The Speech Learning Model (SLM) account of how Japanese speakers learn English /r/ and /l/

James Emil Flege

Professor Emeritus
School of Health Professions
University of Alabama at Birmingham

Sophia University, Tokyo
July 21, 2018

*The research reported here was supported by grants from the National Institute for Deafness & Other Communicative Disorders (NIDCD) over a 25 year period.
Abstract

This presentation reviews research examining the perception and production of English /r/ and /l/ by native Japanese (NJ) speakers first exposed to English as adults. The problems that such “Late learners” have with English liquids is not the result of exposure to English after the close of a “critical period” nor to an irreversible “filtering out” of phonetic properties needed to distinguish one English liquid from the other or from Japanese sounds. The results instead support the Speech Learning Model (SLM), which posits that all capacities used in L1 acquisition remain intact and accessible for L2 learning, even by Late learners.

From the SLM prospective, it is impossible for NJ speakers’ production and perception of English liquids to be identical to English monolinguals’. NJ speakers obtain quantitively less native-speaker input and are exposed to substantially more foreign-accented input than children who learn English as an L1. As well, NJ speakers need to establish and maintain phonetic categories in two phonetic subsystems, some of which are based on input from two languages. NJ speakers who receive abundant native speaker input may approach native-like accuracy for English /r/ due to the formation of a new phonetic category for that sound. However, the perceived phonetic distance between English /l/-Japanese /R/ is too small to trigger new category formation for English /l/. As a result, NJ speakers develop a composite phonetic category based on the aggregated distribution of tokens that are tacitly classified as English /l/ and Japanese /R/. The lack of native-like production and perception of English /l/ by highly experienced Late learners, therefore, does not demonstrate an incapacity to learn. Rather, it provides evidence of ongoing reorganization of bilingual phonetic subsystems in response to new phonetic input.
Aims of this talk

This talk focuses on how native Japanese (NJ) speakers learn to produce and perceive /r/ and /l/. This topic is of interest given the widespread view that it is difficult, even impossible for NJ adults to learn these sounds. Work focusing on /r/ and /l/ thus provides an ideal way to test “general principle of learning” [as well as] “claims about adult neural plasticity” (Bradlow, 2008, p. 294).

Hattori & Iverson (2009: 477) referred to the English /r/-/l/ distinction as “one of the most difficult” examined in the now vast L2 speech learning literature. Takagi & Mann (1995: 387) suggested that “even extensive natural exposure does not ensure perfect perceptual mastery” of English /r/ and /l/. In the 12 studies summarized by Flege et al. (1996), NJ adults identified English /r/ and /l/ correctly just 69% of the time, on average. They made bi-directional errors: labelling /r/ as /l/ and vice-versa. Researchers have endeavored, with varying degrees of success, to train NJ adults to perceive English liquids (e.g., Logan, Lively & Pisoni, 1991) English liquids. When positive results are obtained, the effects of perceptual training transfer to production (Bradlow, Akahane-Yamada, Pisoni, Tohkura, 1999). For those who focus on training in the classroom, /r/ is considered to be a greater challenge than /l/ (e.g., Brown, 1970. Saito & Lyster, 2012).
Aims of this talk

A substantial amount of research over the past 50 years has derived from the fact that Japanese has just a single liquid consonant whereas English has two.

The Japanese liquid is a tap that partially resembles both English /d/, in which the tongue tip contacts the alveolar ridge, and English /l/ (Vance, 2008; Arai, 2013). In keeping with a convention in the speech learning literature, I will symbolize the liquid consonant of Japanese as “R” and the English liquids as “r” (phonetically [ɹ]) and “l”.

The focus of this talk is motivated by the fact that although progress has been made, we do not yet have clear understanding of the “/r/-/l/ problem”. Hoping to shed light on the situation, I will:

1. Review a few of the many studies examining the perception and production of English /r/ and /l/ by NJ speakers;
2. Explain how these data can be interpreted within the framework of the revised Speech Learning Model (Flege, 1995, 2005, 2007, Flege & Bohn, in preparation)
The SLM: basic concepts

The aim of the SLM is to account for how phonetic systems reorganize in the face of new phonetic input that is received over the course of the life span. The SLM is normally applied to the learning a second language (L2) but may also be applied to the learning of a second dialect (D2) of the native language.

The SLM proposes that learning a native language (L1) and a second language (L2) proceed in the same way, using the same mechanisms and processes. However, the outcomes of L1 and L2 speech learning may differ, sometimes drastically.

Whereas L1 learning is thought by many to occur rapidly and perfectly, L2 speech learning is typically seen as a slow, labored process that, for many individuals, will ultimately be unsuccessful. Lenneberg (1967) noted, for example, that nearly everyone who learns an L2 after “puberty” (roughly 12-13 years of age) speaks it with a foreign accent. This observation led to the proposal that a “critical period” exists for L2 learning. By hypothesis, first exposure to an L2 after the closure of the critical period results in incomplete and/or imperfect learning.
The SLM: basic concepts

The SLM rejects the view that persistent differences between L2 learners and individuals who learned the target L2 as young children and speak no other language (monolinguals) are due to the closure of a critical period.

The model proposes that persistent native vs non-native differences in the production and perception of sounds (position sensitive allophones of vowels and consonants) arise from two primary factors:

1. L2 learners typically obtain far less native-speaker input than do monolingual children. The revised Speech Learning Model (Flege & Bohn, forthcoming) defines L2 input pertaining to segment-sized units of speech that is “seen and heard in conversations and other informational exchanges using the L2, including bilingual conversations in which the L1 and L2 are mixed and code-switching occurs”. For the SLM, input is as crucial for the learning of vowels and consonants in an L2 as it is for the attunement to the native language phonetic system in infancy and childhood.

2. Once L2 learners begin using their L2 regularly to communicate they are, by definition, bilinguals. As bilinguals, they face the need to establish and maintain contrast between elements of both their L2 phonetic subsystem and that of the previously established L1 phonetic subsystem.
The SLM: basic concepts

The SLM focuses on the learning of position-sensitive vowels and consonants ("sounds") rather than phonemes. This is because the sounds to be learned in an L2 may vary according to position and phonetic context. In this talk I will focus exclusively on word-initial singleton consonants. The findings for these consonants is unlikely to generalize completely to other positions and contexts.

Consider, for example, the learning of English /r/ and /l/ by native Japanese (NJ) speakers.

Takagi & Mann (1995) found the NJ adults living in the US were less able to identify English liquids in word-initial than -final position (see also Mochizuiki, 1981; Pisoni et al., 1994).
The SLM: basic concepts

The SLM focuses on language-specific phonetic categories found in the L2 and L1 and also on the phonetic realization rules used to implement those categories motorically. It does not focus on phonological “contrasts” used to distinguish words in the mental lexicon. In other words, the SLM focus is the learning of what sounds are rather than what sounds are not.

Phonetic categories are multi-dimensional, cue-weighted representations in long-term memory that are based on input distributions, that is, the sensory stimulation available when speech sounds are seen and heard. Such categories have traditionally been thought of as points in an n-dimensional “phonetic space”.

From the SLM perspective, there is no “/r/-/l/ problem”, only “/r/ problems” and “/l/ problems”. Analyses that combine results pertaining to the production and perception of /r/ and /l/ are likely to be misleading. NJ speakers learn to produce and perceive /r/ and /l/ to varying extents, not a contrast between these sounds. Phonetic contrasts are important, of course, but they emerge through segmental learning and are epiphenomenal units of description. As discussed later, the SLM predicts that NJ speakers, even Late learners, will produce and perceive English /r/ but not /l/ accurately if they have been exposed to a sufficient amount of (mostly) native speaker input over a sufficient number of years.

Research showing that highly experienced NJ speakers are only partially successful in learning /r+/l/ combined would not falsify the SLM because the native vs non-native difference might be due mostly or exclusively to /l/. On the other hand, showing mastery of /l/ but not /r/ would falsify the model, as would the demonstration of incomplete learning of English /r/ by NJ adults who had received abundant native-speaker input for decades.
The SLM: basic concepts

When native speakers hear foreign accent in the speech of nonnatives (Lenneberg, 1967) or detect an error in the production of a specific vowel or consonant, even a purely phonetic change that does not change meaning (Flege, 1984; Flege & Munro, 1994), they normally think that the inaccurate production arose from an inability on the part of the nonnatives to articulate specific sounds.

The SLM takes a different approach. It proposes that the process of L2 speech learning recapitulates what occurs in L1 acquisition, where perception normally reaches adult-like levels sooner than segmental production does.

In the first formal presentation of the SLM, Flege (1995, p. 239) proposed that L2 perception “leads” L2 production, noting that “the production of a sound will eventually correspond” to the properties represented in its phonetic category representation.

If native-like segmental perception of a sound is a necessary but not sufficient condition for its accurate production, then many detectable errors in the production of L2 sounds have a perceptual origin: the absence of native-like phonetic categories defining how a segment-sized unit of speech “ought” to sound in a particular position and context. According to the SLM, an unspecified amount of time may elapse between the establishment of a (perceptual) phonetic category and the creation of phonetic realization rules needed to motorically output the category.
The SLM: basic concepts

The “production before perception” hypothesis of the original version of the SLM (Flege, 1995) is consistent with the findings of Bradlow, Akahane-Yamada, Pisoni & Tohkura (1999) who provided high-variability identification training to adult NJ speakers. As expected, the training led to more accurate perception. It also resulted in improved articulation of /r/ and /l/, which was not explicitly trained. Moreover, the improvements in both segmental perception and production were still evident three months after the laboratory training had been completed.

The “production before perception” hypothesis was also supported by the results of a training study. Shinohara and Iverson (2013) administered feedback training to NJ speakers in the United Kingdom and Japan. Five sessions of perceptual training led to improved production. Acoustic measurements revealed that the F3 frequency of /r/ productions had become significantly more English-like. However, as predicted by the SLM (see below) the F3 values for /l/ productions did not change significantly.

It is important note, with reference to L1 acquisition, a recent study by Idemaru & Holt (2013). These authors examined in detail the development of /r/ and /l/ by 48 monolingual children learning English. They found little evidence of a correlation between improvements in specific aspects of production and perception in monolingual children ranging in age from 4.0 to 8.5 years.
The SLM: basic concepts

It is also important to note that more native-like perception than production has not always been noted in L2 learning (e.g., Goto, 1971). The lack of correspondence between production and perception, or one going in the opposite than expected direction, may have methodological bases (e.g., the inherent incommensurability of measures of accuracy in two domains) and may occur more often in complex late-developing sounds than in less complex sounds that develop early in L1 acquisition and are found in the majority of human languages (Flege, 1999).

Evaluating the production-perception relation is further complicated by the fact that the “/r/-/l/ problem” is so well known to Japanese students and teachers.

Watch this funny YouTube video featuring Japanese speakers who try gamely but unsuccessfully to say “refrigerator”

https://www.youtube.com/watch?v=Ids82u3qJgE
The SLM: basic concepts

A continuing challenge for L2 researchers is to obtain speech samples that adequately reflect how NJ adults actually produce English liquids in everyday speech communication.

Awareness that a study focusses on the “/r/-/l/ problem” may evoke the use of explicit articulation techniques that were learned at school. Many NJ speakers have been taught, for example, that it is possible to produce a recognizable English /r/ by rounding the lips while replacing the the English /r/ with Japanese /w/; see Yamada & Tohkura, 1990; Goto, 1971).

Perceptual support for the use of a conscious articulation strategy in a laboratory experiment derives from the fact that NJ adults with little conversational experience are better able to discriminate /r/ from /w/ than to discriminate /r/ from /l/ (Best and Strange, 1992).

Use of conscious articulation strategies are probably most likely in a standard word list reading task, and least likely in real conversations involving a topic of great interest. Careful L2 production studies therefore feature multiple elicitation tasks supplemented by acoustic analyses.
So what’s the problem?

Children who learn English as their L1 take years to establish language-specific phonetic categories for English /r/ and /l/ (see Idemaru and Holt, 2013, for data and review).

The learning of the singleton word-initial /R/ by monolingual NJ children also takes many years (Arai, 2013).

When native NJ speakers who learn English as an L2 are first exposed to English—in childhood, adolescence or adulthood—they usually judge the two English liquids as being instances of the one Japanese liquid, that is the two sounds are “perceptually assimilated” by Japanese /R/ (see Best & Strange, 1992, for an early application of what became the Perceptual Assimilation Model) and produced accordingly.
So what’s the problem?

From an acoustic standpoint, the most important (of several) characteristic of English /r/ is third formant (F3) onset frequency. The frequency of F3 starts out low in /r/ (top), close to F2 frequency, and then rises rapidly when constriction is released. For English /l/, on the other hand, F3 frequency starts out much higher and so is more distant from the F2 frequency than is the case for /r/.

According to Miyawaki et al. (1975, p. 332), the onset frequency of F3 in Japanese /R/ varies “unsystematically over a range of values sufficient to distinguish American English [r] and [l]” in word-initial singletons.

Lotto, Sato and Diehl (2004) noted that, in addition to the fact that the F3 values in Japanese /R/ occur near the “boundary” between the distribution of values for English /r/ and /l/, the F3 and F2 values of Japanese /R/ tend to covary whereas, in English liquids, F2 and F3 values are largely independent.
So what’s the problem?

Iverson et al. (2003) identified the key perceptual problem facing NJ learners of English (see also Gordon, Keyes & Yung, 2001). These authors created a set of synthetic speech “ra” and “la” syllables in which F2 and F3 frequencies varied orthogonally. By design, the auditory distances (in Mel units) between all pairs of stimuli in the grid were equal.

*Synthetic /r/ and /l/ stimuli used by Iverson et al. (2003)*
So what’s the problem?

Twenty-four native English (NE) adults and 24 NJ adults tested in Tokyo (none of whom had lived abroad) identified the randomly presented synthetic stimuli using phonetic categories found in their native language. The NE speakers labelled the stimuli as “r” or “l”, the NJ speakers as “R” or “w”. The stimuli were then rated in terms of goodness of fit to the category used in labelling. This scale ranged from 1 (bad) to 7 (good).

All possible pairings of the stimuli were then rated for “acoustic similarity” using a scale that ranged from 1 (dissimilar) to 7 (similar). The similarity ratings were then examined in multi-dimensional scaling (MDS) analyses.

*Synthetic /r/ and /l/ stimuli used by Iverson et al. (2003)*
So what’s the problem?

A MDS analysis revealed that the NE speakers gave greater perceptual weight to the F3 dimension when rating the similarity of pairs of stimuli than might have been expected based on purely auditory differences between the stimulus pairs. The F3 dimension was “stretched” due to its key role in distinguishing English /r/ from /l/.

Above: The percentages in the circles at left indicate mean % identification as /r/ (black) or /l/ (white). The values above the percentages indicate the mean goodness of fit rating of the stimuli as classified. So, for example, the top-left stimulus was classified as /r/ 92% of the time and received an average rating (as /r/) of 3.8. The bottom-right stimulus was classified as /l/ 100% of the time, and received a rating (as /l/) of 6.0.
So what’s the problem?

Although the NJ speakers were given two response alternatives to use when identifying the stimuli, they labelled most stimuli as /R/ (black circles at left). What one might consider an “overuse” of F2, from an English perspective, may have resulted from the use that is made of F2 in Japanese to specify /w/. The NJ speakers made less use of the F3 dimension than the NE speakers did and so showed no evidence of perceptually “stretching” the F3 dimension that was evident for the NE speakers.
So what’s the problem?

The MDS analyses carried out by Iverson et al. (2003) were based on ratings obtained for the two stimuli occurring in pairs, something that does often occur in everyday speech perception. The “stretching” observed for NE speakers was likely the consequence of juxtaposing in pairs sounds that were categorized differently. Differential categorization is known to augment the perceived dissimilarity of sounds presented in pairs (see, e.g., Flege, Munro & Fox, 1994, Figure 5).
So what’s the problem?

Iverson et al. (2003) concluded that the native vs non-native differences they observed were not due to a lack of *auditory* sensitivity to the F3 dimension on the part of the NJ speakers.

Their interpretation is consistent with the finding of Miyawaki, Liberman, Jenkins, Strange et al. (1975), who showed that NJ speakers who were unable to discriminate English [ɾa] and [la] syllables were nevertheless able to discriminate isolated third-formant portions extracted from the same syllables.

The conclusion drawn by Iverson et al. (2003) is also congruent with research examining perceptual changes in infants. Kuhl, Stevens et al. (2006) found that whereas both Japanese and American infants were equally able to discriminate /r/ from /l/ at seven months of age, the American infants did so much better than the Japanese infants at 11 months, reflecting an “attunement” to the phonetic system of English. In other words, the cross-language perceptual difference that emerged towards the end of the first year of life was not due to a loss of *auditory acuity* in the normally developing children who were tested.
So what’s the problem?

Iverson et al. (2003) proposed that NJ speakers’ perceptual space may be “miss-tuned” for the acquisition of the phonetic distinction between English /r/ and /l/. As a result, auditory and/or early stage phonetic processing may interfere with NJ adults’ learning of English /r/ and /l/. This interference may be “difficult to reverse”, according to the authors, thereby making it difficult for Japanese learners of English to gain perceptual access to the dimensions they need to learn English /r/ and English /l/.

Of course the inference drawn by Iverson et al. (2003), if true, may not apply to NJ adults who have more experience in English than the 24 NJ adults who were tested in Tokyo.
So what’s the problem?

One frequent aim of L2 speech research is to account for the seeming effect of age of first exposure on phonetic learning. Researchers ask: Why is earlier better than later? Two related questions come to mind regarding the findings obtained by Iverson et al. (2003):

1. Is the low-level “filtering” (alternatively, L1 optimized weighting pattern, see Lotto, 2004) of the acoustic phonetic properties that define the difference between English /r/ and /l/ stronger for NJ Late than Early learners?

2. Is the filtering (L1 optimized cue weighting) irreversible for Late learners?
Perception of /r/ and /l/ 

Goto (1971) was the first to show that NJ speakers have difficulty discriminating English /r/ from /l/. MacKain, Best & Strange (1981) later evaluated NJ speakers’ ability to discriminate and identify members of a synthetic /r/-/l/ continuum. These authors compared relatively “inexperienced” and “experienced” Late learners who differed in length of residence (LOR) in the United States ($M = 0.7$ vs $2.3$ years) and percentage English use ($M = 29\%$ vs. $55\%$). Their aim was to determine if the NJ speakers perceived the stimuli “categorically”.

The members of a speech continuum are said to have been perceived “categorically” if pairs of stimuli straddling a the phoneme boundary defined by identification responses are discriminated more accurately than pairs of stimuli manifesting the same physical difference but located one side or the other of the phoneme boundary.

Six of the seven inexperienced NJ speakers tested by MacKain et al. (1981) showed “near-chance” performance when identifying and discriminating the /r/-/l/ stimuli. However, all five experienced NJ speakers resembled NE speakers in showing categorical perception. This suggested that NJ adults having sufficient conversational experience in English can establish a phonemic category for English /r/, English /l/, or potentially both English liquids.
Perception of /r/ and /l/

Yamada (1995) tested 276 NJ speakers having a mean age of 20 years. More than half ($n = 156$) were tested in Tokyo soon after returning to Japan following a period of residence in the United States. The “returnees” were assigned to one of five “JE” groups based on LOR in the United States. The remaining participants were NJ speakers who had never lived abroad and NE speakers who were residing in Japan at the time of the study.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>AOA</th>
<th>LOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>NJ</td>
<td>120</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>JE 1</td>
<td>46</td>
<td>15.3</td>
<td>1.0</td>
</tr>
<tr>
<td>JE 2-3</td>
<td>33</td>
<td>11.7</td>
<td>2.5</td>
</tr>
<tr>
<td>JE 4-5</td>
<td>34</td>
<td>9.6</td>
<td>4.6</td>
</tr>
<tr>
<td>JE 6-7</td>
<td>31</td>
<td>6.7</td>
<td>6.4</td>
</tr>
<tr>
<td>JE 8+</td>
<td>12</td>
<td>5.4</td>
<td>12.0</td>
</tr>
<tr>
<td>NE</td>
<td>42</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>
Perception of /r/ and /l/

The NJ and NE speakers identified members of a synthetic speech continuum ranging from “right” to “light”. Both F2 and F3 values in the stimuli varied, but not independently (see also Yamada & Tokhura, 1992). The randomly presented stimuli were identified as “r”, “l” or “w”.

![Diagram showing synthetic stimuli in a "right" to "light" (/r/-/l/) continuum (Yamada & Tohkura, 1992)]
Perception of /r/ and /l/

The /r-/l/ identification functions obtained for the NJ returnees resembled those of the NE speakers more than did those of NJ adults who had never lived abroad. However, the returnees’ identification functions exhibited more variation than those obtained for the NE speakers.
Perception of /r/ and /l/

Yamada (1995) also examined the identification of synthetic stimuli in which F2 and F3 frequencies varied independently (F2 in four steps from 800-1400 Hz, F3 in ten steps from 1200-3000 Hz). NE speakers rarely heard any of these stimuli as “w”. They usually labelled stimuli having F3 frequency values higher than 2000 Hz as “l” and stimuli having lower F3 frequencies as “r”.

Identification of synthetic stimuli by NE speakers

![Identification of synthetic stimuli by NE speakers](image)
Perception of /r/ and /l/

NJ adults who had never lived abroad (below left) gave far more “w” responses than the NE speakers did. They made little use of F3 to distinguish English /r/ from English /l/. The responses of this group might be considered indicative of young NJ adults having very little or no conversational experience in English.

The returnees (below right) all had conversational experience in English while living in the United States. They resembled the NE speakers to a greater extent than did the NJ adults who had never lived abroad. However, some of the returnees made far less use of F3 frequency than the NE speakers did.
Perception of /r/ and /l/

The returnees’ use of F3 when identifying English liquids in synthetic stimuli varied as a function of LOR (below, left). The longer they had lived in the United States, the more they used F3. The percentage of correct identifications of /r/, /l/ and /w/ in naturally produced words (below, right) also varied as a function of LOR. The longer the US residence, the more often the returnees correctly identified the three English consonants.

As shown later, NJ speakers have no difficulty identifying English /w/, and so we can infer that the percent correct scores seen at the right reflected identifications of /r/ and /l/. We can also infer that perceptual use of F3 influenced identification of the English liquids.
Perception of /r/ and /l/

What explains the effect of LOR on the NJ speakers’ perception of English /r/ and /l/? Did members of returnee groups JE 6-7 and JE 8+ resemble NE speakers more than members of the other returnee groups because all of them had lived in the US for at least 6 years? Because all had arrived in the US before 8 years of age? Both?

Unfortunately, no certain conclusion can be drawn because AOA (age of arrival) and LOR were confounded in the sample of NJ speakers tested by Yamada (1995).

Yamada (1995) observed that both AOA and LOR were important predictors of performance. She hypothesized that two factors influenced the returnees’ perception of English /r/ and /l/. One was an endogenous (or “biological”) factor that decreases the plasticity of neutral systems used in speech learning as a function of age of first exposure to an L2. The other was an exogenous factor related to overall amount of L2 input received, here defined by LOR (a rough indicator of quantity but not by quality of L2 input).
Perception of /r/ and /l/: the SLM

The SLM takes a different approach. It posits that there is no “critical period” for L2 speech because the capacity to learn speech remains *intact* across the life span.

Differences do of course exist between Early and Late learners, even after years of daily use of the L2 in a predominantly L2-speaking country. However, according to the SLM, such differences are not due to a loss or diminution of the capacities that children exploit when learning their L1 in a monolingual environment. On this view, if a low-level phonetic filter exists, its influence on NJ speakers’ learning of English liquids will not be “irreversible” (Iverson et al., 2003).

The SLM posits that when learners of any age are first exposed to a foreign or second language they automatically relate the L2 sounds they hear to sounds found in their L1 (a process called “interlingual identification”). This form of cross-language mapping is hypothesized to occur automatically and rapidly, even when the L2 sounds are quite different phonetically from those in the L1 inventory.
Perception of /r/ and /l/: the SLM

According to the SLM, the learning of L2 sounds depends importantly on the *perceived relationship* between each “sound” (position sensitive vowels and consonants) encountered on the phonetic surface of an L2 and the perceptually closest sound(s) in the L1 inventory. Segmental-level learning also depends importantly on the quantity and quality of L2 *input* that has been received.

The magnitude of perceived L1-L2 phonetic differences must be measured empirically for individual learners, not groups, because the unit of analysis in speech learning is the individual. Perceived phonetic differences cannot be derived from textbook descriptions of how the L1 and L2 differ at a phonetic and/or phonological level of organization.

The SLM predicts that NJ speakers who are first exposed to English in an immersion context, regardless of age of first exposure to English, are able to develop a new phonetic category for English /r/ if the English /r/ tokens they encounter fall outside the distribution of tokens that define their Japanese /R/ category.
The Speech Learning Model (SLM)

For English /l/, on the other hand, NJ speakers may develop a “composite” phonetic category that aggregates the phonetic properties of English /l/ and Japanese /R/. This will happen if NJ speakers continue to perceive English /l/ tokens to be instances of their Japanese /R/ category. In the SLM framework, the term “equivalence classification” denotes a permanent pattern of interlingual identification.

At the time of first exposure to English (Time 1), most if not all NJ speakers will identify tokens of both the English /r/ and /l/ categories as being instances of their Japanese /R/ category. For L2 speech learning to occur, they must begin to differentiate English /r/ from /l/, not with respect to one another but in terms of how tokens of the two English liquid categories are related perceptually to the Japanese /R/ category (Time2).
The results of Takagi and Mann (1995) indicated that NJ Late learners can gain access to the F3 dimension. Two groups of NJ adults living in the United States identified the members of a synthetic /r/-/l/ continuum. The NJ speakers generally used F3 less than the NE speakers did. However, the NJ speakers with greater conversational experience (i.e., longer LOR) in English showed significantly greater use of F3 than those with less conversational experience (shorter LOR).

Perception of /r/ and /l/
Perception of /r/ and /l/

Changes in the perceptual use of the F3 dimension might be attributed to the fact that tokens of English /r/ are perceived to be phonetically less similar to Japanese /R/ than are tokens of English /l/ (Sekiyama & Tohkura, 1993). Takagi (1993) asked the same NJ speakers to rate natural tokens of /r/, /l/ and Japanese /R/ as instances of the Japanese /R/ category. The English /r/ and /l/ tokens were presented in separate counterbalanced sessions along with Japanese /R/ tokens.

The highest ratings of stimuli as instances of the /R/ category were, of course, obtained for the Japanese /R/ tokens.

The English /r/ tokens were judged to be significantly poorer instances of the Japanese /R/ category than were the English /l/ tokens.
Perception of /r/ and /l/  

Hattori & Iverson (2009) used an identification test rather than a rating procedure to evaluate the perception of cross-language phonetic differences but nonetheless obtained results similar to that of Takagi (1993). NJ adults having a median LOR of 3 months in London identified naturally produced English and Japanese CV syllables using three category labels. They misidentified English /l/ tokens as Japanese /R/ significantly more often ($p<.01$) than they misidentified English /r/ tokens as Japanese /R/.

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Eng /l/</th>
<th>Eng /r/</th>
<th>Jap /l/</th>
</tr>
</thead>
<tbody>
<tr>
<td>English /l/</td>
<td>82%</td>
<td>16%</td>
<td>2%</td>
</tr>
<tr>
<td>English /r/</td>
<td>22%</td>
<td>58%</td>
<td>19%</td>
</tr>
<tr>
<td>Japanese /r/</td>
<td>6%</td>
<td>17%</td>
<td>77%</td>
</tr>
</tbody>
</table>
Perception of /r/ and /l/

Guion et al. (2003) tested two groups of NJ speakers in Japan and a group of NJ speakers living in the United States. The NJ adults tested in the US, group “J1” (red arrow below), were able to discriminate English /r/ from the Japanese /R/ at a significantly above-chance rate. However, despite their conversational experience in English while living in the United States, the members of group J1 were unable to discriminate English /l/ tokens from the Japanese /R/ (blue arrow below) at a significantly above-chance rate.
Perception of /r/ and /l/

Many researchers have concluded that NJ adults are unable to learn English /r/ and /l/. Before reaching such a conclusion, however, it is necessary to consider L1 learning.

Arai and Mugitani (2016) reviewed four studies examining the acquisition of /R/ by monolingual Japanese children. The age at which 90% of the Japanese children could produce /R/ correctly varied from 4;0 to 6;0 years.

According to Arai (2013), /R/ is difficult for children to acquire owing to its articulatory nature and also because Japanese /R/ is produced in diverse ways. Diverse “targets” for an L1 sound may prolong the time needed to reach mature, adult-like levels of performance. Differences in the kind productions to which children are exposed may lead to lead to differences in perceptual cues and cue weightings that are deployed in the identification of L1 categories (see Idemaru & Holt, 2013, for English). In later life, these differences may in turn influence how L2 sounds are perceptually related to the L1 sound.
Perception of /r/ and /l/

Research with monolingual NE children in the United States indicates that children learn to produce /l/ somewhat before /r/. Of the boys tested by Smit, Hand, Freilinger, Bernthal and Bird (1990) 90% could produce /l/ “acceptably” by 6;0 years of age, but this benchmark was not reached for /r/ until two years later. (Girls met both benchmarks somewhat earlier than boys.)

As will be shown later, NJ adults who learn English in the United States are more successful in learning to produce and perceive /r/ than /l/, the opposite of what might be expected from a consideration of how monolingual children learn English as an L1. This indicates the importance of perceived L1-L2 mapping relations.
Perception of /r/ and /l/

The fact that monolingual American children learn /l/ sooner than /r/ might be attributed to the greater allophonic variation in the production of American English /r/ than /l/.

English /l/ is produced with “dark” and “light” variants that are readily detectable by listeners. Dark /l/ is more vowel-like than the light (“clear”) variant and tends to occur in post-vocalic position whereas the light (“clear”) variant tends to occur more often in pre-vocalic position (Mielke et al., 2016, p. 123). Ladefoged and Maddieson (1996) noted that the articulation of English /r/ varies considerably among monolingual in North American. It can be produced as an alveolar retroflex or as a post-alveolar approximant with lower pharyngeal constrictions and lip rounding. Some talkers produce /r/ with a “bunched” tongue in which in the tongue tips is not raised, with constriction along the hard palate as well as in the lower pharynx.

Substantial individual differences exist in the articulation of American English /r/ (e.g., Westbury et al., 1998). Mielke et al. (2016) examined the production of /r/ by 27 speakers of American English, 20 of whom appear to have been monolinguals (see Table 2). Of the 37 speakers examined, two produced only retroflexed /r/s, 16 only bunched tongue /r/s and nine produced both types. The retroflex variant was far more common in pre- than post-vocalic position and when preceding the vowels /ɑ/ and /o/ than /i/. Differences the articulation of /r/ variants are much less evident to listeners than are difference in dark and light productions of /l/ because substantial differences in /r/ production give rise to much the same acoustic output.

Two important questions arise regarding the acquisition of /r/ and /l/ by NE children. Do differences in allophony contribute to the later acquisition of /r/ than /l/? And how do individual differences in the articulation of /r/ arise in the first place? According to McGowan et al. (2004) this difference has a purely articulatory basis, namely the need for two lingual constrictions for initial /r/, “one in the mouth and the other in the pharynx” but only one oral construction for /l/ in word-final position.

Another possibility is that the allophony of /r/ is far more complex than that of /l/ (Song et al., 2015). Mielke et al. (2016, p 101) speculate that a monolingual American child who succeeds in producing an adult-like /r/ in word-initial position will then “attempt to use the same strategy in increasingly challenging contexts until eventually either succeeding in all [contexts] or beginning to use bunched /r/.” The learning of /r/ in American English, for these authors, arises from “idiosyncratic phonetic pressures and acquisition paths” (Mielke et al., 2016, p. 133) given that there is no “input motivating a particular allophony pattern” (p. 132). However, auditory input cannot be ruled out inasmuch as /r/ allophones are distinguished on the basis of F4 and Frequencies (Zhou et al., 2008).
Perception of /r/ and /l/

Importantly, a few monolingual English boys tested by Smith et al. (1990) had not yet learned to produce American English /r/ acceptably at the age of 9 years. As we will see later, some monolingual NE speakers may never learn English /r/ (see also Goto, 1971, p. 318, for Japanese /R/).

Idemaru & Hold (2013) examined monolingual children’s perception of English and provided detailed acoustic analyses of their production of /r/ and /l/. This research suggests that the attunement to the properties of English /r/ and /l/ continues until 8-9 years of age.

adapted from Smit et al. (1990), Figs. 6 and 17
Perception of /r/ and /l/

Returning to the Guion et al. (2003) results:

I calculated years of **equivalent full-time English input** (“English input” for short) for the three groups of NJ speakers in the Guion et al. (2003) study, multiplying length of residence (LOR) in the United States, in years, by self-reported % English use.

Group J1 had a mean of 3.1 years of English input (range = 1.8 to 5.5 years) at the time of testing, about half the input that most monolingual American children need to acquire English /l/ (Smit et al., 1990; see also Idemaru & Holt, 2013). As well, members of Group J1 certainly heard English produced with a Japanese foreign accent more than the children tested by Smit et al. (1990) and Idemaru and Hold (2013).

These considerations lead us to wonder if the apparent inability discriminate English /l/ from Japanese /R/ by the US-resident NJ adults tested by Guion et al. (2003) was permanent, or if it should it be considered *learning in progress*. The theoretical prediction of the SLM is that this limitation on the learning of English /l/ is permanent.
Perception of /r/ and /l/

How much conversational experience do NJ adults need in order to show a difference in the perceived phonetic “distances” between English /r/-Japanese /R/ and English /l/-Japanese /R/? To address this question, Aoyama & Flege (2011) tested 50 NJ speakers living in the United States having these characteristics:

<table>
<thead>
<tr>
<th>%</th>
<th>Mean(SE)</th>
<th>min-max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronological age</td>
<td>33.2(.8)</td>
<td>21-48</td>
</tr>
<tr>
<td>Age of arrival in US</td>
<td>29.5(.8)</td>
<td>19-39</td>
</tr>
<tr>
<td>Length of residence in US</td>
<td>3.7(.7)</td>
<td>0.1-25</td>
</tr>
<tr>
<td>% weekday use of English</td>
<td>55.0(5.0)</td>
<td>0-100</td>
</tr>
<tr>
<td>% weekend use of English</td>
<td>36.0(5.0)</td>
<td>0-99</td>
</tr>
</tbody>
</table>

The NJ Late learners rated the degree of similarity of various English and Japanese consonants to Japanese /R/. The consonants were presented in pairs along with Japanese /R/ tokens for ratings on a 7-point scale. As in Takagi (1993), the perceived relation between English /r/-/R/, and between English /l/-/R/ were evaluated in separate counter-balanced blocks.
Perception of /r/ and /l/

LOR showed weak but nonetheless significant correlations with both the English /r/-Japanese /R/ dissimilarity ratings, \( r(48) = -0.38, p = .007 \), and the English /l/-Japanese /R/ ratings, \( r(48) = -0.29, p = .039 \). The longer the NJ adults had lived in the United States the less similar they judged both English liquids to be with respect to Japanese /R/.

Aoyama & Flege (2011) examined the data obtained for only one of the two testing conditions, the Five-talker condition. To better understand the Aoyama & Flege (2011) finding, I re-examined the data for both that condition and the One-talker condition.

The stimuli in the “One-talker” condition, not previously analyzed, consisted of 5 tokens each of English CV syllables produced by an adult male native speaker of English, and Japanese CV syllables spoken by one adult male NJ speaker.

The CV syllables were randomly presented six times each for ratings on a scale ranging from 1 (not similar at all to Japanese /R/) to 7 (very similar to Japanese /R/). Responses to the first randomization were discarded. The median of five judgements of each token was determined, and then the mean of the five median values for each phonetic category was calculated.

The stimuli used in the following “Five-talker” condition consisted of the same English and Japanese CV syllables produced by five different native speakers of English and Japanese (all male). As for the “One-talker” condition, the ratings obtained for each NJ speaker for all phonetic categories being examined were based on a total of 25 judgments.
Perception of /r/ and /l/

For this re-analysis I assigned 16 NJ speakers each to subgroups based on LOR: relatively Short, Mid and Long. In addition to differing in LOR, the LOR-defined groups differed in AOA and percentage use of English (weighted average of weekday and weekend use).

<table>
<thead>
<tr>
<th>LOR groups</th>
<th>LOR</th>
<th>AOA</th>
<th>Age</th>
<th>%English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td>0.6</td>
<td>33.8</td>
<td>34.3</td>
<td>26.5</td>
</tr>
<tr>
<td>Mid</td>
<td>2.1</td>
<td>28.2</td>
<td>30.4</td>
<td>52.7</td>
</tr>
<tr>
<td>Long</td>
<td>8.4</td>
<td>26.8</td>
<td>35.1</td>
<td>63.8</td>
</tr>
</tbody>
</table>

$F(2,45) = 24.2$, $prob = .0000$,
$\text{Tukey .05} \ S,M<L \ S>M,L \ -- \ S<M,L$

$prob = .0004, .1144, .0026$
Perception of /r/ and /l/

Here are the mean ratings obtained in the One- and Five-Talker conditions (“a” and “b” respectively). In both, the degree of perceived phonetic dissimilarity of English /r/ and /l/ tokens with respect to Japanese /R/ was intermediate to the ratings obtained for actual tokens of Japanese /R/ and for tokens of Japanese /w/. This ordering was evident for all three LOR-defined groups.

As expected from previous research, English /r/ was judged to be less similar to the Japanese /R/ than English /r/ was.
Perception of /r/ and /l/

The dissimilarity ratings were examined in a (3) Group x (4) Consonant x (2) Condition ANOVA. To explore the three-way interaction it yielded, $F(6,135)=3.04$, $p=.008$, I tested the difference between the three LOR-defined groups for all four consonants (Japanese /R/ and /w/, English /r/ and /l/). Of the four consonants examined, the simple effect of Group was significant only for English /r/. This held true for both the One-talker $F(2,159) = 8.7$, Bonferroni adjusted $p<.05$, and the Five-talker condition, $F(2,159) = 7.8$, $p<.05$.

I tested the difference between English /r/-Japanese /R/ and English /l/-Japanese /R/ ratings for all combinations of Group and Condition (6 tests in all). The /r/ tokens received significantly lower ratings than the /l/ tokens did in both conditions for the Mid- and Long-LOR group, $F(1,15)=12.5$-$20.4$, Bonferroni adjusted $p<.05$. The ratings for /r/ and /l/ did not differ significantly for the Short-LOR group ($p>.10$) however.
Perception of /r/ and /l/

These results suggest that NJ adults living in the United States may not immediately detect phonetic differences between English /r/ and Japanese /R/. According to the SLM, the tacit awareness of cross-language phonetic differences develops gradually through experiences speaking and hearing the sounds that make up the L2 phonetic inventory.

Members of the Short-LOR group had lived in the United States for 0.1 to 1.1 years, those in the Mid- and Long-LOR groups for 1.2 to 24.6 years. Perhaps NJ adults need more than about one year of conversational experience in a predominantly English-speaking country in order to note the acoustic phonetic differences between English /r/-Japanese /R/. Of course, such tacit awareness may develop more rapidly for NJ speakers who use English frequently and are usually exposed to the correct productions of NE speakers rather than to the often incorrect productions of other NJ speakers.

For those who focus on classroom pronunciation training (e.g., Saito & Lyster, 2012), a year may seem like a very long time. However, it is a very short period in comparison to the time that monolingual NE children need to master English /r/ and /l/ through exposure to NE speakers (Smit et al., 1990).
Perception of /r/ and /l/

We have seen that the perceived relationship of the English consonants /r/ and /l/ with respect to Japanese /R/ changes as a function of conversational experience in English. Does this mean that NJ adults can gain access to the F3 dimension?

The results obtained by Hattori & Iverson (2009) suggest a positive response to this question. These authors tested 36 NJ speakers in London, using a method of adjustment procedure to evaluate the perception of /r/ and /l/.

The synthetic stimuli differed along five dimensions: F1, F2, F3, closure duration, transition duration. The authors developed software that enabled listeners to quickly search through the complex stimulus array to find the “best exemplars” of both English /r/ and English /l/. 
Perception of /r/ and /l/

At left we see the mean F3 values in the stimuli selected by NJ and NE adults as the best exemplars of the English /r/ and English /l/ categories. The mean F3 values for both liquids were much the same for the NE and NJ adults, indicating that NJ adults can gain perceptual access to the F3 dimension.

At right we see that the more the NJ adults’ preferred F3 values approximated those of NE adults for /r/ and /l/, the more accurately ($r = -0.46, p<.01$) they identified English /r/ and /l/ (see also Gordon et al., 2001). But what explains the enormous variability among NJ speakers?

Use of F3 dimension
Native Japanese  Native English

adapted from Hattori & Iverson (2009), Fig. 4

Hattori & Iverson (2009), Fig. 5
Perception of /r/ and /l/

Variation in use of the F3 dimension may have been influenced by how much English input the NJ adults had received, by differences in their Japanese /R/ category, or both.

I calculated years of full-time equivalent English input based on the LOR and % English use values in Appendix B of Hattori’s dissertation. Of the 36 NJ adults tested in London, four reported never using English while in London. These participants presumably had little or no conversational experience in English despite living at the time of test in a predominantly English-speaking country. Of the remaining participants, only seven had the equivalent of one full year of English input.
Perception of /r/ and /l/

Substantial inter-subject variable is often evident in research examining both the production and the perception of /r/ and /l/.

As mentioned earlier, MacKain et al. (1981) examined the identification and discrimination of members of a synthetic /r/-to-/l/ continuum by groups of NJ adults having mean LORs of 0.7 and 2.3 years. Most “experienced” but not “inexperienced” NJ participants showed an English-like pattern of “categorical” perception.

After data analysis had been completed, MacKain et al. (1981) were able to test an additional NJ speaker who had arrived in the US just two weeks earlier and reported using English 25% of the time. This new NJ participant showed categorial perception despite very little experience speaking and hearing English in the United States.

Ingvalson, Holt ad McClelland (2012) recruited 42 NJ adults who had lived in the United States for about 1.5 years on average for an innovative /r/-/l/ perceptual training study. Of the 42 NJ speakers who were tested, 14 (33%) had to be excluded from the study because they correctly discriminated /r/ and /l/ stimuli at rates that were so high that it would have made it difficult to observe improvements in perception, had the training procedure proved effective (it did not).
Perception of /r/ and /l/

Gordon, Keyes, & Yung (2001) examined use of F3 onset frequency and F1 transition duration as cues to the identity of synthetic speech stimuli as /r/ or /l/. All 12 NJ speakers who participated were living in the United States when tested. Some used both cues, some just F3 or just F1 transition duration, and some used neither cue. Use of the two cues under examination correlated with the NJ speakers’ ability to identify liquids in naturally produced English words.

Jun Yamada (1991) administered one hour of identification training to 152 college students in Tokyo. Feedback (correct, incorrect) as the students attempted to identify the initial consonant in minimally paired non-words beginning with /r/ and /l/.

Several students could already identify /r/ and /l/ correctly in all four minimal pairs before training began. Following the training, an additional 6% of the students could correctly identify /r/ and /l/ in all four minimal pairs and 35% could do so for several minimal pairs. However, 51% of the students remained at chance for all four minimal pairs.
Perception of /r/ and /l/

The inter-subject variability seen in the J. Yamada (1991) study can probably not be attributed to differences in conversational experience. The amount of conversational experience of the NJ speakers in the Gordon et. al. (2001) study did vary but the effect of such variation was not examined, probably because of the small number of NJ speakers tested.

The inter-subject variability seen in the two studies may have arisen from: an undefined individual difference(s) in auditory and/or cognitive capacity; from undocumented differences in English input; or from differences in the NJ speakers’ /R/ categories.

Given that Japanese /R/ can be articulated in diverse ways (Best and Strange, 1992; Arai, 2013), NJ speakers’ /R/ phonetic categories may differ. If so, the perceived differences between Japanese /R/ and English /r/ (and /R/-/l/) may be greater for some individual NJ speakers than for others. The same explanation might also account also for the inter-subject variability seen in the Hattori and Iverson (2009) study. Additional research is clearly needed.
Perception of /r/ and /l/

To summarize so far: NJ adults perceive English /r/ to be phonetically more dissimilar from Japanese /R/ than English /l/; and they can gain perceptual access to the F3 dimension, the most important acoustic phonetic dimension needed to specify the English /r/ category.

That being the case, will NJ adults who have substantial conversational experience in English create new phonetic categories for English /r/? The results obtained by Flege, Takagi and Mann (1995, 1996) suggest a positive response to this question.
Perception of /r/ and /l/  

To avoid the selection bias that occurs when participants are recruited on a university campus, Flege et al. (1995, 1996) recruited NJ residents of Irvine, California on three successive Saturday mornings at a market that sells Japanese food and cosmetics.

Selection of the NJ speakers was based exclusively on length of residence (LOR) and gender (4 males and 8 females per group). Members of the relatively “experienced” group (EJ) had lived in the United States for an average of 20.8 years, those in the relatively “inexperienced” group (IJ) for just 1.6 years.

Flege et al. (1995, 1996) were aware of the short-comings of LOR as an index of quantity of L2 input. They reasoned, however, that NJ speakers who had lived in the US for an average of 20.8 years would surely have had more experience hearing English spoken than those who had lived in the US for just 1.6 years, on average.

The first 12 persons who appeared to be Japanese and who indicated having lived in the United States for three years or less, and the first twelve Japanese adults who indicated an LOR greater than 12 and met gender requirements were invited to participate in a study that involved “the learning of English words”. Given that all accepted, this approach comes close to a random sampling of the Japanese population in Irvine.
Perception of /r/ and /l/

Given that only LOR was used in selecting participants, it is not surprising that the two groups differed significantly in other ways. Compared to members of the IJ group, EJ members were older, arrived in the US at a younger age, and used English more. One or more of these variables may have affected the outcome measures in addition to, or instead of LOR.

<table>
<thead>
<tr>
<th></th>
<th>NE</th>
<th></th>
<th>EJ</th>
<th></th>
<th>IJ</th>
<th></th>
<th>EJ vs. IJ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>range</td>
<td>M</td>
<td>range</td>
<td>M</td>
<td>range</td>
<td>F(1,22)</td>
</tr>
<tr>
<td>LOR</td>
<td>36.1</td>
<td>29-47</td>
<td>20.8</td>
<td>12-29</td>
<td>1.6</td>
<td>1-3</td>
<td>186.5</td>
</tr>
<tr>
<td>Age</td>
<td>36.1</td>
<td>29-47</td>
<td>43.7</td>
<td>40-47</td>
<td>35.0</td>
<td>29-44</td>
<td>34.8</td>
</tr>
<tr>
<td>AOA</td>
<td>--</td>
<td>--</td>
<td>23.4</td>
<td>18-31</td>
<td>33.9</td>
<td>28-42</td>
<td>31.3</td>
</tr>
<tr>
<td>% Eng</td>
<td>--</td>
<td>--</td>
<td>5.1</td>
<td>4-7</td>
<td>3.4</td>
<td>1-4</td>
<td>10.4</td>
</tr>
</tbody>
</table>

Note: Percent English use was the average rating, on a scale ranging from 1 (never) to 7 (frequently) of use of English at home, at work, and socially. None of the NJ speakers were married to a NE speaker, and all had begun to study English at school in Japan at about the age of 12 years.
Perception of /r/ and /l/

The perceptual stimuli used by Flege et al. (1996) were naturally produced English words and non-words: 12 real words beginning with /w/ or /d/, 19 real words beginning with /r/ (e.g., right), 19 minimally paired words beginning with /l/ (e.g., light), four non-words beginning with /r/ (e.g., *ruck) and four non-words beginning with /l/ (e.g., *lun).

The NJ speakers and a NE comparison group identified the initial consonants in the stimuli using a 4-alternative forced-choice test. The NJ speakers had no difficulty identifying English /w/ and /d/ but, as expected, they had difficulty correctly identifying /r/ and especially /l/.

![Graph showing % Correct ID for /l/, /r/, /w/, /d/ with bars for NE, Experienced Japanese adults, and Inexperienced Japanese adults. The graph indicates that NE and Inexperienced Japanese adults had lower accuracy compared to Experienced Japanese adults, with differences being statistically significant.](chart.png)
**Perception of /r/ and /l/**

A two-way ANOVA examining the percent correct identification scores obtained for the two NJ groups yielded significant main effects of Group, $F(1,22) = 17.0, p<.05$, and Consonant, $F(1,22) = 23.3, p<.05$, and a non-significant interaction, $F(1,22) = 0.06, p>.10$. The higher scores for /r/ than /l/ and the lack of a significant interaction were predicted theoretically by the SLM.

Ingvalson et al. (2014) compared groups of NJ adults who had lived in the United States for than less than 2 years or more than 10 years. The higher scores for /r/+/l/ combined replicated the significant Group effect obtained by Flege et al. (1996).

The Experienced NJ group tested by Flege et al. (1996) obtained a mean score of 93% correct for /r/. The question that arises is why they did not obtain perfect scores like the NE speakers if, as proposed by the SLM, they were able to establish new phonetic categories for /r/ and had lived in the United State for an average of 21 years.
Lexical bias

Lexical bias may have influenced the percent correct identification scores obtained for the NJ speakers tested by Flege et al. (1996).

Classification at a phonetic level depends on the acoustic properties of phonetic segments and on how well those properties conform to long-term memory representations ("phonetic categories").

However, the conscious judgments obtained in a forced-choice identification test are subject to the influence of phonological codes in lexical representations.

Lexical bias can, as the name implies, shift conscious judgments away from judgments made at a prior phonetic processing level.
Perception of /r/ and /l/

As noted by R. Yamada et al. (1997, p. 103) “linguistic context” affects NJ adults’ identifications of English /r/ and /l/. Lexical status is one such “context”.

In a study by Yoshida & Hirasaka (1983) NJ students living in Japan identified /r/ and /l/ in words whose lexical status varied. The minimal pairs consisted of two real English words (e.g. rock, lock), two non-words (*remp, *lemp) or one word and one non-word (e.g., run-*lun, *rike-like). The % correct identification scores obtained for all three “lexical status” types differed significantly.
Perception of /r/ and /l/

The NJ speakers recruited by Flege et al. (1996) were asked to participate in research that would “test their knowledge of English words”. After responding to a language background question, an English lexical knowledge test was administered.

A total of 46 words and non-words were presented aurally and in writing. The NJ speakers were told that not all items were real English words. Five possible responses were offered:

1. the correct definition of a word;
2. an incorrect definition of a word;
3. the definition of the minimally paired word;
4. “not sure of meaning”
5. “never read or heard this word”

When participants thought they knew an item they were asked to define it, pressing button #1, #2 or #3 and then to rate the item for subjective familiarity using a scale ranging from 1 (never heard and said) to 7 (very often heard and said). A familiarity rating of 1 was assigned to items that were defined incorrectly or for which button #4 or #5 was selected.
Perception of /r/ and /l/

Members of the IJ group judged the English words to be significantly less familiar than the NE and experienced Japanese (EJ) groups \((p<.05)\); these latter two groups did not differ significantly \((p>.10)\). The IJ participants likely judged the words to be less familiar than member of the other two groups because they had heard and said the English words less often (Auer, Bernstein & Tucker, 2000). Importantly, the familiarity ratings obtained for words beginning with /r/ and /l/ did not differ significantly \((p>.10)\) for any of the three groups (NE, EJ, IJ).
Perception of /r/ and /l/

The ratings of word familiarity obtained from both the EJ and IJ groups correlated strongly with the NE speakers’ ratings (Spearman \( \rho = 0.90 \) and 0.92, respectively). Just like NE speakers, for example, the NJ speakers judged the word room to be more familiar than the word rook, and look to be more familiar than loom.
Perception of /r/ and /l/ 

The subjective familiarity ratings provided an indirect assessment of English input (Auer et al., 2000). Here we see correlations between the subjective familiarity ratings and the LOR and English use ratings that were obtained for all 24 NJ speakers together, and for the two NJ groups separately.

As LOR increased, the perceived subjective familiarity of the English words increased. For the inexperienced Japanese group (IJ), a significant correlation existed between the familiarity ratings and English use, but not between the ratings and LOR. This suggested that self-estimated frequency of use may be a better index of quantity of L2 input than LOR is.

<table>
<thead>
<tr>
<th></th>
<th>ALL (df=46)</th>
<th>Experienced NJ (df=22)</th>
<th>Inexperienced NJ (df=22)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOR</td>
<td>0.759</td>
<td>0.666</td>
<td>0.173</td>
</tr>
<tr>
<td>English use</td>
<td>0.656</td>
<td>0.437</td>
<td>0.482</td>
</tr>
</tbody>
</table>

“English use”, the average of ratings on a scale ranging from 1 (never) to 7 (frequently) for English use at home, at work or school, and in social settings. Boldfaced values are significant at the .05 level.
Perception of /r/ and /l/

Flege et al. (1996) compared the identification of /r/ and /l/ tokens in three sets of minimally paired English words defined on the basis of relative subjective familiarity:

1. For the words with **positive lexical bias**, the word beginning with /r/ was judged to be more familiar than the /l/ word (e.g., *room* vs *loom*).

2. For **neutral** words, the words beginning with /r/ and /l/ were judged to be equally familiar (e.g., *rate* vs *late*).

3. For words with a **negative lexical bias**, the /r/ word was judged to be less familiar than the /l/ word (e.g., *rook* vs *look*)
Perception of /r/ and /l/

The NE speakers showed no lexical bias. However, the NJ speakers were more likely to correctly identify an English liquid if it occurred in the more familiar of two words making up a minimal pair. The lexical bias effects were significantly stronger for /l/ than /r/, and significantly stronger for members of the inexperienced Japanese group (IJ) than for the experienced Japanese group (EJ).
Perception of /r/ and /l/

To evaluate perception at a phonetic category level, we need to focus on responses that have not been influenced by lexical bias, that is, responses to minimal pairs consisting of equally familiar words. Here we see the number of correct identifications of English liquids in the three (of 19) minimal pairs that were balanced for subjective familiarity (rock-lock, right-light and read-lead).

For /r/, perfect scores were obtained for 9/12 experienced and 6/12 inexperienced NJ speakers. For /l/, perfect scores were obtained for just four experienced and no inexperienced NJ speakers.
Perception of /r/ and /l/

An additional finding of Flege et al. (1996) confirmed that lexical bias affected the NJ speakers’ identification of /r/ and /l/. In a follow-up experiment, word and non-word stimuli were presented in two conditions:

- Unedited condition: Word vs nonword status was evident (e.g., ripe vs *lipe)
- Edited condition: The lexical status differences were reduced or eliminated altogether by removing the final consonants. For example, ripe and “*lipe became /rai/ and /lai/, making two English words (i.e., rye and lie).

<table>
<thead>
<tr>
<th></th>
<th>unedited stimuli</th>
<th>edited stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>/r/</td>
<td>ripe</td>
<td>/rai/ rye</td>
</tr>
<tr>
<td></td>
<td>lipe*</td>
<td>/lai/ lie</td>
</tr>
<tr>
<td>/r/</td>
<td>roof</td>
<td>/ru/ rue</td>
</tr>
<tr>
<td></td>
<td>loof*</td>
<td>/lu/ loo</td>
</tr>
<tr>
<td>/r/</td>
<td>wrote</td>
<td>/ro/ row</td>
</tr>
<tr>
<td></td>
<td>lote*</td>
<td>/lo/ low</td>
</tr>
<tr>
<td>/l/</td>
<td>like</td>
<td>/lai/ lie</td>
</tr>
<tr>
<td></td>
<td>rike*</td>
<td>/rai/ rye</td>
</tr>
<tr>
<td>/l/</td>
<td>loose</td>
<td>/lu/ loo</td>
</tr>
<tr>
<td></td>
<td>roose*</td>
<td>/ru/ rue</td>
</tr>
<tr>
<td>/l/</td>
<td>lone</td>
<td>/lo/ low</td>
</tr>
<tr>
<td></td>
<td>rone*</td>
<td>/ro/ row</td>
</tr>
</tbody>
</table>
Perception of /r/ and /l/

As expected, the percent correct identification scores was higher in the Unedited condition for words than for non-words. The lexical status effect disappeared in the Edited condition where lexical status differences did not exist.
Perception of /r/ and /l/  

Why should the NJ speakers, even those who were relatively experienced in English, show lexical bias effects when the NE speakers did not?  

Word recognition depends on both “top down” processes, including effects of lexical context, and “bottom up” (auditory and phonetic) effects.  

The “Ganong effect” is a well-known top-down effect which influences how phonetically ambiguous stimuli are identified.
Perception of /r/ and /l/

Consider, for example, the identification of the members of a VOT continuum. NE speakers’ identification of stimuli with ambiguous VOT values may shift depending on the lexical status of the endpoints (e.g., Ganong, 1980). When NE speakers are not sure, they tend to give a response that corresponds to a word in their mental lexicon, for example, “dash” (a word) rather than “dask” (a non-word).

The absence of lexical bias effects for the NE speakers in the Flege et al. (1996) study indicated that, for them, the phonetic identity of the initial consonant in the naturally produced words was never ambiguous. The same might not have held true for some NJ adults, however.
Perception of /r/ and /l/

By hypothesis: Some of the NJ adults tested by Flege et al. (1996) created a new phonetic category for English /r/ whereas all of the NJ adults used a “composite” Japanese-English category to process Japanese /R/ and English /l/ tokens.

If so, then the distributions of phonetic qualities the NJ speakers encountered in everyday conversations—which is what ultimately defines both “one language” and “composite” (two-language) phonetic categories—differed. Specifically: the distribution defining the NJ speakers’ composite /R/-/l/ category may have been more broadly tuned than the distribution defining the NJ speakers’ new English /r/ category.
Perception of /r/ and /l/

Flege et al. (1996) offered an account of the lexical bias effects observed for NJ but not NE speakers. This account, derived from the Theory of Signal Detection, assumes the difference in distributions just described.

---

FIG. 5. Illustration of a signal detection theory account of the data obtained in experiment 2.
Production of /r/ and /l/

Now let’s consider the production of English /r/ and /l/.

Bradlow, Akahane-Yamada, Pisoni and Tohkura (1997) used listener judgments to evaluate the accuracy with which word-initial singleton tokens of /r/ and /l/ had been produced by 12 young Japanese adults in Kyoto, Japan. These NJ speakers were found to have produced English /r/ significantly more accurately than English /l/, both before perceptual training (means = 68% vs 60% correctly identified) and after the training (means = 81% vs 68%).

Saito and Munro (2014) recruited 60 NJ adults who were studying English in Canada. The students were assigned to four groups differing in LOR in Canada (means = 1.0, 2.5, 5.2 and 10.1 months). The authors examined /r/ but not /l/ because English /r/ has traditionally been considered to present a greater learning challenge for NJ speakers than English /l/ does. (This is likely true if one considers the full range of articulations used for the production of English /r/ across all positions and contexts.)

Production accuracy was assessed acoustically, via measurements of initial /r/ token in words elicited in three speaking conditions. Values became more English-like as LOR increased for F2 frequency, F3 frequency and formant transition durations.
Production of /r/ and /l/

The results obtained by Saito and Munro (2014) suggested that progress in speech production can occur in a relatively short period of time in an immersion setting. The study also raised four questions:

1. Would similar results be obtained for NJ adults who were less “highly motivated”, that is, individuals who had not travelled to a predominantly English-speaking country to study English?

2. Would the observed acoustic changes influence how NE-speaking listeners evaluated the /r/ productions?

3. Would even greater progress be noted in /r/ production by NJ adults who had lived longer in a predominantly English-speaking country?

4. Would highly experienced NJ speakers produce English /r/ more accurate than /l/?

These four questions were addressed by Flege, Munro and MacKay (1995) for the same 24 NJ adults whose lexical knowledge and perception of /r/ and /l/ were described earlier (Flege et al., 1996).
Production of /r/ and /l/

Flege et al. (1995) examined word-initial /r/ and /l/ tokens that were elicited in three speaking tasks in the following fixed order:

1. **Definition task.** Test words beginning with /w/, /d/, /r/ and /l/ were prompted by a brief definition in English of the target word (e.g., “What we get from the sun” for light) and also the Japanese translation equivalent of the target word (hikari). Each word was said twice. The analyzed /r/ and /l/ tokens were the second productions, said at the end of a carrier phrase (The next word is_).

2. **Word list reading task.** This is the elicitation method used most commonly in L2 speech research. The test words were read in isolation from a written list. This method runs the risk both of “spelling pronunciations” and by differences arising from variation in reading proficiency.

3. **Unscripted (spontaneous) speech production.** Words obtained in spontaneous speech are ultimately the most indicative of L2 production, but phonetic context can not be easily controlled and obtaining data this way is time consuming inasmuch as the target words have to be located in the speech stream and edited out.
Production of /r/ and /l/

The test words were edited so that just the initial consonant (/r/ or /l/) and roughly half of the following vowel remained. This was done to prevent errors in the production of the following vowel or word-final consonant from influencing listener judgments.

The resulting CV “demi-syllables” were randomly presented to 10 NE-speaking listeners who classified each token using one of six buttons: (1) definitely L , (2) probably L, (3) possibly L, (4) possibly R, (5) probably R (6), or definitely R.

The intelligibility scores obtained for /r/ and /l/ in Exp. 1 for each of the 36 participants were based on 190 judgments (10 listeners x 19 words). The /l/ productions was counted as correct (i.e., heard as intended) when button #1, #2 or #3 was selected; /r/ tokens were correct if button #4, #5 or #6 was selected.
Production of /r/ and /l/

The intelligibility scores obtained for IJ and EJ groups, on the one hand, and the NE group, on the other hand, were evaluated by six independent t-tests (following arcsine transformation), one for both between-group comparisons for tokens produced in three speaking tasks.

For /r/, the NE-EJ and NE-IJ differences were non-significant in all three speaking tasks (Bonferroni corrected $p>.10$).

For /l/, on the other hand, the EJ and NE groups did not differ significantly for any speaking task but the IJ-NE difference reached significance (Bonferroni $p<.05$) for the definition and spontaneous speech tasks.
Production of /r/ and /l/

The intelligibility test was insufficiently sensitive to reveal if the NJ speakers produced English /r/ and /l/ accurately because listeners can correctly identify distorted tokens. Here we see the percentage of liquids elicited in the reading task (Exp. 1) that can be considered to be undistorted realizations of English liquids, operationalized as tokens identified as “definitely” /l/ or “definitely” /r/ (buttons #1 and #6, respectively).

An ANOVA examining these scores yielded a significant effect of Group, $F(2,33) = 11.8, p<.05$, and Consonant, $F(1,33) = 21.5, p<.05$) but not a significant GxC interaction, $F(2,33) = 0.1, p>.10$.

Post-hoc tests indicated that both English liquids produced by members the NE and EJ groups were more accurate than those produced by the IJ group ($p<.05$) and that the EJ group did not differ significantly from the NE group for either /r/ or /l/.

Contrary to expectation, the /l/ tokens were produced more accurately than the /r/ tokens were.
Production of /r/ and /l/

Accordingly, CVs derived from eight words (right, light, rock, lock, read, lead, rate, late) were presented to a new group of 12 NE-speaking listeners in Exp. 3. As before, all listeners heard all stimuli.

The /r/ and /l/ tokens were presented in separate counterbalanced blocks. The NE listeners were informed as to the identity of the intended initial consonant in each block before stimulus presentation began. They rated the initial consonants on a scale ranging from 1 (strong foreign accent) to 7 (no foreign accent).

Ten extra items at the beginning of each block were for practice, and so not analyzed. The mean ratings obtained for each of the 36 participants’ production of both /r/ and /l/ were based on 48 ratings (CVs derived from 4 words x 12 listeners).

Independent t-tests evaluated differences between the NE and EJ groups, and between the NE and IJ groups for all three speaking tasks (6 tests in all).
Production of /r/ and /l/

For /r/, the inexperienced Japanese speakers (IJ) differed significantly from the NE speakers for all three speaking tasks (Bonferroni adjusted $p<.05$) whereas the experienced Japanese speakers (EJ) and the NE speakers did not differ significantly for any speaking task.

For /l/, on the other hand, IJ-NE differences were significant for all three tasks, and EJ-NE differences were significant for the reading and definition tasks (Bonferroni adjusted $p<.05$).

These findings confirm the SLM prediction that greater progress would be made for /r/ than for /l/ and that, with sufficient experience in English, many NJ speakers can produce English /r/ accurately.
Production of /r/ and /l/

A question often asked in L2 research is whether Late learners can manage to produce and perceive L2 phonetic segments in a “native-like” fashion.

For the SLM, the unit of analysis is the individual learner. This is especially important in the present context because the perceived phonetic distance between Japanese /R/ and English /r/ is likely to differ substantially among NJ speakers when they are first exposed to spoken English (see Lotto et al., 2004, Figure 1b).

That being the case, between group differences (or their absence) do not provide a sure way to evaluate SLM predictions. A more relevant question is “How many, if any, NJ speakers differed from NE speakers in producing /r/ and /l/?” Addressing the “How many” question requires defining the phonetic “norm” for English based on the data obtained from a NE “comparison” (control) group.

Various standards have been used to operationalize “native-like”. One is a value or score falling, for example, within 2 SDs of the mean value obtained for a native-speaker comparison group. Another is to determine if individual L2 learners have obtained a value/score falling within the range of values observed for members of a native-speaker comparison group.
Production of /r/ and /l/

Here we see a problem for both approaches to defining “native-like” L2 performance. Among the NE speakers tested by Flege et al. (1995) there was one clear “outlier” for /r/ but no outliers for /l/. This was not entirely unexpected given that /r/ develops more slowly than /l/, and some NE nine year-olds do not yet produce English /r/ correctly (Smit et al., 1991). As we will see later, the outlier produced English /r/ with F3 values that were highly unusual for English.

If the outlier is excluded, /r/ productions were native-like for 10/24 NJ speakers and /l/ productions were native-like for just 5/24 NJ speakers, a non-significant difference between the two English liquids. However, if all 12 NE speakers are used to specify the NE “range”, the /r/ vs /l/ difference (21 vs 5) is significant, $X(1) = 9.8$, $p < .001$, confirming the SLM prediction.
Production of /r/ and /l/

Flege et al. (1995) performed acoustic analyses of 180 /r/ tokens produced in the reading task (3 groups x 12 members x 5 words).

In a regression analysis, four acoustic measures (F2 and F3 starting frequencies, F2 and F3 frequencies at the end of rapid spectral change) accounted for 65% of variance in listener ratings of /r/ production accuracy.

The beta weight of the F3 onset frequency variable was substantially larger (-.965) than those of the other three variables (-.467 to .237). This leads to the conclusion that, in agreement with previous research, F3 onset frequency is the crucial acoustic phonetic dimension specifying English /r/. 
Production of /r/ and /l/

This figure plots the relation between mean F3 onset frequency values and listener ratings of /r/ tokens produced by all 36 participants. The two variables were correlated, $r(34) = -0.83$, $p<.001$. The lower the F3 onset frequencies were, the more accurately NE listeners judged the /r/ tokens to be. A NE speaker who produced English /r/ with unusually high F3 values, the “outlier”, was judged to have produced English /r/ inaccurately. Not all NE adults, it seems, have “mastered” English /r/!

Another finding pointed to the importance of input. The mean F3 values obtained for the 24 NJ speakers and their average self-estimates of frequency of English use correlated, $r(22) = -.41$, $p<.05$.a This modest correlation might have been stronger had information been available pertaining to how much of the English input the NJ speakers had received over the course of their lives was Japanese-accented (expected in a conversation involving a NE speaker and more than one NJ speaker).

aThe English use variable was the average of six 7-point scale estimates of English use (at home, at work, in social settings, with children, and over the past 5 years).
Production of /r/ and /l/

Larson-Hall (2004) attempted to replicate the Flege et al. (1995) production results for NJ Late learners by comparing NJ speakers who had lived in the US for 1.1 years (IJ) and 23.2 years (EJ).

The NJ speakers and NE speakers (15 per group) read an 8-paragraph story and then a list of words. No definition or spontaneous speaking task was included. The accuracy of word-initial /r/ and /l/ tokens in eight minimal pairs were evaluated by 12 NE-speaking university students, who each heard tokens produced by about one-third of the 45 participants.

The test stimuli were not edited as in Flege et al. (1995). The initial consonant in the eight words each beginning with /r/ and /l/ was judged to be (1) definitely L, (2) probably L, (3) probably R or (4) definitely R. In addition to elicitation and scaling techniques, the 1995 and 2004 studies differed in other potentially important ways. Given NJ adults’ sensitivity to the “/r/-/l/ problem”, NJ speakers in the 1995 study were invited to take part in a study examining “knowledge of English words”. And, in fact, an English vocabulary test was administered before speech production was elicited. NJ speakers in the 2004 study, on the other hand, were invited to participate in a “pronunciation” study. The first elicitation task involved reading a text containing many /r/ and /l/ tokens in the presence of the author, with no opportunity to read the text before beginning.

All NJ speakers in the 1995 study were tested on a university campus by a research assistant who tested listeners and made acoustic measurements and later became the third author. In the 2004 study, most members of the EJ group were tested in their home or office by the author, most of those in the IJ group on a university campus in Pittsburgh, PA.
Production of /r/ and /l/

This figure shows the mean percentage of times that the NE-speaking listeners correctly identified English /r/ and /l/ tokens produced by NJ speakers in the Larson-Hall (2004) study.

The substantial inter-subject variation in the original data, which may have derived in part from errors in the production of the vowels or final consonants in the test words, was reduced in data analysis by the elimination of an unspecified number of outliers. It appears that English /l/ was produced more successful than /r/, but this was not tested directly. Crucially, the EJ and IJ groups did not differ significantly.

At right: The mean percentage of /r/ and /l/ tokens spoken in two reading tasks by relatively experienced and inexperienced NJ speakers of English (EJ, IJ) in a study by Larson-Hall (2004). The means are based on 120 forced choice judgments (15 talkers per group x 8). The error bars bracket +/- 1 SE.
Production of /r/ and /l/

Larson-Hall (2994) had the same NE-speaking listeners rate /r/ and /l/ tokens from the word-list reading task using a scale that ranged from 1 (strong foreign accent) to 7 (no foreign accent). The stimuli were edited, as in Flege et al. (1995), to prevent errors in the following vowels or final consonants from influencing listener judgments. Once again, the EJ and IJ groups differed from NE, but not from one another. The /r/-/l/ difference was non-significant.

One unexpected finding reported by Larson-Hall (2004) was a significant inverse correlation for members of the EJ group: the longer was their residence in the United States, the poorer was their production of English liquids. This was attributed to the chronological age of the experienced NJ speakers.

Exp. 3, which examined overall degree of foreign accent in sentences, yielded the same pattern of results: both NJ groups differed from the NE speakers but did not differ from one another.
Production of /r/ and /l/

Grosjean (1998, p. 131) noted that “research dealing with bilinguals has often produced conflicting results” due to a lack of standardization. Studies often differ in terms of methodology (e.g., materials, method of elicitation, measurement procedures, etc.) and in terms of the participants who are studied, who are often poorly described.

The 2004 study of Larson-Hall differed substantially from the 1995 study of Flege et al., and so the 2004 study cannot be viewed as a replication. The fact remains, however, that whereas the 1995 study showed better performance (segmental production and perception) for /r/ than /l/, and better performance for NJ speakers with a long than short residence in the United States, the 2004 study showed neither.

The two negative findings of the 2004 study are likely to be related to one another. According to the SLM, NJ Late learners are expected to make greater progress learning /r/ than /l/ and, if they have received abundant native speaker input, to produce and perceive /r/ in a native like fashion. It is therefore important to understand why the “experienced” NJ speakers tested in 2004 did not learn English /r/ whereas most of those tested by in 1995 did manage to learn /r/.
Production of /r/ and /l/

For the SLM, the two most important parameters for L2 speech learning are perceived cross-language phonetic differences between L1 and L2 sounds and input. The crucial difference between the 1995 and 2004 studies may have been the input received by the two nominally “experienced” groups of NJ adults who were recruited.

Participants in the two studies were selected solely on the basis of LOR. However, as noted by Piske, Flege and MacKay (2001, p. 197) LOR provides only a “rough index” of input. In fact, LOR is of limited use for two reasons.

First, as an index of quantity of input, LOR can be highly misleading. The LOR variable specifies an interval of time. In studies of immigrants, this is the interval between arrival in a predominantly L2 speaking country and the time of testing.

Research studies using the LOR variable have usually provided little information concerning what actually happened during the LOR interval. This is a potentially serious limitation inasmuch as not all immigrants receive abundant and equal L2 input.
Production of /r/ and /l/

Specifically, not all immigrants begin using their L2 immediately after arriving in the host country (e.g., Flege, Munro & MacKay, 1995a, Table I) or use their L2 on a regular basis (Moyer, 2009, p. 162). The results of Flege and Liu (2001) suggested that LOR may provide a useful estimate of quantity of L2 input only for immigrants who have the opportunity and need to use the L2 on a daily basis.

The second serious limitation of LOR as an index L2 input is that it provides no insight whatsoever into quality of L2 input. As far as I know, quality of input has never been assessed in research with NJ learners of English. We can be certain, however, that NJ speakers living in the United States speak English in the presence of other NJ speakers and so hear English spoken with a Japanese foreign accent, including the mispronunciation of /r/ and /l/. The question is, how often?

Imagine two Japanese managers working at Japanese-owned manufacturing plants in the US. One works in a small plant along with two other Japanese employees, the other in a larger plant with 20 NJ employees. The former hypothetical individual is likely to hear far less Japanese accented English than the latter. If this situation persists over time, it might well influence the distribution of English /r/ and /l/ tokens that have been heard and seen. By hypothesis, this will change phonetic category representations and then segmental production.
Production of /r/ and /l/

We have no way to evaluate the quality of input obtained by the nominally “experienced” NJ speakers tested in the 1995 and 2004 studies. However, we can infer that they differed in quantity of input.

As discussed earlier, a difference in input for the relatively experienced and inexperienced groups (EJ, IJ) in the 1995 study was demonstrated by significant differences in the perceived familiarity of English words beginning in /r/ and /l/.

Members of EJ reported using English significantly more often than members of IJ did, $F(1,22) = 10.4$, $p=.004$. This difference arose because immigrants in a predominantly L2-speaking country typically use their L2 more over time as they integrate into the local culture. An increase in percentage L2 use occurs within a few years. In the Aoyama and Flege (2011) study, NJ adults who had lived in the US for averages of 8.4 and 2.1 years reported using English significantly more often ($p<.01$) those who had lived in the U.S. for just 0.6 years.

The EJ and IJ groups tested in 2004, on the other hand, did not differ significantly in terms of self-rated frequency of English use ($p=0.54$). From this we might infer that the nominally “experienced” NJ speakers tested in 2004 did not integrate into the English-speaking community to the same extent as those tested in the 1995 study.
Production of /r/ and /l/

Larson-Hall (2004, p. 541) observed a negative correlation between LOR and L2 segmental production accuracy and overall foreign accent for 15 “experienced” NJ speakers (the EJ group). This counter-intuitive finding was attributed to the chronological age of members of the EJ group, whose ages ranged from 38 to 66 years ($M = 52.5$).

Only one study to date has examined the effect of chronological age on foreign accent. MacKay, Flege and Imai (2006) tested 138 native Italian immigrants who arrived in Canada between the age of 7 and 36 years ($M = 18$ years) and were 40 to 72 years of age ($M=57$) when tested.

At right we see the mean foreign accent ratings obtained for six groups of 23 Italian immigrants each who differed in chronological age at the time of testing. The participants who were older than about 60 years of age had stronger foreign accents than those who were younger. However the effect of age seen here arose from factors correlated with chronological age at the time of testing, not age per se.
Production of /r/ and /l/

MacKay et al. (2006) carried out a principle components analysis of questionnaire data, which indicated three underlying factors. The factor scores were then regressed onto the foreign accent ratings obtained for the 138 Italian immigrants.

A factor that might be called “bilingual balance” (high loadings on age of arrival, bilingual dominance, years of education in Italy and proficiency in Italian) accounted for 45.1% of the variance in the foreign accent ratings.

A factor with high loadings on frequency of English and Italian use accounted for an additional 13.9% of the variance.

Crucially, a factor defined by LOR and chronological age at the time of testing (which were strongly correlated) did not account for a significant amount of additional variance.

These results suggests that the age of the nominally experienced NJ speakers tested by Larson-Hall (2004) can not explain why their production of phonetic segments and sentences deceased in accuracy as their LOR in the United States increased.
Production of /r/ and /l/

Given the lack of significant differences between the EJ and IJ groups in Larson-Hall (2004), it is relevant to consider how input influenced the foreign accent ratings obtained for Italian immigrants in MacKay et al. (2006).

Italians who had lived longest in Canada ($M = 48$ years) had a significantly better pronunciation than those having shorter LORs ($p<.05$). Those who used English more than 60% of the time had a better pronunciation than those who used English less. Those who had more than 29.0 years of full-time English input pronounced English better than those with 20.3 to 28.8 years of input, who pronounced English better than those with 3.3 to 20.1 years of input. The ratings showed a significantly strongly correlated with LOR*% English (years of input) than LOR ($p<.05$) because the former variable indicates something about what happened during the LOR interval.

At right. Mean foreign accent ratings in English sentences spoken by 138 Italian immigrants in Canada who have been assigned to groups of 23 each on the basis of LOR, % use of English, and years of full-time equivalent English input (% English * LOR). The error bars bracket +/- 1 SE-
Summary

The results reported here do not support the view that NJ Late learners have persistent difficulty producing and perceiving English /r/ and /l/ as the result of passing a critical period, nor because they persistently filter out the acoustic phonetic properties needed to define English /r/ and distinguish it from English /l/. The results are, however, readily understandable within the framework of the SLM (Flege, 1995, 2007, Flege & Bohn, in prep.).

Three phases in the learning of English /r/ and /l/ can be identified.

At Time 1, NJ adults fail to distinguish English /r/ and /l/ from one another because both English liquids are identified as being instances of Japanese /R/, which is used to produce both English liquids. It is uncertain how long this phase lasts, but it may be just a matter of weeks or months for NJ speakers who are exposed daily to English in an immersion setting.
Summary

At Time 2, NJ speakers have noted that a phonetic difference exists between English /r/ and Japanese /R/ and have begun to establish a new phonetic category for the English /r/ (but not /l/). English /r/ is produced with increasing accuracy as the new English /r/ category develops.

The extent to which the new /r/ category resembles that of NE monolinguals depends on the distribution of tokens identified as being instances of the new category. The /r/ categories of NJ speakers who frequently use English in the presence of other NJ speakers (whose pronunciation may well be Japanese-accented) will be less English-like than the /r/ categories developed by NJ speakers who use English with equal frequency but only hear English spoken by native speakers.
Summary

At Time 3, the NJ speakers have a fully specified English /r/ category that remains, however, subject to modification if the input distributions upon which it is based changes (as is the case for all phonetic categories). The accuracy with which individual NJ speakers are judged by NE-speaking listeners to produce English /r/ at this stage will depend on the extent to which the NJ speakers’ /r/ categories match those of the NE-speaking listeners (which in turn depends on input distributions).

Productions of English /l/ will continue to differ measurably (acoustically, or in terms of listener judgments) from those of English monolinguals. This is because production is guided by a composite English /l/-Japanese /R/ category based on the distribution of English /l/ and Japanese /R/ tokens that have been heard. Production of /l/ may be slightly more English-like at Time 3 than Time 1 but, if so, a measurable effect due to learning English will be evident in production of the Japanese /R/.

![Diagram showing changes over time for /r/ and /l/ categories]
Where to from here?

We have learned much in the past 50 years regarding how NJ speakers learn English /r/ and /l/. However, much additional work is needed to fully understand the phenomena discussed here.

Much of the research cited here have been “experiments of nature”, not true experimental designs. Such research can yield important results, such as the study by Yamada (1995) of NJ speakers who were tested in Japan soon after returning from a stay in the United States.

One study of this type that might have been undertaken in the 1980s, but was not, could have decisively answered many questions addressed here. This never-undertaken study would have examined some of the nearly 45,000 Japanese “war brides” who came to the US with an American soldier husband in the early 1950s. The SLM would predict highly accurate production and perception of English /r/ but not /l/ for these Japanese-American women. Brides could have been tested after 30 years of US
Where to from here?

Brides could have been tested after 30 years of US residence, and subgroups defined on the basis of the number of other Japanese speakers with whom the Brides spoke Japanese.

The SLM predicts that Brides who were effectively isolated from Japanese, but not those who often heard English produced by others NJ speakers, would have shown native-like production and perception of English /r/ but that all Brides, regardless of their social networks, would have produced and perceived English /l/ inaccurately.

At right: photo from a Washington Post article entitled “The untold stories of Japanese war brides” by Kathryn Tolbert.
More work needed

Rather than wait for another opportunity like this (which may never come) we should begin concentrating now on aspects of experimental design and participant selection that will increase the reliability of our research and the quality of the results and conclusions it yields.

To begin, it is essential to document in detail the nature of L1 phonetic categories that are likely to interact with L2 sounds as a new phonetic subsystem is created for the L2. Information regarding L1 sounds must be obtained for the individuals who learn the L2, not an L1 comparison group.

Better methods are needed for quantifying the perceived relationship between sounds found in the L1 and L2. These measures much be taken for each individual who learns an L2, not groups.

The role of input in L2 research is unclear because at present we “simply do not know how much input [L2] learners actually get [nor] how much exposure a learner requires” due to the fact that “only recently [has] input begun to receive consideration” (Piske & Young-Scholten, 2008, pp. 12-13).
Where to from here?

Future research needs to provide far more information than at present regarding the individuals who learn an L2. Why, where, when and with whom do they use the L2? How often do they use the L1 and L2, and how often to they use the L2 in “mixed company”, that is, in conversations in which they are likely to hear the L2 produced with an L1-inspired foreign accent.

As shown earlier in this presentation, length of residence (LOR) in a predominantly L2-speaking country provides an inadequate index of L2 input. It needs to be replaced by more robust measures such as those suggested by Flege (2019) and Flege and Wayland (2019).

Only when we have more accurate information regarding input, and have a better understanding of how the perceived relationships between L1 and L2 sounds change over time, will we fully understand L2 speech learning.


References


References


Flege, J., & Wayland, R. (2019). The role of input in native Spanish Late learners’ production and perception of English phonetic segments. J. of Second Language Studies, 2(1), 1-


References


References


References


