

English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy

Kota Hattori^{a)} and Paul Iverson

Division of Psychology and Language Sciences, University College London, 2 Wakefield Street, London WC1N 1PF, United Kingdom

(Received 1 November 2007; revised 7 October 2008; accepted 8 October 2008)

Native speakers of Japanese often have difficulty identifying English /r/ and /l/, and it has been thought that second-language (L2) learning difficulties like this are caused by how L2 phonemes are assimilated into ones native phonological system. This study took an individual difference approach to examining this relationship by testing the category assimilation of Japanese speakers with a wide range of English /r/-/l/ identification abilities. All Japanese subjects were assessed in terms of (1) their accuracy in identifying English /r/ and /l/, (2) their assimilation of /r/ and /l/ into their Japanese flap category, (3) their production of /r/ and /l/, and (4) their best-exemplar locations for /r/, /l/, and Japanese flap in a five-dimensional set of synthetic stimuli (F1, F2, F3, closure duration, and transition duration). The results demonstrated that Japanese speakers assimilate /l/ into their flap category more strongly than they assimilate /r/. However, there was little evidence that category assimilation was predictive of English /r/-/l/ perception and production. Japanese speakers had three distinct best exemplars for /r/, /l/, and flap, and only their representation of F3 in /r/ and /l/ was predictive of identification ability.

© 2009 Acoustical Society of America. [DOI: 10.1121/1.3021295]

PACS number(s): 43.71.Hw, 43.71.Es [AJ]

Pages: 469–479

I. INTRODUCTION

Adult second-language (L2) learners can have difficulties learning L2 phonemes. For example, native speakers (L1) of Japanese are often poor at identifying and producing the English consonants /r/ and /l/ (e.g., Goto, 1971; Miyawaki *et al.*, 1975), although they can somewhat overcome these problems by receiving phonetic training with talker and stimulus variability (e.g., Hazan *et al.*, 2005; Iverson *et al.*, 2005; Lively *et al.*, 1993; Logan *et al.*, 1991) or by having long periods of exposure to English-speaking environments (e.g., Yamada, 1995). However, even highly experienced Japanese speakers can have difficulty identifying English /r/ and /l/ at a native speaker's level (e.g., Yamada and Tohkura, 1992).

Some current L2 speech perception models attribute such learning difficulties to the similarity between L1 and L2 phonetic or phonemic categories. For example, Best's perceptual assimilation model (PAM) (Best, 1995; Best *et al.*, 2001) states that L2 phonemes are perceived based on their articulatory similarity to the listener's closest L1 phonemes; pairs of L2 phonemes are particularly hard to distinguish when they both sound like equally good examples of the same L1 category. Likewise, Flege's speech learning model (SLM) (Flege, 1995) states that the learner's L1 and L2 categories interact in a common phonological system; learning L2 categories is particularly difficult when there is an existing L1 category that is similar, and learning is easier when the new L2 categories fall into relatively unoccupied regions of the listener's phonological space.

Previous research has suggested that these kinds of similarity relationships may be the cause of Japanese learners' problems with English /r/-/l/. Japanese has an apicoalveolar tap /r/ (Vance, 1987) that has been described as being related to English /r/ and /l/. Best and Strange (1992) hypothesized that both English /r/ and /l/ are perceived as poor exemplars of this Japanese phoneme (or as the Japanese /w/); according to PAM, English /r/ and /l/ are hard to perceptually distinguish because they are the same with regard to the Japanese phonological system. Subsequent assimilation rating experiments have generally supported this view (Guion *et al.*, 2000; Iverson *et al.*, 2003); when Japanese listeners hear English /r/ and /l/ stimuli and are asked how close these are to Japanese phonemes, they rate that both are poorly related to Japanese /r/ or /ur/ (i.e., the Japanese phoneme preceded by a vowel). Aoyama *et al.* (2004) noted that there are trends for Japanese speakers to rate English /l/ as being somewhat closer to Japanese /r/, even though both are poor exemplars, and also found that Japanese children were better at learning English /r/ than /l/. This fits with SLM's claim that learning difficulties are caused by an interaction between L1 and L2 categories; English /r/ is thought to be somewhat easier to learn because the formation of this category is less affected by the listener's existing Japanese /r/ category.

The present study further investigated the role of assimilation by testing Japanese adults who have a wide range of English /r/-/l/ identification abilities and by determining whether these individual differences are related to the degree that they assimilate these phonemes into their L1 /r/ category. If there is a causal relationship between L1 assimilation and L2 category learning, it seems likely that individuals who assimilate these phonemes less strongly into their L1 /r/ category would be better at identifying and producing Eng-

^{a)}Author to whom correspondence should be addressed.

lish /ɹ/ and /l/. That is, inexperienced Japanese learners of English have been claimed to assimilate all of these phonemes into the same category (e.g., Best and Strange, 1992). Individuals who have learned to accurately identify English /ɹ/ and /l/ may thus have learned to dissimilate them from Japanese /r/ (e.g., Flege, 1995; Flege *et al.*, 2003), allowing these individuals to create three independent categories within the same phonological space. Alternatively, it is possible that individuals who are more accurate with English /ɹ/ and /l/ have changed their assimilation patterns asymmetrically. For example, auditory training results (Iverson *et al.*, 2005) suggest that Japanese listeners who are learning English /ɹ/-/l/ may actually increase their assimilation of /l/ into the Japanese /r/ category, but not change their assimilation of /ɹ/. Successful learners may thus merge English /l/ and Japanese /r/ into a single compromise category rather than dissimilate (see Flege, 2003; MacKay *et al.*, 2001), and learn a separate category for English /ɹ/.

Previous work has quantified assimilation by having individuals identify and rate which L1 categories are most similar (e.g., Guion *et al.*, 2000; Iverson *et al.*, 2003), but these types of subjective ratings may not be appropriate for individual difference studies. For example, if a Japanese listener rates that /l/ is more similar to Japanese /r/ than does another Japanese listener, it is not clear whether the difference between their rating scores would reflect a real difference in the degree of assimilation or a superficial difference in their use of the scale (e.g., a bias to use the middle of the scale rather than the extremes). Such biases are essentially factored out in traditional studies that average across individuals. Alternatively, acoustic analyses of productions (e.g., Lotto *et al.*, 2004) could be used to measure category assimilation (e.g., one could measure whether an individual's production of English /l/ is acoustically distinct from their Japanese /r/). However, such a measure has at least three disadvantages. One is that English /ɹ/, /l/, and Japanese /r/ cannot always be compared on common acoustic dimensions (e.g., Japanese /r/ productions can sometimes have a burst that is not typical in English /ɹ/ and /l/). Another disadvantage is that between-talker differences (e.g., vocal tract length and speaking rate) introduce variability that is hard to control. Finally, speech production does not always correlate strongly with speech perception (e.g., Bradlow *et al.*, 1997), so it is not clear whether acoustic analyses of speech production would be relevant for perception.

We instead took two new approaches to measuring category assimilation. Our first approach was to use a bilingual identification task including English /ɹ/, English /l/, and Japanese /r/ as response alternatives on the same trial. This allowed us to determine, for example, the extent to which individuals actually confused English /l/ with Japanese /r/, rather than measuring whether they are subjectively related. If English /l/ and Japanese /r/ were frequently confused, this would indicate that the stimuli mapped onto the same underlying phonetic categories (i.e., strong category assimilation). This method, however, may be less effective for measuring weaker degrees of assimilation, where stimuli are subjectively judged to be a poor exemplar of an L1 category without being similar enough to actually sound like that category.

Our second approach was to have individuals find best exemplars for /ɹa/, /la/, and /ra/ in a large five-dimensional acoustic space that varied F1, F2, and F3 frequencies during the initial closure, and the durations of the closure and transition, using efficient search methods previously developed for vowels (Evans and Iverson, 2004, 2007; Iverson and Evans, 2003, 2007; Iverson *et al.*, 2006). The approach is similar to acoustically measuring productions, but it has the advantages of mapping all ratings onto the voice of a single speaker and a set of common dimensions, as well as being a perceptual task that may thus be more relevant to identification accuracy. If listeners have similar best exemplars for English /l/ and Japanese /r/, for example, this would again suggest that they have a single category representation for each phoneme.

The same set of Japanese adults participated in a series of three experiments. Experiment 1 measured their baseline identification ability for English /ɹ/-/l/, and used a three-alternative bilingual identification task to measure assimilation. Experiment 2 assessed the spoken accent of these subjects. Experiment 3 mapped best exemplars of the L1 and L2 phonemes in a five-dimensional space. The aim was to determine whether the degree to which Japanese adults assimilate English /ɹ/ and /l/ into their Japanese /r/ category is related to their ability to identify and produce these English phonemes.

II. EXPERIMENT 1: PHONEME IDENTIFICATION

The baseline English /ɹ/-/l/ identification abilities of subjects were assessed by eliciting forced-choice /ɹ/-/l/ judgments for initial-position minimal-pair words (e.g., *rock* or *lock*) that were recorded from multiple native speakers of British English.

The degree of assimilation between the English and Japanese phonemes was measured by having listeners make a three-way bilingual forced-choice judgment between English /ɹ/, English /l/, and Japanese /r/. In the interests of removing any cues for whether the stimuli were from English or Japanese that were not related to the consonant, CV syllables were used with vowels that are common to both languages (i.e., /i/, /e/, /a/, /o/, and /u/), and highly skilled Japanese-English bilingual speakers recorded all phonemes, rather than having separate groups of Japanese and English speakers. Listeners heard all stimuli mixed in the same block, and judged which of the three consonants they heard.

A. Method

1. Subjects

Thirty-nine adult native speakers of Japanese were tested; 3 were omitted from the data due to computer problems, leaving a total of 36 participants. Their ages ranged from 19 to 48 years (median=25 years), they started learning English between 6 and 13 years (median=13 years), and had received English instruction for 7–25 years (median=9 years). All participants were brought up in monolingual environments in Japan, and all participants were tested in London. They had lived in English-speaking countries be-

tween 1 month and 13 years (median=3 months). None of the participants reported having hearing problems.

2. Stimuli and apparatus

Stimulus recordings were made in an anechoic chamber with 44 100 16 bit samples/s. The stimuli were recorded with Radio Spares (RS) 249-946 microphone, Edirol USB Audio Capture UA-25, and Dell Optiplex GX 260. The stimuli were played to subjects using Sennheiser HD 280 headphones while subjects sat in a sound-treated room. Responses were recorded using PRAAT (Boersma and Weenik, 2008).

The English stimuli were initial-position /ɹ/-/l/ minimal-pair words (e.g., rock and lock) used previously by Iverson *et al.* (2005). Four British English speakers (two male and two female) recorded a total of 120 minimal-pair words (words were not repeated between speakers).

The bilingual stimuli were 15 CV syllables created by combining three consonants (i.e., /ɹ/, /l/, and /r/) and five vowels (i.e., /i/, /e/, /a/, /o/, and /u/). Five English-Japanese bilingual speakers made the recordings. Two were native speakers of British English (one male and one female) who had majored in Japanese at UK universities and had studied abroad in Japan. The others (two male and one female) were native speakers of Japanese who had lived in the UK for more than 15 years; two of them moved to the UK before age 5, and the other speaker immigrated to the UK in his 30 s. All speakers recorded two repetitions of each syllable. All stimuli were screened by native British and Japanese phonetically trained listeners, to ensure that all consonants sounded like normal native versions. One speaker (native Japanese) was unable to correctly produce the Japanese /r/ followed by /i/; these recordings were omitted, leaving a total of 148 stimuli (15 syllables × 5 speakers × 2 repetitions – 2 errors).

B. Procedure

1. English /ɹ/-/l/ identification

On each trial, the participants heard one stimulus word and clicked on a button to indicate whether it began with /ɹ/ or /l/. They heard each stimulus once and did not receive feedback. The experiment began with a practice block of 10 trials, and was followed by 120 experimental trials that were blocked by talker but were otherwise in a random order.

2. Bilingual /ɹ/-/l/-/r/ identification

On each trial, the participants heard a syllable and clicked on a button to indicate whether it began with English /ɹ/, English /l/, or Japanese /r/. These options were spelled out on the screen as “R,” “L,” and “JR;” the categories were described to subjects to ensure that they understood which categories were indicated. They heard each stimulus once and did not receive feedback. The experiment began with a practice block of 30 trials, and was followed by 148 trials that were blocked by talker but were otherwise in a random order.

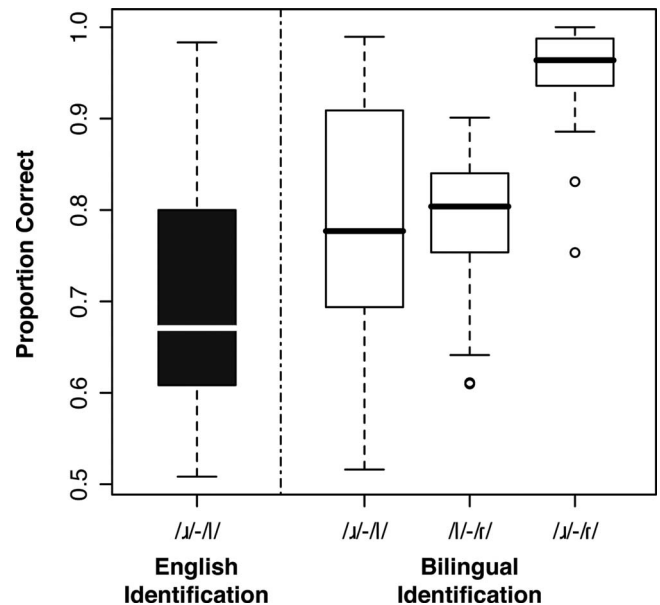


FIG. 1. Boxplots of English /ɹ/-/l/ identification and bilingual /ɹ/-/l/-/r/ identification. The bilingual identification results are presented as pairwise confusions for /ɹ/-/l/, /r/-/l/, and /l/-/ɹ/, with a higher proportion correct indicating less category assimilation. Boxplots display the quartile ranges of scores, with outliers marked by circles.

C. Results

Figure 1 displays the percent-correct English /ɹ/-/l/ identification. As intended, there was a wide range of scores for individual speakers, ranging from near chance to near 100% correct. On average, /ɹ/ was identified correctly on 71% of trials and /l/ was identified correctly on 67% of trials, demonstrating that listeners did not have a strong bias to identify these stimuli as one phoneme or the other.

Table I displays the mean confusion matrix for the bilingual identification task. Japanese listeners demonstrated confusion between English /ɹ/ and /l/; they misidentified /ɹ/ as /l/ on 16% of the trials and /l/ as /ɹ/ on 22% of the trials. They also demonstrated category assimilation between /l/ and Japanese /r/; they identified /l/ as Japanese /r/ on 19% of the trials and Japanese /r/ as /l/ on 17% of the trials. However, they rarely made errors between /ɹ/ and Japanese /r/; they identified /ɹ/ as Japanese /r/ on 2% of the trials and Japanese /r/ as /ɹ/ on 6% of the trials.

In order to examine individual differences in these confusions, we calculated the similarity between individual phoneme pairs (i.e., /ɹ/-/l/, /l/-/r/, and /ɹ/-/r/) from these 3 × 3 confusion data. The calculation of pairwise similarity from larger matrices has been a common issue when analyzing consonant confusion data using multidimensional scaling (e.g., Shepard, 1972; Walden *et al.*, 1980) and hierarchical

TABLE I. Confusion matrix for bilingual identification.

Stimulus	Response		
	/ɹ/	/l/	/r/
/ɹ/	82%	16%	2%
/l/	22%	58%	19%
/r/	6%	17%	77%

cluster analysis (e.g., Iverson *et al.*, 1998) or when applying models of audiovisual integration (e.g., Braidá, 1991; Iverson, 2002). Researchers have mostly used straightforward averages of the confusion data to measure similarity [e.g., the distance between /ɪ/ and /ɪ/ can be measured by summing the number of trials in which an /ɪ/ stimulus was identified as /ɪ/ or an /ɪ/ stimulus was identified as /ɪ/; e.g., see Shepard (1972) and Walden *et al.* (1980)]. Alternatively, more theoretical decision models, based on signal detection theory (e.g., Braidá, 1991) or Luce's choice theory (e.g., Iverson, 2002), can be applied to correct for response bias. We tried both approaches for the present data and they produced results that were substantively the same. The data reported here are from the more straightforward measure. Specifically, a pairwise percent-correct measure was calculated by summing the correct responses for a pair (e.g., /ɪ/ identified as /ɪ/, and /ɪ/ identified as /ɪ/) and dividing by the total number of correct and incorrect responses for that pair (e.g., /ɪ/ identified as /ɪ/ and /ɪ/ identified as /ɪ/), omitting all trials in which the stimulus or response was from the other category (e.g., /ɪ/).

As displayed in Fig. 1, Japanese speakers demonstrated substantial variability in /ɪ/-/ɪ/ identification of bilingual stimuli, as was found for English /ɪ/-/ɪ/ identification. Identification accuracy was significantly higher for the bilingual than the English stimuli, $t(35)=-7.37$, $p<0.01$, but there was a high correlation between the individual differences for both measures, $r=0.84$, $p<0.01$. Thus, the bilingual stimulus materials seemed to capture most of the individual differences that were measured in the monolingual English task, although the stimuli may have been somewhat easier to identify, at least when those that sounded like /ɪ/ had been excluded.

Individuals confused /ɪ/-/ɪ/ at a rate that was not significantly different than their bilingual /ɪ/-/ɪ/ confusions, $t(35)=-0.03$, $p>0.05$, confirming that English /ɪ/ is perceived as similar to the Japanese /ɪ/. However, there was no significant correlation between individual differences in English /ɪ/-/ɪ/ identification and the degree of /ɪ/-/ɪ/ assimilation, $r=0.22$, $p>0.05$, suggesting that /ɪ/-/ɪ/ category assimilation is not predictive of English /ɪ/-/ɪ/ identification. An inspection of the scatterplot for these two measures in Fig. 2 confirms this view; individuals who were relatively poor at recognizing English /ɪ/ and /ɪ/ had a range of /ɪ/-/ɪ/ category assimilation scores, as did individuals who were better at English /ɪ/-/ɪ/ identification. It is thus questionable whether the degree of assimilation between /ɪ/-/ɪ/ is the cause of difficulties in English /ɪ/-/ɪ/ identification, despite the fact that Japanese adults evidently perceive /ɪ/ and /ɪ/ to be very similar.

Japanese speakers infrequently assimilated English /ɪ/ to Japanese /ɪ/, doing so at a significantly lower rate than for /ɪ/-/ɪ/, $t(35)=-9.25$, $p<0.01$. However, the degree of assimilation was significantly correlated with English /ɪ/-/ɪ/ identification, $r=0.44$, $p<0.01$, suggesting that Japanese listeners whose English /ɪ/ category was more similar to the /ɪ/ category had more problems with English /ɪ/-/ɪ/ identification. Although this may seem to indicate that /ɪ/-/ɪ/ assimilation is the cause of /ɪ/-/ɪ/ identification difficulties, an inspection of the scatterplot in Fig. 2 weakens this view. That is, individuals who were poor at identifying English /ɪ/ and /ɪ/ had a

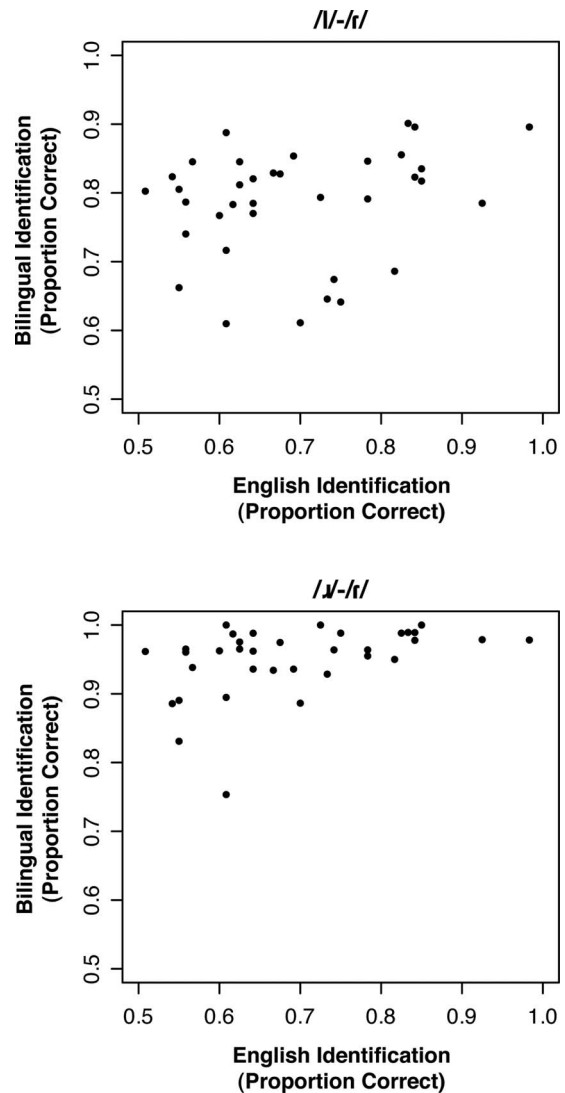


FIG. 2. Scatterplots of the relationships between English /ɪ/-/ɪ/ identification accuracy with the identification of bilingual stimuli for /ɪ/-/ɪ/ and /ɪ/-/ɪ/. Higher bilingual identification scores indicate less confusion between the categories, and a lower degree of category assimilation.

range of /ɪ/-/ɪ/ scores, with many of these individuals rarely confusing /ɪ/ and /ɪ/; it is hard to argue that /ɪ/-/ɪ/ assimilations cause their English /ɪ/-/ɪ/ problems given that these assimilations were not made very frequently, and the more strongly assimilated pair, /ɪ/-/ɪ/, was not correlated with identification accuracy. The frequency of /ɪ/-/ɪ/ assimilation may indicate instead that there is a problem with the underlying /ɪ/ representation, without this problem actually being caused by the proximity of /ɪ/.

Given that English /ɪ/ is similar to both English /ɪ/ and Japanese /ɪ/, it may seem paradoxical that /ɪ/ and /ɪ/ are relatively dissimilar. Part of the answer may be that Japanese listeners use temporal cues to distinguish these consonants (e.g., transition duration) more than they use spectral cues (e.g., F3) (Iverson *et al.*, 2005; Yamada, 1995), and these temporal cues have natural variation, with /ɪ/ having short transitions, /ɪ/ having longer transitions, and /ɪ/ being in between (e.g., see Experiment 3 results). That is, /ɪ/ and /ɪ/ may sometimes be confused because some of these phonemes have overlapping transition durations (e.g., some /ɪ/ have

short enough transitions to sound like /l/, as do /l/ and /r/ (e.g., some /r/ have long enough transitions to sound like /l/), but there may still be little overlap between /ɹ/ and /r/ (i.e., few /ɹ/ have short enough transitions to sound like /r/, and few /r/ have long enough transitions to sound like /ɹ/). In other words, it is possible that /ɹ/ and /r/ are dissimilar even though both are close to /l/, because of the way in which their acoustic distributions overlap; the result is only paradoxical if one conceives of these phonemes simplistically, as single points in a geometric space.

III. EXPERIMENT 2: RATINGS OF SPOKEN ACCENT

Experiment 1 demonstrated that the relationships between category assimilation and English /ɹ/-/l/ identification were weak. However, it is possible that category assimilation could have a greater role in affecting /ɹ/-/l/ production; much of the data supporting SLM has come from production data (e.g., [Flege, 1987, 1995, 2003](#); [Bohn and Flege, 1992](#)), and it has often been found that production and perception accuracy for L2 phonemes are only weakly correlated (e.g., [Bradlow et al., 1997](#)). We thus decided to include a short test of the subjects' production ability; subjects read an accent-revealing sentence (*The red robin looked across the lovely lake*) and phonetically trained listeners rated the accent of the /ɹ/ and /l/ consonants, as well as the degree of contrast that the talker made between /ɹ/ and /l/.

A. Method

1. Subjects

Same as Experiment 1.

2. Apparatus

Same as Experiment 1.

3. Procedure

Each subject was recorded reading the sentence “The red robin looked across the lovely lake.” After all recordings had been made, three phonetically trained British English speakers made three ratings for each recording: degree of /ɹ/ accent, degree of /l/ accent, and the degree of contrast that the speaker made between /ɹ/ and /l/. We have used the contrast rating in previous work ([Iverson et al., 2008](#)), and it is useful in cases where speakers do not produce phonemes in a nativelike way, but nonetheless produce a phonetic contrast that could make their speech intelligible to a listener. The ratings were made on an integer scale from 1 (*poor*) to 7 (*good*). The raters were allowed to repeatedly listen to the stimuli before giving their judgment. The rating scores were averaged across the raters for each recording to provide a measure of the degree of accent.

B. Results

As displayed in Fig. 3, the accents of the Japanese speakers varied across the entire range in terms of the degree of contrast and the /ɹ/ accent. They were, as a group, more consistently able to produce /l/ in a way that was closer to the “good” end of the scale.

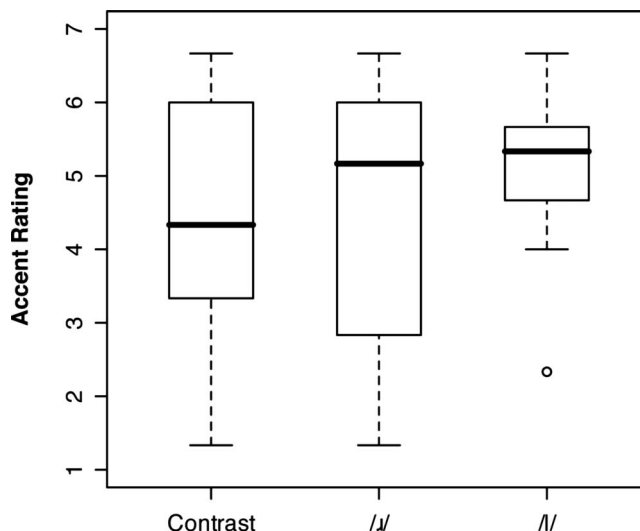


FIG. 3. Boxplots of the rated accent for Japanese speakers saying “The red robin looked across the lovely lake.” Three ratings were made for each sentence: the degree of /ɹ/-/l/ contrast, /ɹ/-accent, and /l/-accent. Ratings were made on a scale from 1 (poor, indicating a strong non-native accent or poor contrast) to 7 (good, indicating little non-native accent or high contrast).

The baseline English /ɹ/-/l/ identification accuracy scores of Experiment 1 were significantly correlated with the accent ratings for /ɹ/ accent, $r=0.53$, $p<0.01$, and /ɹ/-/l/ contrast, $r=0.42$, $p<0.01$. However, there was no significant correlation between /l/ accent and English /ɹ/-/l/ identification, $r=0.20$, $p>0.05$. This pattern of results is in accord with previous findings that perceptual identification is moderately related to the degree of spoken accent (e.g., [Bradlow et al., 1997](#)).

The degree of assimilation for /l/-/r/ (Experiment 1) was not significantly correlated, $p>0.05$, with /ɹ/ accent, $r=0.13$, /l/ accent, $r=0.19$, or /ɹ/-/l/ contrast, $r=0.08$. Likewise, the degree of assimilation for /ɹ/-/r/ was not significantly correlated, $p>0.05$, with /ɹ/ accent, $r=0.29$, /l/ accent, $r=0.17$, or /ɹ/-/l/ contrast, $r=0.23$. Thus, similar to Experiment 1, there is little evidence that the degree of perceptual assimilation between English /ɹ/ and /l/ with Japanese /r/ has a substantial role in causing the individual differences in /ɹ/-/l/ production abilities.

IV. EXPERIMENT 3: PERCEPTUAL MAPPING OF BEST EXEMPLARS

This experiment measured the cues that the Japanese individuals use to represent /ɹ/, /l/, and /r/, in order to better understand the assimilation process, as well as attempt to measure assimilation in a way that better predicts the ability of individuals to identify English /ɹ/ and /l/. The experiment adapted the goodness optimization method that we have used recently to map best-exemplar locations for vowels in high-dimensional, phonetically detailed, stimulus spaces ([Evans and Iverson, 2004, 2007](#); [Iverson and Evans, 2003, 2007](#); [Iverson et al., 2006](#)). On each trial, subjects see a target syllable on the computer screen, hear a synthesized stimulus embedded in a natural carrier sentence, and rate on a continuous scale how close the stimulus that they hear is to

being a good exemplar of the syllable on the screen. The algorithm then adjusts the stimulus that they hear on successive trials to iteratively find a good exemplar of that syllable. In this way, we have been able to find best exemplars in very large sets of stimuli after 35 trials.

This approach has been useful for assessing cross-language differences in vowel spaces (Iverson and Evans, 2007). The best exemplars for 14 English vowels were found for L1 speakers of Spanish, French, German, and Norwegian, and the accuracy with which their best exemplars represented static formant targets and formant movement were highly correlated with their ability to recognize natural vowel stimuli.

The present experiment used a similar method to find best exemplars of /ɹ/, /l/, and /r/ in a five-dimensional stimulus space that included F1, F2, and F3 frequencies during the closure, and the closure and transition duration. Although F3 is the primary cue that native speakers use to distinguish /ɹ/ and /l/, F2 and transition duration are often considered to be secondary cues (e.g., Underbakke *et al.*, 1988); these secondary cues can be particularly important to Japanese adults, who typically have difficulty hearing F3 variation (e.g., Iverson *et al.*, 2003; Underbakke *et al.*, 1988; Yamada, 1995). Closure duration and F1 were included to allow good examples of /r/ to be found; /r/ is often described as being much more rapid than /ɹ/ or /l/ (e.g., Miyawaki *et al.*, 1975), and occasionally has a stoplike quality that could possibly lower F1. We did not include a burst dimension in this space, because our pilot investigations showed that many speakers do not produce /r/ with a burst, and it was possible to synthesize a subjectively good exemplar of /r/ without it.

A. METHOD

1. Participants

The Japanese speakers were the same as in Experiments 1 and 2. We additionally tested 13 native speakers of British English to provide normative data. Their ages ranged between 18 and 62 years (median=24), and all were born and raised in southern England. No listeners reported having hearing problems.

2. Apparatus

The stimuli were played to subjects using Sennheiser HD 280 headphones while subjects sat in a sound-treated room. Responses were entered and recorded using a custom-written program which also controlled the presentation of stimuli.

3. Stimuli

The stimuli were synthetic C-/a/ syllables embedded in naturally spoken English and Japanese carrier sentences (i.e., say [] again, and *mata [] to itte kudasai*). The same female speaker was used for both carrier sentences, an English/Japanese bilingual who spent her childhood in both the UK and Japan, and is highly fluent in both languages.

The synthetic syllables were modeled based on this natural talker, using a Klatt synthesizer (Klatt and Klatt, 1990). A five-dimensional set of stimuli was created by or-

thogonally varying F1, F2, F3, closure duration, and transition duration (i.e., from the consonantal articulation to the following vowel). The values were chosen so that they would span an acoustic space that included /ɹ/, /l/, and /r/, as well as match the target speaker's voice. F1 varied from 123 to 603 Hz. F2 was always at least 1 ERB (Glasberg and Moore, 1990) greater than F1, and less than 2489 Hz. F3 was always at least 1 ERB greater than F2, and less than 3951 Hz. The closure and transition durations varied from 5 to 290 ms. The frequency values were quantized in 1-ERB steps and the duration values were quantized with a log spacing (12 steps for each). All other synthesis parameters were the same for all stimuli, and were set so that they closely matched the original F0 contour, amplitude envelope, long-term average spectra, and vowel formants of the original syllables produced by this talker. There were a total of 60 660 stimuli for each carrier sentence.

B. Procedure

On each trial, subjects saw a consonant printed on the computer screen (e.g., R) and heard a stimulus (synthesized CV embedded in a natural carrier sentence). They rated on a continuous scale how far away the CV that they heard was from being a good exemplar of the printed consonant. Their ratings were given by mouse clicking on a continuous bar presented on a computer screen. Japanese subjects found best exemplars for all three consonants and English subjects found best exemplars for only /ɹ/ and /l/. All subjects performed this task in both the Japanese and English carrier sentences, even though the English speakers did not understand Japanese. Both languages were included in case there were highly language-specific best-exemplar locations (e.g., it was possible that Japanese subjects would have trouble finding a good /r/ when listening to English). However, this kind of language variation was not found (i.e., people appeared to choose the same kinds of best exemplars in both sentences), so the results were averaged across carrier sentence for later analysis.

A goodness optimization procedure (Evans and Iverson, 2004, 2007; Iverson and Evans, 2003, 2007; Iverson *et al.*, 2006) was used to iteratively change the stimuli that subjects heard on each trial, to search through the multidimensional stimulus space for better exemplars. This procedure involves searching along individual vectors (i.e., one-dimensional straight-line paths crossing through the five-dimensional stimulus space) and finding the best exemplar on each vector. There were a total of 7 search vectors and 5 trials per vector for each consonant.

Vector 1 covaried all five dimensions along a straight-line path that passed through two points: (1) measurements based on the natural productions of each consonant, and (2) a middle point in the stimulus space (i.e., values averaged for all three stimuli). Vectors 2–6 varied one dimension at a time, in a descending order based on their presumed importance to /ɹ/-/l/ identification (F3, F2, transition duration, closure duration, and F1). Vector 7 varied all dimensions, passing through the best value found thus far on each dimension and the middle point in the space. The endpoints of all vec-

tors were constrained by the boundaries of the stimulus space so that listeners were able to choose stimuli over a wide range.

For each vector, subjects first heard stimuli at the two endpoints (i.e., at the boundaries of the stimulus space), with the order of these two trials randomized. The selection of stimuli on the remaining trials was based on the subjects' judgments, using formulas that were designed to find stimuli along the path that would be perceived as better exemplars. On the third trial, subjects heard a stimulus that was selected by a goodness-weighted average of the first two stimuli. On the fourth and fifth trials, the stimuli were selected by fitting a parabola to the goodness data and finding the minimum of this function. See Iverson and Evans (2003) for exact formulas and procedures. At the completion of the fifth trial, subjects were allowed to repeat the search if it had produced a poor exemplar (i.e., they made an explicit judgment about whether the stimulus was or was not close to sounding good). If the best exemplar was judged to be close, the parameters of the best stimulus found thus far were passed onto the next stage of the search algorithm (i.e., to search along the next vector).

C. Results

Figure 4 displays the best exemplars of English /ɪ/, /ɪ/, and Japanese /ɪ/. The results demonstrate that /ɪ/ is more similar to /ɪ/ than /ɪ/ in terms of F3 and transition duration, in accord with the asymmetric pattern of assimilation found in Experiment 1. However, /ɪ/ was different from both /ɪ/ and /ɪ/ in terms of F2, and was possibly closer to /ɪ/ in terms of closure duration. One-way repeated-measures analyses of variance (ANOVAs) were run separately for each dimension to determine whether these three best exemplars were significantly different for the Japanese listeners. The dependent variables were the five acoustic dimensions (i.e., F1, F2, F3, closure duration, and transition duration), and consonant was coded as a three-level within-subject variable (i.e., English /ɪ/, /ɪ/, and Japanese /ɪ/). There were no significant differences for F1, $p > 0.05$. However, the consonants differed significantly in F2, $F(2, 70) = 28.48$, $p < 0.01$, F3, $F(2, 70) = 129.28$, $p < 0.01$, closure duration, $F(2, 70) = 6.36$, $p < 0.01$, and transition duration, $F(2, 70) = 68.45$, $p < 0.01$. Tukey HSD comparisons revealed that /ɪ/ was significantly different from both /ɪ/ and /ɪ/ in terms of F2, F3, and transition duration, $p < 0.05$, but not F1 or closure duration; /ɪ/ and /ɪ/ were significantly different in terms of F2, F3, closure duration, and transition duration, $p < 0.05$.

The results thus demonstrate that Japanese adults were able to have three distinct representations for these categories (e.g., not assimilating /ɪ/ and /ɪ/), despite the fact that many of these listeners had difficulty recognizing natural /ɪ/ and /ɪ/ stimuli. However, it is clear that there were substantial individual differences in their best exemplars. For example, the closure durations of best exemplars for English /ɪ/ and /ɪ/ varied widely and overlapped with the range of Japanese /ɪ/. In order to examine whether their degree of separation between L1 and L2 categories is predictive of their English /ɪ/-/ɪ/ identification accuracy, the absolute values for their

difference for /ɪ/-/ɪ/ and /ɪ/-/ɪ/ were computed for the five acoustic dimensions. On all acoustic measures, there were no significant correlations between /ɪ/-/ɪ/ difference and English /ɪ/-/ɪ/ identification (F1, $r = -0.15$; F2, $r = -0.17$; F3, $r = 0.10$; closure duration, $r = 0.08$; transition duration, $r = -0.09$, $p > 0.05$), and no significant correlations between the /ɪ/-/ɪ/ distance and English /ɪ/-/ɪ/ identification (F1, $r = -0.11$; F2, $r = -0.03$; F3, $r = 0.25$; closure duration, $r = -0.30$; transition duration, $r = 0.12$, $p > 0.05$). The results are thus in accord with the findings of Experiment 1; there is little evidence that the similarity of a listener's L1 and L2 categories causes individual differences in /ɪ/-/ɪ/ identification accuracy.

The best exemplars chosen by Japanese speakers for /ɪ/ and /ɪ/ were surprisingly similar to those of English speakers, considering their levels of /ɪ/-/ɪ/ identification accuracy. ANOVA analyses tested whether these best exemplars varied with language group for each acoustic dimension. The acoustic values were entered in separate analyses as dependent measures, consonant (/ɪ/ or /ɪ/) was a within-subject factor, and language group (Japanese or English) was a between-subject factor. There was a significant interaction between consonant and language group for F3, $F(1, 94) = 5.03$, $p < 0.05$; English speakers systematically chose higher F3 frequencies for /ɪ/ than did Japanese speakers, but their values were similar for /ɪ/. This indicates that Japanese adults did not have completely nativelike representations for F3. There was no significant interaction between language and consonant for the other acoustic dimensions, $p > 0.05$, and there were no main effects of language. Unsurprisingly, there were main effects of consonant for F3, $F(1, 94) = 351.85$, $p < 0.01$, closure duration, $F(1, 94) = 12.73$, $p < 0.01$, and transition duration, $F(1, 94) = 97.83$, $p < 0.01$, demonstrating that English /ɪ/ and /ɪ/ differ on these dimensions. There were no significant main effects of consonant for F1 and F2, $p > 0.05$.

Pearson correlations were used to determine whether the accuracy with which Japanese listeners represent /ɪ/ and /ɪ/ (i.e., their similarity to the average best exemplars chosen by native English speakers) was predictive of identification accuracy. The accuracy of the representations was measured for each of the five acoustic dimensions, combining /ɪ/ and /ɪ/ using a Euclidean metric. For example, we calculated how far each individual's /ɪ/ and /ɪ/ best exemplars were from the English averages on the F3 dimension, and then combined these two values by calculating the square root of the sum of squares. There was a significant correlation between F3 accuracy and /ɪ/-/ɪ/ identification, $r = -0.46$, $p < 0.01$; Japanese individuals who had F3 representations that were similar to those of native English speakers were better at English /ɪ/-/ɪ/ identification. However, there were no significant correlations between the other four dimensions and /ɪ/-/ɪ/ identification, $p > 0.05$.

Figure 5 displays the scatterplot of F3 accuracy versus /ɪ/-/ɪ/ identification. The significant correlation occurred because individuals who were accurate in their /ɪ/-/ɪ/ identification tended to have nativelike F3 values, whereas less accurate subjects were more variable in their F3 values, with some of these individuals having nativelike F3 values and

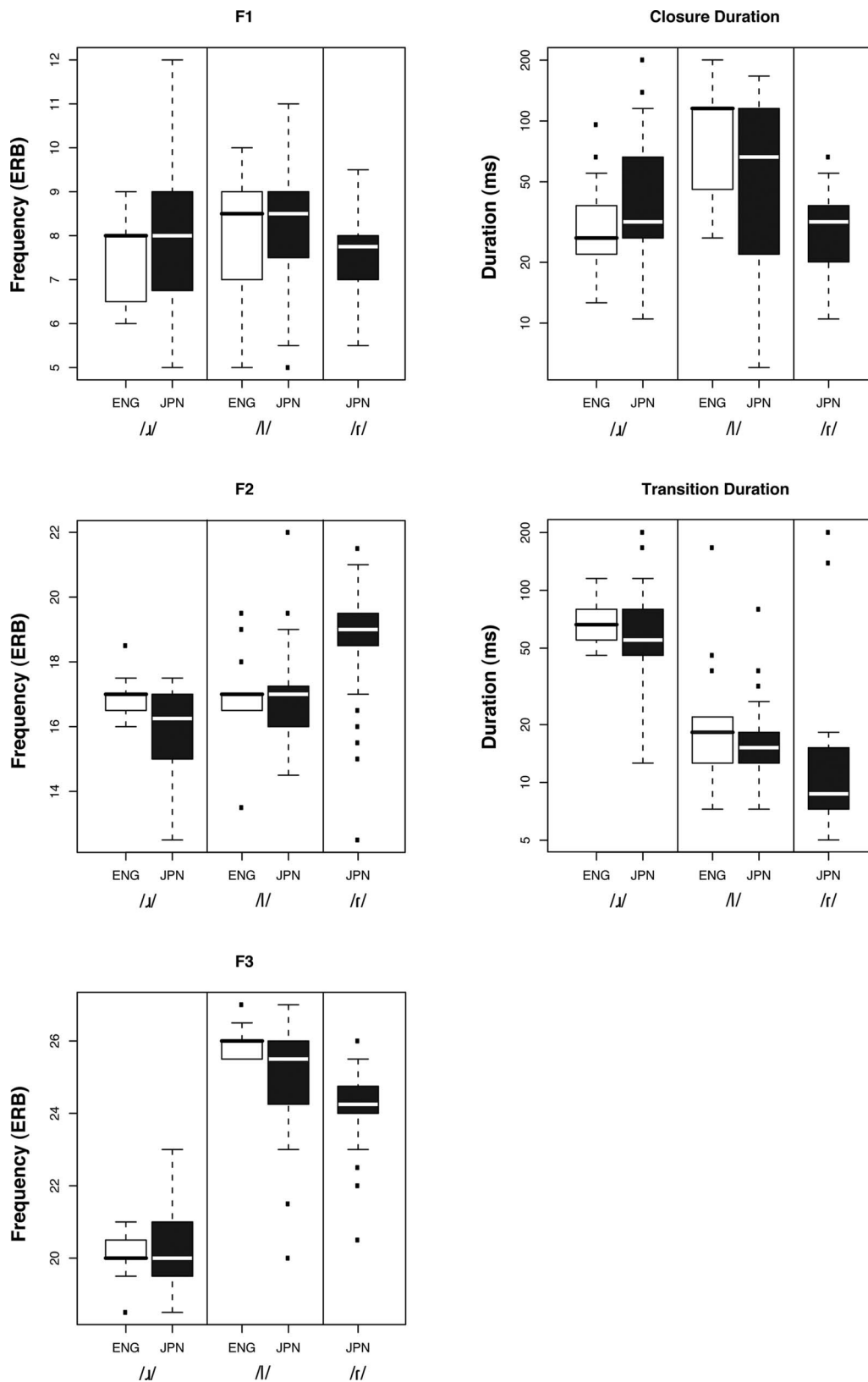


FIG. 4. Boxplots of the acoustic parameters of the best exemplars for /ɹ/, /l/, and /r/ by L1 speakers of English (ENG) and Japanese (JPN).

others having ones that were far away. This may have occurred because the less accurate subjects either had an unclear representation of F3 or had a perceptual difficulty hearing variation along this dimension (e.g., Iverson *et al.*, 2003).

That is, less accurate subjects may have simply been more random about which F3 values they said they preferred, rather than systematically preferring values that were far from those of native listeners.

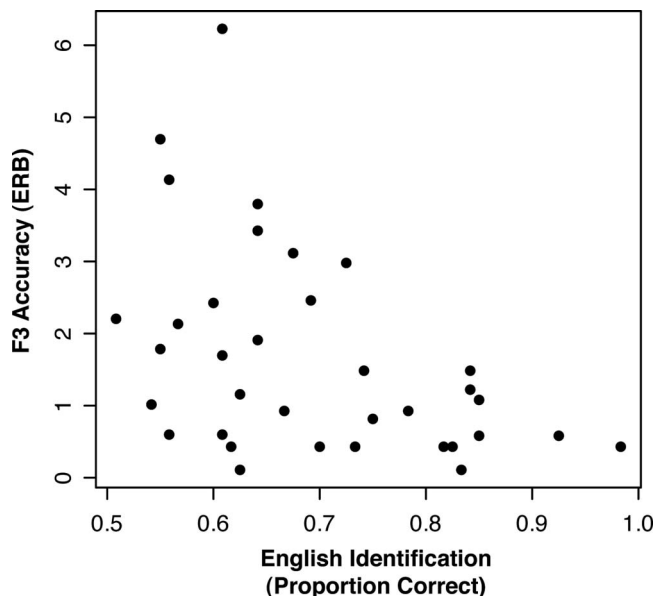


FIG. 5. Scatterplot of the relationship between the accuracy of the English /r/ and /l/ best exemplars for Japanese listeners along the F3 dimension (i.e., distance from English averages) and their English identification accuracy.

V. GENERAL DISCUSSION

The present results confirm that English /l/ is assimilated into the Japanese /r/ category, at least to some degree. That is, listeners confused /l/ and /r/ in a forced-choice task as much as they confused /ɹ/ and /l/, but their best exemplars for /l/ were not identical to those for /r/, even though they were broadly similar on several dimensions (e.g., F3 and transition duration). The assimilation of /ɹ/ and /r/ was comparatively weak; individuals confused /ɹ/ and /r/ infrequently in a forced-choice task and chose best exemplars for /ɹ/ that were different from /r/ in terms of F3 and transition duration. The results thus suggest that /ɹ/ and /l/ differentially assimilate into the Japanese /r/ category, with /l/ being closer, supporting the claims of Aoyama *et al.* (2004). It is possible that previous work using subjective ratings of category goodness (e.g., Guion *et al.*, 2000; Iverson *et al.*, 2003) did not find a clearer asymmetry because it is hard for individuals to rate differences in category goodness when both tokens are poor exemplars overall.

Within the classification scheme of PAM, this appears to be a Category Goodness (CG) difference or an Uncategorized versus Categorized type (UC) (Best, 1995; Best *et al.*, 2001), not the single category assimilation type hypothesized by Best and Strange (1992). The problem for PAM is that CG and UC contrasts are predicted to be relatively easy to discern; in both cases the two phonemes should be perceived as being different because they differentially assimilate into the L1 phonological system. However, the English /ɹ/-/l/ contrast perceived by Japanese listeners is one of the most difficult contrasts reported in the literature (e.g., Goto, 1971; Miyawaki *et al.*, 1975). In combination with previous work, demonstrating that listeners can hear acoustic variation in /ɹ/ and /l/ that does not affect perceived goodness (Iverson *et al.*, 2003), it thus seems clear that PAM does not explain why Japanese listeners have difficulty with English /ɹ/-/l/.

The results of this study are also problematic for SLM. Previous work supporting SLM has suggested that language learners merge (Flege, 2003; MacKay *et al.*, 2001) or dissimilate (Flege, 1995; Flege *et al.*, 2003) L1 and L2 phonetic categories that are close enough to interact, so we expected that Japanese learners who were better at identifying /ɹ/ and /l/ would have assimilation patterns that would be distinct from those who identified /ɹ/ and /l/ less accurately. We found little evidence that such individual differences in assimilation were related to identification performance. There was a significant correlation between /ɹ/-/r/ assimilation and identification performance in Experiment 1, so the possibility that this assimilation makes it hard to learn to recognize these phonemes cannot be completely discarded. However, it seems unlikely that /ɹ/-/r/ assimilation is the actual cause of these learning problems because very few listeners made this confusion frequently. Moreover, the strength of /l/-/r/ assimilation was stronger than /ɹ/-/r/, so SLM predicts that /l/-/r/ assimilation should have an even bigger effect on category learning (Aoyama *et al.*, 2004). Instead, we found no reliable correlation of /l/-/r/ assimilation with the individual differences in identification performance.

In addition, the results of Experiment 3 suggest that listeners were able to maintain three separate categories for /ɹ/, /l/, and /r/, despite the fact that the subject group as a whole was not particularly skilled at identifying /ɹ/ and /l/. For example, their best exemplars demonstrated that they knew that /l/ has a longer initial closure and a lower F2 than /r/; the closeness of /l/ and /r/ did not cause them to assimilate these phonemes into a single category. The only evidence for problems caused by /l/-/r/ similarity was the fact that the best exemplars for /l/ by Japanese speakers were less nativelike along the F3 dimension, compared to their best exemplars for /ɹ/. This fits SLM's predictions (Aoyama *et al.*, 2004) that assimilation causes more problems for learning /l/ than /ɹ/. However, it is unexplained why this should affect only the F3 dimension; the best exemplars of /l/ for Japanese listeners were not significantly different from those of native speakers in any other respect.

If assimilation does not explain individual differences in identification performance, what does? The only positive conclusion from the present study is that the representation of F3 is important; individuals who had more nativelike best exemplars along the F3 dimension were more accurate at identifying /ɹ/ and /l/. F3 has been known to be problematic for Japanese speakers for a long time (e.g., Miyawaki *et al.*, 1975), but this is perhaps the first study to demonstrate that such results on synthetic speech relate to how individuals identify real speech, produced by multiple talkers in multiple word contexts. It is worth noting that this correlation was not particularly high, $r = -0.46$, but this may have occurred because individuals with poor identification ability were fairly random in the F3 values that they preferred.

It is plausible too that this correlation was low because there are other factors that affect individual differences in the ability to identify /ɹ/ and /l/. For example, previous work (Iverson *et al.*, 2003) has suggested that central auditory processing, at a stage prior to phonemic categorization, affects the ability of Japanese listeners to identify /ɹ/ and /l/; Japa-

nese adults are much more sensitive to acoustic variation that is irrelevant to /ɹ/-/l/ categorization (e.g., F2 variation, and within-category F3 variation) than they are to the critical F3 differences that distinguish these categories for native speakers. Such sensitivities could interfere with category learning, making it harder for them to learn which F3 values are correct for each phoneme, as found in the best exemplars measured in Experiment 3. In addition, this pattern of auditory processing could interfere with online speech perception even if listeners manage to form nativelike category representations; it could make it harder for them to focus on the acoustic variation that is important even though, at some level, they may represent that /ɹ/ has a lower F3 than /l/.

The present best-exemplar assessment of English and Japanese categories for /ɹ/ and /l/ may seem at odds with previous work on category boundaries. That is, previous work (e.g., Iverson *et al.*, 2005; Yamada, 1995) has shown that Japanese and English speakers use very different cues for /ɹ/-/l/ identification (F3 for English speakers; F2, transition, and closure duration for Japanese speakers), but the present results suggest that their best exemplars are broadly similar. Part of this difference may occur because the best exemplars better demonstrate sensitivity to secondary cues. For example, English speakers in the present study showed that they knew that /ɹ/ and /l/ differ in transition and closure duration, even though English identification decisions are mostly driven by F3. Likewise, Japanese listeners appear to have broadly nativelike representations for secondary cues in their best exemplars, but they may rely on them more for identification because their perception of F3 is relatively poor.

To summarize, the results of this study suggest that Japanese adults have an asymmetric pattern of category assimilation for the /ɹ/-/l/ contrast, with /l/ assimilating stronger into the /r/ category than does /ɹ/, but the degree to which individuals assimilate /ɹ/ and /l/ into their L1 categories is not strongly predictive of their ability to identify these phonemes accurately. Rather, their ability to identify /ɹ/ and /l/ is more directly related to their representation of F3 for these phonemes, which may be caused by their auditory/phonetic sensitivities along this dimension.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., and Yamada, T. (2004). "Perceived phonetic distance and L2 learning: The case of Japanese /r/ and English /l/ and /r/," *J. Phonetics* **32**, 233–250.

Best, C. T. (1995). "A direct realist view of cross-language speech perception," in *Speech Perception and Language Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), pp. 171–204.

Best, C. T., McRoberts, G. W., and Goodell, E. (2001). "American listeners' perception of nonnative consonant contrasts varying in perceptual assimilation to English phonology," *J. Acoust. Soc. Am.* **109**, 775–794.

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

Boersma, P., and Weenik, D. (2008). "Praat: Doing phonetic by computer," retrieved from <http://www.praat.org> (Last viewed September, 2008).

Bohn, O.-S., and Flege, J. E. (1992). "The production of new and similar vowels by adult German learners of English," *Stud. Second Lang. Acquis.* **14**, 131–158.

Bradlow, A. R., and Pisoni, D. B., 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.

Braida, L. D. (1991). "Crossmodal integration in the identification of consonant segments," *Q. J. Exp. Psychol. A* **43**, 647–677.

Evans, B. G., and Iverson, P. (2004). "Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences," *J. Acoust. Soc. Am.* **115**, 352–361.

Evans, B. G., and Iverson, P. (2007). "Plasticity in vowel perception and production: A study of accent change in young adults," *J. Acoust. Soc. Am.* **121**, 3814–3826.

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

Flege, J. E. (1995). "Second language speech learning: Theory, findings, and problems," in *Speech Perception and Language Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), pp. 233–277.

Flege, J. E. (2003). "Assessing constraints on second-language segmental production and perception," *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, edited by A. Meyer and N. Schiller (Mouton de Gruyter, Berlin), pp. 319–355.

Flege, J., Schirru, C., and MacKay, I. (2003). "Interaction between the native and second language phonetic subsystems," *Speech Commun.* **40**, 467–491.

Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138.

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., and Pruitt, J. C. (2000). "An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants," *J. Acoust. Soc. Am.* **107**, 2711–2724.

Hazan, V., Sennema, A., Iba, M., and Faulkner, A. (2005). "Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English," *Speech Commun.* **47**, 360–378.

Iverson, P. (2002). "A choice theory method of evaluating audiovisual integration," *Speech, Hearing, and Language: Work in Progress* **14**, 85–92.

Iverson, P., Bernstein, L. E., and Auer, E. T., Jr. (1998). "Modeling the interaction of phonemic intelligibility and lexical structure in audiovisual word recognition," *Speech Commun.* **26**, 45–63.

Iverson, P., Ekanayake, D., Hamann, S., Sennema, A., and Evans, B. G. (2008). "Category and perceptual interference in second-language phoneme learning: An examination of English /w/-/v/ learning by Sinhala, German, and Dutch speakers," *J. Exp. Psychol. Hum. Percept. Perform.* **34**, 1305–1316.

Iverson, P., and Evans, B. G. (2003). "A goodness optimization method for investigating phonetic categorization," *Proceedings of the 15th International Conference of Phonetic Sciences, Barcelona*.

Iverson, P., and Evans, B. G. (2007). "Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration," *J. Acoust. Soc. Am.* **122**, 2842–2854.

Iverson, P., Hazan, V., and Bannister, K. (2005). "Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults," *J. Acoust. Soc. Am.* **118**, 3267–3278.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (2003). "A perceptual interference account of acquisition difficulties for non-native phonemes," *Cognition* **87**, B47–B57.

Iverson, P., Smith, C. A., and Evans, B. G. (2006). "Vowel recognition via cochlear implants and noise vocoders: Effects of formant movement and duration," *J. Acoust. Soc. Am.* **120**, 3998–4006.

Klatt, D. H., and Klatt, L. C. (1990). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *J. Acoust. Soc. Am.* **87**, 820–857.

Lively, S. E., Logan, J. S., and Pisoni, D. B. (1993). "Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories," *J. Acoust. Soc. Am.* **94**, 1242–1255.

Logan, J. S., Lively, S. E., and Pisoni, D. B. (1991). "Training Japanese listeners to identify English /r/ and /l/: A first report," *J. Acoust. Soc. Am.* **89**, 874–886.

Lotto, A. J., Sato, M., and Diehl, R. L. (2004). "Mapping the task for the second language learner: The case of Japanese acquisition of /r/ and /l/," in *From Sounds to Sense: 50+ Years of Discoveries in Speech Communication*, edited by J. Slifka, S. Manuel, and M. Matthies (MIT Research Laboratory in Electronics, Cambridge, MA), pp. C-181–C186.

MacKay, I. R. A., Flege, J. E., Piske, T., and Schirru, C. (2001). "Category

- restructuring during second-language (L2) speech acquisition," *J. Acoust. Soc. Am.* **110**, 516–528.
- Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J., and Fujimura, O. (1975). "An effect of language experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English," *Percept. Psychophys.* **18**, 331–340.
- Shepard, R. N. (1972). "Psychological representation of speech sounds," in *Human Communication: A Unified View*, edited by E. E. David and P. B. Denes (McGraw-Hill, New York), pp. 67–133.
- Underbakke, M., Polka, L., Gottfried, T. L., and Strange, W. (1988). "Trading relations in the perception of /r/-/l/ by Japanese learners of English," *J. Acoust. Soc. Am.* **84**, 90–100.
- Vance, T. (1987). *An Introduction to Japanese Phonology* (State University of New York Press, Albany, NY).
- Walden, B. E., Montgomery, A. A., Prosek, R. A., and Schwartz, D. M. (1980). "Consonant similarity judgments by normal and hearing-impaired listeners," *J. Speech Hear. Res.* **23**, 162–184.
- Yamada, R. A. (1995). "Age and acquisition of second language speech sounds: Perception of American English /r/ and /l/ by native takers of Japanese," in *Speech Perception and Language Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), pp. 305–320.
- 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.