Journal of Phonetics 39 (2011) 571–584

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Predicting native English-like performance by native Japanese speakers

Erin M. Ingvalson [a,*], James L. McClelland [b], Lori L. Holt [a]

[a] Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, United States
[b] Department of Psychology, Stanford University, Palo Alto, CA, United States

ARTICLE INFO

Article history:
Received 28 July 2010
Received in revised form
14 March 2011
Accepted 17 March 2011
Available online 18 May 2011

ABSTRACT

This study tested the predictions of the Speech Learning Model (SLM, Flege, 1988) on the case of native Japanese (NJ) speakers' perception and production of English /ɹ/ and /l/. NJ speakers' degree of foreign accent, intelligibility of /ɹ–l/ productions, and ability to perceive natural speech /ɹ–l/ were assessed as a function of length of residency in North America, age of arrival in North America, years of student status in an English environment, and percentage of Japanese usage. Additionally, the extent to which NJ speakers' utilized the F3 onset cue when differentiating /ɹ–l/ in perception and production was assessed, this cue having previously been shown to be the most reliable indicator of category membership. As predicted, longer residencies predicted more native English-like accents, more intelligible productions, and more accurate natural speech identifications; however, no changes were observed in F3 reliance, indicating that though performance improves it does so through reliance on other cues.

1. Introduction

Many frameworks have been proposed as a means of accounting for the patterns of data seen in second-language speech sound learning in adulthood. One framework that we have found to be particularly helpful when developing hypotheses is the Speech Learning Model (SLM, Flege, 1995, 2003). The SLM is based on the assumption that the capacity for the successful acquisition of speech sounds is never lost. However, though the capacity for successful acquisition is never lost, adult and child learners may not become equally proficient in a second language. Crucial to accounting for these discrepancies is *assimilation*, which reflects the SLM's assumption that first- and second-language phonetic subsystems are not fully separated and that second-language (L2) speech sounds may be judged to be instances of first-language (L1) speech sound categories (Flege, 2003; see also the Perceptual Assimilation Model, Best, 1995; Best, McRoberts, & Goodell, 2001; Best & Tyler, 2007). In order for new L2 speech sound categories to be formed, listeners must detect sufficient phonetic difference between existing L1 speech sound categories and to be acquired L2 speech sound categories.[1] The development of new L2 speech sound categories is more likely in children, whose L1 speech sound categories are not as well developed as adults' and therefore are less likely to assimilate new L2 sounds (Flege, 2003).

Two factors the SLM considers important to speech sound acquisition are the *amount* and *type* of experience with the L2. The more often an individual uses the L2 (*amount* of experience), the more frequently the individual will hear exemplars of L2 speech sounds that are sufficiently distinct from L1 speech sound categories to warrant the formation of new speech sound categories, thereby increasing the likelihood or extent of new category formation. Similarly, the more an individual interacts with native speakers of the L2 (*type* of experience), the more likely the individual is to hear good exemplars of L2 sounds that are distinct from L1 categories. This too should enhance the probability or extent of new category formation. Operationally, amount of experience is often defined as length of residency in the L2 environment (LOR) or percentage of language usage that is L2. Flege and Liu (2001) reasoned that students may be more likely to interact with native speakers of the L2 whereas non-students may be more likely to use the L2 with other non-native speakers, such as when doing business in a multiethnic environment or when conversing with another native L1 speaker and a third individual who does not speak the L1; student status was thereby a means of assessing type of experience. These factors (LOR, amount of L2 usage, and student status) are often contrasted with age of arrival in a new language environment (AOA). Proponents of the SLM have demonstrated that duration of immersion in a L2 environment better accounts for performance than the age at which the learner was immersed (MacKay, Flege, Piske, & Schirru, 2001; Oyama, 1976); amount of L2 usage accounts for L2 performance (Flege, Frieda, & Nozawa, 1997; Flege & MacKay, 2004;

* Correspondence to: Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, United States. Tel.: +1 847 491 2430; fax: +1 847 491 2429.
E-mail address: ingvalson@northwestern.edu (E.M. Ingvalson).

[1] Unfortunately, the SLM does not specify sufficient phonetic distance, stating only that it must be perceptually discernable to the language learner (Flege, 2003).

0095-4470/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2011.03.003

Oyama, 1976); interactions with native speakers (as assessed via student status) interacts with duration of immersion such that longer immersions that emphasize interactions with native speakers are associated with the most native-like performance (Flege & Liu, 2001).

A classic example of the difficulty L2 learners can have mastering the sounds of a new language is the well-documented finding that native Japanese (NJ) speakers fail to perceive and produce the English /ɹ/ and /l/ as reliably as native English (NE) speakers (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Goto, 1971; Guion, Flege, Akahane-Yamada, & Pruitt, 2000; Iverson, Hazan, & Bannister, 2005; Iverson et al., 2003; Lively, Logan, & Pisoni, 1993; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Logan, Lively, & Pisoni, 1991; Lotto, Sato, & Diehl, 2004; McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002; Miyawaki et al., 1975; Strange & Dittman, 1984; Yamada & Tohkura, 1990). Applications of the SLM to NJ speakers' perception and production of /ɹ–l/ have revealed that perception and production do improve as a function of LOR (Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004; Flege, Munro, & MacKay, 1995; Flege, Takagi, & Mann, 1996). The aim of the present work is to test the SLM in the context of NJ speakers' perception and production of English /ɹ–l/, expanding upon previous efforts by examining speakers' sensitivity to the acoustic cues to /ɹ–l/ category membership.

The difficulty NJ speakers have with /ɹ/ and /l/ is associated with lesser reliance by NJ speakers on the primary acoustic cue NE speakers use to differentiate /ɹ/ and /l/ (Lotto et al., 2004). O'Connor, Gertsman, Liberman, Delattre, and Cooper (1957, pp. 28–29) observed that the primary cue to the /ɹ–l/ distinction is the onset frequency of the third formant, or F3; "for /ɹ/ the third formant begins at a point just slightly above the second formant and then rises to its steady-state level, while for /l/ the third formant starts at a level at least as high as the steady-state". Systematically manipulating F1, F2, and F3 onset frequencies in sounds presented to NE listeners revealed that a change in F3 onset frequency from low to high is sufficient to result in a categorical response shift from /ɹ/ to /l/ (O'Connor et al., 1957), demonstrating NE listeners have learned to make use of the most reliable acoustic cue when making perceptual category judgments. As demonstrated by Lotto et al. (2004) NE speakers' productions of /ɹ/ and /l/ are also separated along the F3 dimension, indicating the salience of this cue in both perception and production. Conversely, NJ speakers appear to rely on F2 onset frequency more heavily than F3 onset frequency in perception (Iverson et al., 2003, 2005; Yamada & Tohkura, 1990) and also to differentiate /ɹ/ and /l/ on F2 more than on F3 in production (Lotto et al., 2004), possibly accounting for their less reliable English /ɹ–l/ performance (Iverson et al., 2005; Miyawaki et al., 1975). NJ speakers' reliance on the F2 cue likely stems from the extensive variability in F3 onset frequencies for instances of the Japanese /ɾ/ – typically described as a tap or flap – making F3 a poor predictor of Japanese category membership and presumably leading the NJ speakers' to down weight F3 as /ɹ–l/ are assimilated to the single Japanese category (Miyawaki et al., 1975).

The SLM predicts, and indeed we have seen, that longer residency in an L2 environment leads to lower degrees of accentedness, better identification of L2 speech contrasts, and more intelligible productions of L2 speech contrasts (Flege, MacKay, & Meador, 1999; MacKay et al., 2001; Munro, Flege, & MacKay, 1996). Also as predicted, NJ speakers have more reliable identifications and NJ speakers' productions of /ɹ–l/ are more intelligible as a function of LOR (Aoyama et al., 2004; Flege et al., 1995, 1996).

What is less clear, and what we hope to elucidate here, is whether increasing LOR might lead to greater reliance on the F3 onset cue to differentiate /ɹ/ and /l/. Interestingly, while there is good reason to believe that LOR is associated with functional improvements in natural settings, it is not clear whether or not this is based on learning to rely on F3 cues. Yamada and Tohkura (1990) asked NJ listeners who were proficient in English to judge synthetic instances of /ɹ/ and /l/, finding non-categorical performance that depended at least partially upon F2 onset frequency; Iverson et al. (2003) found that NJ listeners weight F2 most heavily in the /ɹ–l/ context. Iverson et al., 2005 found improvement in /ɹ–l/ minimal pair identification following training, but no changes in listeners' weightings of F3 onset frequency. Ingvalson, Holt, and McClelland (under review) utilized training stimuli that differentiated only on F3 onset frequency and found minimal success on the trained context and no generalization to novel contexts or natural speech, emphasizing how difficult the F3 cue is for NJ listeners. Thus, it seems possible that increasing LOR might result in improved /ɹ–l/ perception and production without a corresponding increase in F3 reliance. However, Hattori and Iverson (2009) have noted that weighting of F3 is the best predictor of /ɹ–l/ natural speech identification, indicating that though /ɹ–l/ differentiation may proceed without reliance on F3, usage of the F3 cue maximizes NE-like performance.

The SLM also predicts an effect of years of student status in an English environment (henceforth, EngEd; Flege & Liu, 2001). We therefore assessed the effect of EngEd, both alone and in conjunction with LOR, on reliance on F3 in /r–l/ perception and production. We anticipated that those individuals with longer LORs and more EngEd would show more native-like performance than their counterparts with shorter LORs or less EngEd (Flege & Liu, 2001). We further predicted, in line with earlier findings of Flege and MacKay (2004), that percentage of Japanese usage (henceforth, JUse) would account for differences in performance such that those with higher JUse levels would show less NE-like performance than those with lower JUse levels. We expected JUse alone to have an effect but we also anticipated a possible interaction of JUse and AOA (as seen in earlier work, e.g., Flege et al., 1997), such that those individuals with early AOAs and small amounts of Japanese usage would show the most NE-like performance.

Participants completed a language usage questionnaire in which they reported LOR, AOA, years of education in English, and estimates of weekly Japanese usage and English usage. Estimates of Japanese usage were divided by the sum of the estimates of Japanese usage and English usage to estimate the proportion of language use that was Japanese (Flege & MacKay, 2004). We used these self reports to divide the participants into non-overlapping LOR, AOA, EngEd, and JUse groups. We collected participants' identifications of both natural speech and synthetic speech /ɹ–l/; the latter judgments were used to determine their reliance on F3 onset frequency when perceiving /ɹ–l/. Participants produced a series of English sentences and /ɹ–l/ minimal pairs. Monolingual English listeners judged the degree of foreign accent in the sentences and identified the minimal pairs. We measured the F2 and F3 onset frequencies in the syllable-initial minimal pairs (e.g., "rock–lock"). We then entered all variables into a series of between subjects ANOVAs to determine the extent to which differences in performance on the above tasks could be attributed to differences in LOR, AOA, EngEd, and/or JUse.

The description of the methods used is divided into two sections: those portions of the experiment that tested NJ speakers as the participants and those that tested NE listeners.

2. Native Japanese: methods

2.1. Participants

Fifty seven native Japanese (NJ) speakers participated. Two participants were unable to complete the experiment due to an equipment failure. The remaining fifty five participants were divided into three groups based on length of residency in North America (LOR): less than two years (< Two, 15 participants), two to five years (Two–Five, 15 participants), or ten or more years (Ten+, 25 participants).[2] All participants reported normal or corrected to normal vision and normal hearing. With one exception, all participants reported Japanese was the only language spoken in the home and claimed English as their second language; the one exception was a member of the < Two Group, raised Chinese–Japanese simultaneously bilingual in Japan. All participants were raised in Japan and began learning English between 10 and 13 years of age, when it was introduced in the Japanese education system.

Participation was limited to those individuals who had not studied English abroad or in an international school in Japan before age 18. These individuals were also required to hold a job outside the home or to be married to a NE speaker, to ensure regular use of English. Thirty five NJ speakers were recruited from the Pittsburgh, Pennsylvania area and tested at Carnegie Mellon University; an additional sixteen were recruited from the Vancouver, British Columbia area and tested at the University of British Columbia; the remaining six were recruited from the Palo Alto, California area and tested at Stanford University.

2.2. Materials

2.2.1. Questionnaire of immersion

The questionnaire was based on a language history questionnaire designed by Tokowicz, Michael, and Kroll (2004; see Ingvalson (2008) for the full questionnaire). Only those aspects of the questionnaire used in the analyses are reported here. The initial portion dealt with individuals' experiences learning English and included questions such as the age at which participants began learning English and the age at which they immigrated to North America. This portion also asked participants to detail the time and duration of each visit to Japan since immigrating to North America. The final portion of the questionnaire queried participants about their daily English and Japanese experience.

2.2.2. Natural speech perception

The test for the perception of natural speech was a direct replication of tests used in earlier work training NJ listeners to differentiate /ɹ–l/ (Bradlow et al., 1997, 1999; Lively et al., 1993, 1994; Logan et al., 1991). The stimuli consisted of 121 English /ɹ–l/ minimal pairs spoken by three NE speakers (two male; not all words were produced by all talkers). The words were broken down into six categories: syllable-initial (e.g., *rock–lock*), initial cluster (e.g., *clack–crack*), intervocalic (e.g., *allay–array*), final cluster (e.g., *halt–heart*), word final (e.g., *mile–mire*), and other (e.g., *swimming–swinging*). The stimuli were recorded at a 22,050 Hz sampling rate, low pass filtered at 4.8 kHz, and normalized for peak amplitude.

2.2.3. Synthetic speech perception

The synthetic speech perception stimuli were a replication of stimuli used by Yamada and Tohkura (1990), henceforth called YT stimuli. These stimuli consisted of 37 synthetic realizations of the words "*right*" and "*light*" created using the cascade branch of the Klatt synthesizer (Klatt & Klatt, 1990). The f0 and amplitude of voicing contours for all stimuli were based on a NE male producing /rait/; the final /t/ was edited from this production and placed at the end of each synthesized /rai/ and /lai/ stimulus. All stimuli had a total duration of 580 ms, with the /rai/–/lai/ portion lasting 360 ms. There were 20 ms of silence between the offset of the diphthong and the onset of the burst; there were 140 ms between the onset of the burst and when the signal amplitude reached 0 dB. The final 60 ms of the stimulus were silence. All stimuli were sampled at 11,025 Hz and RMS amplitude matched in energy.

Yamada and Tohkura (1990) designed their stimuli to examine how listeners categorize English syllable-initial /ɹ/ and /l/ across changes in F2 and F3 onset frequencies. The 37 stimuli formed a two dimensional grid sampling a factorial combination of F2 and F3 onset frequencies, excepting those combinations for which F3 onset frequency was less than or equal to the F2 onset frequency. F2 onset frequency was manipulated in four steps of 200 Hz from 800 to 1400 Hz. F3 onset frequency was manipulated in 10 steps of 200 Hz from 1200 to 3000 Hz. In accord with the findings of Yamada and Tohkura, pilot data indicated NE listeners separated their judgments of these stimuli into /ɹ/ and /l/ categories on the basis of F3 onset frequency.[3]

For all stimuli, F3 onset frequency was held at the initial steady-state for 80 ms. It then linearly transitioned to the /a/ portion of the diphthong, 2465 Hz, reached at 180 ms. This value was maintained until 240 ms, at which point it linearly transitioned to the /i/ portion of the diphthong, 2735 Hz, reached at 300 ms and held for the remaining 60 ms of the syllable. F2 onset frequency steady-state duration varied in conjunction with the onset frequency to keep slope constant from consonant to vowel states in all stimuli. The onset–duration pairings were as follows: 800 Hz and 80 ms, 1000 Hz and 105 ms, 1200 Hz and 130 ms, 1400 Hz and 150 ms. For all stimuli, F2 reached 2350 Hz at 280 ms and maintained this value until the sounds' end.

As in the original Yamada and Tohkura stimuli, F1 onset steady-state duration systematically covaried with F3 onset frequency. With each 200 Hz increase in F3 onset frequency, F1 onset steady-state duration increased by 6 ms, ranging from 80 ms to 134 ms with the constraint that the total duration from stimulus onset to the initiation of the /a/ portion of the diphthong be 150 ms. Thus, in those stimuli with a F3 onset frequency of 1200 Hz, the F1 initial steady-state was 80 ms and the transition duration to the vowel state was 70 ms; whereas in those stimuli with a F3 onset frequency of 3000 Hz, the F1 initial steady-state was 134 ms and the transition duration was 16 ms. For all stimuli, the F1 onset frequency was 400 Hz. Following the steady-state, F1 frequency transitioned linearly to the /a/ portion of the diphthong, reaching 750 Hz at 150 ms. It maintained this value until 200 ms, at which point it transitioned linearly to the /i/ portion of the diphthong, reaching 410 Hz at 250 ms and held for the remainder of the sound. For all stimuli, F4 and F5 frequencies were set at constant values of 3400 and 3950 Hz, respectively. Graphical representations of the series' extremes can be seen in Fig. 1.

2.2.4. Speech production

A list of English /ɹ–l/ minimal pair words was created based on stimuli reported by Bradlow et al. (1997). Our list deviated from the original in that it was biased toward initial singleton pairs to maximize number of syllable-initial /ɹ–l/ productions available for acoustic analysis (21 initial singleton pairs, 7 initial cluster pairs, 5 intervocalic pairs, 9 final singleton pairs, 8 final cluster pairs, and 100 pairs total); a list of all pairs can be found in

[2] The < Two group had a mean age of 29.80 years and included 4 females; the Two–Five group had a mean age of 31.00 years and included 13 females; the Ten+ group had a mean age of 43.24 years and included 21 females.

[3] Twenty-one monolingual NE listeners participated in this task. Participants heard 10 repetitions of each of the 37 stimuli (370 trials) and responded /ɹ/, /l/, or /w/.

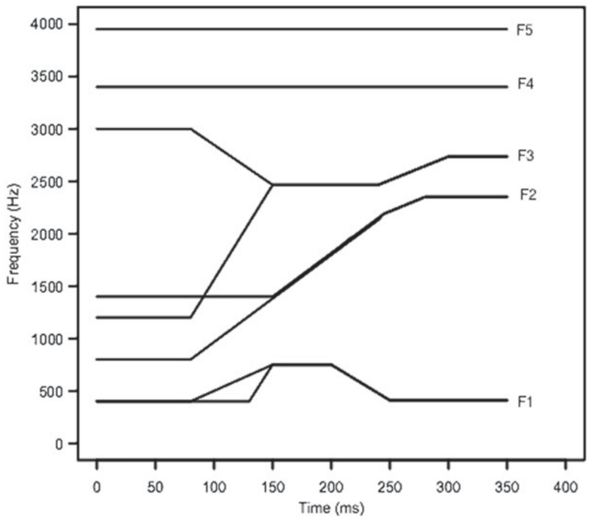

Fig. 1. Endpoints of the synthetic /rai–lai/ stimuli. F2 and F3 onset frequencies were manipulated in 200 Hz steps. Stimuli where F2 onset frequency was equal to or greater than F3 onset frequency were not created. F2 initial steady-state was manipulated to keep the slope of the transition to the vowel constant. F1 initial steady-state increased by 5 ms with each 200 Hz increase in F3 onset frequency with the constraint that steady-state plus transition duration sum to 150 ms.

Appendix A. A male NE speaker with no discernable regional accent made two recordings of each word within the carrier sentence, "The next word is _____, _____." The second of these productions was used as an auditory prompt for both NJ participants' and NE controls' productions.

The same NE speaker also made two recordings of Flege et al. 1995) sentences:

The good shoe fits Sue.
The red book was good.
I can read this for you.
He turned to the right.
Paul ate carrots and peas.

These sentences had been previously used to collect accentedness ratings from non-native speakers of English and were constructed so as to not emphasize any particular English phonetic distinction that may be more difficult for a given non-native population. The second of the two utterances of each sentence served as an auditory prompt for both NJ participants' and NE controls' productions.

All stimuli were recorded at 22,050 Hz and low-pass filtered at 9 kHz. All prompts were matched for RMS amplitude.

2.3. Procedure

The tests were always presented in the same order: questionnaire, synthetic speech perception, natural speech perception, and production. In most cases, a Japanese–English interpreter was present for all portions of the experiment; all instructions, with the exception of those for the production test, were given in Japanese. In those instances when an interpreter was not available, instructions were given in English; in each of these cases, the participant had resided in North America for 10+ years and was highly proficient in English. The entire experiment lasted approximately 3 h.

With the exception of the questionnaire, all experiments were run and all data was recorded using E-Prime (Psychological Software Tools, Pittsburgh PA). With the exception of the production task, all auditory stimuli were presented diotically over linear headphones (Beyer DT-150) at a constant volume of approximately 70 dB (SPL); auditory prompts in the production tasks were presented over the laptop speakers. A Marantz PMD 670 recorder was used in conjunction with a microphone to record productions digitally at a sampling rate of 22,050 Hz. Visual stimuli were presented in the center of the laptop monitor at viewing distance determined by the participant.

In both the synthetic and natural speech perception tests, the response options were presented in colored boxes on the screen. The synthetic speech test used a red box on the left, blue on the right, and yellow centered above the other two; the natural speech test used only the red and blue boxes. One response option was presented in each box. The position of the boxes did not change over the course of the experiment but the response options randomly changed positions on each trial. On the keyboard were red, blue, and yellow stickers laid out linearly. Listeners were told their responses were color coordinated: if they heard the word in the red box, press the red key, etc. Each trial began with a 500 ms central fixation cross. The colored response options were then presented for 1000 ms. Listeners heard each stimulus only once. Listeners were given an infinite response time but were encouraged not to deliberate on trials.

2.3.1. Synthetic speech perception

The YT replication used 37 stimuli each presented 10 times, for 370 trials. Participants were invited to take a short break after 74 trials. We collected responses (/ɹ/, /l/, or /w/) and reaction times (RTs) for each conjunction of cues.

2.3.2. Natural speech perception

The words were blocked by talker. The two male talkers were presented first, followed by the female talker. The first block consisted of the 48 pairs from Strange and Dittman (1984), consistent with Lively et al. (1993, 1994) and Logan et al. (1991). All words were presented in random order within blocks and each block was repeated twice. Participants were invited to take a short break after each block. Accuracy and RT were collected.

2.3.3. Speech production

The instructions for the speech production test were given in English, though an interpreter was present (if available for that session) to answer any clarification questions.

Following Flege, Yeni-Komshian, and Liu (1999), each trial began with a 500 ms orthographic presentation of the word or sentence to be produced. While the prompt remained on the screen an auditory prompt was presented over speakers. Following the first auditory prompt, there was a 700 ms delay after which participants were visually cued by the appearance of the word or sentence to speak. Participants then heard the auditory prompt again, followed by a 3000 ms delay before a second visual cue to speak. Thus, participants produced each utterance twice per trial. All words appeared in two trials, for four productions of each minimal pair word, followed by the sentences, which appeared in three trials, for six productions of each sentence.

All utterances were recorded at 22,050 Hz. These digital recordings were analyzed via linear predictive coding (LPC) algorithms using Multi-Speech (Kay Elemetrics, Lincoln Park NJ). Starting at the beginning of the waveform, the LPC analysis was performed at every 0.01-s interval until the algorithm was able to detect F3 onset frequency. The algorithm's judgment was checked against a visual inspection of the spectogram; all F3 values were consistent with a visual inspection of the formants and no hand corrections were needed. At this point, F1, F2, and F3 onset frequency, amplitude, and bandwidth were recorded.

3. Native English: methods

3.1. Participants

Six adult native English (NE) monolingual speakers participated as control subjects. Three of these participants were female; the mean age was 26.17 years. Forty-seven additional NE speakers participated in tasks exclusive to NE speakers. Twelve of these participants were randomly assigned to rate the accentedness of NJ speakers, the remaining 34 judged the intelligibility of NJ speech. All participants reported normal or corrected-to-normal vision and normal hearing. All participants had been exposed to only English in the home, had had most of their education in English (with the exception of foreign language classes and study-abroad programs of not more than a year), and reported English as their first and dominant language. No exposure to Japanese was reported. All NE participants were recruited from the Carnegie Mellon community.

3.2. Materials

3.2.1. Accentedness

The first two sentence recordings were discarded as practice. The remaining four were separated into individual recordings and divided into three tests. NE speakers 1–3, NJ speakers 1–8 from the Ten+ group, NJ speakers 1–8 from the Two–Five group, and NJ speakers 1–8 from the <Two group were placed in test A; NE speakers 4–6, NJ Ten+ speakers 9–15, NJ Two–Five speakers 9–15, and NJ <Two speakers 9–15 were placed in test B; NE speaker 5, NJ Ten+ speakers 16–25, NJ Two–Five speaker 9, and NJ <Two speaker 10 were placed in test C. All utterances within a test were RMS amplitude matched.

3.2.2. Intelligibility

The first utterance of each word was discarded as practice and the remaining three utterances were divided into individual sound files. The recordings from one NE speaker and six NJ speakers (three from each group, Ten+, Two–Five, <Two) were combined to make one intelligibility test, resulting in eight total tests. All words from the nine talkers in a test resulted in a total set of approximately 2700 words per test (100 words × 3 productions × 9 talkers; less some due to poor recording quality of some samples). For each experimental session, 900 of these words (100 from each talker) were randomly selected, balanced for /ɹ–l/ category membership. All productions in a test were RMS amplitude matched.

3.3. Procedure

We used the same equipment described in Section 2.3.

3.3.1. Accentedness

Productions from different speakers were intermixed and randomly presented. Each production was presented once. Participants were told to rate the degree of foreign accent in each sentence on a scale from 1 to 9, 1 being, "No discernable foreign accent" and 9 being "Heavily accented." This methodology and rating scale were based on those used by Flege (1988). The scale was present on the screen throughout the experiment to serve as a prompt. Following the auditory presentation of the sentence, participants were given 3 s to respond. The accentedness ratings given by NE listeners were inverted for the analyses – making 9 "least accented" and 1 "most accented" – then averaged for each speaker.

To avoid the use of different rating schemes amongst listeners, anchors were provided. We created the anchors by presenting one instance of each sentence as spoken by a native Bulgarian speaker to 10 NE listeners with no training in linguistics but who interacted with non-native English speakers regularly. These 10 listeners judged the accentedness of each sentence; their ratings were averaged to create a score for each sentence. Listeners who were to judge the accentedness of the NJ productions heard each sentence spoken by the Bulgarian speaker while simultaneously seeing its accentedness rating. The listeners were told to map the presented rating onto their own judgments of the speaker's accentedness and to use this mapping when making their judgments of the upcoming sentences.

3.3.2. Intelligibility

The productions from all talkers – eight NJ talkers and one NE control talker – in a test were randomly intermixed. NE listeners saw the /ɹ–l/ minimal-pair word on the monitor with the phoneme of interest removed (e.g., if the word to be heard was "rude" on the monitor would be "_ude"). Visual presentations were written to be phonetically consistent with the word and not constrained for spelling accuracy (e.g., both "rude" and "lewd" would be presented on the monitor as "_ude"). While this may have biased literate listeners to a particular response, the differences in intelligibility on the basis of LOR reported below lead us to believe these effects were minimal. Below the word were the letters "D", "B", "T", "R", "W", "L", these being the sounds NE listeners often perceive instead of intended /ɹ–l/ in NJ English speech (Miyawaki et al., 1975). Participants were asked to select which letter should be placed in the blank to make the visual presentation consistent with the auditory presentation; they were told answers may not be real English words. Listeners saw the incomplete word and response choices for 100 ms prior to the onset of the auditory presentation. Listeners had the option to replay a word once before making their decision. After the auditory presentation, participants had 3 s to make a response. If the participants did not make a response within 3 s, the program logged a non-response and advanced to the next trial; fewer than five non-responses were logged for any experiment trial. Participants were invited to take a short break every 300 trials.

3.4. Data transformations

3.4.1. Questionnaire measures

The LOR groupings (Section 2.1) were less than 2 years of residency (<Two), between 2 and 5 years of residency (Two–Five), or more than 10 years of residency (Ten+). All participants had AOAs of 18 or older (Section 2.1). The median AOA was 26 years (range 18–60 years) and participants were divided into those arriving prior to age 26 ($N=24$) and those arriving after age 26 ($N=27$). In earlier work, AOA grouping are non-overlapping. Thus, for consistency, we excluded participants who arrived at age 26 ($N=4$) in analyses utilizing AOA.

EngEd was assessed by asking participants how many years they had attended schools where most of the courses were in English. This value was cross-checked against participants' yearly estimates of English usage and the smallest value was used in the case of discrepancies (one instance). The median length of EngEd was 2 years (range 0–13 years). Participants therefore were divided into those who had fewer than two years of education ($N=26$) and those who had more than two years ($N=20$). Those individuals with two years of education in English ($N=9$) were not included in analyses utilizing this variable.

In the questionnaire, participants completed a chart that indicated the number of minutes during a typical week that they engaged in an English conversation with given individuals. They completed an identical table in for Japanese conversation. These values were summed to determine the minutes of English and Japanese conversation per week. We then divided the number of minutes in Japanese conversation per week by the total number of minutes of conversation per week to find the percentage Japanese usage (JUse). The median usage was 25% of total conversational minutes in Japanese (range 0–86%). Participants were divided into those with JUse values of less than 25% ($N=27$) and those with JUse values of more than 25% ($N=28$).

Recruitment efforts were limited to ensure participants were a member of one of the target LOR groups and were older than 18 when they arrived in North America. All other group divisions (AOA, years of English-based education, and amount of Japanese usage) were done post-hoc. We therefore ensured that there were no differences in AOA, EngEd, or JUSe across LOR groups. There was no main effect of AOA, $F(2, 52)=0.52$, $p=0.60$, Ten+ $M=26.56$ years, $SD=6.23$ years; Two–Five $M=27.87$ years, $SD=6.71$ years; $<$Two $M=29.07$ years, $SD=10.17$ years. There was also no main effect of EngEd, $F(2, 52)=1.25$, $p=0.29$, Ten+ $M=3.16$ years, $SD=2.53$ years; Two–Five $M=1.87$ years, $SD=2.07$ years; $<$Two $M=2.07$ years, $SD=3.75$ years. Finally, there was no main effect of JUse, $F(2, 52)=0.04$, $p=0.96$, Ten+ $M=28\%$, $SD=25\%$; Two–Five=29%, SD=23%; $<$Two=30%, $SD=23\%$.The number of individuals in each tested interaction cell can be seen in Fig. 2.

3.4.2. Experimental measures

The dependent measure from the identification of natural speech was the proportion correct response, regardless of contrast position or sentence type. Intelligibility was the proportion of words heard as intended by the NE listeners.

Perceptual measures relating to F3 performance (YT and production analyses) were assessed using logistic regression slopes. The slope of the logistic regression function provides a measure of the degree to which the two categories – /ɹ/ and /l/ – are separated, with larger numbers indicating greater separation. Robust logistic regressions (Yohai, 1987, 1997; Yohai, Stahel, & Zamar, 1991) were used for the production analyses, because some values appeared to not be in line with an individual's overall pattern.

4. Results

We begin by assessing differences in global measures of English proficiency, including degree of foreign accent, intelligibility of bilingual speech, and perception of natural speech as a function of LOR, AOA, years as a student in a predominantly English environment (EngEd), and percentage of Japanese language use (JUse). We tested the hypotheses that longer LORs would predict more NE-like performance, whereas AOA would fail

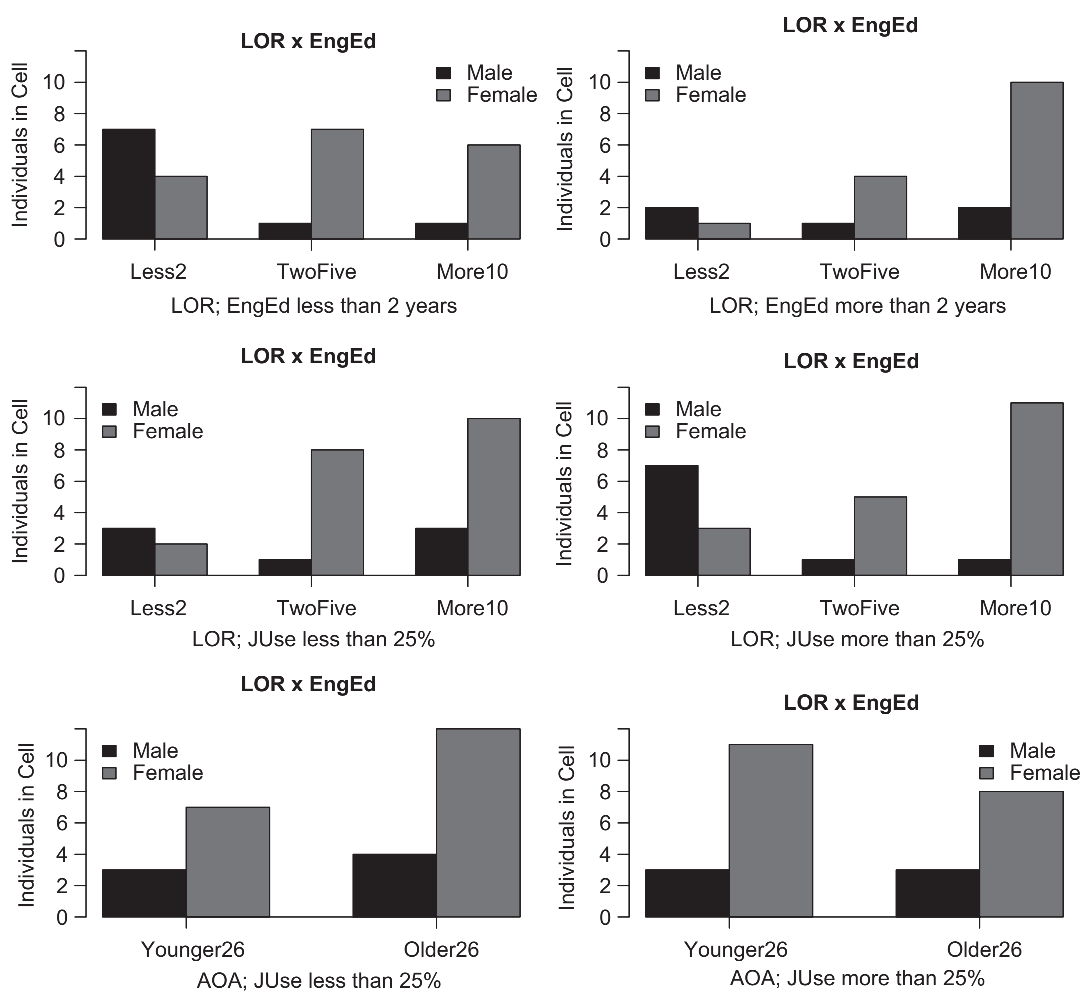

Fig. 2. Number of data points in the tested LOR × EngEd (years of English-based education, top), LOR × JUse (percentage of conversations in Japanese, middle), and AOA × JUse (bottom) interactions. LOR and AOA groupings are on the *x*-axis in each panel. Individuals with EngEd of less than two years are represented in the top right panel, those with more than two years are represented in the top left. Individuals with JUse values of less than 25% are represented in the middle and bottom right panels, those with JUse values of more than 25% are represented in the middle and bottom left panels. In the initial ANOVAs we collapsed across gender. Follow up analyses tested a LOR × EngEd × Gender interaction when assessing intelligibility.

to be a significant predictor when entered simultaneously with LOR (Flege et al., 1995; Flege & MacKay, 2004; MacKay et al., 2001), that greater EngEd would predict more NE-like performance alone and in an interaction with LOR (Flege & Fletcher, 1992; Flege & Liu, 2001), and that less JUse would predict more NE-like performance alone and in interactions with AOA and LOR (Flege et al., 1997; Flege & MacKay, 2004).

We then turn our focus to the extent to which NJ speakers express reliance on the F3 onset cue when perceiving and producing English /ɹ/ and /l/. We assessed whether more NE-like patterns of reliance on the F2 and F3 cues in /ɹ–l/ perception and production could be predicted by LOR, AOA, EngEd, and/or JUse. Based on the predictions of the SLM (Flege, 2002, 2003), we anticipated that increases in LOR and EngEd would predict greater reliance on the F3 cue whereas greater JUse would predict greater reliance on the F2 cue. We did not expect to find an effect of AOA (Flege, 2002, 2003), although we anticipated the possibility of AOA and JUse might interact to predict performance (Flege et al., 1995).

Before presenting the details we summarize our main findings, seen in Table 1.

(1) Consistent with the SLM, accentedness, /ɹ–l/ intelligibility, and /ɹ–l/ natural speech perception all varied with LOR; longer LORs corresponded to more NE-like performance. LOR also interacted with EngEd when assessing intelligibility of NJ productions such that, within the longest LOR grouping, more years of EngEd predicted more words heard as intended. There was a main effect of EngEd on intelligibility of NJ productions, where more years of EngEd predicted more words heard as intended. Inconsistent with our hypotheses, there was a main effect of AOA when assessing accentedness, where (relatively) early arrivals had less pronounced foreign accents than (relatively) late arrivals.
(2) Analyses focused on reliance on F2 and F3 onset frequency in /ɹ–l/ perception and production found little support for the SLM. We did observe a main effect of AOA on the degree of /ɹ–l/ category separation along the F2 dimension for /ɹ–l/ production. Late arrivals showed more separation along this dimension than early arrivals, indicating more NJ-like productions by the *later* arrivals. Thus, whereas analyses that replicated earlier work were generally consistent with this work and with the predictions of our hypotheses, analyses at the level of acoustic cues tended not to be significant or were not consistent with our hypotheses.

4.1. Predicting performance on speech perception and production

The first set of analyses attempted to replicate the findings of earlier tests of the SLM among NJ adults learning English /ɹ–l/. We use categorical variables and ANOVA here to maintain continuity with previous work, where listeners are typically divided by high vs. low L2 usage and early vs. late arrival in the L2 country. In order to maximize the number of data points per cell, we limited the factors to those that were motivated by the hypotheses: LOR, AOA, EngEd, JUse, an interaction of AOA and JUse, an interaction of LOR and JUse, and an interaction of LOR and EngEd. In all the ANOVAs presented below, the independent measures were LOR, AOA, EngEd, JUse, AOA × JUse, and LOR × JUse. Number of individuals per cell for each tested interaction can be seen in Fig. 2. As can be seen in the figure, there were many more males in the < Two group than in the other two LOR groupings (Section 2.1). Perceived foreign accent (accentedness), proportion of /ɹ–l/ words heard as intended by NE listeners (intelligibility), and proportion correctly identified natural speech /ɹ–l/ (natural speech perception) were the dependent measures, each submitted to an independent ANOVA with the above factors.

As anticipated, there was a main effect of LOR on degree of perceived foreign accent, $F(2, 30)=5.89$, $p=0.007$, Fig. 3. Overall, speakers with LORs of ten or more years were judged to have less pronounced foreign accents than those with fewer than two years of residency ($M=4.57$ and 3.15 accentedness ratings, respectively); no significant differences were found between members of the Ten+ and Two–Five groups ($M=3.84$ accentedness rating) or between the < Two and Two–Five groups. There was also a main effect of AOA, $F(1, 30)=6.10$, $p=0.02$, Fig. 3. Early arrivals were judged to have less pronounced foreign accents than late arrivals ($M=3.99$ and 3.73 accentedness ratings, respectively). Inclusion of the NE controls revealed a main effect of Group, $F(3, 57)=37.03$, $p<0.001$, where the accents of the NE controls were judged to be significantly more native-like than any NJ LOR grouping. There were no significant effects of EngEd ($F(1, 30)=0.07$, $p=0.79$) or JUse ($F(1, 30)=1.68$, $p=0.21$), nor any significant interactions.

As with accentedness, there was a main effect of LOR on intelligibility, $F(2, 30)=3.44$, $p=0.04$, Fig. 4; NE listeners correctly identified more productions by speakers in the Ten+ group ($M=86\%$ correctly identified) than speakers in the < Two group ($M=76\%$ correctly identified); no significant differences were found between the Two–Five group ($M=82\%$ correctly identified) and the < Two group or Ten+ group. The productions of NE controls ($M=92\%$ correctly identified) were correctly identified more often than those of the NJ speakers, $F(3, 57)=5.14$, $p<0.01$. There was a main effect of EngEd, $F(1, 30)=5.09$, $p=0.03$, Fig. 4. Individuals with more than two years of EngEd had more intelligible productions than those with fewer than 2 years ($M=85\%$ and 78% correctly identified, respectively). LOR and EngEd interacted, $F(2, 30)=3.40$, $p=0.04$, Fig. 4. Within the Ten+ group, speakers who had more than two years of EngEd produced more intelligible words ($M=92\%$ correctly identified) than those with fewer than two years of EngEd ($M=75\%$ correctly identified). In sum, for proportion of words heard as intended, individuals with longer LORs or more years of EngEd had more intelligible productions of /ɹ–l/. Also as anticipated, years of EngEd interacted with LOR such that those with the longest residencies and the

Table 1
Summary of the significant main effects and interactions.

	LOR	AOA	EngEd	JUse	AOA × JUse	LOR × JUse	LOR × EngEd
Accentedness	X	X					
Intelligibility	X		X				X
Natural speech perception	X						
Perceptual F2							
Perceptual							
Production F2		X					
Production F3							

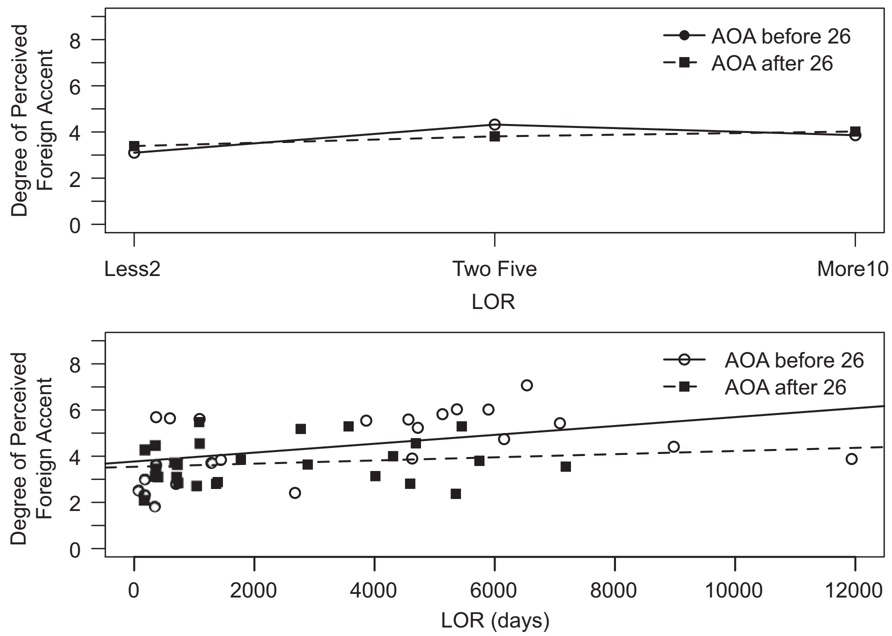

Fig. 3. Degree of perceived foreign accent as a function of LOR and AOA. The median AOA was 26 years; participants were divided into those who arrived before age 26 and those who arrived after 26. Higher accentedness ratings indicate a less pronounced (more native English-like) foreign accents. Scatterplots of the data used to find the interaction are below.

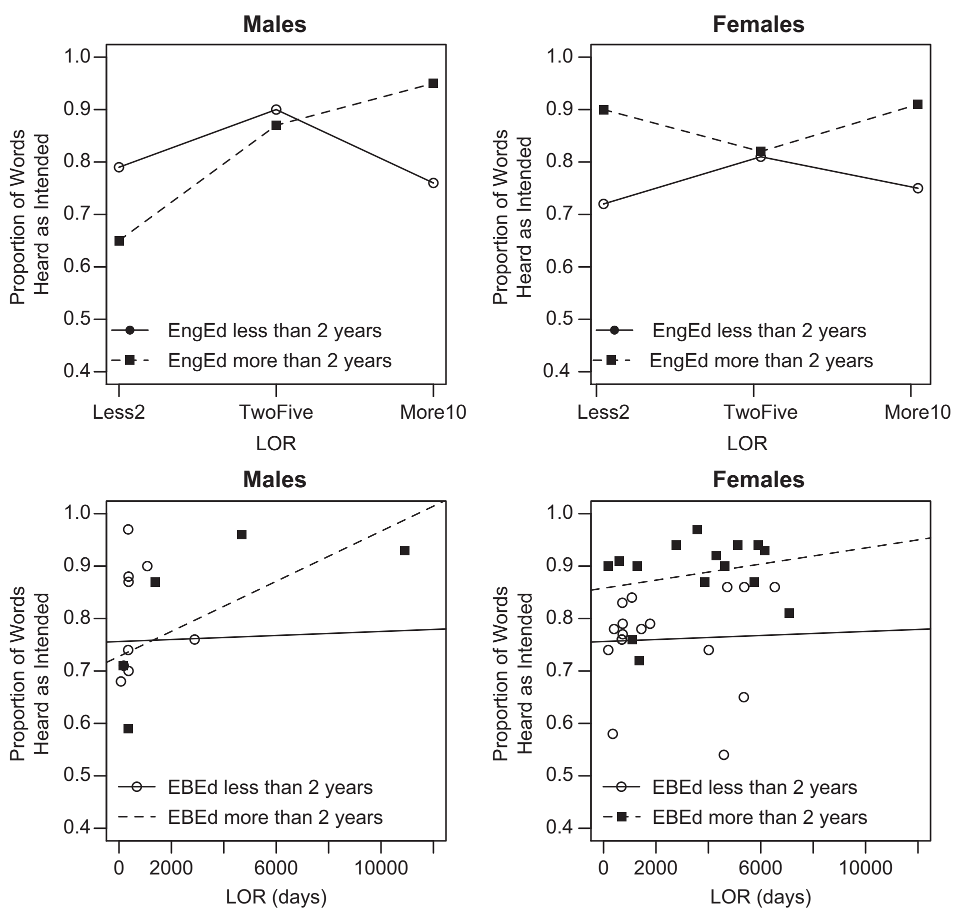

Fig. 4. Proportion of words heard as intended as a function of EngEnd (years of English-based education) and LOR separated into males (left) and females (right). Scatterplots of the data used to find the interaction are below.

most education showed better performance than their less-English-educated counterparts (consistent with Flege & Liu, 2001). We found no main effects of AOA ($F(1,\ 30)=0.71$, $p=0.41$) or of JUse ($F(1,\ 30)=0.62$, $p=0.44$) nor interactions involving these terms.

LOR was again found to be significant when assessing natural speech perception, $F(2,\ 30)=4.63$, $p=0.02$. Listeners in the <Two groups showed the worst identification performance ($M=68\%$ correct), being significantly below listeners in the Two–Five group ($M=79\%$) and listeners in the Ten+ group ($M=82\%$ correct). However, the Two–Five and Ten+ groups were not significantly different from one another. NE listeners ($M=98\%$ correct) correctly identified more /ɹ–l/ words than any group of NJ listeners, $F(3,\ 57)=14.95$, $p<0.001$. There were no main effects of AOA ($F(1,\ 30)=1.91$, $p=0.18$), EngEd ($F(1,\ 30)=1.17$, $p=0.29$), or JUse ($F(1,\ 30)=0.60$, $p=0.45$), nor any significant interactions.

Thus, the overall pattern of results supported our hypothesis that increased experience with English, as measured by LOR, would result in more NE-like levels of performance and that AOA would have little, if any, effect. Flege and Liu (2001) found student status and LOR to interact to account for differences in degree of perceived foreign accent whereas we only found an effect of EngEd when assessing intelligibility. We also failed to find any effect of JUse (cf. Flege et al., 1997), possibly due to the fact that speakers in the present study had later AOAs ($M=27.6$ years) than those in earlier work that found interactions between AOA and amount of first-language use (e.g., $M=5.75$ years; Flege et al., 1997).

It should be noted that though the lack of an effect of AOA was as predicted it may not have been due to the hypothesized importance of LOR. Our sample was limited to individuals with AOAs of 18 years or more to better assess speech sound learning in adulthood. Consequently, a lack of an AOA effect might be attributable to all participants being equally limited in their speech sound learning abilities. However, the fact that our findings replicate previous support of the SLM (Flege et al., 1997) leads us to believe this is not the case.

4.2. Predicting performance on F3 cue weighting

The second set of analyses focused on the amount of separation between /ɹ/ and /l/ along the F2 and F3 dimensions as a function of LOR, AOA, EngEd, and JUse. Though we refer only to F3 onset frequency, the reader should bear in mind that this cue covaried with F1 transition duration (Yamada & Tohkura, 1990). We present first differences in the amount of separation along the F3 dimension, then differences in the amount of separation along the F2 dimension; perception data are presented before production data. NE speakers typically show separation between /ɹ/ and /l/ categories along the F3 dimension whereas NJ speakers show separation along the F2 dimension (Lotto et al., 2004; Yamada & Tohkura, 1990). We therefore looked for increased separation along the F3 dimension and decreased separation along the F2 dimension as a function of LOR, AOA, EngEd, and JUse. As in the above analyses, we limited the number of interactions to those motivated by the hypotheses: AOA × JUse, LOR × JUse, and LOR × EngEd.

There were no differences as a function of LOR in the amount of separation along the F3 dimension judged from perceptual responses ($F(2,\ 30)=0.87$, $p=0.43$). Similarly, there were no differences as a function of AOA ($F(1,\ 30)=1.97$, $p=0.17$), EngEd ($F(1,\ 30)=0.17$, $p=0.68$), or JUse ($F(1,\ 30)=0.10$, $p=0.75$). There were also no significant interactions. Measuring the amount of separation along the F3 dimension among NJ /ɹ–l/ productions, there was no influence of LOR ($F(2,\ 30)=1.55$, $p=0.23$), AOA ($F(1,\ 30)=0.79$, $p=0.38$), EngEd ($F(1,\ 30)=1.26$, $p=0.27$), JUse ($F(1,\ 30)=0.55$, $p=0.46$), nor any significant interactions.

There were no significant differences in the amount of separation along the F2 dimension when making /ɹ–l/ perceptual judgments on the basis of LOR ($F(2,\ 30)=0.69$, $p=0.51$), AOA ($F(1,\ 30)=3.15$, $p=0.09$), EngEnd ($F(1,\ 30)=0.34$, $p=0.56$), or JUse ($F(1,\ 30)=0.16$, $p=0.70$). Amount of separation along the F2 dimension when producing /ɹ–l/ differed significantly on the basis of AOA, $F(1,\ 30)=5.03$, $p<0.05$. Later arrivals (M slope$=0.005$) showed significantly more separation along the F2 dimension than earlier arrivals (M slope$=-0.002$). We found no main effects of LOR ($F(2,\ 30)=3.14$, $p=0.06$), EngEd ($F(1,\ 30)=0.90$, $p=0.35$), or JUse ($F(1,\ 30)=1.36$, $p=0.25$), nor any significant interactions.

Our investigations into the separation along the F2 and F3 dimensions when perceiving and producing /ɹ–l/ did not reveal the expected outcomes. We found no differences in separation along the F3 dimension as a function of any of our variables of interest; we had predicted longer LORs would result in more separation along this dimension, which would indicate that more years of immersion in English resulted in more NE-like reliance on this cue. However, we did find significant differences in the degree of separation along the F2 dimension when producing /ɹ–l/, where greater separation along this dimension is indicative of more NJ-like performance. Problematic for the SLM was the finding that later arrivals had more separation along the F2 dimension when producing /ɹ–l/ than early arrivals. The SLM would predict LOR and JUse to account for differences in separation along the F2 dimension instead of the variance being accounted for by AOA (Flege, 2002, 2003).

The lack of a relationship between LOR and F3 when paired with the positive relationship between LOR and natural speech perception and production, in conjunction with data indicating that NJ listeners who are sensitive to F3 best identify natural speech /ɹ–l/ (Hattori & Iverson, 2009), leads us to question whether the more NE-like performance of the long-term residents may be a function of changing communication strategies and assimilation patterns without a corresponding change in category structure or cue weighting (Iverson & Evans, 2009). We therefore sought to determine if degree of separation along the F2 and F3 dimensions when perceiving or producing /ɹ–l/ might predict performance on natural speech /ɹ–l/ perception or production. We performed two regressions. The first of these predicted natural speech by LOR, EngEd, JUse, slope of the F2 logistic regression function for /ɹ–l/ perception, and slope of the F3 logistic regression function for /ɹ–l/ perception. The second predicted intelligibility by LOR, EngEd, JUse, slope of the F2 logistic regression function for /ɹ–l/ production, and slope of the F3 logisitic regression function for /ɹ–l/ production. When predicting natural speech, LOR was significant, $b=1.04\text{e}-5$, $t(49)=2.40$, $p=0.02$, as was F3 onset frequency, $b=14.65$, $t(49)=6.65$, $p<0.001$; the overall model accounted for a fair amount of the variance, $R^2=0.64$, $F(5,\ 49)=17.32$, $p<0.001$. For intelligibility, LOR was significant, $b=4.72\text{e}-06$, $t(49)=2.17$, $p=0.03$, as was EngEd, $b=4.79\text{e}-03$, $t(49)=2.19$, $p=0.03$; the overall model had an $R^2=0.22$, $F(5,\ 49)=2.75$, $p=0.03$. The effects of LOR and EngEd were anticipated based on the above ANOVAs. The relationship between F3 onset frequency and natural speech perception is also not surprising inasmuch that F3 onset frequency is the most reliable cue to /ɹ–l/ category membership (O'Connor et al., 1957; Hattori & Iverson, 2009; Lotto et al., 2004; Yamada & Tohkura, 1990). However, it is interesting that increasing LOR predicts better natural speech /ɹ–l/ perception and greater F3 reliance predicts better natural speech /ɹ–l/ perception but that LOR and F3 separation are not related (the correlation matrices for the natural speech and intelligibility regressions can be seen in Appendix B and C, respectively). The relationships among LOR, F3, and natural speech perception are presented in Fig. 5. As can be seen in the figure, though LOR and natural speech perception are moderately correlated ($r=0.37$, $p=0.005$) and natural speech perception and F3 separation are well correlated ($r=0.72$, $p<0.001$), LOR and F3 separation are not related ($r=0.06$, $p=0.60$). Thus, it appears

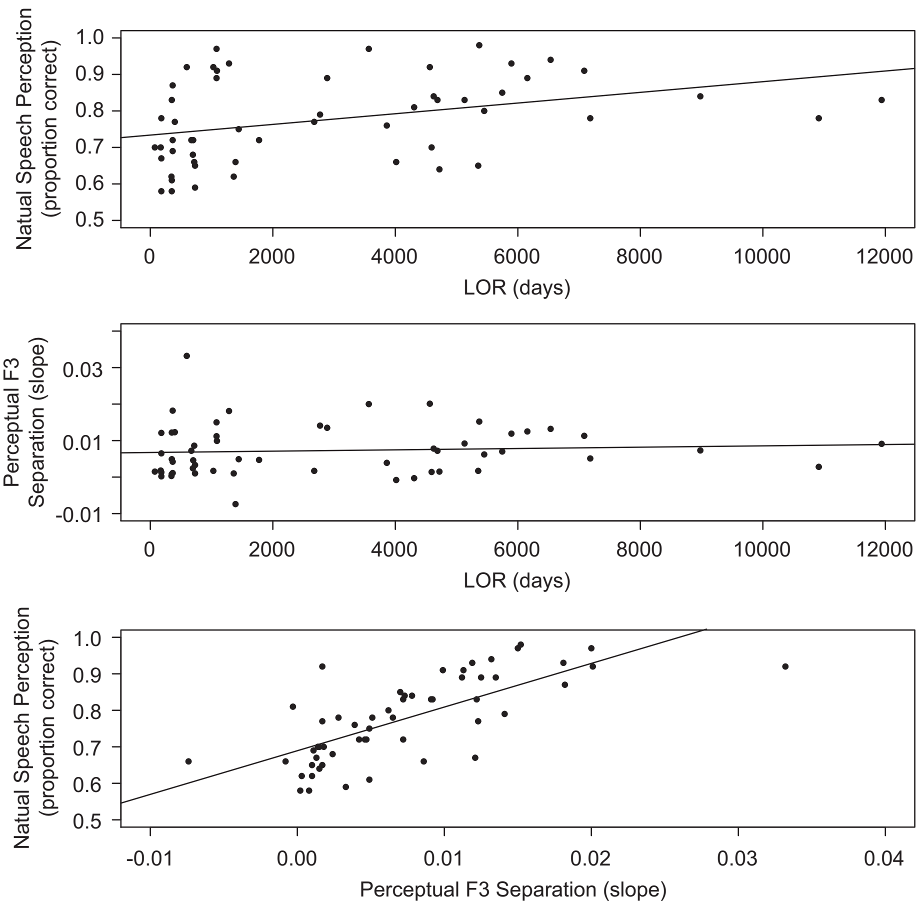

Fig. 5. Scatterplots of the relationships between LOR and natural speech /ɹ–l/ perception, LOR and degree of separation on F3 when perceiving /ɹ–l/, and natural speech /ɹ–l/ perception and degree of F3 separation when perceiving /ɹ–l/.

that the changes in NJ speakers' perceptions and productions of /ɹ–l/ improve over the course of experience with English, but this does not happen as a function of changes in F3 weighting.

4.3. The role of gender in speech perception

An examination of the data indicated that most of the participants in the <Two group were male ($N=10$), but that the majority of participants in both the Two–Five and Ten+ groups were female ($N=13$ and 21, respectively). Previous work has demonstrated mixed effects of gender on non-native speech sound acquisition (Flege et al., 1995; Flege & Fletcher, 1992; Purcell & Suter, 1980; Thompson, 1991), leading us to investigate the possibility of gender effects in the current data.

We first verified that there were no differences between male and female participants on any of the predictor variables other than LOR ($F(1, 53)=11.41$, $p<0.01$). Indeed, no differences between males and females was found for AOA ($F(1, 53)=0.113$, $p=0.74$), EngEd ($F(1, 53)=0.008$, $p=0.92$), or JUse ($F(1, 53)=0.13$, $p=0.72$). Inasmuch as only LOR differed as a function of gender, we assessed only those results for which LOR was found to be a significant main effect or part of a significant interaction. There was a significant main effect of LOR on accentedness, intelligibility, and natural speech perception; LOR interacted with EngEd on intelligibility.

For accentedness, entering gender, AOA (previously significant) and LOR as the factors in an ANOVA revealed no effect of gender, $F(1, 44)=0.66$, $p=0.42$. LOR ($F(2, 44)=5.30$, $p=0.01$) and AOA ($F(1, 44)=5.72$, $p=0.02$) continued to be significant. Similarly, for natural speech perception there was no effect of Gender, $F(1, 49)=0.24$, $p=0.62$, though LOR continued to be significant, $F(2, 49)=8.77$, $p<0.001$.

Proportion of words heard as intended had been found to differ on the basis of LOR, EngEd, and an interaction of LOR and EngEd. We therefore entered LOR; EngEd; Gender; and interactions of LOR and EngEd; LOR and Gender; and LOR, EngEd, and Gender into the ANOVA. There was no main effect of Gender, $F(1, 34)=0.79$, $p=0.38$. LOR continued to be significant, $F(2, 34)=4.94$, $p=0.04$, as was EngEd, $F(1, 34)=7.03$, $p=0.03$, and the interaction of LOR and EngEd, $F(2, 34)=4.56$, $p=0.03$. The interaction of LOR, EngEd and Gender was not significant, $F(3, 34)=2.53$, $p=0.07$.

Overall, these results support previous work indicating no effect of gender on speech sound learning (Andreou, Andreou, & Vlachos, 2006; Andreou, Vlachos, & Andreou, 2005; Flege & Fletcher, 1992; Piske, MacKay, & Flege, 2001; Purcell & Suter, 1980).

5. Discussion

The aim of our investigation was to test the predictions the Speech Learning Model (SLM, Flege, 2002, 2003) in the context of native Japanese (NJ) speakers perceiving and producing English /ɹ/ and /l/. Based on the SLM, we expected that increasing lengths of residency in North America (LOR), more years of being a student in an English environment (EngEd), and/or decreasing amounts of Japanese usage (JUse) would predict more NE-like performance on degree of perceived foreign accent, proportion of NJ-spoken /ɹ–l/ words heard as intended by NE listeners, and proportion of correctly identified naturally produced /ɹ–l/ by NJ in perception tests (Flege et al., 1997; Flege & MacKay, 2004; MacKay et al., 2001). Further, given that F3 onset frequency is the most reliable cue of /ɹ–l/ category membership (Hattori & Iverson, 2009; Lotto et al., 2004; O'Connor et al., 1957) we anticipated that increases in the above measures would predict greater separation between /ɹ/ and /l/ categories on the basis of F3 onset. We also anticipated no main effect of AOA when entered simultaneously into analyses with LOR, though we expected that AOA may interact with JUse such that earlier arrivals using Japanese less frequently would show more NE-like performance (Flege & MacKay, 2004).

Consistent with earlier work testing the SLM and with the predictions of the SLM, we found differences in accentedness, intelligibility, and proportion of correctly identified natural speech /ɹ–l/ as a function of LOR; in all cases NJ speakers with longer residencies showed more NE-like performance than those with shorter durations of residency. We also found the expected interaction of LOR and EngEd (Flege & Liu, 2001), though only on intelligibility. We did not, however, find an effect of JUse, either alone or in an interaction with AOA. We suspect this may be due to the fact that all our participants would traditionally be categorized as late arrivals (Flege et al., 1997; Flege & MacKay, 2004; MacKay et al., 2001). In those studies that have found an effect of L1 usage the effect is generally most prominent among early learners who have AOAs of 13 years or earlier (e.g., Flege & MacKay, 2004). More work is needed to determine whether amount of L1 usage influences L2 mastery differently in early and late learners.

Earlier work has indicated NJ listeners' difficulty categorizing /ɹ–l/ is due to reliance on a less reliable acoustic cue, F2 onset frequency. However, we found little evidence that increased experience using English changes NJ listeners' cue usage. The accentedness, intelligibility, and natural speech perception data indicate that NJ speakers – including late arrivals – improve their mastery of English speech sound categories, but the lack of an effect at the level of F2 and F3 onset frequencies suggests that a shift in perceptual cue weightings might not be the source of this learning. Iverson et al. (2005) have found that NJ listeners shift their perceptual weightings of non-F3 cues in conjunction with identification improvements following training. Other investigations into cross-linguistic speech perception have found that listeners' L1 backgrounds influence identification and discrimination patterns (e.g., Ingram & Park, 1998), suggesting that patterns of assimilation to existing L1 categories determine performance (Flege, 2003). Iverson and Evans (2009); see also Lengeris and Hazan (2010) have recently suggested that training does not alter listeners' phonetic categories but instead improves efficiency of categorization within the bounds of the existing L1 and L2 categories, highlighting the inability to change cue weightings over training as an indicator of unchanging category structure. If that is the case, it may be that something similar happens with immersion, where listeners do not fundamentally alter their category structure but instead become better at using their L1 and L2 categories to identify native and non-native speech sounds. Though this account does describe the pattern of data seen here, it is somewhat problematic from the perspective of the SLM, where L1 and L2 speech sound categories are hypothesized to exist in a shared phonological space and to shift as a function of language experience (Flege, 2003; see also Antoniou, Best, Tyler, & Kross, 2010), although it might be possible to argue that subtle changes in the structure of these categories do mediate the performance improvements. Further research is necessary to better understand the relationship between L1 and L2 phonetic categories and the extent to which language context might influence listeners' and speakers' categorization and cue weighting strategies.

Though we found no effect of our predictor variables on F3 reliance, we did find that F3 reliance was a significant predictor of natural speech /ɹ–l/ perception, in line with Hattori and Iverson (2009). The fact that F3 reliance and natural speech performance were related should not be surprising given how reliable the F3 onset cue is (Hattori & Iverson, 2009; Lotto et al., 2004; O'Connor et al., 1957; Yamada & Tohkura, 1990), but it is interesting that increasing LOR is related to improved natural speech performance while being unrelated to F3 reliance. We found similar effects in our earlier work, where individuals excluded from training on the basis of existing F3 sensitivity were found not to differ from eligible individuals on the basis of LOR, frequency of English use, or amount of education (Ingvalson et al., under review). It appears some individuals are better able to make use of the F3 cue in the /ɹ–l/ context. Further research is necessary to determine what factors predict this ability (Chandrasekaran, Sampath, & Wong, 2010).

Clearly, a full understanding of second-language speech sound learning remains an elusive goal. We suggest that the field may be best advanced by identifying the characteristics that exemplify late language acquisition – such as the role of internal motivation, the relative benefits of immersion vs. classroom instruction, and the importance of interaction with native speakers – as a means of understanding what characterizes one who is successful at L2 speech sound learning from an individual who persists in demonstrating non-native-like performance following long-term immersion or training.

To this end, the fact that adults have more difficulty than children learning the speech sounds of a new language appears incontrovertible (Baker, Trofimovich, Flege, Mack, & Halter, 2008; Flege, 2002, 2003; MacWhinney, 2007; Werker & Tees, 1984). Whether this difficulty stems from the closing of a sensitive period (Oyama, 1976) or from the entrenchment of existing categories (Flege, 2003) cannot be determined from the data at hand. We can attempt to identify possible mechanisms for learning via the use of computational modeling (Vallabha & McClelland, 2007; Vallabha, McClelland, Pons, Werker, & Amano, 2007), which may answer this question, but our current methods of speech assessment limit our ability to determine the source of learning difficulties. Understanding what facilitates language learning in older learners and the extent to which attainment is limited may help to clarify how strong an influence age of acquisition exerts, and should ultimately inform our efforts to understand the nature and acquisition of second-language speech sound processing skill.

Experience also appears to play a large role in category formation and refinement. Though it may be that early speech experience exerts a greater influence on category formation than late speech experience, data presented earlier under the framework of the SLM and the data presented here indicate that communicating with native speakers (such as receiving schooling in a predominantly English environment) predicts more native-like performance and therefore may be a contributing factor. Emphasizing not only *when* experience occurs, but also *what kind* of experience it is and *with whom* the experience occurs might enable us to better predict how a learner will perform on a given task. Recognizing whether there are limits on learners' ability to change category structure or shift cue weightings will also allow us to better predict perception and production performance. Similarly, cognitive factors that are not generally assessed as part of language learning appear to be related to ultimate performance (e.g., working memory, MacDonald, 2008). For example, Hakuta, Bialystok, and Wiley (2003) and Wiley, Bialystok, and Hakuta (2005) have noted that amount of formal education, whether in the L1 or the L2, is a strong, independent predictor of English proficiency. Observations of the interplay between age of acquisition, duration of immersion, type of speech experience, and cognitive factors such as processing ability and degree of education may provide us with a sense of how novel sounds are incorporated into an existing category space.

It may be the case that not all possible factors that influence L2 speech sound learning are equally efficacious for all L2 learners. One obvious place to identify possible limitations on L2 speech sound learning is through a better understanding of L1 assimilation patterns. Multi-language investigations allow us to characterize how a particular L2 is assimilated to categories in different L1s. Observations of differential assimilation of a particular phoneme on the basis of L1 background provide insight into what phoneme properties might be particularly salient for a given listener on the basis of her L1 background. Better awareness of what makes a particular cue to category membership salient to a given L2 learner enables better predictions of L2 speech sound differentiation. Having knowledge of listeners' sensitivity to particular L2 cues and having the capability to predict L2 speech sound differentiation on the basis of cue saliency would then in turn enable more rigorous assessments of which aspects of L2 learning are particularly efficacious for learners with a given L1 background.

Finally, in our data as well as in those of previous researchers, all non-native speakers showed significantly worse performance than native speakers, and there was extensive variability among participants. Data such as these have often been used to suggest the existence of a sensitive period, as there is less variability among native speakers and a truly plastic system should be able to fully acquire the new speech categories (Flege, 1988; Takagi, 2002). The SLM has tried to counteract these claims by suggesting late L2 learners could become native-like given sufficient L2 experience (Flege, 2002, 2003). We suggest an alternative possibility. If, as the SLM suggests, a bilingual individual cannot separate the two phonetic categories, then being bilingual implies having a different phonological space from a monolingual in each of the two languages. If this space is indeed different, then we should not expect performance in either language to be identical to monolinguals (Flege, 2003). We suggest that the performance of the individuals in the Ten+ group may be indicative of the construction of a special hybrid perceptual space within which an approximation of the English /ɹ–l/ distinction has been constructed. We anticipate that their performance on Japanese sounds would also differ from Japanese monolinguals, though perhaps to a lesser degree than their English performance differs from English monolinguals due to the greater experience with Japanese. Similarly, individuals raised as simultaneous English Japanese bilinguals with relatively balanced proficiency across the two languages would also differ from both groups (Pallier, Bosch, & Sebastián-Gallés, 1997; Sebastián-Gallés, Echeverría, & Bosch, 2005; Sebastián-Gallés & Soto-Faraco, 1999). Indeed, we suspect that two languages are not represented identically in any two individuals – possibly due to differences in usage, motivation, experience with each language, or the age of first exposure, among other factors – and that these differences in representation may be the locus of the individual differences. If we are able to understand what makes language learners differ from one another, we will simultaneously be more aware of what makes them similar, moving us closer to the goal of understanding the speech learning mechanism.

Acknowledgments

The authors wish to thank Alisa Brown Kuppe, Laurel Fais, Kaori Idemaru, Kjell Ericson, and Daniel Dickison for translating materials, serving as interpreters, and answering questions on the Japanese language and Japanese culture. We thank Janet Werker, Laurel Fais, Laura Hope, Vivian Pan, Christi Gomez, and David Klionsky for assistance with participant recruitment. This work was supported by NIH grant 2-R01DC004674-04A2 and NSF grant BCS-0746067 to the third author and NIMH grant P50-MH64445 to the second author.

Appendix A

See Table A1.

Appendix B

See Table B1.

Table A1
Words used in the production task, separated by /ɹ–l/ position.

Initial singleton		Initial cluster		Intervocalic		Final singleton		Final cluster	
late	rate	bleach	breach	allay	array	feel	fear	shield	sheared
lock	rock	blade	braid	pilot	pirate	tail	tear	halt	heart
lewd	rude	clock	crock			while	wire	cold	cord
lice	rice			pallet	parrot	soul	sore		
lip	rip	blues	bruise	elect	erect	dial	dire	tiled	tired
link	rink	blink	brink			seal	sear	pealed	peered
lope	rope			alight	aright	ball	bar	bald	bard
list	wrist					foal	four	colt	cart
lent	rent	bland	brand						
lamp	ramp	gland	grand			call	car	wild	wired
lace	race								
lair	rare								
lake	rake								
lane	rain								
leap	reap								
led	red								
leer	rear								
life	rife								
limb	rim								
lime	rhyme								
lit	writ								

Table B1
Correlation matrix for the variables entered into a regression to predict intelligibility.

	LOR	EBEd	JUse	Production F2	Production F3
LOR					
EBEd	−0.16				
JUse	−0.04	−0.08			
Production F2	0.04	−0.18	0.03		
Production F3	−0.27	0.09	0.13	−0.33	

Appendix C

See Table C1.

Table C1
Correlation matrix for the variables entered into a regression to predict natural speech perception.

	LOR	EBEd	JUse	YTF2	YTF3
LOR					
EBEd	−0.14				
JUse	0.00	−0.07			
YTF2	0.07	0.03	0.07		
YTF3	−0.06	−0.10	−0.08	−0.06	

References

Andreou, G., Andreou, E., & Vlachos, F. (2006). Individual differences in second language learning among university students. In S. N. Hogan (Ed.), *Trends in learning research*. New York: Nova Science Publishers.

Andreou, G., Vlachos, F., & Andreou, E. (2005). Affecting factors in second language acquisition. *Journal of Psycholinguistic Research*, *34*, 429–438.

Antoniou, M., Best, C. T., Tyler, M. D., & Kross, C. (2010). Language context elicits native-like stop voicing in early bilinguals' production in both L1 and L2. *Journal of Phonetics*, *38*, 640–653.

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

Baker, W., Trofimovich, P., Flege, J. E., Mack, M., & Halter, R. (2008). Child–adult differences in second-language phonological learning: The role of cross-language similarity. *Language and Speech*, *51*, 317–342.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language research*. Timonium, MD: York Press.

Best, C. T., McRoberts, G., & Goodell, E. (2001). American listeners' perception of nonnative consonant contrasts varying in perceptual assimilation to English phonology. *Journal of the Acoustical Society of America*, *1097*, 775–794.

Best, C., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn, & M. Munro (Eds.), *Language experience in second language speech learning. In honor of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). HUTraining Japanese listeners to identify English /ɹ/ and /l/: Long term retention of learning in perception and productionUH. *Perception & Psychophysics*, *61*, 977–985.

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

Chandrasekaran, B., Sampath, P., & Wong, P. C. M. (2010). Individual variability in cue weighting and lexical tone learning. *Journal of the Acoustical Society of America*, *128*, 456–465.

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America*, *84*, 70–79.

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

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

Flege, J. E. (2003). Assessing constraints on second language segmental production and perception. In N. O. Schiller, & A. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 319–355). Berlin: Mouten de Gruyter.

Flege, J. E., & Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *Journal of the Acoustical Society of America*, *91*, 370–389.

Flege, J. E., Frieda, E. M., & Nozawa, T. (1997). Amount of native-language (L1) usage affects the pronunciation of an L2. *Journal of Phonetics*, *25*, 169–186.

Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, *23*, 527–552.

Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, *24*, 1–34.

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

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of a foreign accent in a second language. *Journal of the Acoustical Society of America*, *97*, 3125–3134.

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

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, *41*, 78–104.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "l" and "r". *Neuropsychologica*, *9*, 317–323.

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

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science*, *14*, 31–38.

Hattori, K., & Iverson, P. (2009). English /ɹ/–/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *Journal of the Acoustical Society of America*, *125*, 469–479.

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

Ingvalson, E. M. (2008). *Predicting F3 usage in /ɹ–l/ perception and production by native Japanese speakers*. Unpublished doctoral dissertation. Pittsburgh, PA: Carnegie Mellon University.

Ingvalson, E. M., Holt, L. L., & McClelland, J. L. (under review). Can native Japanese listeners learn to differentiate /ɹ–l/ on the basis of F3 onset frequency? *Bilingualism: Language and Cognition*.

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language systems, II: Auditory training for native Spanish and German speakers. *Journal of the Acoustical Society of America*, *126*, 866–877.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /ɹ/–/l/ to Japanese adults. *Journal of the Acoustical Society of America*, *118*, 3267–3278.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., & Ketterman, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phones. *Cognition*, *87*, B47–B57.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among male and female talkers. *Journal of the Acoustical Society of America*, *87*, 820–856.

Lengeris, A., & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *Journal of the Acoustic Society of America*, *128*, 3757–3768.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /ɹ/ and /l/, II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, *94*, 1242–1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /ɹ/ and /l/, III: Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America*, *96*, 2076–2087.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /ɹ/ and /l/: A first report. *Journal of the Acoustical Society of America*, *89*, 874–885.

Lotto, A.J., Sato, M., & Diehl, R.L. (2004). Mapping the task for the second language learner: Case of Japanese acquisition of /ɹ/ and /l/. In *Proceedings of the from sound to sense: 50+ years of discoveries in speech communication* (pp. 1–6). Cambridge: MIT Press.

MacDonald, J. L. (2008). Grammaticality judgments in children: The role of age, working memory and phonological ability. *Journal of Child Language*, *35*, 247–268.

MacKay, I. R. A., Flege, J. E., Piske, T., & Schirru, C. (2001). Category restructuring during second-language speech acquisition. *Journal of the Acoustical Society of America*, *110*, 516–528.

MacWhinney, B. (2007). A unified model. In N. Ellis, & P. Robinson (Eds.), *Handbook of cognitive linguistics and second language acquisition*. Hillsdale, NJ: Lawrence Erlbaum.

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]–[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective and Behavioral Neuroscience*, *2*, 89–108.

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

Munro, M. J., Flege, J. E., & MacKay, I. R. A. (1996). The effect of age of second-language learning on the production of English vowels. *Applied Psycholinguistics*, *17*, 313–334.

O'Connor, J. D., Gertsman, L. J., Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1957). Acoustic cues for the perception of initial (w, j, r, l) in English. *Word*, *13*, 24–43.

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, *5*, 261–283.

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, *64*, B9–B17.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*, 191–215.

Purcell, E. T., & Suter, R. W. (1980). Predictors of pronunciation accuracy: A reexamination. *Language Learning*, *30*, 271–287.

Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language*, *52*, 240–455.

Sebastián-Gallés, N., & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, *72*, 111–123.

Strange, W., & Dittman, S. (1984). Effects of discrimination training on the perception of /ɹ–l/ by Japanese adults learning English. *Perception and Psychophysics*, *36*, 131–145.

Takagi, N. (2002). The limits of training Japanese listeners to identify English /ɹ/ and /l/: Eight case studies. *Journal of the Acoustical Society of America*, *111*, 2887–2894.

Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning*, *41*, 177–204.

Tokowicz, N., Michael, E. B., & Kroll, J. F. (2004). The roles of study-abroad experience and working-memory capacity in the types of errors made during translation. *Bilingualism: Language and Cognition*, *7*, 255–272.

Vallabha, G. K., & McClelland, J. L. (2007). Success and failure of new speech category learning in adulthood: Consequences of learned Hebbian attractors in topographic maps. *Cognitive, Affective and Behavioral Neuroscience*, *7*, 53–73.

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Science*, *104*, 13273–13278.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49–63.

Wiley, E., Bialystok, E., & Hakuta, K. (2005). New approaches to using census data to test the critical-period hypothesis for second-language learning. *Psychological Science*, *16*, 341–343.

Yamada, R. A., & Tohkura, Y. (1990). Perception and production of syllable-initial English /ɹ/ and /l/ by native speakers of Japanese. In *Proceedings of the 1990 international conference on spoken language processing* (pp. 757–760).

Yohai, V. J. (1987). High breakdown point and high efficiency robust estimates for regression. *Annals of Statistics*, *15*, 642–656.

Yohai, V. J. (1997). Local and global robustness estimators. *Journal of Statistical Planning and Inference*, *57*, 73–92.

Yohai, V. J., Stahel, W., & Zamar, R. (1991). A procedure for robust estimation and inference in linear regression. In W. Stahel, & S. Weisberg (Eds.), *Directions in robust statistics and diagnostics, Part II* (pp. 365–374). New York: Springer-Verlag.