The role of phonetic category formation in second language speech acquisition*

James Emil Flege, PhD

Professor Emeritus
School of Health Professions
University of Alabama at Birmingham

New Sounds 2016

8° International Conference on Second Language Speech
10-12 June 2016, Aarhus University, Denmark**

*The research reported here was supported for 25 years by the NIH, mostly the National Institute for Deafness & Other Communicative Disorders (NIDCD).

**This is the full version of the talk I intended to present in Denmark, which had to be shortened due to time constraints. References can be downloaded from: www.jimflege.com\research\conferences\aarhusrefs.pdf
Introduction

It’s great to be back.

A few of you here today will recall that the second «New Sounds» conference, held in Amsterdam in 1992, concluded with a panel discussion focusing on the construction of an «acquisition model» for L2 speech.

Three years later, after 15 years of preparatory work, I formally launched the Speech Learning Model (Flege, 1995).

It’s aim was to provide the needed acquisition model.
Introduction

The SLM drew on findings from diverse areas of research, including:

- **Experimental Phonetics**: for variation between and within languages, key concepts such as that of “phonetic space”, and so on;
- **Speech Science**: speech motor control, listener evaluation of speech errors, etc.;
- **L1 acquisition research**: establishment of contrastive phonemic level representations, attunement to the L1 phonetic system, etc.;
- **Infant development**: the precursors to speech, influence of the ambient language, etc.;
- **Experimental Psychology** (speech perception, models of language comprehension, etc.);
- **Psycholinguistics** (models of speech production, interactions between levels of organization, between production and perception, etc.

*Note*: See Flege (1998, 1992) for reviews of research that contributed to the specification of the SLM.
Introduction

The existing research indicated the importance of phonetic category formation in L1 speech development. In formulating the SLM, I «imported» phonetic categories from L1 to L2 speech research.

And then I did the unthinkable: proposed that L2 learners of any age, even late learners, retain the capacity to create new phonetic categories for certain L2 sounds. This was a daring move in 1995 considering that:

The general public and most L2 researchers were (are?) convinced that, as humans age, the capacity for learning new forms of speech is either:

- Lost;
- Diminished;
- Or becomes so difficult that, well, why bother?

**Wisdom from the internet.** The author says that it is «...plain to see that the human brain demonstrates a systematic decline in the acquisition of new information. No scientist will dispute this curve.»

[https://dodoresurrected.wordpress.com/2012/03/13/what-goes-on-inside-a-childs-brain/](https://dodoresurrected.wordpress.com/2012/03/13/what-goes-on-inside-a-childs-brain/)
Introduction

When I first began talking about L2 phonetic category formation at ASA meetings in the mid 1980s I recall that people in the audience rolled their eyes. This initial incredulity was followed by a period of indifference and then, quite miraculously, I discovered that many researchers seemed to think that the formation of new phonetic categories for certain L2 sounds was quite obvious, even banal.

I continue to think that phonetic category formation is a crucial aspect of L2 speech development. However, it remains important to evaluate this construct critically, with neither blind rejection or acceptance.

My goal today is to convince at least some of you that there is need for additional research focusing on L2 phonetic category formation. There are several motives for doing so. To name one: without understanding category formation in an L2 it will be impossible to account for the substantial inter-subject variability we often see in studies of L2 learning.

To illustrate this last point, consider some VOT data we (Flege et al., 1998) obtained from native Spanish adults – all late learners – who were learning English in the US.

Details regarding the next slide, a subset of the originally published data: VOT was measured in word-initial stops found in ten English words that were highly familiar to all of the participants (e.g., talk, teach, tall), 35 native speakers of Spanish living in Birmingham USA. The Ss differed substantially in AOA (mean age of arrival = 25 years, range: 6-50), LOR (mean residence = 7 years, range: 0.2-38), and self-reported % use of English (mean = 49%, range: 4-100)
Introduction

Some of these late learners seem to have used an unmodified Spanish /t/ to produce familiar English words. Others produced English /t/ with native-like VOT. Most, however, were somewhere in between.

The inter-subject variability you see here could not be explained by information drawn from a Language Background Questionnaire.*

To understand variability like this, we need to know which Ss have established a phonetic category for an L2 sound and upon what input the category was based.

*Data in the figure were drawn from Flege et al. (1998). Only a very modest correlation existed between VOT and AOA (r = -0.38, p < 0.05: the later, the more Spanish-like). The correlations between VOT and self-reported English use (r = -0.13, ns) and LOR (r = -0.09) were non-significant.
Organization of today’s talk

In the remainder of this talk I will:

1. State the SLM position on category formation;
2. Describe general characteristics of phonetic categories;
3. Delineate specific properties of phonetic categories;
4. Provide examples of category formation in L2 learning;
5. Offer several brief conclusions.
1. The SLM position on category formation

The SLM (Flege, 1995 ff.) differed substantially from previous approaches to L2 speech learning. It posits that

- L2 learners – even late learners – maintain all of the capacities used in successful L1 speech acquisition;

- The L1 phonetic system continues to influence L2 speech, even that of highly experienced learners, because the developing L2 phonetic system co-exists in a common «space» with the L1 phonetic system.
1. The SLM position on category formation

According to the SLM, two factors determine if a new phonetic category will be established for an L2 sound:

**Factor 1** (endogenous) pertains to an aspect of normal cognitive development.* By hypothesis: As L1 categories develop, they become stronger «attractors» of the closest L2 sound, making it less likely that a category will be established for the L2 sound.

**Factor 2** (exogenous) pertains to an interaction between the learner and the ambient language. By hypothesis: The greater is the perceived dissimilarity of an L2 sound from the closest L1 sound, the more likely it is that a category will be established for the L2 sound.**

*The state of development of L1 categories is likely to be correlated with chronological age, at least in a sample of learners having ages of L2 learning ranging from approximately 2 to 15-18 years, that is, the age range in which L1 categories gradually develop to mature levels.

**Note: The SLM does not specify whether Factors 1 and 2 interact. This represents one of several important ways in which the SLM remains incomplete.
The SLM position

Before continuing, a brief historical note: In 1992 I proposed that sounds encountered in an L2 could be typologized as: identical, similar or new. Like Henning Wode (1978), I thought that it would eventually be possible to establish an a priori metric to distinguish «new» from «similar» sounds.

As a preliminary, I proposed two crude methods for identifying «new» L2 sounds. For both vowels and consonants, an IPA symbol not used for any L1 sound. For vowels: a relatively great distance in an F1-F2 acoustic space from the closest L1 vowel.

This approach proved to be problematic and was abandoned when the SLM was formally launched in Flege (1995).
The SLM position

According to the SLM, identical L2 sounds are those sounds whose differences from the closest L1 sound, if any, can not be detected at a sensory level.

The boundary between «new» and «similar» L2 sounds must be established empirically based on measures of degree of perceived cross-language phonetic dissimilarity.

The SLM hypothesis regarding perceived dissimilarity is testable, but no one seems to have gotten around to doing so. Let’s pause here for some methodology.
The SLM position

Let’s consider how to go about evaluating perceived cross-language phonetic dissimilarity for a hypothetical L1 having five vowels and a hypothetical L2 with seven vowels.

**Step 1:** Obtain natural tokens of vowels spoken by representative monolingual speakers of the L1 and the L2.

**Step 2:** Present all possible pairings of L1-L2 sounds to native speakers of L1 who intend to learn the L2. This listeners’ task is to rate the degree of perceived dissimilarity of the vowels in each pair using a 7-point equal appearing interval scale.*

*Time needed in an unspeeded task in which participants can listen again if they want: about 15 min per participant. 7 x 5 = 35 pair types (a-1, a-2 ... e-7) x 10 monolingual talkers per language (5m, 5f) = 350 trials.
The SLM position

Here’s some hypothetical data for the 35 L1-L2 pairs. The data is plotted separately for the seven L2 vowels (L2-1 to L2-7).

The next step is to identify the highest cross-language dissimilarity ratings obtained for each of the seven L2 vowels (shown above in blue) from among the L1-L2 pairings.
The SLM position

This graph shows the rank-ordered maximum perceived dissimilarity ratings. A consideration of these data can be used to generate predictions within the SLM framework.

The SLM predicts that phonetic category formation will be most likely for L2-4 (i.e., L2 vowel #4) because this vowel differs more from the closest L1 than any other L2 vowel. Category formation will be least likely for L2 vowel #1.

Category formation for one L2 vowel (e.g., L2-4) but not another (e.g., L2-1) leads to the prediction of differing L2 effects on L1 production:

- **Deflection** of an L1 vowel(s) that are adjacent in vowel space to the L2 vowel for which a category has been formed;

- **Merger** of phonetic properties of L1 vowels adjacent in phonetic space to L2 vowels for which a new category has not been formed.
Now let’s consider to whom the predictions just laid out pertain. In L2 research, it is common to see comparisons between groups. However, assessing perceived L1-L2 differences via grouped data might lead to problems if the L1 categories of individual Ss differ.

Consider, for example, a study which examined perception of the vowel /i/ by monolingual English adults. Frieda et al. (2000) used a method of adjustment procedure to identify Ss’ «best example» of English /i/ from a grid of stimuli differing orthogonally in F1 and F2. These Ss differed considerably.

If Ss like these were included in research examining the acquisition of Spanish /i/, Ss whose preferred English /i/ is fairly low in vowel space might be erroneously interpreted to have learned more than Ss whose preferred /i/ is higher in vowel space.

That being the case: within the SLM framework, predictions need to be generated for individual subjects because L1 phonetic categories – hence cross-language phonetic dissimilarity ratings – may vary across individuals.
3. Phonetic categories: General properties

So what exactly are phonetic categories?

Think of them as «containers» in which learners store and structure information relevant to classes of speech sounds. To develop phonetic categories the learner must:

- Know what are the phonetic categories of the language being learned;
- Gain access to the properties defining those phonetic categories.

My working assumption is that L2 phonetic categories are formed in the same way as native language categories.* It is important to note, however, that L2 learners face certain difficulties that infants and young children who are learning an L1 do not.

*I refer to this as an “assumption” because I know of no research which has directly compared category formation in an L1 to that in an L2.
For infants, sorting a wide array of ambient-language phones into «sound types» is analogous to sorting shells found on the beach.

Knowing beforehand how many categories the L1 possesses would greatly simplify the infant’s speech learning task.
Of course, infants can’t know this. Fortunately, variations in the statistical distributions of tokens to which infants are exposed reduces the learning problem they face (e.g., Maye et al., 2002; Werker et al., 2012).

Infants enjoy two sources of information unavailable to most L2 learners:

1. The phonetic enhancements of infant directed speech (IDS) help infants distinguish the members of categories.

2. Somewhat later, children begin rapidly adding words to their lexicon. This forces them to attend closely to the phonetic differences distinguishing those words:

   \[ bat \neq pat \neq cat \neq sat \neq fat \] etc.
When L1 learning children misarticulate sounds, they get feedback. L2 learners, especially adults, rarely get this kind of feedback.

Consider the American tourist looking for a bathroom in Rome who asks:

Dove si trova la porta del bagno? (Where is the door to the bathroom?)

An incorrect rendition of the /p/ in «porta» is unlikely to trigger useful feedback from the barista (Eh? )

The identity of the target word («porta» door) is perfectly obvious from context
Another problem for L2 learners: orthography. At times, orthography is helpful. For example, it informs the learner of Italian L2 that /n/ ≠ /nn/ (bel cane, nice dog vs. canne alte, tall canes).

Just as often, orthography creates problems that have nothing to do with speech learning. Problems that are apt to persist (Flege, 1991).

Example: Italians need to learn English /ɪ/, a vowel not found in their L1. The closest Italian vowel is written with «i». Alas, in English, /ɪ/ is written with «i» (bit, pit, etc.) and /i/ is written with «ee» «ea» (beet, beat). It is no wonder that Italians complain constantly about English!
Phonetic categories (X) are perceptual in nature. They are represented in long-term memory, and are intermediate in degree of abstractness to:

- The **phonological codes** – phonemes – that differentiate words in the mental lexicon;
- The **sensory input** obtained via the ears and eyes.

As phonetic categories develop, they are defined via

- **Acquired distinctiveness**: which augments sensitivity to differences localized at the **boundary** between categories;
- **Acquired similarity**: which reduces sensitivity to differences between members of the **same** category.
Now let’s consider phonetic categories in more detail. To begin, phonetic categories are **language specific** representations that specify all properties needed to distinguish:

- The speech sounds in a single language *from one another*;
- Any *pair* of sounds drawn from two languages – even those traditionally represented by the same IPA symbol.

A phonetic category specifies:

- All of the properties (cues) used to identify a speech sound as being a member of the category;
- An organizational framework that elaborates how those properties are to be integrated with one another in real time (cue weighting).
Example: VOT is a property specified in phonetic categories developed for voiceless stops in both Spanish and English. However, it’s parameter values in those languages differ.

As a result: English monolinguals must hear longer VOT to identify a stop as voiceless than do Spanish monolinguals.
Another example: Word-final stops are released more frequently in French than English. As a result: final release bursts have greater perceptual «weight» for French than English listeners.

Identification of word-final stops as /g/ or /k/

adapted from Flege & Hillebrand (1987)
L1 learners eventually talk like the people they have heard most often.

As L1 phonetic categories develop they guide the parallel development of language-specific implementation rules for production.
**Example:** In a 2-dimensional acoustic space, the /i/ vowels of Spanish and English differ in F1 frequency, an index of vowel height.

Over time, children learning Spanish and English accurately assess the properties of /i/ in their native language. This information, stored in phonetic categories, guides the development of articulatory motor plans.

As adults, the children who learned English will produce /i/ with a slightly tighter constriction between the tongue and hard palate than those who learned Spanish as an L1.*

*Flege (1989) examined 8 talkers per language. The between language articulatory difference averaged about 1.4 mm at the point of maximum constriction (at Sensors 2 and 3).
Phonetic categories: guide production

The example just provided pertains to L1 speech development. But what about L2 learning? The SLM proposes that all of the capacities that were used in successful L1 speech development remain intact and accessible to learners of an L2.

A very different approach was suggested by Tom Bever.

Bever (1981, p. 193) proposed that once the production and perception of L1 phonetic segments have been successfully “aligned” via a “psychogrammar” there is no further need for the “internal communication” between production and perception. As a result, the psychogrammar which served to align production and perception in L1 speech acquisition “falls into disrepair because of disuse”. Use it or lose it.

Bever suggested, however, that losing the capacity to align production and perception is not inevitable so long as “one is continually learning a new language.”
Flege (1999) summarized research in which the relation between segmental production and perception was evaluated through correlational analyses. All of the studies yielded moderate correlations of about $r = 0.50$, including these:

- Flege (1993) examined Chinese Ss’ production and perception of vowel duration as a cue to the word-final distinction between English /t/-/d/, yielding a correlation of $r = 0.54$, $p > 0.01$;

- Research examining word-initial stops (Flege & Schmidt, 1995; Schmidt & Flege, 1995) focused on native Spanish Ss’ production of VOT in English /p/ and the location of the “best” /p/ in continua differing in VOT; $r = 0.54$, $p < 0.01$. In a post-hoc analysis, Ss were divided according to overall degree of foreign accent. A significant production-perception correlation was obtained for “proficient” Ss (i.e., those with relatively mild foreign accents, $r = 0.49$, $p > 0.01$) but not for less proficient Ss ($r = -0.004$);

- Flege et al. (1997) examined the production and perception of English vowels by 20 native speakers each of German, Spanish, Korean and Mandarin. The measure of perception was the size of the shift from one vowel category to another based on changes in F1 frequency. The measure of production was the size of F1 differences produced in pairs of English vowels. A production-perception correlation of $r = 0.53$ was obtained for English /i/ and /ɪ/, a correlation of $r = 0.52$ was obtained for /ɛ/ and /æ/. 
Phonetic categories: guide production

Flege (1999) offered several explanations for why higher production-perception correlations are not obtained in L2 research:

1. Not all aspects of perception are “transported” (Bever’s term) from perception to production. Example: I can perceptually distinguish Italian trilled /r/ from other variants of /r/ but, to my great embarrassment, can not produce trills. (I take consolation in the fact that trills are learned late by most Italian children and not at all by some Italians, including my late father-in-law who used a uvular /r/ instead of a trilled Italian /r/.)

2. The “transportation” of properties from (perceptual) phonetic categories to phonetic implementation rules takes time. This observation implies that if two groups differed in perception but not production at Time 1 of a longitudinal study, they may differ in production at Time 2.

3. The measure of segmental production and/or perception submitted to correlation analyses are limited by ceiling effects.

4. Production and perception are inherently incommensurable, making comparison difficult. Flege (1999) cited a study examining the perception and production of phonemic length contrasts in Swedish. The phonetic dimension of interest in both domains was overall vowel duration, seemingly a commensurable dimension. A correlation of $r = 0.70$ was obtained.
Phonetic categories: guide production

These considerations suggest that correlation analysis does not provide the best method to evaluate the contingency at the heart of the SLM hypothesis. The underlying logic is that accurate perception does not entail accurate production whereas accurate production requires accurate perception.

Flege (1999) cited a study which made use of a more appropriate form of analysis, a study examining the production and perception of English vowels by native speakers of Italian living in Canada. Most of the Ss examined succeeded in producing most English vowels in a readily identifiable fashion.

One exception was English /ʌ/, which was produced inaccurately by 31 of the 72 native Italian Ss. Given that the Italian vowel that is perceptually closest to English /ʌ/ is /a/, a categorial discrimination test was used to evaluated the discrimination of English /ʌ/ from Italian /a/.

The Ss who produced English /ʌ/ accurately were found to discriminate English /ʌ/-Italian /a/ significantly better than the 31 Italian Ss who produced /ʌ/ poorly. A difference in production accuracy was not obtained, however, when the Ss were divided into subgroups based on the ability to discriminate English /ʌ/-English /æ/ or the ability to discriminate English /ʌ/-English /ɑ/.
Relevant phonetic input is needed to establish phonetic categories in an L2.

The needed input:

- Can’t be accessed via the written word.
- Derives from language use: talking and trying to understand what others are saying.

Unlike young L1 learners, most learners of an L2:

- Can’t just sit there and listen. They have to make sense of what they’re hearing, and respond! Moreover, they don’t always get nice bi-modal distributions of input that highlight category contrasts.
- In fact, the may get input distributions that blur the difference between certain categories.
Phonetic categories: based on input distributions

**Example:** A child learning English in a monolingual community will likely be exposed to a distribution of VOT values for /t/ like this. (The distribution will shift, of course, as a function of factors such as speaking rate, degree of stress, etc.)

Things are different for native speakers of Spanish who immigrate to the US. They hear English spoken by:

- Native speakers of English;
- And also by fellow native speakers of Spanish.

**Above:** 1,200 tokens of American English /t/ as spoken by 20 adult monolingual native speakers of American English; the data from Flege et al. (1998)
Here I juxtapose the VOT distributions obtained from three groups (all data from Flege et al. 2001):

- English monolinguals;
- Native Spanish speakers who learned English relatively early in life;
- Native Spanish speakers who learned English somewhat later in life.

Spanish speakers learning English in the US will likely be exposed to all of these VOT values.

Unfortunately, we don’t know the details because no study of naturalistic learning has ever managed to define the input actually received by L2 learners.
However, we can be confident that:

The VOT distributions that Spanish speakers hear while they are learning English L2 in the US tend to fill in the phonetic space between the distributions typically heard by monolingual Spanish and English children.
Both the *quantity* and *quality* of input matter.

We don’t yet know the relative importance of the two.

However, to understand how the *quality* of input affects L2 learning, let’s consider the acquisition of word-initial English stops by native speakers of Spanish.
Phonetic categories: input distributions matter!

We (Flege & Eefting, 1986) tested two groups of university students in Puerto Rico.

Both groups consisted of native Spanish adults who had learned English in childhood, that is, early learners.*

Both groups of early learners produced English stops with VOT values that were significantly shorter than those of native English speakers.**

*Early learner-1: Had lived in the US for 9.7 years, attended public schools there for \( M = 6.4 \) years before returning to Puerto Rico. The members of Early learner-2 had never lived in the US but had attended a bilingual school in Mayagüez for \( M = 7.1 \) years beginning at the age of 5-6 years.

**Important, but off topic for the moment: Both early learner groups produced English stops with VOT values that were longer than the VOT values produced by Spanish monolinguals.
Phonetic categories: input distributions matter!

This (by now) surprising finding was almost certainly due to the fact that the early learners we tested in Puerto Rico had frequently been exposed to Spanish-accented English.

Several years later (Flege, 1991) recruited another group of early learners in Austin USA.

This time I made sure to recruit Ss who had been exposed frequently to native-produced English*

The early learners from Texas produced English /t/ just like English monolinguals.

*The Early learners in Austin reported having been unable to speak English when they began English speaking public schools at the age of 5-6 years. All had NE teachers in the first three grades and majority native English classmates. On average, they had received 13 years of formal education in the US and used English 82% of the time.
Phonetic categories: input matters

Getting the right input won’t matter, of course, if learners stop paying attention to it.

Some believe that phonetic categories are unlikely to change once they have been defined by ample input distributions.

I call this «rut theory».

According to Saito (2013), L2 learners usually reach an asymptote in L2 speech learning after 6 years of residence in a primarily L2 speaking country.

I believe, on the other hand, that input continues to matter well beyond 6 years.
In 2003 we (Ian MacKay and myself) had 160 Italian immigrants return to re-record English sentences they had recorded 10 years earlier.

When first tested in 1992 our native Italian Ss had already been living in Canada for an average of 33 years.

As you can see, pairs of English sentences produced a decade apart received very similar foreign accent ratings.

This indicated little change in the overall pronunciation of English.
We also elicited word production, measuring VOT in /p t k/ just as we had done a decade earlier.

There was little overall change in VOT in the two samples, which supports the «rut theory» of L2 speech learning.

However, when we took a closer look at our longitudinal VOT data, a different pattern emerged.

Traditions die hard just as much in the real world as in the world of L2 research.

Above: The ruts in Roman roads in England are separated by 4 feet 8.5 inches. This forced wagon makers to space wheels by the same measure. Years later, when wagons were converted for use as the first railroad cars, the ruts in Roman roads defined the spacing between railroad tracks.
Phonetic categories: input matters

From our sample of 160 we selected three subgroups of 24 native Italian Ss each based on changes in self-reported percentage use of English between 1992 and 2003. **Group 1:** less use in 2003; **Group 2:** no change; **Group 3:** more use in 2003.

The VOT values obtained for the three subgroups differed significantly ($p < 0.01$)

The Ss who reported using English more in 2003 than in 1992 showed an average 8 msec increase in VOT.
Phonetic categories: develop slowly

By school age, most children have established a **phonemic inventory** for their native language and are ready to learn how to read.

**Phonetic categories**, however, take far longer to develop fully.

The time course for **L1 category formation** is shorter for vowels than consonants ... 

For which an adult-like level of performance may not be evident until the late teens.*

**Phonetic categories: develop slowly**

**Example:** We found that phoneme boundaries, in both Spanish and English, occurred at significantly shorter VOT values for 8-9 year-old children than for adults.

Indeed, we found (Flege & Eefting, 1986) a significant difference between native English adults and 17 year-olds.*

*Comment: this study should be replicated with edited natural stimuli and an extended age-range.
Phonetic categories: develop slowly

Why so slow? Phonetic categories takes many years to fully develop because children learning an L1:

- Must identify the multiple properties defining each phonetic category;
- Then optimally integrate those properties to permit the rapid and reliable identification of phonetic segments in real time.

Both processes (identification, integration) are slowed by variability in the input received:

- The parameter values of phonetically relevant properties shift in a non-linear fashion as a function of phonetic context, speaking rate, degree of stress, etc.
- They may also differ as a function of talker, gender, age and dialect.
Phonetic categories: develop slowly

Phonetic categories permit the rapid and reliable identification of segments in non-ideal listening conditions. I suspect that the time needed to reach adult-like levels in the specification of phonetic categories will ultimately depend on the complexity of the information encoded. If so, the time needed for category formation may differ across languages. I’ll illustrate this idea with reference to the /p/ of Spanish and English.

In a magnitude production task we (Schmidt & Flege, 1996) had monolingual English and Spanish adults produce sentences at three speaking rates:

- Normal;
- half as fast as normal;
- twice as fast.

The two groups of monolinguals modified speaking rate in a similar way.
However, the Spanish and English monolinguals showed a different effect of speaking rate on the production of VOT in words found within the sentences.

When the English monolinguals shifted from a slow to a fast rate, VOT in /p/ decreased. Spanish monolinguals showed a non-significant trend in the opposite direction.
Phonetic categories: develop slowly

The speaking rate effect on VOT production averaged 15 msec for the English monolinguals.

A similar effect of speaking rate was also observed for their perception of /p/. We (Flege et al., 1996) had English monolingual English adults rate stops differing in VOT for «goodness». Two VOT continua simulated stops produced at a fast vs. slow rate. The VOT of their «best /p/» decreased by an average of 13 msec from the slow-rate to the fast-rate continuum.
Phonetic categories: develop slowly

As expected, Spanish monolinguals did not show rate dependent processing. Given that VOT varies as a function of speaking rate in English but not Spanish I hypothesize that it will take monolingual English children longer to fully establish a /p/ category in their native language than monolingual Spanish children.

But it’s not only that: The child who learns English as a native language must integrate rate-induced effects on VOT just mentioned with the effects:

- height of the following vowel;
- number of syllables in the word;
- degree of stress and/or emphasis;
- speaking clarity;
- burst intensity;
- aspiration intensity.

Doing all this takes many years.
Phonetic categories: are accessible

As for category development in an L2: None of this would be possible if learners were unable to access the phonetically relevant information that specifies L2 sounds. Can they?

There have been proposals that listeners may “filter out” information that is phonetically relevant to phonetic categories in an L2:

- They might do so from the top-down, discarding information not relevant to phonemic contrasts in the L1.

- Or, information may be excluded from the bottom-up. As a result of attunement to the native language sound system information relevant to phonetic categories of the L1 may be privileged at a pre-attentive stage of processing, so that certain properties that important in the L2 don’t get incorporated into L2 phonetic categories.
I agree that in early stages of learning some L1-L2 phonetic differences may go unnoticed. However, I think that over time, learners can gain access to all phonetically relevant information in L2 sounds. At least if they have ample opportunity to do so.

Let me explain why I think this.

The examples to follow involve foreign accent (FA), which is perceived when listeners detect divergences in a speech sample from their own perceptual representations for speech.
Flege (1988) assessed overall degree of perceived foreign accent in English sentences spoken by Taiwanese late learners differing in L2 experience:

- Relatively inexperienced, LOR = 1.2 years
- Relatively experienced, LOR = 5.1 years

A four-year LOR difference between the groups did not result in a significantly improved global pronunciation of English.

Both groups of late learners had significantly stronger foreign accents than did native English speakers and a group of early learners (p >0.05)
However, a four-year difference in LOR did affect perception. The late learners’ ability to gauge degree of foreign accent, which depends on the ability to note pronunciation differences between natives and non-natives, increased significantly with added L2 experience. In the graph below NE > LOR 5.1 years > LOR 1.2 years (p > 0.05).

This finding suggests that the Taiwanese late learners were detecting cross-language phonetic differences that were not (yet) evident in their production of English sentences.
Phonetic categories: are accessible

Flege (1984) used a paired-comparison task to evaluate listeners’ ability to detect cross-language phonetic differences.

Native English listeners heard two English stimuli per trial, one produced by a fellow native speaker, the other by a native speaker of French. The listeners’ task: decide which member of a pair of stimuli had been produced by a non-native (French) speaker.

Significantly above-chance performance for whole syllables (u/, t/); for the vowels /u/ and /i/; for cross-spliced /t/ tokens, which consisted of a release burst and a variable VOT interval; and for isolated “t-burst” stimuli, all 30 msec long.

Notes: Cross-editing made it possible to evaluate syllables in which talker background (English, French) varied in a single segment (/i, /u/, /t/). Chance performance was 50% correct.
Results for the “/t/” and “t-burst” stimuli indicated that the native English listeners were able to detect VOT differences between English and French-accented English, and even the tiny place of articulation difference between French and English.

One might reasonably ask if L2 learners can exploit the auditory sensitivity evident in a controlled laboratory experiment.

The results obtained by Flege & Hammond (1982) suggest that they can.
Participants in our 1982 study were native English students taking a 1st year Spanish class at the University of Florida. We could be sure that they had all been exposed to Spanish-accented English because

- All were from Florida, where many native Spanish speakers reside;
- All were currently taking a Spanish class taught by a native Spanish instructor who spoke English with a strong foreign accent

The task was simple: read English sentences of the form “The X is on the Y” with a “Spanish accent”. One test sentence was:

*The vase is on the pig*

Which some Ss rendered as:

*dʌ bes iz an dʌ pig*

*Note:* no instructions were given regarding how to speak with a «Spanish accent». The Ss found the task wildly entertaining which was likely due, at least in part, to the presence of their native Spanish teacher.
VOT was measured in 3 test words beginning in /t/ (*tape, tube, toad*)

We estimated our Ss’ exposure to Spanish-accented English by counting the number of “Spanish accent substitutions” they produced.

<table>
<thead>
<tr>
<th>Substitution</th>
<th>Total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>nose, cheese, hose</td>
<td>z → s</td>
</tr>
<tr>
<td>vice, veil, vase</td>
<td>v → b</td>
</tr>
<tr>
<td>fig, pig, wig</td>
<td>i → i</td>
</tr>
<tr>
<td>book, hook, crook</td>
<td>u → u</td>
</tr>
<tr>
<td>shell, sheet, sheep</td>
<td>j → č</td>
</tr>
<tr>
<td>bean, phone, bone</td>
<td>n → ŋ</td>
</tr>
</tbody>
</table>
The Ss most familiar with Spanish-accented English significantly reduced VOT by 30 msec, to values often seen in Spanish-accented English.

This finding indicates that our participants were able to detect VOT differences between English and Spanish-accented English; store the detected information in long-term memory; and later use that information in production.

In so doing our young adult monolingual English Ss demonstrated the core capabilities needed for L2 speech learning.
Phonetic categories: are accessible

What about cases in which a property (dimension) that helps define an L2 sound isn’t found in the L1?

To address this question, let’s consider a study examining the use by Taiwanese late learners of vowel duration as a perceptual cue to the distinction between word-final English /t/-/d/.

Taiwanese has words ending in /p t k/ but none ending in /b d g/. Thus, the Taiwanese late learners examined by Flege (1993) had no prior experience using vowel duration as a cue to voicing distinction in word-final stops.

Design of Flege (1993)

**Stimuli:** Natural edited *beat-bead* and *bat-bad* continua in which vowel duration varied in 17 steps

**Method of adjustment task:** chose the “best example” of /t/ and /d/ (separate runs) from among the stimuli differing in vowel duration

**DV:** the size of the vowel duration differences for the “preferred” /t/ vs. /d/-final stimuli
The early learners tested by Flege (1993) showed an English-like use of vowel duration as a cue to the /t/-/d/ distinction.

Both groups of Taiwanese late learners, on the other hand, made significantly less use of vowel duration than did the native English speakers.

The late learners differing in LOR did not differ significantly in their use of vowel duration (as they did in their perception of foreign accent), but a trend in the expected direction exists.
Another important aspect of phonetic categories is that:

phonetic categories are not phonemes

Let’s now consider some of the ramifications of this obvious fact.
Phonetic categories: are not phonemes

To begin, listeners have decisive «all or none» judgments of phonemic identity that is evident at a conscious level.

The seeming ambiguity of phonemic judgments obtained in certain testing conditions, e.g., a two-alternative forced-choice test examining VOT stimuli ranging from /t/ to /d/, is an artefact of the test. In experiments like this, listeners are told to «guess if you are not sure. They do not perceive a stimulus located near the “phoneme boundary” as being «half /t/, half /d/.

Results such as those obtained by Flege (1993), where listeners’ judgments of preferred vowel durations was examined (see preceding slide), demonstrate a partial awareness of phonetic dimensions This level of awareness, sometimes referred to as “inner speech”, is less decisive than the awareness associated with phonemic judgments.
Second, phonetic categories often outnumber phonemes.

We see this clearly in a study in which monolinguals and bilinguals were asked to imitate the randomly presented members of a VOT continuum (Flege & Eefting, 1988).

The Ss’ vocal output did not linearly track VOT values in the stimuli. Instead …
Phonetic categories: are not phonemes

When imitating the VOT stimuli:

• **Spanish monolinguals** (children, adults) tended to produce stops with lead or short-lag VOT;

• **English monolinguals** (children, adults) tended to produce stops with short-lag or long-lag VOT values.
Phonetic categories: are not phonemes

Two groups of native Spanish early learners, on the other hand, imitated the stimuli with values in all three modal VOT ranges: lead, short-lag and long-lag.

This pattern indicated that although the bilinguals used two phonemes to label the stimuli in an identification experiment, they possessed three phonetic categories.
Phonetic categories: are not phonemes

Second, unlike «one-size-fits all» phonemes, phonetic categories are position sensitive. They correspond roughly to what was once designated the “major allophones” of a phoneme in pre-SPE phonological analyses.

Consider, for example, the identification of English liquids by native Japanese adults.

As had others before them, Takagi & Mann (1995) found that native Japanese adults are less able to identify English liquids in word-initial than word-final position.
Phonetic categories: are not phonemes

Third, phonetic categories are less abstract than phonemes.

This has important ramifications for L2 speech research. It means that to test for the presence of new phonetic categories in an L2, it is crucial to use a task focusing on the phonetic level of organization.
I’ll illustrate this by presenting two experiments from a study examining the perception of English vowels by native speakers of Italian.

One experiment evoked phonemic level representations; The other focused on the phonetic level of organization.

Phonetic categories: less abstract than phonemes
Phonetic categories: less abstract than phonemes

Flege & MacKay (2004) recruited four groups of Italians who had lived in Canada for decades. The groups differed orthogonally in:

Age of L2 learning (AOA):

- **Early** learners       $M = 8$ years, range: 3-13
- **Late** learners       $M = 19$ years, range 15-28

Self-reported % use of English*

- **High use** of English $M = 92\%$  range: 85-98%
- **Low use** of English  $M = 54\%$  range: 25-71%

*The Ss actually estimated their % use of **Italian**. Given that they were bilinguals, the English values shown here were obtained by subtracting the Italian estimates from 100%
Both experiments examined the learning of English /ɪ/, a vowel of interest because Italian has an /i/ but no /ɪ/. Italians classify English /i/ tokens as Italian /i/, English /ɪ/ tokens as «poor» instances of the same Italian vowel.*

The first experiment to be presented here made use of an error detection task. The stimuli were short English phrases spoken extemporaneously by native speakers of Italian:

- Half of the phrases contained the target vowel /i/, produced correctly, or as /ɪ/. Examples: *with p*ple, *to sp*k English*

- The remaining half of the test phrases contained the target vowel /ɪ/, produced correctly or as /i/. Examples: *to l*ve in, *big c*ty

*We obtained these data from Italian university students who had been in Canada for 3 months at the time of testing. See Flege & MacKay (2004), Experiment 1.*
Phonetic categories: less abstract than phonemes

The task of native English listeners – drawn from the same location where the production data had been gathered – was to decide if the target vowel in each phrase – indicated by an asterisk – was «correct» or «incorrect».

The experiment yielded strong effects of age of L2 learning (AOA) and % English use. As expected, both groups of late learners differed significantly from the NE group.
Differing results were obtained for the two groups of early learners.

Early learners who used English relatively seldom – in this case, slightly more than 50% of the time on average – differed significantly from the NE group.

However, the early learners who used English more often showed native-like performance.
Phonetic categories: less abstract than phonemes

It is important to remember that the «Early low» group consisted of native Italian speakers who:

• Had arrived in Canada as children;
• Had lived there for 40 years, on average;
• Used English more than half the time, on average.

Does our finding for the Early-low group indicate that they were unable to establish a new phonetic category for /ɪ/?
Phonetic categories: less abstract than phonemes

Probably not. The results we obtained using an error detect task showed the influence of phonemic level representations on the listeners’ judgments.

We got different – and more readily interpretable – results when we used a task focusing on phonetic level representations.
The stimuli used in our \textbf{categorial discrimination test} were multiple natural tokens /i/ and /ɪ/.

To get a high score the Ss had to

- Detect between-category differences (trials with one /i/, one /ɪ/ token)
- Ignore irrelevant within-category variation (in catch trials having either two /i/ tokens or two /ɪ/ tokens).
We again obtained strong effects of age of L2 learning (AOA) and % English use.

Neither group of early learners differed significantly from the NE group.

Importantly: Ss in all four native Italian groups obtained scores falling within the NE range (+/- 2 SDs).

This criterion was met more often by early than late learners (86% vs. 44%), and by more Ss who reported using English relatively frequent than infrequently (83% vs. 47%).
Now let’s consider research examining the ability of learners to establish new phonetic categories for L2 sounds.

In this section I’ll present research focusing on the perception in English of:

1. /p/ - by native speakers of Spanish
2. /s/ - by native speakers of Italian
3. /r/ - by native speakers of Japanese
Category formation examples: English /p/

1. English /p/ - native speakers of Spanish

We (Schmidt & Flege, 1995) examined the production of English /p/ by native Spanish early and late learners.

Here I have arranged the VOT values obtained from individual Ss in ascending order.

Note the greater variability among late than early learners.
Category formation examples: English /p/

Of the 10 late learners we examined:

(a) Four produced English /p/ with Spanish-like VOT values;

(b) Four others produced /p/ with values that equaled or exceeded those of NE speakers.

Why the difference?
Here is a possible interpretation provided by the SLM:

The Ss in subgroup “a” seem to have used a Spanish /p/ in English, exhibiting no learning so far;

Ss in subgroup “b” established a phonetic category for English /p/, which guided the development of language-specific realization rules for use in English.
Perceptual results support this interpretation.

We (Flege et al., 1996) used a goodness rating task to examine Ss’ perception of English /p/.

The procedure enabled us to find each S’s “best” /p/ from an array of VOT stimuli.

As in previous research, native English monolinguals exhibited “rate dependent” processing.

This indicated that the influence of speaking rate on VOT was incorporated into their phonetic category representation for /p/.
Category formation examples: English /p/

The late learners who produced English /p/ with Spanish-like VOT values (subgroup “a”) did not show rate dependent processing because they didn’t have an English /p/ category.

On the other hand, those who produced English /p/ accurately (subgroup “b”) did show rate-dependent processing because they had established a phonetic category for English /p/.
2. English /ɜ/ - native speakers of Italian

Italian has no vowel resembling the «r-colored» /ɜ/ of English. This English vowel differs substantially from any Italian vowel in terms of F2 and, especially, F3.
Category formation examples: English /ɜ/ 

Not surprisingly: Italians rate English /ɜ/ tokens as being substantially more dissimilar from the closest Italian vowel than any other English vowel.

Given a high degree of perceived cross-language phonetic dissimilarity, the SLM predicts that many Italian learners will establish a phonetic category for English /ɜ/.

I have no relevant perceptual data to report today in support of this prediction.

The problem is: English /ɜ/ is simply too easy for Italians to distinguish from other English vowels.

(from Fleget & MacKay (2004), Exp. 2)
Instead, I’ll present some production data.

We (Flege et al., 2003) used a delayed repetition task to elicit the production of /ɜ/ by Italians differing in age of L2 learning (AOA: early vs. late) and % English use (relatively high vs. low).*

Nearly all of the Italians’ /ɜ/ productions were identified as intended by native English listeners.

The listeners were somewhat less likely to rate the Italians’ productions of /ɜ/ as «good», that is, more than simply «acceptable»

*We actually had Ss estimate % Italian use. However, for communicative ease, I have again switched the valence of these estimates.
Individual Ss in all four native Italian groups obtained «good» scores falling within the NE range. This held true for most early learners (92%) and many (53%) late learners.

Acoustic analyses yielded similar results.

**Background:** A correct articulation of /ɜ/ lowers F3 frequency. We examined the difference between F2 and F3 frequencies (following conversion to the Bark scale).
Category formation examples: English /ɜ/ 

For this analysis we added productions of English /ɜ/ by Italians who had never lived outside Italy (arrow).

All four groups of bilinguals in Canada showed evidence of learning, producing smaller F2-F3 differences than did the Italians in Italy.
Category formation examples: English /ɜ/ 

Most (72%) early learners and many (19%) late learners produced F2-F3 differences that fell within the NE range.

These data conformed to our expectations, however:

1. Only production data were considered. Perceptual measures that evaluate the integration of multiple cues to the /ɜ/ category are needed.

2. Even when/if such data are obtained, we need to develop a clear measure that will enable us to distinguish Ss who have established an /ɜ/ category from those who have not.
3. English /r/ - native speakers of Japanese

In this last section I’ll consider a few of the many studies examining the learning of English liquids by native speakers of Japanese. It is widely agreed that English liquids are difficult – perhaps even impossible – for Japanese adults to learn.*

Flege et al. (1996) summarized the results of 12 prior studies examining the identification of /r/ and /l/ by native speakers of Japanese. The mean % correct identification of /r/ and /l/ in those studies was just 69%; Japanese adults were found to make bi-directional errors: labelling /r/ as /l/ and vice-versa.

*Takagi & Mann (1996: 387) “… even extensive natural exposure does not ensure perfect perceptual mastery of /r/ and /l/ by adult Japanese learners of English”. According to Bradlow (2008: 294) examining /r/-/l/ is a good way to test “general principle of learning and claims about adult neural plasticity”. Hattori & Iverson (2009: 477) “… one of the most difficult contrasts reported in the literature”
Category formation examples: English /r/

Why so difficult?

The two English liquids are usually identified as being instances of the one liquid found in Japanese.*

However, Japanese learners of English do not judge tokens of the two English liquids as being equally poor instances of their /R/ category.

They perceive English /r/ to be more dissimilar phonetically from /R/ than English /l/.

*Note: for convenience, I’ll refer to the Japanese liquid as “R”
Takagi (1993)* showed that, for Japanese adults who have lived in the US, /r/ is a poorer “fit” to Japanese /R/ than English /l/ is.

*See Hattori & Iverson (2009) for use of a different technique leading to the same conclusion.
Guion et al. (2003) found that Japanese adults who had conversational experience in English were able to discriminate English /r/ from /R/ at significantly above-chance rates. However, even Japanese adults who had conversational experience in English were unable to discriminate English /l/ from /R/ (arrow).

---

**Graphical Data**

<table>
<thead>
<tr>
<th>Groups of Japanese adults (n=10)</th>
<th>Mean LOR in US = 3.1 years (range = 1.8 - 5.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>J1</td>
<td>Use English at work in Japan; no foreign residence</td>
</tr>
<tr>
<td>J2</td>
<td>University students in Japan; no foreign residence</td>
</tr>
</tbody>
</table>

---

How much conversational experience is needed by Japanese adults in the US to show a difference between /r/ and /l/? To address this question, I re-analyzed data presented by Aoyama & Flege (2011).

From the original sample of 50 native Japanese adults living in Birmingham, Alabama, I selected 3 groups of 15 Ss each having non-overlapping values for a variable I'll call “conversational input”, calculated by multiplying length of residence (in years) by self-estimated percentage use of English.

<table>
<thead>
<tr>
<th>Group</th>
<th>Input</th>
<th>AOA</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>0.01(0.08)</td>
<td>34.7(3.4)</td>
<td>35.5(3.7)</td>
</tr>
<tr>
<td></td>
<td>0.0-0.2</td>
<td>26-39</td>
<td>26-40</td>
</tr>
<tr>
<td>Med</td>
<td>1.16(0.47)</td>
<td>27.9(4.0)</td>
<td>30.4(5.2)</td>
</tr>
<tr>
<td></td>
<td>0.5-2.1</td>
<td>20-32</td>
<td>21-43</td>
</tr>
<tr>
<td>High</td>
<td>6.46(4.40)</td>
<td>25.3(5.1)</td>
<td>33.4(7.1)</td>
</tr>
<tr>
<td></td>
<td>2.4-14.4</td>
<td>18-36</td>
<td>23-48</td>
</tr>
</tbody>
</table>

Means, SDs, and ranges for three groups. The groups did not differ significantly in chronological age, F(2,42)=3.13 p>0.05) but did differ in age of arrival in the US, F(2,42)=19.7, p > .001. Tukey’s tests revealed that Low had arrived in the US later in life than the Ss in the Medium and High input groups (p> .05).
In separate counter-balanced blocks, the native Japanese Ss used a 7-point EAI scale to rate the perceived dissimilarity of English /r/ and /l/ to Japanese /R/.

The difference between the ratings obtained for /r/-/R/ and /l/-/R/ were evaluated by two-tailed t-tests.

Results: The Medium and High input groups rated English /r/ as more distant from the Japanese liquid than English /l/ (p > 0.05) whereas Ss in the Low input group did not. Based on group characteristics, we infer that Japanese adults need roughly 6 months of conversational experience with native English speakers to begin assimilating English and /l/ differently. Obviously, more fine-grained measures will be necessary.
To better understand how perceptual assimilation develops, I carried out simple and partial correlation analyses. Examining the results tabled below suggests that as Japanese adults gain conversational experience in English they note differences between English /r/-Japanese /R/ but not between English /l/-Japanese /R/. Importantly, the difference in the perceived dissimilarity of /r/ and /l/ vis-à-vis Japanese /R/) predicted the categorial discrimination of English /r/-/l/.

<table>
<thead>
<tr>
<th></th>
<th>Age of arrival</th>
<th>/r/-/R/ ratings</th>
<th>/l/-/R/ ratings</th>
<th>/r/-/l/ difference</th>
<th>/r/-/l/ Discrimination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conversational input</td>
<td>(a) -0.52**</td>
<td>(a) -0.60**</td>
<td>(a) -0.38</td>
<td>(a) 0.53**</td>
<td>(a) 0.33</td>
</tr>
<tr>
<td></td>
<td>(b) -0.49**</td>
<td>(b) 0.08</td>
<td>(b) 0.49**</td>
<td></td>
<td>(b) 0.42*</td>
</tr>
<tr>
<td>Age of arrival in US</td>
<td>(a) 0.43*</td>
<td>(a) 0.31</td>
<td>(a) -0.24</td>
<td>(a) -0.39*</td>
<td>(a) 0.03</td>
</tr>
<tr>
<td></td>
<td>(c) 0.17</td>
<td>(c) 0.15</td>
<td>(c) 0.05</td>
<td></td>
<td>(c) 0.26</td>
</tr>
<tr>
<td>/r/-/R/ dissimilarity ratings</td>
<td>(a) 0.88**</td>
<td>(a) -0.39*</td>
<td></td>
<td></td>
<td>(a) -0.08</td>
</tr>
<tr>
<td>/l/-/R/ dissimilarity ratings</td>
<td></td>
<td></td>
<td>(a) 0.08</td>
<td></td>
<td>(a) 0.15</td>
</tr>
<tr>
<td>/r/-/l/ rating difference</td>
<td></td>
<td></td>
<td></td>
<td>(a) 0.51**</td>
<td>(a) 0.53**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(b) 0.53**</td>
<td>(b) 0.41*</td>
</tr>
</tbody>
</table>

Simple correlations, df = 43; (b) Variation in AOA partialled out, df = 42; (c) Variation in Conversational input partialled out, df = 42.

*significance at p > 0.01, **significance at p > .001
What these findings suggest, when taken together, is that:

As Japanese adults gain experience in English, they continue to judge /l/ to be similar to /R/;

However, as they begin to discover properties that define /r/, Japanese adults begin to perceive this English sound to be less similar to /R/. This leads to a better discrimination of English /r/-/l/.

Why is this important?

The SLM predicts that a difference in perceived dissimilarity between the two English liquids will lead to important learning differences.

Specifically: native Japanese adults should be better able to identify English /r/ than /l/.
Category formation examples: English /r/

To test this, we (Flege et al., 1996) examined the identification of English /r/ and /l/ by Japanese adults living in the US.

Our study differed from previous research in two important ways:

1. It compared groups of Japanese adults having substantial differences in conversational experience.

   Experienced Japanese (EJ): mean LOR = 21 years (range 12-29)
   Inexperienced Japanese (IJ): mean LOR = 2 years (range 1-4)

2. It examined the subjective lexical familiarity of the /r/- and /l/-initial words used as stimuli.
As expected, we found that the Japanese Ss had no difficulty identifying English /w/ and /d/ but generally had difficulty identifying English liquids.

Also as expected: the experienced Japanese Ss outperformed those who were relatively inexperienced in English; and both Japanese groups obtained higher scores for /r/ than /l/.

Above: identification data obtained by Flege et al. (1996)
Scores obtained for the **experienced** Ss for /r/ were slightly lower (*mean* = 93% correct) than the perfect scores obtained by the NE group.

The experienced Ss' scores were, of course, higher than those obtained by the relatively inexperienced Ss examined in this study and by Japanese late learners examined in prior and subsequent studies.

Still, one might reasonably ask: Why, after 21 years of residence in the US, didn’t the experienced group show **perfect performance** like the NE group?

Before answering, I’d like to remind you of something important.
We used a two-alternative forced choice test to examine the identification of English liquids in words like lead and rode.

The Ss were required to identify initial liquids in the words using one of two overt category labels, “r” or “l”.

In a procedure like this, results are subject to the influence of phonological codes in lexical representations.

The NE Ss showed no effect of lexical bias in labelling the stimuli. Their phonetic categories for /r/ and /l/ were perfectly attuned to the stimuli, yielding 100% correct scores for both /r/ and /l/.

However, many Japanese Ss did show effects of lexical bias, tending to respond with a word that was more familiar than its minimal pair.
Category formation examples: English /r/

The lexical bias effects observed for Japanese Ss were: significantly stronger for the inexperienced than the experienced group; and significantly stronger for /l/ than /r/.

To see results at the phonetic category level, that is, results uncontaminated by lexical bias, we need to consider scores for /r/ and /l/ tokens drawn from minimal pairs consisting of equally familiar words.
Category formation examples: English /r/

At the appropriate level of analysis we see better performance on /r/ than /l/ for nearly all of the Japanese Ss (10/12 in both the inexperienced and experienced groups) …

And perfect performance on /r/ by 8 Japanese Ss (7/12 EJ, 1/12 IJ).

To understand this finding better, let’s turn briefly to production.
In a companion study, we (Flege et. al., 1995) elicited production of /r/ and /l/ by the same Japanese Ss.

We used three elicitation tasks: Definition, Word-list reading, and Extemporaneous. Production accuracy was assessed by listener judgments (ratings). The productions elicited by many native Japanese adults fell within the range of values obtained for the native English group.
Production elicited using the “Definition” task is of special interest because this method avoided unwanted effects of orthography.* Let’s consider these results in further detail.

**All** experienced Japanese late learners obtained scores falling within the NE range* for both /r/ and /l/;

**Most** of the relatively inexperienced late learner obtained scores within the NE range for /r/ but not /l/.

<table>
<thead>
<tr>
<th></th>
<th>/r/</th>
<th>/l/</th>
</tr>
</thead>
<tbody>
<tr>
<td>NE</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>EJ</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>IJ</td>
<td>11</td>
<td>4</td>
</tr>
</tbody>
</table>

*On each trial of the Elicitation task (the first of the three administered) the NJ Ss heard the Japanese translation equivalent of one of the 36 English words to be elicited. Then they saw a definition of the English words to be produced. For example, “What we get from the sun” to elicit the target word *light*. 
The listener ratings suggested that all experienced and most relatively inexperienced Japanese late learners managed to produce English /r/ just like the NE speakers.

How can that be? The Japanese late learners could not possibly have received the same input as the NE Ss. And they surely heard other native Japanese speakers of English mispronouncing English liquids.

Perhaps the listeners missed subtle differences between the /r/ tokens that had been spoken by the native English and Japanese participants.

Given this possibility, we carried out acoustic analyses.
Category formation examples: English /r/

The experienced Japanese Ss produced /r/ with significantly lower (more English-like) F3 values than did those having less experience in English, but with significantly higher values than the NE group.

The F3 values in /r/ tokens produced by the three groups differed significantly, $F(2,147) = 27.7, p < 0.0001$. A Tukey test revealed that all 3 groups differed from one another ($p< 0.0001$).

<table>
<thead>
<tr>
<th></th>
<th>F3 onset frequency (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native English (NE)</td>
<td>1,750</td>
</tr>
<tr>
<td>Experienced (EJ)</td>
<td>2,106</td>
</tr>
<tr>
<td>Inexperienced (IJ)</td>
<td>2,261</td>
</tr>
</tbody>
</table>

These differences in F3 are important. In a stepwise multiple regression analysis, variation in F3 frequency accounted for 67% of the variance in the listeners’ ratings.*

*Note: Flege et al. (1995) measured four acoustic dimensions in /l/ and six dimensions in /r/ tokens. These values were regressed, in separate analyses, onto the listener ratings for /l/ and /r/. The analysis reported above differed slightly from the original analysis for /r/. Here I examined: F2 frequency at release and at the instant of rapid spectral change, F3 frequency at the same two points, and F3-F2 difference at the same two points. Of these six variables, only the F3 values at release was identified as a significant predictor of the listener ratings.
The production results lead us to wonder if many Japanese late learners are unable to discover that a low F3 frequency is crucial to the definition of English /r/.

Work by Paul Iverson suggests that acoustic properties not used to define L1 categories may be inaccessible for late learners. If so, Japanese learners of English may be unable to use F3 when/if they establish a phonetic category for English /r/.

Iverson et al. (2003) examined listeners ratings of an array of stimuli differing in F2 and F3.
MDS analyses of the ratings obtained for pairs of stimuli drawn from the stimulus array yielded very different perceptual spaces for native English and Japanese adults.

Iverson et al. suggested (2003, p. B53) Japanese adults’ perceptual processing is “miss-tuned” for the identification of English liquids because they tend to respond more to an acoustic property that is helpful in Japanese but not English, F2, while tending to ignore a property that is “more salient” in English, namely F3.

Iverson was referring to a level of processing that is not consciously available to listeners, that is, processing that occurs before phonetic categories are accessed.
Category formation examples: English /r/

The Japanese adults tested by Iverson et al. (2003) had never lived outside of Japan. Experiments with Japanese adults who did have conversational experience in English while living abroad, on the other hand, suggests that Japanese learners of English can gain access to F3.

Hattori & Iverson (2009) tested 39 Japanese adults who had been living in London for an average of 3 months at the time they were tested.

A method of adjustment procedure was used to determine the F3 values in the «best» exemplars of the English /r/ and Japanese /R/ categories.

The Japanese Ss selected significantly lower F3 values for English /r/ than for the Japanese /R/. However, they did not rely on F3 when identifying liquids as /r/ or /l/. Perhaps, having resided for only 3 months in London, these Japanese adults were in very early stages of forming a category for English /r/.
Takagi and Mann (1995) examined the identification of /r/ and /l/ by Japanese adults who had far more conversational experience in English (many of whom also participated in the research reported by Flege et al., 1995, 1996).

Overall, the native Japanese adults used F3 less than the NE Ss did. However, the experienced Ss showed significantly greater use of F3 than did the relatively inexperienced Japanese Ss.

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

The SLM framework provides a way to interpret the data just presented.

Late native Japanese learners of English who do not establish an category for English /r/ identify both liquids of English using a single phonetic category, the one established in the L1 for /R/.

Those who do establish an category for English /r/, on the other hand, use two distinct phonetic categories when identifying English liquids.
Category formation examples: English /r/

There is no time available to describe the additional research needed to confirm or disconfirm the interpretation just offered. However, before concluding I want to speak briefly about factors that contribute to inter-subject variability among late learners.

The amount and kind of input received is clearly important. But there is more to it than that, as we can see in the results of an /r/-/l/ training experiment.

Training study:

Jun Yamada (1991) administered 1 hour of identification training to 152 college students in Japan.

Stimuli: Four naturally produced non-word minimal pairs (e.g., rosti vs. losti).
The effect of training varied dramatically. Of the 152 adults who received /r/-/l/ training in Japan:

- 1% could differentially identify /r/ and /l/ before training;
- 6% could do so in all 4 minimal pairs after training;
- 35% did so for several minimal pairs;
- However, 51% of the Ss remained at chance for all four pairs.

Differences in input probably can not explain the inter-subject variability yielded by the Yamada (1991) study. The Japanese students received the same input in the laboratory, and all had very similar prior instruction in English.

More likely, the inter-subject variability was due to: undefined individual differences*, including motivation; or to individual differences in the Ss’ Japanese /R/ category, which could influence the initial perceptual mapping between the one liquid of Japanese and the two liquids of English.

*On this point, see, e.g., Ingvalson et al. (2011), p. 581
5. Summary & conclusions

In this presentation I have reviewed:

- Research that helps define the properties of L1 phonetic categories;
- Research indicating that, under certain conditions, learners of an L2 establish new phonetic categories for L2 sounds.

I have concluded that:

- The likelihood of phonetic category formation for an L2 sound increases as a function of the sound’s degree of perceived phonetic dissimilarity from the closest L1 sound(s);
- Early learners are more likely to form categories than late learners are;
- Individuals who use the L2 relatively often (and the L1 relatively seldom) are more likely to establish new L2 phonetic categories than those who use the L1 less frequently (and, of course, the L1 more frequently).
- At any age: the formation of phonetic category in the L2 requires, just as in L1 acquisition, many years of adequate input.
Summary & conclusions

There are many important things we still don’t know about the formation of phonetic categories for L2 sounds. For example:

1. What triggers the formation of a new category?

2. How much time/input is needed for L2 categories to become functionally equivalent to native speakers’ categories?*

3. Is more input needed by learners whose L1 phonetic system is fully established when L2 learning commences than by those whose L1 system is still “under development”.**

4. Is some special skill or aptitude needed by late learners to establish phonetic categories for L2 sounds?

*In my writings over the years I referred to the “mastery” of the production or perception of L1 phonetic segments. I now regret using this term for it implies that bilinguals might develop L2 phonetic categories that are identical to those of monolingual native speakers of the target L2. This is not possible inasmuch as bilinguals possess two phonetic systems that necessarily interact, and cannot have received the same input as L2 native speakers.

**More specifically, do Factors 1 and 2 (see slide 9, this talk) interact? Unless and until this interaction terms is specified the SLM will remain a framework rather than an actual model of L2 speech development.
Summary & conclusions

In preparing this talk, I was surprised to see that little research in the past decade has focused on phonetic category formation. Why not? I suspect that research is being hindered by:

1. Difficulty recruiting highly experienced Ss;

2. A «litmus test» that can indicate whether a phonetic category has or has not been formed for an L2 sound;

3. Precise measures of the amount and kind of input L2 learners receive.

It is nevertheless crucial that we learn more about L2 category formation, which will necessarily play a central role in any acquisition model of L2 speech learning.

To underscore this point – and in conclusion – let’s return to the slide I showed at the beginning of this talk.
Summary & conclusions

Most late learners examined by Flege et al. (1998) produced English /t/ with VOT values that were intermediate to the phonetic norms of Spanish and English.

The two subgroups you see here produced /t/ with differing VOT values but did not differ in age of L2 learning, years of L2 use or self-reported % use of English.
Summary & conclusions

The SLM framework might be used to explain a data pattern like this.

The «lower VOT» group had not established a phonetic category for English /t/;

Their production in English was based on the distribution of VOT values to which they have been exposed in both Spanish and English. The SLM predicts that their Spanish VOT will lengthen as a result of the English input they have received.

The «higher VOT» group, on the other hand, had established a phonetic category for English /t/. Their production in English is based on the distribution of VOT values to which they have been exposed just in English, including both native- and foreign-accented input. The SLM predicts that their Spanish VOT will shorten as a result of the English input they have received.
Summary & conclusions

How might this explanation be tested? There is no time here for details. However, I will outline an approach that might be considered in future research.

1. Identify a group of native speakers of Spanish living in the US who have characteristics that are similar to those of the 35 Ss seen here. From these new Ss obtain two production samples, one at the beginning of Day 1, the other after other portions of the protocol have been completed on Day 2.

2. Administer tests of category formation to determine if the Ss have or have not established a new phonetic category for English /t/.*

*Two possible tests have been mentioned in today’s talk; a test evaluating rate-dependent processing of VOT in continua simulating a fast vs. slow rate of speech; and a task in which Ss imitate the randomly presented members of VOT continua. Other possibilities exist.
3. Determine not only how much English input the Ss had received, but also what kind. As described in a 2009 chapter entitled *Give input a chance!* (pp. 188-189), I suggested that the best way to obtain measures of L2 use is to create and install on Ss’ cell phones an application that can be used to administer a variant of the Experience Sampling Method.

4. Administer tests of individual differences.*

If this research is extended to include early as well as late learners, and if the measure(s) of individual differences is found to account for a significant amount of variance in the performance of late but not early learners, it would support the existence of a «critical period» for L2 speech learning.

*MacKay et al. (2001) found that differences in Phonological Short Term Memory predicted the identification of word-final English consonants by long-time Italian residents of Canada after variables such as age of learning and L2 use had been partialed out. Other measures of individual differences are of course available and should be evaluated.
Summary & conclusions

But if, on the other hand, the results show that the inter-subject variability can best be accounted for by knowing

1. **How much and what kind** of input the Ss have received, and

2. **Where** they stored it – in a composite Spanish-English category or in a new English phonetic category …

Then the SLM would be supported.
I hope that some of you here today will be inspired to continue developing an acquisition model for L2 speech. Please keep me informed!

And if you find yourself in central Italy, please consider stopping by for a visit. You’ll likely find me working in my new field of research in Tuscania.

This talk, along with references, will be posted at on www.jimflege.com