5 slow) each of the 15 consonant types (8 English and 7 Japanese), for a total of 1350 responses. The average identification and goodness-of-fit rating for each consonant type was based on 90 responses (9 participants $\times$10 tokens).

B. Results

Table I presents the results for the Japanese consonant stimuli. Two types of data are presented. The percentages indicate the frequency with which various Japanese consonant categories were used to classify the Japanese stimuli. The numbers in parentheses indicate the average ratings for the stimuli receiving a particular classification. The boldfaced percentages indicate the modal (i.e., most frequently used) classification. For example, the Japanese /b/ stimuli were classified as /b/ in 81% of instances (the modal classification), as /p/ in 8% of instances, and as /v/ in 11% of instances. The average ratings given to stimuli receiving the modal classification were always higher than the ratings obtained for stimuli receiving other classifications.

The Japanese consonant stimuli were not identified as intended in 100% of instances. However, the correct identification rates were high for most of the Japanese consonants, which indicates that the participants understood the task and could reliably perform it. The Japanese listeners identified the Japanese consonants correctly (i.e., as intended) 88% of the time, on average. The correctly identified Japanese consonants were given an average goodness rating of 5.1. As shown in Table I, /t/, /ɾ/, and /h/ were identified correctly at near-perfect rates. The consonants /b/, /d/, and /s/ were identified correctly in more than 80% of instances. The Japanese consonant correctly identified least often (62%) was /ɰ/ (an unrounded velar approximate), which was frequently identified as /ɾ/.

Table II presents the results for the English consonants. Unlike the Japanese consonants, there was not a ‘‘correct’’ classification for the English stimuli. However, the data in Table II allow us to determine which Japanese consonant category (or categories) would be used most often to classify the English consonant stimuli. Of the eight English consonants examined, five were consistently (>75%) classified as instances of a single Japanese consonant category. English /t/ was heard as Japanese /t/ in 91% of instances, English /s/ as Japanese /s/ in 87% of instances, English /b/ as Japanese /b/ in 84% of instances, English /v/ as Japanese /v/ in 80% of instances, and English /w/ was heard as Japanese /ɰ/ in 79% of instances. These consonants received a mean goodness rating of 4.3 for the corresponding Japanese category.

The grouped data in Table II indicate that some English consonants, on the other hand, were identified in terms of two Japanese consonants. It was *not* the case that some participants consistently classified an English consonant type in terms of one Japanese category, while other participants consistently classified the same consonant type in terms of another Japanese category. A close inspection of the data revealed that individual listeners identified these consonants as examples of two Japanese sounds. That is, individual participants heard these consonants as *intermediate* between two Japanese categories. English /θ/ was heard as Japanese /s/ 39% of the time, and as /ɸ/ (a labial fricative) 38% of the time. English /ɹ/ was heard as Japanese /ɾ/ 46% of the time

TABLE II. Mean percent identification and goodness rating (in parentheses) of English consonant stimuli in terms of Japanese categories. Boldfaced values indicate the modal identification response. The goodness ratings are based on a scale that ranged from ‘‘bad example’’ (1) to ‘‘very good example’’ (7).

Consonant Stimuli	Percent Identification and Rating																
	ワ	ラ	ウラ	リャ	ヤ	ハ	パ	バ	ヴァ	ファ	タ	ダ	サ	ザ	シャ	ツァ	ア
	ɰa	ɾa	ɰɾa	ɾʲa	ja	ha	pa	ba	va	ɸa	ta	da	sa	za	ʃa	tsa	a
/b/								**84**	16								
								(5.3)	(4.8)								
/v/	2						1	17	**80**								
	(4.5)						(6)	(4.2)	**(4.4)**								
/w/	**79**	10	6			2			3								
	(3.5)	(3.0)	(2.2)			(3.0)			(2.0)								
/θ/								2	2	38			**39**	7		12	
								(2.0)	(4.5)	(3.4)			**(3.8)**	(2.7)		(3.3)	
/t/											**91**					8	
											(3.9)					(2.4)	
/s/													**87**			13	
													(4.5)			(3.6)	
/ɹ/		46	**50**	4													
		(3.4)	**(3.3)**	(1.2)													
/l/	7	**50**	37	1					5								
	(3.5)	**(3.2)**	(3.0)	(1.0)					(3.6)								

and as /ɰɾ/ (a high back unrounded vowel+tap) 50% of the time. Finally, English /l/ was heard as Japanese /ɾ/ 50% of the time and as /ɰɾ/ 37% of the time. As mentioned above, the /ɰɾ/ response alternative was determined by free transcription of the stimuli in the pilot experiment. The /ɰ/ before the /ɾ/ might be the result of the approximate production of English /ɹ/, which is more vowel-like than a tap. The identifications of the /θ/, /ɹ/, and /l/ stimuli received an average goodness rating of 3.3.

The responses to the English stimuli were analyzed further in terms of overall fit to a Japanese category in order to provide what will be called a ‘‘fit index.’’ The fit index used here combined both the identification and the goodness-of-fit data into a single metric.[3] The fit indexes calculated here will be used in Sec. III to investigate the relations between cross-language mapping and discrimination. For the five English consonants that were consistently classified (>75%) as one Japanese consonant (*viz.*, /b/, /v/, /w/, /t/, /s/), the modal response was considered. For the three English consonants that were classified in terms of two Japanese consonants (*viz.*, /θ/, /ɹ/, /l/), both Japanese classifications were considered. The proportion of classification as a particular Japanese consonant was weighted by the mean goodness rating for stimuli receiving that identification. For example, the fit index for English /b/ stimuli was obtained by multiplying the proportion of responses receiving the modal identification (0.84) by the goodness rating of that identification (5.3). This resulted in a fit index of 4.5 for English /b/ to Japanese /b/. To take another example, the English /θ/ stimuli were classified in terms of two Japanese categories, /s/ and /ɸ/. Fit indexes were calculated for both Japanese response categories. The proportion of /s/ identifications (0.39) was multiplied by the mean rating for the /s/ response (3.8), giving a fit index of 1.5. The proportion of /ɸ/ identifications (0.38) was multiplied by the mean rating for the /ɸ/ response (3.4), giving a fit index of 1.3.

The fit indexes derived for the eight English consonants are shown in Table III. (As mentioned earlier, two fit indexes were derived for the three English consonants that were identified in terms of two different Japanese consonants.) The fit indexes spanned a wide range, from a low value of 1.3 (the fit of English /θ/ to Japanese /ɸ/) to a high of 4.5 (the fit of English /b/ to Japanese /b/). It seems reasonable to suppose that the English consonants with relatively high fit indexes would be readily accepted as instances of a Japanese consonant category, whereas those with relatively low fit indexes would be heard either as ‘‘foreign’’ or as distorted instances of a Japanese category.

An important empirical question is whether variation in the fit indexes just described is relevant to the discriminability of English consonants (see experiment 2 below). As a working hypothesis, we divided the English consonants into subclasses based on the fit indexes using a standard deviation (s.d.) criterion. The mean fit index obtained for the seven Japanese consonants was 4.5 (s.d.=1.1). The English consonants receiving a fit index that fell within 1.0 s.d. of the mean fit index obtained for the Japanese consonants (not shown in Table III) were classified as ‘‘good’’ instances of a Japanese category. (Thus, a good fit index for an English consonant was considered to be 3.4 and over.) The good-

TABLE III. Fit indexes derived for English consonants in terms of Japanese categories. The fit index was derived from the proportion of identifications and goodness ratings (see the text). Only identifications that were more than 30% are included.

English consonant	Most common identification	Proportion of identifications	Goodness rating	Fit index	
/b/	/b/	0.84	5.3	4.5	good /b/
/s/	/s/	0.87	4.5	3.9	good /s/
/t/	/t/	0.91	3.9	3.5	good /t/
/v/	/v/	0.80	4.4	3.5	good /v/
/w/	/ɰ/	0.79	3.5	2.8	fair /ɰ/
/ɹ/	/ɯɾ/	0.50	3.3	1.7	poor /ɯɾ/
	/ɾ/	0.46	3.4	1.6	poor /ɾ/
/l/	/ɾ/	0.50	3.2	1.6	poor /ɾ/
	/ɯɾ/	0.37	3.0	1.1	poor /ɯɾ/
/θ/	/s/	0.39	3.8	1.5	poor /s/
	/ɸ/	0.38	3.4	1.3	poor /ɸ/

fitting English consonants were /b/, /s/, /t/, and /v/. English consonants that received fit indexes within 2 s.d.'s of the Japanese mean (2.3 to 3.3) were considered to have a ''fair'' fit index. English /w/ was considered to have fair fit index to Japanese /ɰ/. Three other English consonants, /ɹ/, /l/, and /θ/, received a fit index that was more than 2.0 s.d.'s below the mean fit index computed for the Japanese consonants (*viz*., values<2.3). These consonants have been designated the ''poor''-fitting English consonants.[4] These classifications will be used in later sections to evaluate the relation between perceived cross-language phonetic distance and the discrimination of L2 consonants.

II. EXPERIMENT 2

The purpose of this experiment was to examine the discrimination of word-initial consonants by native speakers of Japanese and English. The stimuli were the Japanese and English consonants used in the cross-language mapping experiment (experiment 1). These stimuli were presented in a categorial discrimination test examining three types of contrasts. As shown in Table IV, four contrasts between two English consonants (the ''$E-E$'' contrasts) were examined. Seven contrasts between an English and a Japanese consonant (the ''$E-J$'' contrasts) were examined. Finally, three contrasts between two Japanese consonants (the ''$J-J$'' contrasts) were examined. As discussed in the Introduction, there was evidence that the $E-E$ contrasts /ɹ/–/l/, /s/–/θ/, and /ɹ/–/w/ would prove difficult for at least some of the native Japanese participants. As far as we know, no previous study has examined $E-J$ contrasts.

TABLE IV. Three kinds of consonant contrasts examined in experiment 2. The English consonant is listed first, the Japanese consonant second for all English–Japanese contrasts.

English–English	English–Japanese	Japanese–Japanese
/b/–/v/	/ɹ/–/ɾ/	/ɾ/–/d/
/ɹ/–/l/	/ɹ/–/ɰ/	/s/–/h/
/s/–/θ/	/l/–/ɾ/	/s/–/d/
/ɹ/–/w/	/θ/–/s/	
	/v/–/b/	
	/t/–/t/	
	/b/–/b/	

Four groups of individuals participated. One group consisted of native speakers of English; the other three consisted of native speakers of Japanese with varying amounts of English-language experience. We had two general expectations concerning how participants in these groups would perform. The first expectation was that the native speakers of Japanese would discriminate some $J-J$ contrasts more successfully than the native speakers of English, whereas the reverse would hold true for the $E-E$ contrasts. The second expectation was that the Japanese participants who were relatively experienced in English would discriminate at least some of the $E-E$ and $E-J$ contrasts more successfully than those who were relatively inexperienced in English.

A. Method

1. Speech materials

The stimuli used in experiment 1 were used again here.

2. Participants

As summarized in Table V, 30 native speakers of Japanese varying in experience with English participated. Ten native Japanese speakers living in the United States comprised the ''high-experience'' group. These participants had

TABLE V. Characteristics of the four groups of ten participants in experiment 2.

Group	Age[a]	AOA[b]	LOR[c]
Native English	24.6 (3.1)[d]	⋯	24.6 (3.1)
High-experience Japanese	29.7 (5.1)	12.3 (0.9)	3.1 (1.4)
Mid-experience Japanese	28.5 (4.3)	11.8 (1.4)	⋯
Low-experience Japanese	19.5 (1.2)	12.5 (0.5)	⋯

[a]Age=mean chronological age at the time of testing in years.
[b]AOA=mean age of acquisition of English.
[c]LOR=mean length of residence in an English-speaking country in years.
[d]Standard deviations are in parentheses.

resided in the United States for an average of 3.1 years (range 1.8 to 5.5 years). Ten native Japanese speakers matched to the United States group for age and education made up the ''mid-experience'' group. The participants in this group had never lived outside of Japan, but used English often in their jobs. The ''low-experience'' group consisted of ten Japanese college students who had never lived outside of Japan. Most of their exposure to English had taken place in the classroom and consisted largely of written English. (In Japan, English education focuses on reading and writing, while conversational practice with a native English speaker is rare.) Ten monolingual native speakers of American English made up the comparison group.

There were five male and five female participants in each group. All of the native Japanese participants had begun learning English at about 12 years of age, and had studied English through middle school and high school. All of the participants had at least some college and many (in the high- and mid-experience groups) had advanced degrees. No participant reported a history of hearing or speech disorders; all passed a pure-tone hearing screening from 250 to 4000 Hz at 15 dB SPL in both ears. All of the participants were paid.

3. Procedure

A categorial discrimination test was used to assess consonant perception. The test used here is similar to an ABX or AXB discrimination test (Gottfried, 1984; Best *et al.*, 1988) in that listeners heard three stimuli per trial and were asked to pick the odd item. However, the test used here differed from traditional oddity tasks in that it incorporated ''catch'' trials consisting of three physically different tokens of the same consonant. This encouraged the participants to respond only to phonetically relevant differences, not to any auditorily detectable difference. To successfully discriminate consonants, the participants had to recognize the *categorical* identity of a set of physically different tokens of the same consonant category while ignoring acoustic/auditory differences among instances of the category, which were phonetically irrelevant to their categorical identity.

Each consonant contrast investigated was tested by eight catch trials (comprised of three physically different tokens of the same stimulus type) and by eight ''different'' trials (in which there was an odd item among the three stimuli). The odd item appeared equally in all three possible positions. There were four catch trials for each of the two consonants being contrasted. In addition, an instance of each consonant was the odd item four times each in the different trials. For example, to test the /ɹ/–/l/ contrast, four /ɹ/–/ɹ/–/ɹ/ and four /l/–/l/–/l/ catch trials were administered. Four different trials in which /ɹ/ was the odd item, and four trials in which /l/ was the odd item, were also presented. All trials consisted of tokens spoken by three different speakers. The three tokens were played at a 1.2 s ISI. The participants could replay a trial as often as they wished, but could not change a response once given.

The participants were tested individually in a sound booth and heard the stimuli at a comfortable listening level over headphones. They were told that the three stimuli in each trial were always spoken by different talkers, and that they should ignore differences in speakers' voices as much as possible. The participants selected ''1,'' ''2,'' or ''3'' if they judged a stimulus in one of those three serial positions to be different from the other two stimuli. They selected ''no'' if all three examples were considered to be instances of the same consonant.

All 448 trials were presented in two blocks during a 1-h session. The rate at which the stimuli had been produced (i.e., fast or slow) was counterbalanced across the listeners. To begin, the participants were familiarized with the task using catch trials and different trials made up of stimuli drawn from the experiment (but which did not test any contrasts found in the experiment). The participants received feedback during the practice, but not during the experiment.

4. Analysis

A-prime (A') scores were calculated for each of the 14 consonant contrasts examined. These scores were derived from the proportion of ''hits'' (correct selection of the odd item in different trials) and ''false alarms'' (incorrect selections of an odd item in catch trials) obtained for each contrast, using the formula provided by Snodgrass, Levy-Berger, and Haydon (1985).[5] The A' scores provide an unbiased measure of perceptual sensitivity by taking into account the responses to the different trials and the catch trials. An A' score of 1.0 indicates perfect discrimination of a contrast, and an A' score of 0.5 or lower indicates insensitivity to a contrast.

B. Results

A series of analyses revealed that discrimination of the faster-vs-slower-rate stimuli did not vary according to group. A rate (2)×group (4)×contrast (4) analysis of variance (ANOVA) examined A' scores obtained for the $E-E$ contrasts. Similar ANOVAs examined the $E-J$ and the $J-J$ contrasts. The overall effect of rate was not significant for any of the three contrast types (F-values$=$0.6 to 0.9). Nor did the rate factor interact significantly with group in any of the two-way interactions (F-values$=$0.9 to 2.2) or three-way interactions (F-values$=$0.4 to 1.5). This indicated that stimulus rate had a similar (non-)effect for the native English group and the three Japanese groups. Accordingly, the decision was made to pool the results obtained for the two sets of stimuli. New A' scores were computed based on the 16 catch and 16 different trials available for each contrast. These new A' scores represent a robust measure of the participants' perception of the contrasts investigated.

Figure 1 displays the results of the four $E-E$ consonant contrasts. There was a general trend for the native English group to receive the highest scores, followed by the high-experience Japanese, then the mid-experience Japanese, and finally the low-experience Japanese. In addition, the /v/–/b/ and /ɹ/–/w/ contrasts showed more between-group differences for the three Japanese groups than the other two $E-E$ contrasts.

The A' scores for the $E-E$ contrasts were submitted to a group (4)×contrast (4) ANOVA. The main effects of group, $F(3,26)=25.9$, $p<0.01$, and contrast, $F(3,108)$

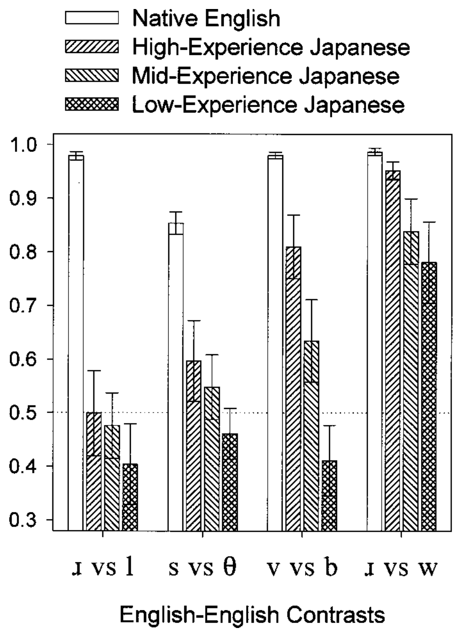

FIG. 1. Mean *A*-prime (A') scores obtained in experiment 2 for the four groups for four English–English contrasts. Error bars represent standard errors. A score of 1.0 indicates perfect discrimination, and a score of 0.5 or below indicates insensitivity to a contrast.

$=35.4$, $p<0.01$, as well as the two-way interaction, $F(9,108)=4.7$, $p<0.01$, were significant. The simple effect of group was significant for all four contrasts [/ɹ/–/w/:$F(3,36)=19.6$, $p<0.01$; /s/–/θ/:$F(3,36)=10.5$, $p<0.01$; /v/–/b/: $F(3,36)=19.2$, $p<0.01$; and /ɹ/–/w/: $F(3,36)=4.2$, $p=0.01$]. A Tukey's test ($\alpha=0.01$) revealed that the native English group received higher scores than all three Japanese groups for /ɹ/–/l/ and /s/–/θ/. On these contrasts, none of the Japanese groups differed significantly from one another. For /v/–/b/, the native English and high-experience Japanese groups did not differ significantly from one another, but received higher scores than both the mid-and low-experience Japanese groups. In turn, the mid-experience group received higher scores than the low-experience group. Finally, for /ɹ/–/w/, the native English group received significantly higher scores than the low-experience group. The other two Japanese groups did not differ from the native English listeners.

Figure 2 shows the results from the $E-J$ consonant contrasts. For three consonant contrasts, /v/–/b/, /θ/–/s/, /ɹ/–/ɾ/, the more experienced Japanese groups tended to receive higher A' scores than the less experienced groups. Between-group differences for the remaining four consonant contrasts, /b/–/b/, /l/–/ɾ/, /t/–/t/, /ɹ/–/ɰ/, were smaller.

The A' scores for the $E-J$ contrasts were submitted to a group (4)×contrast (7) ANOVA, which yielded significant main effects of group, $F(3,26)=7.3$, $p<0.01$, and contrast, $F(6,216)=51.3$, $p<0.01$, and a significant two-way interaction, $F(18,216)=7.3$, $p<0.01$. The simple effect of group was significant for three contrasts [/θ/–/s/:[6] $F(3,36)=9.1$, $p<0.01$; /v/–/b/: $F(3,36)=15.4$, $p<0.01$; /ɹ/–/ɾ/: $F(3,36)=9.5$, $p<0.01$], but not for four others (/b/–/b/, /l/–/ɾ/, /t/–/t/, /ɹ/–/ɰ/, $p>0.01$). All four groups received scores of about 0.5 for /b/–/b/, /l/–/ɾ/, and /t/–/t/ ($M=0.43$, 0.54, and 0.59, respectively), indicating a lack of sensitivity to these contrasts. Relatively high scores ($M=0.87$) were obtained for /ɹ/–/ɰ/ from all four groups, indicating a high level of discriminabilty for this contrast.

A Tukey's test ($\alpha=0.01$) revealed that the native English and high-experience Japanese groups did not differ significantly from one another for the /v/–/b/ contrast. However, both of these groups received significantly higher scores than the mid-and low-experience Japanese groups for /v/–/b/. The mid-experience group, in turn, received significantly higher scores than the low-experience group. The native English, high-experience, and mid-experience groups received significantly higher scores than the low-experience group for /ɹ/–/ɾ/. For /θ/–/s/, the native English group received significantly higher scores than all three Japanese

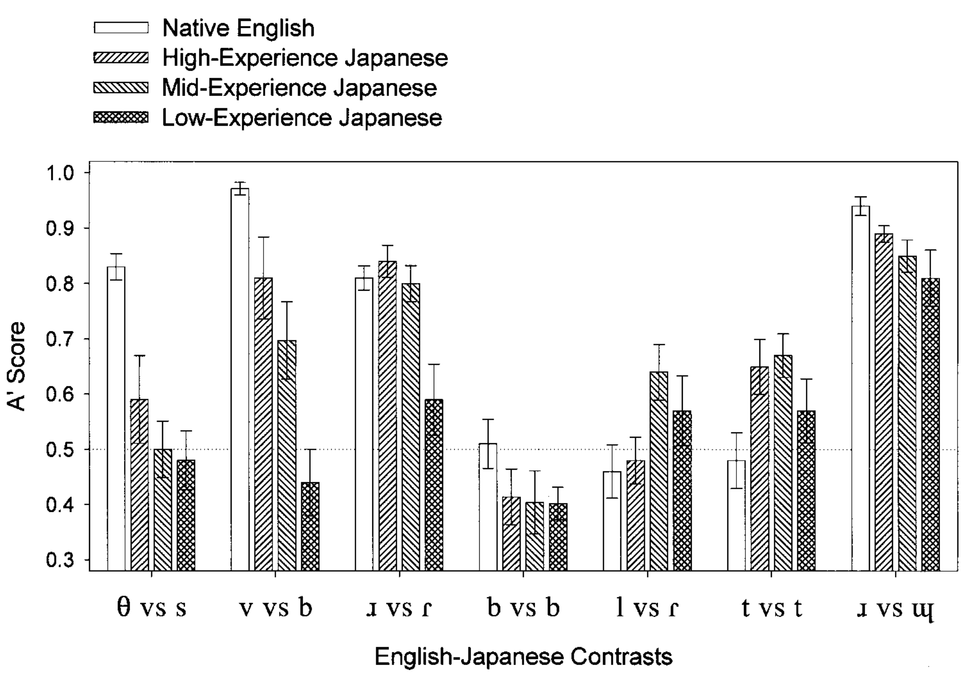

FIG. 2. Mean A' scores obtained in experiment 2 for four groups for the eight English–Japanese contrasts. Error bars represent standard errors. A score of 1.0 indicates perfect discrimination, and a score of 0.5 or below indicates insensitivity to a contrast.

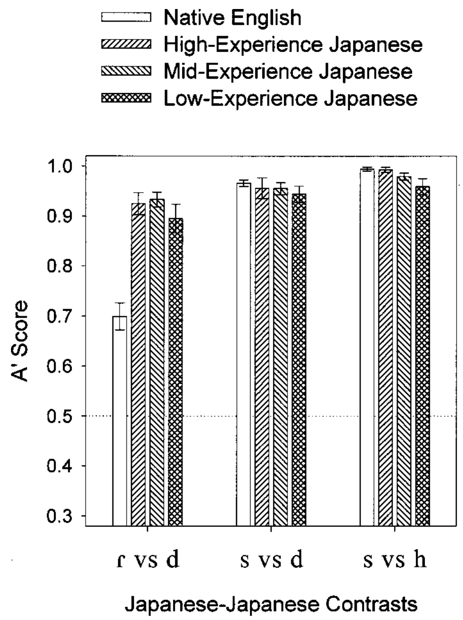

FIG. 3. Mean A' scores obtained in experiment 2 for four groups for the three Japanese–Japanese contrasts. Error bars represent standard errors. A score of 1.0 indicates perfect discrimination, and a score of 0.5 or below indicates insensitivity to a contrast.

groups, whereas the Japanese groups did not differ from one another.

Figure 3 shows the results of $J-J$ contrasts. All four groups received relatively high A' scores for the /s/–/d/ and /s/–/h/ contrasts. However, the native English group received lower scores for /ɾ/–/d/ than the three Japanese groups did.

The A' scores for the $J-J$ consonants were submitted to a group (4)$\times$contrast (3) ANOVA, which yielded significant main effects of group, $F(3,26)=6.5$, $p<0.01$, and contrast, $F(2,72)=98.8$, $p<0.01$, and a significant two-way interaction, $F(6,72)=28.3$, $p<0.01$. The simple effect of group was significant for /ɾ/–/d/, $F(3,36)=22.5$, $p<0.01$, but not for /s/–/h/ or /s/–/d/, $p>0.01$. A Tukey's test ($\alpha=0.01$) revealed that the native English listeners received significantly lower scores than all three native Japanese groups for /ɾ/–/d/. For /s/–/h/ and /s/–/d/, all four groups received high scores ($M=0.98$ and 0.96, respectively) and no significant between-group differences were obtained.

C. Discussion

The results obtained here suggest that some $E-E$ and $E-J$ consonant contrasts were more learnable than others. The Japanese participants with more English-language experience obtained higher discrimination (A') scores than the Japanese participants with less English experience for two $E-E$ contrasts (/v/–/b/ and /ɹ/–/w/), but not for two other $E-E$ contrasts (/ɹ/–/l/ and /s/–/θ/). Of the eight $E-J$ contrasts examined, two (/v/–/b/ and /ɹ/–/ɾ/) showed comparable effects of experience. It is interesting to note that the /ɹ/–/ɾ/ contrast showed improvement with experience, whereas the /l/–/ɾ/ contrast did not. All four groups received low A' scores for /l/–/ɾ/, suggesting that English /l/ is closer to (and thus more difficult to discriminate from) Japanese /ɾ/ than English /ɹ/.

All three native Japanese groups, and also the native English group, obtained very low A' scores for /b/–/b/ and /t/–/t/. This indicated a lack of sensitivity for these $E-J$ contrasts. However, the findings for the $J-J$ contrasts revealed that the Japanese participants could perform the discrimination task at high levels of accuracy. In fact, the Japanese participants obtained higher A' scores for some $J-J$ contrasts than the native English participants. This demonstrates that the lower A' scores obtained by the Japanese participants for some $E-E$ and $E-J$ contrasts was based on discrimination ability, not on test-taking ability or some type of cultural factor.

III. PREDICTION OF DISCRIMINABILITY

In this section we examine the relationship between the cross-language mapping data obtained in experiment 1 and the consonant discrimination scores from experiment 2. The purpose was to determine if PAM and the SLM can be extended to early stages of naturalistic L2 acquisition. (As mentioned in the Introduction, PAM usually focuses on the discrimination of sounds in an unknown foreign language; and the SLM usually focuses on highly experienced learners of an L2.)

A. Perceptual assimilation model

Best (Best *et al.*, 1988; Best, 1993, 1995) formulated the perceptual assimilation model (PAM) of cross-language speech perception. This model predicts the discriminability of non-native sounds based on their perceived relation to sounds in the L1. Best (1995, pp. 194–195) proposes the following patterns of perceptual assimilation for non-native sounds: They can either be (1) assimilated to a native category, (2) assimilated as uncategorizable speech sound, or (3) not assimilated to speech and heard as a nonspeech sound. If (1), then the non-native sound is clearly assimilated to a particular native segmental category, or perhaps to a cluster or string, in which case it may be heard either as a good exemplar of that category, an acceptable but not ideal exemplar of the category, or a notably deviant exemplar of the category. If (2), then the non-native sound is assimilated within native phonological space, but not as a clear exemplar of any particular native category (i.e., it falls within native phonological space, but in between specific native categories). If (3), then the non-native sound is not assimilated into native phonological space at all, but heard, instead, as some sort of nonspeech sound. PAM proposes that the patterns of perceptual assimilation will predict the discriminability of non-native consonants (and vowels).

Each of the assimilation patterns just mentioned makes a specific prediction regarding degree of discriminability. Based on the cross-language mapping data collected in experiment 1, the $E-E$ contrasts provide examples of the following assimilation pattern types: As shown in Table VI, the /ɹ/–/l/ contrast is an example of the ''both uncategorizable'' type. Both consonants fall within native phonological space, but ''in between'' (Best, 1995, p. 194) specific L1 categories. That is, both /ɹ/ and /l/ fall in between Japanese /ɯɾ/ and /ɾ/. The /s/–/θ/ and /ɹ/–/w/ contrasts are examples of the ''uncategorized vs categorized'' type. English /s/ was assimi-

TABLE VI. Categorization of English–English contrasts according to PAM.

Contrast	Fit index	PAM category type	Performance[a] Low-experience	Mid-experience	High-experience
/ɹ/–/l/	poor /ɯɾ/ or /ɾ/ poor /ɯɾ/ or /ɾ/	both uncategorizable	poor	poor	poor
/ɹ/–/w/	poor /ɯɾ/ or /ɾ/ fair /ɰ/	uncategorized vs categorized	moderate	moderate	good
/s/–/θ/	good /s/ poor /s/ or /ɸ/	uncategorized vs categorized	poor	poor	poor

[a]Performance=A' scores obtained by the three Japanese groups (low-experience, mid-experience, and high-experience) on the English–English contrasts tested in experiment 2.

lated to Japanese /s/, and English /θ/ fell in between Japanese /s/ and /ɸ/. English /w/ was assimilated to Japanese /ɰ/, and English /ɹ/ fell in between Japanese /ɯɾ/ and /ɾ/. The remaining $E-E$ contrast, /b/–/v/ is more difficult to interpret in terms of the PAM typology. Both /b/ and /v/ were identified as good examples of Japanese categories, seeming to illustrate a case of ''two-category assimilation.'' However, the phonemic status of /v/ in Japanese is uncertain, rendering it questionable to interpret /v/'s perceptual assimilation pattern as a case of assimilation to a native category.

Given the cross-language mapping data obtained in experiment 1, PAM makes a number of testable predictions. These will be evaluated for the $E-E$ contrast data obtained from the Japanese groups. Discrimination of the both uncategorizable type is expected to range from poor to very good, depending on the proximity of English consonants to each other and to Japanese categories. The /ɹ/–/l/ contrast is predicted to show poor discrimination because both sounds fall in between the same Japanese categories. This was indeed the case. All Japanese groups discriminated the /ɹ/–/l/ contrast at low rates.

Discrimination of the uncategorized vs categorized type was expected to be good. This prediction was borne out in the case of /ɹ/–/w/, which the Japanese participants discriminated with a moderate to high level of accuracy. It is interesting to note that the results of this type of contrast (uncategorized vs categorized), which is predicted to show good discrimination, improved with experience in the L2. In contrast, the results for the both uncategorized type of contrast, which is predicted to be poor, indicated that none of the three groups showed any sensitivity to the contrast.

The results for /s/–/θ/ were not as predicted. All three Japanese groups obtained low A' scores for this contrast. A closer look at the /s/–/θ/ contrast reveals some interesting facts. English /θ/ was classified as uncategorizable because it fell in between two native sounds. This is, it was heard as Japanese /s/ or as Japanese /ɸ/. English /θ/ was contrasted with English /s/ (heard as a good example of Japanese /s/), so there was likely to be overlap in category assimilation between the two English consonants. These results point to a possible revision of PAM in the predicted discrimination of uncategorized vs categorized non-native sounds. It might be useful to make provision for cases where the uncategorized sound is close in phonological space to the categorized sound.

Discrimination of the two-category assimilation type was expected to be excellent. However, as outlined above, the /b/–/v/ contrast did not provide a good test of this prediction. This prediction has nonetheless been tested, and supported, by an investigation into English speakers' assimilation patterns of Ethiopian ejectives /p′/ and /t′/ (Best, 1990). Native English participants identified the ejectives as examples of two English categories /p/ and /t/ (two-category assimilation) and discriminated the Ethiopian contrast at high rates.

In summary, the data from the cross-language mapping allowed us to test some predictions that might be derived for PAM if this model were extended to L2 learning. The predictions were supported with one exception. A contrast containing an uncategorized vs categorized sound was discriminated poorly, contrary to prediction. A revision of PAM was proposed to allow for poor discrimination of this contrast type when the uncategorized sound is in close phonological space to the categorized sound.

B. Speech learning model

The speech learning model (SLM) proposes that the likelihood of category formation for ''sounds'' is affected importantly by at least two factors. The likelihood of category formation is hypothesized to be inversely related to the age of first exposure to the sound system of an L2, but positively related to the perceived phonetic distance of an L2 sound from the closest native sound (Flege, 1995). This leads to the following two predictions for highly experienced speakers of an L2. The first is that, all else being equal (including degree of perceived phonetic distance), individuals learning an L2 early in life will be more likely to have established a category for an L2 consonant than those who began learning the L2 later in life. The native Japanese participants examined in this study were all exposed to English in school at about the same age (12 years old). Therefore, the first prediction, which addresses the effect of age of L2 learning, cannot be tested here.

A second prediction of the SLM is that, all else being equal (including the age of first exposure to the L2), L2 learners will be more likely to have developed a phonetic category for sounds that are perceptually distant from the closest native category than for sounds that are perceptually close to a native category. This prediction seem to holds true for highly experienced speakers of an L2. For example, Flege, Takagi, and Mann (1995, 1996) provided evidence

TABLE VII. Fit indexes derived for the English–Japanese contrasts (see the text).

English–Japanese contrast	Fit index[a]	Learning effect?[b]
/t/–/t/	/t/ 3.5 (good)	No
/b/–/b/	/b/ 4.5 (good)	No
/ɹ/–/ɾ/	/ɹ/ 1.6 (poor)	Yes
/θ/–/s/	/θ/ 1.5 (poor)	No
/l/–/ɾ/	/l/ 1.6 (poor)	No

[a]Fit index=Fit index of the English consonant to the Japanese consonant to which it is being compared (see Sec. I B).

[b]Learning effect?=results of pairwise comparisons for the Japanese groups' A' scores from experiment 1. A ''yes'' indicates that the high-and/or mid-experience groups received significantly higher A' scores than the low-experience group (see Sec. II B).

that category formation may require many years of native speaker input, at least for adult learners. Native Japanese participants showed evidence of category formation for /ɹ/ after 21 years in the United States, whereas those with two years of residence in the United States years did not. Thus, the results obtained for the relatively ''experienced'' group of native Japanese participants in the present study, who had lived in the United States for an average of 3.1 years, may not provide a fair test of the SLM's prediction regarding the effect of perceived phonetic distance on phonetic category formation. However, if category formation is related to perceived phonetic distance from an L1 category, then one might expect to see some evidence of learning, even after a relatively limited exposure to the L2. Evidence for this would be more accurate perception by the relatively experienced than the relatively inexperienced Japanese learners of English examined here. In other words, learning effects should be more likely for contrasts between English and Japanese consonants if the English consonant is phonetically distant from the closest Japanese consonant than if the English consonant is similar to the closest Japanese consonant.

The SLM proposes that L2 learners must detect phonetic differences between L2 and L1 sounds before beginning to establish a new L2 category. This motivated a closer examination of the results obtained for the $E-J$ contrasts. Our specific aim was to investigate a prediction that could be derived from the SLM if it were extended to relatively early stages of L2 acquisition. The prediction is that amount of L2 perceptual learning will depend on the perceived similarity of an English (L2) sound to the closest Japanese (L1) sound.

Table VII summarizes the predictions generated in this way for five $E-J$ contrasts. The table shows the fit index for the English member of each English–Japanese contrast, along with its adjectival classification (see experiment 1). The prediction is that, when the English consonant represents a good fit to the Japanese consonant it was contrasted with, there will be little, if any, evidence of learning. However, evidence of learning should be seen for contrasts in which the English member of the contrast represented a poor fit to the Japanese consonant to which it was contrasted. Table VII also summarizes the effects of English-language experience reported in experiment 2. ''No'' indicates that there was no difference between groups differing in English–language experience, whereas ''yes'' indicates that a relatively experienced group obtained significantly higher A' scores than a less experienced group. As predicted, the two contrasts involving ''good-fitting'' English consonants, /t/–/t/ and /b/–/b/, showed no effect of learning.

Contrary to prediction, just one of the three contrasts involving a ''poor-fitting'' English consonant showed evidence of learning, namely the /ɹ/–/ɾ/ contrast. The Japanese groups did not differ significantly for the /θ/–/s/ contrast; however, a nonsignificant trend in the expected direction is evident (see Fig. 2). The /l/–/ɾ/ contrast showed no effect of learning; all the Japanese participants received low A' scores for this contrast.

The asymmetry between English /ɹ/ and /l/ as compared to Japanese /ɾ/ is of interest. Even though both /l/ and /ɹ/ received poor fit indexes to Japanese /ɾ/, there was some evidence that the English /ɹ/ is more perceptually distant from the Japanese /ɾ/ than the English /l/ is. As mentioned earlier, Takagi (1993) found that inexperienced Japanese listeners judged English /l/ tokens to be more similar to Japanese /ɾ/ than English /ɹ/ tokens. Takagi used a different methodology to assess the perceived cross-language relations. He had one group of listeners identify the English stimuli and then another group of listeners rate those identifications. This procedure was adopted to avoid any biasing effect the identification responses might have on the ratings. Perhaps a rating procedure such as that used by Takagi would have revealed a difference in the /ɹ/ and /l/ ratings in our stimuli, as well. Further investigation into the /ɹ/ and /l/ ratings using Takagi's methodology is warranted.

The remaining two $E-J$ contrasts, /ɹ/–/ɰ/ and /v/–/b/, could not be used to test predictions derived from the SLM. This is because the English consonants were not considered to be an example of the Japanese categories to which they were being compared. The cross-language mapping data showed that English /ɹ/ was mapped to Japanese /ɯɾ/ or /ɾ/ and that English /v/ mapped onto Japanese /v/. At the time the experiment was designed, we thought it possible that English /ɹ/ would be considered an example of Japanese /ɰ/ and that English /v/ would be considered an example of Japanese /b/ (see the Introduction), but that was not the case.

IV. GENERAL DISCUSSION

Experiment 2 examined the discriminability of English–English ($E-E$), English–Japanese ($E-J$), and Japanese–Japanese ($J-J$) consonant contrasts. Three groups of native Japanese speakers, who differed in English–language experience, and a native English group, participated. The Japanese groups obtained high discrimination (A') scores for the $J-J$ contrasts, indicating that they fully understood and were able to perform the discrimination test. However, they received low scores on some $E-E$ and $E-J$ contrasts, indicating that these contrasts were difficult to discriminate. In addition, some contrasts were discriminated more accurately by Japanese groups with more English experience than by groups with less English experience, showing an effect of learning. This indicated that some $E-E$ and $E-J$ contrasts were more learnable than others.

A careful assessment of the experiment 1 results provided insight into how the perceived phonetic distance of English and Japanese consonants affected discrimination of English consonants, as well as how phonetic distance affected learning of English consonants over time. Experiment 1 provided the identification and goodness-of-fit of English to Japanese consonants as heard by native speakers of Japanese. The cross-language mapping data obtained in experiment 1 made it possible to explore possible extensions of current models of non-native discrimination and L2 learning.

The question pertaining to the perceptual assimilation model (PAM) (Best, 1995) addressed here was: Can the PAM framework be extended to the naturalistic acquisition of English as an L2? The results obtained here suggest an affirmative answer to this question. PAM was able to predict the discrimination of L2 consonants based on the perceived relationship between English (the L2) and Japanese (the L1) consonants. The results from our study also indicated a minor revision to PAM.

The question pertaining to the speech learning model (SLM) addressed here was: Can the SLM framework be extended to the acquisition of L2 by relatively inexperienced learners of an L2? The results obtained here suggest that the SLM cannot be extended without revision. The SLM was able to predict learnability effects for some $E-J$ contrasts but not others. Only one out of the three contrasts in which the English consonant was perceptually distant from the Japanese consonant to which it was being compared showed learning effects. Perhaps the amount of English experience of the most experienced group (3 years spent living in the United States) was insufficient for a learning effect to emerge for some contrasts. Or, perhaps a more sensitive cross-language mapping methodology might have supported the predictions. Investigation into whether *degree* of perceptual distance (among sounds considered to be poor fits to the L1 sound) is related to the emergence of learning effects could suggest an extension of the SLM. Perhaps those L2 sounds that are more distant from the L1 sounds will be learned earlier than those sounds that are not as distant (although still considered poor fits to the L1 category).

In summary, the results obtained here indicated that certain English consonant contrasts are more difficult for Japanese adults to discriminate than others. The degree of perceptual difficulty seemed to depend on the extent to which the two members of a consonant contrast would be identified as instances of a single Japanese consonant category. Thus, the results obtained here suggest that the PAM framework can be extended to early stages of naturalistic L2 speech learning. However, the relative ''learnability'' of the consonant contrasts did not seem to depend lawfully on the perceived cross-language similarity of English and Japanese consonants. This finding suggests that the SLM cannot be readily extended to early stages of L2 speech acquisition without further investigation.

ACKNOWLEDGMENTS

This research was supported by a grant from the National Institute for Deafness and Other Communicative Disorders (No. DC00257). The authors thank Thorsten Piske, Ratree Wayland, and Jonathan Loftin for their many helpful comments.

[1]The symbol /ɾ/ will be used to represent the Japanese liquid.

[2]Japanese is not usually analyzed as having a /v/ phoneme; however, there is an orthographic representation for /v/ that is used primarily in writing loan words.

[3]It is possible that some identifications were highly agreed upon because they were in fact good tokens of the Japanese category, whereas others were highly agreed upon because there were no other good competitors. The goodness-of-fit scores were collected to help distinguish between these two possibilities. Weighting the identification scores by the goodness-of-fit data served to raise the scores of those identifications that were indeed considered good tokens of the category and to lower the scores of those identifications that were selected because they had no good competitors. The data in Table III show that the identification ratings are *not* perfectly correlated with the fit indexes; the goodness-of-fit data have done the expected job and modified the identifications.

[4]It is noteworthy that when the same procedure was applied to the Japanese consonants, none of them received the designation of ''poor.''

[5]A' was calculated by the following formula provided by Snodgrass *et al.* (1985, p. 451), where H=hit rate (i.e., the proportion of different trials in which the odd item was correctly selected), FA=false alarm rate (i.e., the proportion of catch trials in which an odd item was incorrectly selected).
(1) If $H>FA$ then $A'=0.5+[(H-FA)(1+H-FA)]/[4H(1-FA)]$;
(2) If $H=FA$ then $A'=0.5$;
(3) If $H<FA$ then $A'=0.5-[(FA-H)(1+FA-H)]/[4FA(1-H)]$.

[6]The English consonant is listed first, the Japanese consonant second for all $E-J$ contrasts.

Best, C. (**1990**). ''Adult Perception of Nonnative Contrasts Differing in Assimilation to Native Phonological Categories,'' J. Acoust. Soc. Am. **88**, S177.

Best, C. (**1993**). ''Emergence of Language-Specific Constraints in Perception of Non-Native Speech: A Window on Early Phonological Development,'' *in Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, edited by B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, and J. Morton (Klewer, Dordrecht, The Netherlands).

Best, C. (**1995**). ''A Direct Realist View of Cross-Language Speech Perception,'' *in Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 171–204.

Best, C., Faber, A., and Levitt, A. (**1996**). ''Assimilation of Non-Native Vowel Contrasts to the American English Vowel System,'' J. Acoust. Soc. Am. **99**, 2602(A).

Best, C., and Strange, W. (**1992**). ''Effects of Phonological and Phonetic Factors on Cross-Language Perception of Approximants,'' J. Phonetics **20**, 305–350.

Best, C., McRoberts, G., and Sithole, N. (**1988**). ''The Phonological Basis of Perceptual Loss for Non-Native Contrasts: Maintenance of Discrimination Among Zulu Clicks by English-Speaking Adults and Infants,'' J. Exp. Psychol. Hum. Percept. Perform. **14**, 345–360.

Bradlow, A., Pisoni, D., Akahane-Yamada, R., and Tohkura, Y. (**1997**). ''Training Japanese Listeners to Identify English /r/ and /l/. IV. Some Effects of Perceptual Learning on Speech Production,'' J. Acoust. Soc. Am. **101**, 2299–2310.

Flege, J. (**1995**). ''Second-Language Speech Learning: Theory, Findings and Problems,'' *in Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 233–273.

Flege, J., Takagi, N., and Mann, V. (**1995**). ''Japanese Adults Can Learn to Produce English /ɹ/ and /l/ Accurately,'' Lang Speech **38**, 35–55.

Flege, J., Takagi, N., and Mann, V. (**1996**). ''Lexical Familiarity and English-Language Experience Affect Japanese Adults' Perception of /ɹ/ and /l/,'' J. Acoust. Soc. Am. **99**, 1161–1173.

Goto, H. (**1971**). ''Auditory Perception by Normal Japanese Adults of the Sounds L and R,'' Neuropsychologia **9**, 317–323.

Gottfried, T. (**1984**). ''Effects of Consonant Context on the Perception of French Vowels,'' J. Phonetics **12**, 91–114.

Lado, R. (**1957**). *Linguistics Across Cultures* (University of Michigan Press, Ann Arbor, MI).

Lambacher, S., Marten, W., Nelson, B., and Berman, J. (**1997**). ''Perception of English Voiceless Fricatives by Native Speakers of Japanese,'' in *New Sounds 97: Proceedings of the Third International Symposium on the Acquisition of Second-Language Speech*, edited by J. Leather and A. James (University of Klagenfurt Press, Klagenfurt, Austria), pp. 186–195.

Lively, S., Logan, J., and Pisoni, D. (**1993**). ''Training Japanese Listeners to Identify English /r/ and /l/. II. The Role of Phonetic Environment and Talker Variability in Learning New Perceptual Categories,'' J. Acoust. Soc. Am. **94**, 1242–1255.

Lively, S., Pisoni, D., Yamada, R., Tohkura, Y., and Yamada, T. (**1994**). ''Training Japanese Listeners to Identify English /r/ and /l/. III. Long-Term Retention of New Phonetic Categories,'' J. Acoust. Soc. Am. **96**, 2076–2087.

Logan, J., Lively, S., and Pisoni, D. (**1991**). ''Training Japanese Listeners to Identify English /r/ and /l/: A First Report,'' J. Acoust. Soc. Am. **89**, 874–886.

MacKain, K., Best, C., and Strange, W. (**1981**). ''Categorical Perception of English /r/ and /l/ by Japanese Bilinguals,'' Appl. Psycholing. **2**, 369–390.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J., and Fujimura, O. (**1975**). ''An Effect of Linguistic Experience: The Discrimination of /r/ and /l/ by Native Speakers of Japanese and English,'' Percept. Psychophys. **18**, 331–340.

Mochizuki, M. (**1981**). ''The Identification of /r/ and /l/ in Natural and Synthesized Speech,'' J. Phonetics **9**, 283–303.

Polka, L. (**1995**). ''Linguistic Influences in Adult Perception of Non-Native Vowel Contrasts,'' J. Acoust. Soc. Am. **97**, 1286–1296.

Pruitt, J., Akahane-Yamada, R., Strange, W., Kubo, R., and Takada, T. (**1999**). ''Perceptual Training with Native Japanese Speakers on Various L2 Sounds,'' J. Acoust. Soc. Am. **105**, 1096(A).

Ritchie, W. (**1968**). ''On the Explanation of Phonic Interference,'' Language Learning **18**, 183–197.

Schmidt, A.-M. (**1996**). ''Cross-Language Identification of Consonants. Part 1. Korean Perception of English,'' J. Acoust. Soc. Am. **99**, 3201–3211.

Sekiyama, K., and Tohkura, Y. (**1993**). ''Inter-Language Differences in the Influence of Visual Cues on Speech Perception,'' J. Phonetics **21**, 427–444.

Snodgrass, J., Levy-Berger, G., and Haydon, M. (**1985**). *Human Experimental Psychology* (Oxford University Press, New York).

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S., Nishi, K., and Jenkins, J. (**1998**). ''Perceptual Assimilation of American English Vowels by Japanese Listeners,'' J. Phonetics **26**, 311–344.

Takagi, N. (**1993**). ''Perception of American English /r/ and /l/ by Adult Japanese Learners of English: A Unified View,'' unpublished Univ. of California, Irvine, Ph.D. dissertation.

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

Yamada, R., and Tohkura, Y. (**1992**). ''The Effects of Experimental Variables on the Perception of American English /r/ and /l/ by Japanese Listeners,'' Percept. Psychophys. **52**, 376–392.

Yoshida, K., and Hirasaka, F. (**1983**). ''The Lexicon in Speech Perception,'' Sophia Linguistica **11**, 105–116.