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 for the National Institute for Deafness & Other Communicative Disorders (NIDCD)*
Aim of this talk

This talk has two aims

1. Review a few of the many studies examining the perception and production of English /r/ and /l/ by native Japanese (NJ) speakers;

2. Explain how these data can be interpreted within the framework of the Speech Learning Model (SM, Flege, 1995, 2007)

Note: For convenience, I will designate the two liquids of English as “r” and “l”, and the single liquid of Japanese, phonetically a flap, as “R”
Aim of this talk

The SLM focuses on the learning of *position-sensitive* phonetic categories, not phonemes, because the learning of L2 speech sounds may vary according to position.

Here I focus exclusively on word-initial consonants.

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.
Why /r/ and /l/?

It is difficult—some would say “impossible”—for Japanese adults to learn to produce and perceived English /r/ and /l/ accurately.

Work that focuses on this issue is nevertheless important because it provides a way to test “general principle of learning and claims about adult neural plasticity” (Bradlow, 2008, p. 294).

Hattori & Iverson (2009: 477) referred to the 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.
So what’s the problem?

Children who learn English as their L1 take years to establish language-specific phonetic representations for English /r/ and /l/.

Monolingual Japanese children don’t do this because Japanese doesn’t have sounds that are phonetically identical to English /r/ or /l/.

When speakers of Japanese—children or adults—are first exposed to English they usually judge the two English liquids to be instances of the one liquid consonant found in Japanese, a flap, symbolized here as “R” for the sake of convenience.
So what’s the problem?

The most important (of several) differences between English /r/ and /l/ is third formant (F3) frequency.

For /r/  
F3 frequency is close to F2 frequency at release.

For /l/  
F3 frequency is much higher — and so more distant — from F2 than is the case for /r/.
So what’s the problem?

Iverson et al. (2003) identified the key perceptual problem facing Japanese learners of English.

The created a set of synthetic “ra” and “la” syllables in which F2 and F3 frequencies varied orthogonally.

The auditory distances (in mel) between all pairs of stimuli were equal.

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

Listeners were asked to identify the synthetic stimuli using phonetic categories found in their L1:

- NE listeners labelled the stimuli as “r” or “l”
- NJ listeners labelled the same stimuli as “R” or “w”

All pairs of stimuli were rated for acoustic similarity (7-point scale). The ratings were then examined in MDS (multi-dimensional scaling) analyses.
So what’s the problem?

NE adults gave greater weight to the F3 dimension when rating the similarity of pairs of stimuli than might have been expected from the auditory differences between the stimuli. The F3 dimension was “stretched” due to its key role in distinguishing /r/ from /l/.

![Diagram showing goodness and identification percentages and multidimensional scaling with labels indicating stretching of F3 dimension.](image)
So what’s the problem?

The NJ listeners, on the other hand, made more use of the F2 dimension than NE speakers did and less use of F3. They showed no evidence of perceptually “stretching” the F3 dimension.
So what’s the problem?

According to Iverson et al. (2003)

1. The NJ adults did not lack **auditory sensitivity** to the F3 dimension;

2. Their perceptual space was “miss-tuned” for the acquisition of the phonetic distinction between English /r/ and /l/;

The authors suggested that

3) Auditory and/or early stage phonetic processing may interfere with NJ adults’ learning of English /r/ and /l/;

4) This interference may be “difficult to reverse” by making it difficult for Japanese learners of English to gain perceptual **access** to the dimensions they need to learn English /r/ and /l/.

Conclusion #4 may be premature, however, inasmuch as the 24 NJ adults who were tested in Tokyo had little if any conversational experience in English.
Early vs late learners

One aim of L2 speech research is to account for the effect of age of first exposure on phonetic learning.

Two questions regarding the Iverson et al. (2003) findings are:

• Is the low-level “filtering” **stronger** for Late than Early learners?
• Is it **irreversible** for Late but not Early learners?

Work by Yamada (1995) sheds light on these questions.
Early vs late learners

Yamada (1995) tested NJ adults in Tokyo who had a mean age of 20 years.

Of the 276 tested, more than half (n = 156) were “returnees” tested soon after returning from a stay in the United States. The returnees were assigned to five “JE” groups according to length of US residence.

The remaining participants were NJ speakers who had never lived abroad and native English (“NE”) speakers who had resided briefly in Japan.

<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>
Early vs late learners

Participants 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”. 
Early vs late learners

The /r/-/l/ identification function obtained for the returnees resembled the NE speakers’ function more than did the function obtained for NJ adults who had never lived abroad.

However, a great amount of variation was evident for the returnees.
Early vs late learners

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 listeners rarely heard 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/.
Early vs late learners

NJ adults who had never lived abroad gave far more /w/ responses than the NE speakers did. They made little used of F3 to distinguish /r/ from /l/. 
Early vs late learners

The returnees resembled the NE speakers to a greater extent than the Japanese adults who had never lived abroad, but some made far less use of F3 frequency than the NE speakers did.
Early vs late learners

The extent to which the returnees used F3 when identifying English liquids in synthetic stimuli depended on how long they had lived in the United States (LOR).

The greater the LOR, the more the returnees used F3.
Early vs late learners

How well the returnees identified naturally produced /r/ and /l/ tokens also depended on how long they had lived in the United States.

The longer the returnees had lived in the US, the more often they correctly identified English liquids.
Early vs late learners

What explains the effect of LOR on perception?
Did members of returnee groups JE 6-7 and JE 8+ resemble native English speakers more than members of the other returnee groups because:

• all of them had lived in the US for at least 6 years?
• all had arrived in the US before 8 years of age?
• both?

Confound between LOR and AOA
Early vs late learners

AOA (age of arrival) and LOR (length of residence) in the US were confounded, and so no certain conclusions could be drawn.

Yamada’s analyses suggested that the two factors were equally important predictors of performance. She hypothesized that two factors influenced the returnees’ perception of English /r/ and /l/:

**endogenous:**

a “biological” factor that decreases the **plasticity** of neutral systems used in speech learning as a function of age of first exposure to an L2;

**exogenous:**

a factor related to overall amount of L2 **input** received, here defined by LOR (a rough indicator of quantity but not by quality of L2).
The Speech Learning Model (SLM)

The SLM takes a different approach.

This theoretical model of L2 speech learning posits that there is no Critical Period for L2 speech because the capacity to learn speech remains intact across the life span.

On this view, if a low-level phonetic filter exists, its influence is not irreversible.
The Speech Learning Model (SLM)

The SLM posits that:

When first exposed to a foreign or second language, learners automatically relate the sounds they hear to sounds found in their L1 (‘interlingual identification’);

This form of cross-language mapping occurs even in cases in which the L2 sounds are quite different from those found in the L1.
The Speech Learning Model (SLM)

According to the SLM, the learning of L2 sounds depends most importantly on:

1. the perceived relationship between each L2 sound and the closest sound in the L1;

2. the quantity and quality of L2 input.
The Speech Learning Model (SLM)

Attention, please! The magnitude of perceived cross-language phonetic differences must be measured empirically.

Perceived phonetic differences cannot be “guessed” or “borrowed” from textbook descriptions.

If English /r/ is judged to be more dissimilar from Japanese /R/ than English /l/ is, the SLM predicts that category formation is more likely to occur for /r/ than /l/. 

[Diagram showing /r/ and /l/ pointing to /R/ with arrows]
The Speech Learning Model (SLM)

**AND IF** a new phonetic category is not established for English /l/ because it is too similar to Japanese /R/, the SLM predicts that a single “composite” phonetic category will develop, one that aggregates the phonetic properties of English /l/ and Japanese /R/.

By hypothesis, such a composite category would be used to process instances of two perceptually linked sounds: /l/ and /R/.

This scenario leads to the expectation that /l/ will be produced and perceived less accurately than /r/ by NJ adults with substantial L2 conversational experience, regardless of the age of first exposure to English.
Measuring cross-language differences

Evidence does indeed exist that English /r/ is perceived to be less similar to Japanese /R/ than English /l/ is.
Cross-language phonetic differences


The English /r/ and /l/ tokens were presented along with Japanese /R/ tokens in separate counterbalanced sessions.
Cross-language phonetic differences

The highest ratings were of course obtained for the Japanese /\textit{R}/ tokens.

The English /\textit{r}/ tokens were judged to be \textit{poorer instances} of the Japanese /\textit{R}/ than the English /\textit{l}/ tokens were.
Cross-language phonetic differences

Guion et al. (2003) found that Japanese adults in the United States – group “J1” – were able to discriminate English /r/ from the Japanese /R/ at significantly above-chance rates.

Group J1 had lived in the US for an average of 3.1 years (range = 1.8 to 5.5 years) at the time of testing.
Cross-language phonetic differences

Despite their conversational experience in English, the members of group J1 were unable to discriminate English /l/ from the Japanese /R/ at a significantly above-chance rate.
Cross-language phonetic differences

Hattori & Iverson (2009) used an identification task to evaluate cross-language differences.

NJ adults who had a median of 3 months of conversational experience in English identified natural English and Japanese CVs using three category labels.

They misidentified the 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 /l/</th>
<th>Jap /r/</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>
Cross-language phonetic differences

How much conversational experience is needed by Japanese adults living in the US to show a difference between /r/ and /l/?

To address this question, I re-examined data published by Aoyama & Flege (2011) for 50 NJ adults living in Birmingham, AL.

I assigned 15 NJ adults each to three non-overlapping subgroups defined on the basis of native speaker input, estimated by multiplying LOR by self-reported % English use.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>English Input</th>
<th>AOA</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>15</td>
<td>0.01</td>
<td>34.7</td>
<td>35.5</td>
</tr>
<tr>
<td>Med</td>
<td>15</td>
<td>1.16</td>
<td>27.9</td>
<td>30.4</td>
</tr>
<tr>
<td>High</td>
<td>15</td>
<td>6.46</td>
<td>25.3</td>
<td>33.4</td>
</tr>
</tbody>
</table>
Cross-language phonetic differences

The NJ adults rated the perceived similarity of the English liquids to Japanese /R/ in separate counter-balanced blocks.

Similarity to Japanese /R/ was significantly greater for /l/ than /r/ for the Japanese adults having 6.5 and 1.2 years of English input, but not for those who had just 0.01 years of English input.
Cross-language phonetic differences

From this we might infer that NJ adults need about 3 months of full-time conversational experience – equivalent to 50% use of English over 6 months of residence – to appraise phonetic differences between the two English liquids with respect to the Japanese /R/.
Cross-language phonetic differences

This estimate must be considered tentative, however, because not all NJ monolinguals are likely to have identical /R/ categories.

That being the case, individual Japanese learners of English may perceive the relationship between /r/-/R/ and /l/-/R/ differently when first exposed to English.
Cross-language phonetic differences

Consider, for example, the results obtained by Jun Yamada (1991)

Yamada administered 1 hour of identification training to 152 college students in Japan. Feedback was provided following the identification of the initial liquid consonant in non-word minimal pairs beginning with /r/ and /l/.

**Before training:**
several students could already correctly identify /r/ and /l/ in all 4 minimal pairs

**Following training:**
• an additional 6% could identify /r/ and /l/ in all 4 minimal pairs;
• 35% did so for several minimal pairs;
• 51% remained at chance for all 4 pairs.
Cross-language phonetic differences

To better understand how perceptual assimilation develops, I computed simple and partial correlations for data taken from the Aoyama & Flege (2011) study.

<table>
<thead>
<tr>
<th></th>
<th>AOA</th>
<th>/r/-/R/ ratings</th>
<th>/l/-/R/ ratings</th>
<th>/r/-/l/ diff.</th>
<th>/r/-/l/ discrim.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Input</td>
<td>(a) -0.52**</td>
<td>(a) -0.60</td>
<td>(a) -0.38</td>
<td>(a) 0.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) -0.49</td>
<td>(b) 0.08</td>
<td>(b) 0.49</td>
<td>(b) 0.42</td>
</tr>
<tr>
<td>2</td>
<td>AOA</td>
<td>(a) 0.43</td>
<td>(a) 0.31</td>
<td>(a) -0.24</td>
<td>(a) 0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(c) 0.17</td>
<td>(c) 0.15</td>
<td>(c) 0.05</td>
<td>(c) 0.26</td>
</tr>
<tr>
<td>3</td>
<td>/r/-/R/ ratings</td>
<td></td>
<td>(a) 0.88</td>
<td>(a) -0.39</td>
<td>(a) -0.08</td>
</tr>
<tr>
<td>4</td>
<td>/l/-/R/ ratings</td>
<td></td>
<td></td>
<td>(a) 0.08</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>/r/-/l/ rating difference</td>
<td></td>
<td></td>
<td></td>
<td>(a) 0.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(b) 0.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(c) 0.41</td>
</tr>
</tbody>
</table>

(a) Simple correlations, df = 43; (b) variation in AOA partialled out, df = 42; (c) variation in input partialled out, df = 42. Boldfaced coefficients significant at p < .001
Cross-language phonetic differences

The more conversational input the 45 NJ adults had received, the less similar to Japanese /R/ they judged the English /r/ tokens to be ($r = -0.60$, $p < .001$).

The correlation between input and the /l/-/R/ ratings did not reach significance.

A correlation existed between the magnitude of /r/-/l/ ratings of similarity to Japanese /R/ and the discrimination of /r/-/l/ ($r = 0.50$, $p < .001$).
Access to the F3 dimension

We have seen that the perceived relationship of English /r/ and /l/ 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?

Yes.
Access to the F3 dimension

Hattori & Iverson (2009) used the method of adjustment to evaluate /r/-/l/ perception.

Their **synthetic stimuli** differed along five dimensions: F1, F2, F3, closure duration, transition duration.

An efficient algorithm enabled listeners to “search through” the complex stimulus array to find the “**best exemplars**” of both English /r/ and English /l/.
Here we see the mean F3 values in the stimuli that were selected by NJ and NE adults as the “best” exemplars of English /r/ and /l/.

The mean F3 values were much the same for the NE and NJ adults. This indicates that NJ adults can gain perceptual access to the F3 dimension in English liquids.

But what explains the enormous variability among the NJ participants?

adapted from Hattori & Iverson (2009), Fig. 4
Access to the F3 dimension

Perhaps variation in conversational input. Of the 36 NJ adults tested

- 4 reported never using English;
- 13 had lived in an English-speaking country for < 1 month;
- only two had more than 2 years of full-time conversational experienced in English.
Access to the F3 dimension


The more the NJ adults’ preferred F3 values approximated those of NE adults for /r/ and /l/, the more accurately they identified English /r/ and /l/ ($r = -0.46$, $p < .01$).
Access to the F3 dimension


The NJ adults generally used F3 less than the NE Ss did.

adapted from Fig. 5 of Takagi & Mann (1995) for stimuli having F2 values of 1000 Hz; not shown are responses for /w/
Access to the F3 dimension

However, the NJ adults with greater conversational experience in English showed significantly greater use of F3 than those with less conversational experience.

adapted from Fig. 5 of Takagi & Mann (1995) for stimuli having F2 values of 1000 Hz; not shown are responses for /w/
Category formation for English /r/

So far we have seen that

- English /r/ is phonetically more distant phonetically from Japanese /R/ than is English /l/;
- NJ adults can gain perceptual access to the F3 dimension in English liquids.

Does this mean the NJ adults who have substantial conversational experience in English are able, as predicted by the SLM, to create new phonetic categories for English /r/?

Yes.
Category formation for English /r/


Both groups consisted of 4 males and 8 females who began to study English at school in Japan at 12 years of age and had arrived in the U.S. as adults.

<table>
<thead>
<tr>
<th></th>
<th>Native English</th>
<th>Exper. NJ</th>
<th>Inexper. NJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronological age in years</td>
<td>M range</td>
<td>M range</td>
<td>M range</td>
</tr>
<tr>
<td>Age of arrival in US</td>
<td>36 29-47</td>
<td>44 40-47</td>
<td>35 29-44</td>
</tr>
<tr>
<td>Length of residence in US</td>
<td>-- --</td>
<td>23 18-31</td>
<td>34 28-42</td>
</tr>
<tr>
<td>Average use of English</td>
<td>-- --</td>
<td>5.7 4.3-7.0</td>
<td>2.9 1.3-3.7</td>
</tr>
</tbody>
</table>
Category formation for English /r/

Compared to member of the Inexperienced group, members of the Experienced group had three advantages:

- an slightly earlier arrival in the US (means = 23 vs 34 years);
- a much longer residence (means = 21 vs 2 years);
- more frequent use English (5.7 vs 2.9 on 7-point EAI scales).
Category formation for English /r/

The naturally produced stimuli identified in a 4-alternative forced-choice test began with /r/, /l/, /w/ or /d/.

Of the stimuli examined:

- 12 words began with /w/ or /d/
- 19 began with /r/ (e.g., right)
- 19 were minimally paired words that began with /l/ (e.g., light)
- 4 were non-words beginning with /r/ (e.g., ruck*)
- 4 were non-words beginning with /l/ (e.g., lun*)
Category formation for English /r/

As expected:

- the NJ adults had difficulty identifying English /r/ and /l/ but not English /w/ and /d/.
- the members of both NJ groups correctly identified /r/ more often than /l/.
Category formation for English /r/

For /r/, mean percent correct identification was higher for the Experienced than Inexperienced group (93% vs 80%).

But why didn’t the Experienced NJ adults obtain perfect scores like the NE speakers? After all, they had lived in the US for an average of 21 years.

Their problem was: lexical bias.
Lexical bias

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

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.
Lexical bias

As noted by Yamada et al. (1997, p. 103) “linguistic context” affects NJ adults identifications of English /r/ and /l/.

Lexical status is one such context. Consider the pioneering study carried out by Yoshida & Hirasaka (1983) here at Sophia University.

Students were asked to identify /r/ and /l/ in words in which lexical status varied. The English minimal pairs consisted of

- two real words (e.g. rock, lock);
- two nonwords (*remp, *lemp);
- one word and one nonword (e.g., run-*lun, *rike-like).

The % correct identification scores obtained for all three “lexical status” types differed significantly.
Lexical bias

Flege et al. (1995) hypothesized that lexical context effects go beyond simple word vs non-word status.

This study examined the effect of subjective familiarity differences between the members of /r/-/l/ minimal pairs.

NJ adults rated the subjective familiarity of English words and non-words beginning with /r/ and /l/ (7-point rating scales).
Lexical bias

Both the native Japanese and English participants judged the words beginning with /r/ and /l/ to be equally familiar.

However, members of the NE and Experienced NJ groups judged the English words to be more familiar than did members of the Inexperienced NJ group.
Category formation

The ratings of word familiarity obtained from both NJ groups correlated strongly with the NE speakers’ ratings.

Just like NE speakers, for example, the NJ participants judged

- *room* to be more familiar than *rook*
- *look* to be more familiar than *loom*
Lexical bias

Flege et al. (1995) compared the identification of /r/ and /l/ tokens in three sets of minimally paired English words defined on the basis of relative subjective familiarity:

- **Positive bias**: /r/ word more familiar than /l/ word (e.g., room vs loom)
- **Neutral**: both words are equally familiar (e.g., rate vs late)
- **Negative bias**: /r/ word less familiar than /l/ word (e.g., rook vs look)
Lexical bias

The NE speakers showed no lexical bias.

However, the NJ adults 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 Inexperienced than Experienced NJ adults.
Lexical bias

To see results at the phonetic category level, we need to focus on results that are not influenced by lexical bias.

This means focusing on the results obtained for minimal pairs made up of equally familiar words.
Lexical bias

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).

Perfect scores for /r/ were obtained for 9/12 Experienced and 6/12 Inexperienced NJ participants (63%). For /l/, on the other hand, just four Experienced (17%) and no Inexperienced participants obtained perfect scores.
Lexical bias

Experiment 2 of Flege et al. (1995) confirmed that lexical bias affected the NJ adults’ identification of English /r/ and /l/.

Words and non-words were presented in two conditions:
- **unedited**: word vs nonword differences were evident (e.g., ripe vs *lipe)
- **edited**: Lexical status difference eliminated by removing the final consonant “ripe” and “*lipe” have become /rai/ and /lai/, which make two English words, “rye” and “lie”.

<table>
<thead>
<tr>
<th>unedited stimuli</th>
<th>edited stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>r</strong> ripe</td>
<td>/rai/ rye</td>
</tr>
<tr>
<td><strong>r</strong> roof</td>
<td>/ru/ rue</td>
</tr>
<tr>
<td><strong>r</strong> wrote</td>
<td>/ro/ row</td>
</tr>
<tr>
<td><strong>l</strong> like</td>
<td>/rai/ rye</td>
</tr>
<tr>
<td><strong>l</strong> loose</td>
<td>/ru/ rue</td>
</tr>
<tr>
<td><strong>l</strong> lone</td>
<td>/ro/ row</td>
</tr>
</tbody>
</table>
Lexical bias

As expected, percent correct identification was higher for words than nonwords in the Unedited condition.

The lexical status effect disappeared in the Edited condition where a lexical status difference did not exist.
Lexical bias

Why should the NJ adults – even those who were highly experienced in English – show lexical bias effects but not the NE speakers?

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

An example of a top-down effect is the well known “Ganong” effect (e.g., Ganong, 1980).

This effect shows that the perception of ambiguous phonetic segments may change as a function of the lexical identity of two responses alternatives.
Lexical bias

Example: NE speakers’ identification of stimuli having ambiguous VOT values may shift depending on the lexical status of the endpoints.

When not sure, NE listeners 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).
Lexical bias

The absence of lexical bias effects for the NE speakers indicates that the phonetic identity of the initial consonant in the naturally produced words was never ambiguous.

That might not have held true for some NJ adults, however.
Lexical bias

By hypothesis:

- Some NJ adults created new phonetic categories for English /r/.
- All NJ adults used a “composite” Japanese-English category to process Japanese /R/ and English /l/ tokens.

If so, then the distributions of phonetic qualities our NJ participants encountered in everyday conversations—which is what ultimately defines both “normal” (one language) and “composite” (two language) phonetic categories, differed.

Specifically: the distribution defining the /R/-/l/ category was broader than the distribution defining /r/.
Lexical bias

Flege et al. (1995) offered an account of the observed lexical bias effects observed for NJ but not NE participants that was derived from the Theory of Signal Detection (TSD).

This account assumes the existence of the difference in distributions just described.

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

Given time limits, I will present only one production study, Flege et. al. (1996) because it appears to be the most thorough study to date.

The NE and NJ adults from the Flege et al. (1995) study produced dour minimally paired English words. The production of these words was elicited three ways:

- **Word-list reading.** The most common task used in L2 speech results
- **Word definition.** The test words were prompted be a brief definition of the target word (e.g., “What we get from the sun” for light + the Japanese translation equivalent, *hikari*)
- **Spontaneous** speech. Very time consuming inasmuch as the target words had to be located and edited out.

---

*a right-light, rock-lock, read-lead, rate-late;*
Production

The final consonants were deleted from the original test words, yielding CV stimuli.

NE-speaking listeners (n = 12) used a 7-point scale to rate syllable-initial consonant in eight counterbalanced blocks (one for each English word).

This listeners always new beforehand the identity of the target words they were rating. a

a Half of the listeners heard the CVs with /r/ first, and half heard the CVs with /l/ first. The 108 stimuli in most blocks was preceded by an extra 36 practice items, which were not analyzed.

The final consonants were deleted from the original test words, yielding CV stimuli. This prevented errors in production of word-final consonants from influencing results.

The scores obtained in an identification were often near ceiling, and so not optimal for assessing production accuracy.

Therefore, we had a panel 12 NE-speaking listeners used a 7-point scale to rate the initial consonant in eight counterbalanced blocks. This listeners always new beforehand the identity of the target words.

Half of the listeners heard the CVs with /r/ first, and half heard the CVs with /l/ first.

To stabilized judgments the 108 stimuli in most blocks was preceded by an extra 36 practice items, which were not analyzed.
Production

The /r/s of the NE and Experienced NJ group were identified significantly more often than the /r/s produced by the Inexperienced group (p < .05). The NE and Experienced groups did not differ significantly.

The effects of Group and Condition were highly significant (p > .0001), but did not interact significantly with one another (p = 0.18). Significantly higher ratings were obtained in the definition task than in the other two tasks (p > .05).

\[\text{Native English} \quad \text{Experienced NJ} \quad \text{Inexperienced NJ}\]

Factors: Group, Condition

Tasks: read, define, spont.

Accuracy Scale: 1 to 7

Significance Levels: p < .05, p > .0001, p = 0.18, p > .05
Production

The results for spontaneously produced /l/ was the same: NE + Experienced > Inexperienced.

For /l/ tokens elicited in the reading and definition tasks, however, all between-group differences were significant:

**NE > Experienced > Inexperienced**

[Bar chart showing /l/ production accuracy across different groups and speaking tasks]
Production

Here are data for individuals. For /r/.

8 members of the Experienced group can be credited with “perfect” production, one fewer than was credited with “perfect” perception of /r/.

2 members of Inexperienced group can be credited with “perfect” production, four fewer than credited with “perfect” perception of /r/.

![Graph showing /r/ and /l/ ratings by group: Native English, Experienced NJ, Inexperienced NJ.](image)
Production

For /l/:

5 members of the Experienced group but no Inexperienced group members can be credited with perfect production of English /l/ (as compared to 4 and 0 “perfect” perceivers).
Summary

The results reported here are inconsistent with the view that native Japanese (NJ) speakers’ well-known difficulty in producing and perceiving English liquids is due to either:

- the passing of a critical period;
- the filtering out of the phonetic information needed to define English /r/ and to distinguish English /r/ from /l/.

The results are, however, readily understandable within the framework of the SLM (Speech Learning Model).

NJ speakers who are first exposed to English as children (Early learners) enjoy greater success in producing and perceiving English liquids than those exposed as adults (Late learners) because they get more and better input from native speakers of English.
Summary

Both Early and Late learners are more successful at perceiving English /r/ than /l/ because the /r/ is perceptually more distant from the closest Japanese sound (/R/)

NJ speakers, even Late learners, manage to establish new phonetic categories for English /r/ if they get abundant native speaker input.

They are limited in the progress they can make for English /l/ because the phonetic properties this sound are merged with the properties of Japanese /R/.

Regardless of exposure age, Japanese-English bilinguals use a composite /R/-/l/ category when processing both the /R/ of their L1 and the /l/ of their L2.
Looking ahead

After more than 50 years of research we now have a fairly good idea of how the phonetic learning of English /r/ and /l/ unfold.

However, much additional work is needed to fully understand the phenomena discussed here.

It will be crucial in future research to evaluate the quantity and quality of the phonetic input that native speakers of Japanese obtain when learning English.
Looking ahead

I have wondered about phonetic learning in women like Hiroko Furukawa (now “Susie”) one of the nearly 45,000 Japanese “war brides” who came to the US in the early 1950s.

I predict that, if tested, Susie would be perfect for /r/ but not /l/ (at least if she can still speak Japanese).

The photos above are from a Washington Post article entitled “The untold stories of Japanese war brides” by Kathryn Tolbert.
Looking ahead

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 be able to fully understand L2 speech learning.

For more on this topic see:

Flege, J. E. L2 speech learning: Time to change the paradigm. Center for Research on Bilingualism, Stockholm University, June 11, 2018
Thanks for your kind attention

Above: my garden in Tuscania (VT), Italy
References


References


