Origins and development of the Speech Learning Model*

James E. Flege
Univ. of Alabama at Birmingham

Keynote lecture presented at the
1st ASA Workshop on L2 Speech Learning
Simon Fraser Univ., Vancouver, BC
April 14-15, 2005

*In May 2015 I corrected errors in a previously published version of this talk and added some of the commentary that I presented vocally when I gave the talk in 2005 in order to make it more comprehensible to those seeing it now. JEF
Outline

1. Purpose of the Speech Learning Model (SLM)
2. Historical background
3. Core aspects of the SLM
4. Some predictions generated by the SLM
5. How to falsify the SLM
6. What is most needed now?
Aim of the SLM

The primary aim of the SLM is to account for variation in the extent to which individuals learn – or fail to learn – to accurately produce and perceive phonetic segments (i.e., vowels & consonants) in a second language (L2)
Aim of the SLM

From the very beginning it was apparent that the age at which L2 learning begins (the “age of L2 learning”) is an important factor.

Much of our work focused on immigrants. For such participants, the age of L2 learning is indexed by participants’ (Ss) age of arrival (AOA) in the host country. In such contexts, learning the target L2 is necessary for every day use. Most Ss begin receiving meaningful L2 input almost immediately. (Another index of the “age of L2 learning” is needed for non-immigrant populations, or for learning that occurs in other contexts.)

Soon thereafter we came to understand that the amount and kind of input received by L2 learners was very important and, alas, confounded with the age factor
SLM Purpose

Our research has focused on questions such as:

• Is it impossible for learners of an L2 to produce certain L2 speech sounds accurately?
• Are there “un-learnable” L2 sounds? If so, just for persons who begin learning the L2 after a certain age (or, more accurately, stage of L1 development)?
• How is the perception & production of L2 phonetic related? Are the links between the two the same as those which exist as the L1 develops in childhood and into adolescence?
Outline

1. Purpose of the Speech Learning Model (SLM)
2. Historical background
3. Core aspects of the SLM
4. Some predictions generated by the SLM
5. How to falsify the SLM
6. What is most needed now?
Zeitgeist: 1979-1984

It is useful to consider the state of research in the first 5 years of development of what later became known as the Speech Learning Model, or SLM. Some topics to be considered today are:

- Critical Period Hypothesis (CPH)
- Contrastive Analysis Hypothesis (CAH)
- Focus on abstract linguistic units, not phonetic substance
- Categorical Perception/filtering via the phonological "grid"
- Unidirectional L1 → L2 interference
Critical Period Hypothesis (CPH)

• Eric Lenneberg (1967) proposed that as humans mature neurologically, their ability to learn the L1 diminishes rather drastically at a certain point.
• Lenneberg casually observed that it is difficult to learn to pronounce an L2 without a foreign accent after onset of puberty (i.e., after the age of 12 years).
• This casual remark profoundly altered L2 research, for it triggered extension of CPH from L1, Lenneberg’s focus, to L2 learning.
Doubts regarding the CPH

• It is certainly true that “earlier is better” as far as learning the sound system of an L2

• However, in the early 1980s no one had produced objective evidence showing:
  1. A sharp drop in learning success when the learning of an L2 began after a particular chronological age (e.g., 12 years), neuro-endocrinological status (e.g., “puberty”), or state of neurological development
  2. Rapid and perfect success in L2 learning for all children
  3. Failure to learn an L2 by all adults
Doubts regarding the CPH

Subsequent empirical research failed to sustain certain predictions generated by the CPH

We will now consider two companion studies that examined relatively large (n = 240) groups of immigrants to North America
Doubts regarding the CPH

In both studies:

• All participants (Ss) were adults at the time of testing
• All were selected on the basis of their age of arrival (AOA) in a country where an L2 had to be learned for everyday uses
• The Ss repeated English sentences after a filled delay.
• The sentences were digitally prepared and then later rated for overall degree of perceived foreign accent (FA) by native English-speaking listeners drawn from the same community where the Ss and members of a native English control group resided
The Ss were 240 Italian who had immigrated to Canada at various ages (AOAs) and had lived there for decades.

FA was rated using a continuous scale by native English-speaking listeners. The ratings obtained for member of the native English control group are shown with unfilled circles.
Flege et al. (1999) examined English sentences spoken by native Korean immigrants to the US. The earlier continuous scale was replaced by a 9-point EAI rating scale.

The results closely replicated those obtained earlier with Italian immigrants despite differences in L1, host country, speech materials and scaling technique.
Index of “accent free”

• As mentioned, Lenneberg cited the presence of foreign accent (FA) as evidence that L2 learning might be limited by a critical period.
• We sought to determine how many of our Italian and Korean Ss managed to produce the English test sentences without an obvious FA.
• To address this question we determined if the sentences spoken by each non-native participant received a mean rating that fell within 2 SDs of the mean rating obtained for the 24 Ss in the native English control group.
Post-hoc analysis of FA ratings in two studies (Flege et al., 1995, 1999)

The percentage (%) of Ss in each AOA-defined subroups (n = 24 each) whose FA ratings fell within 2 SDs of the mean rating obtained for a control group n = 24 of native English speakers.

Non-native Ss meeting this lax criterion were deemed «accent free»
Few Ss who began learning the L2 after the age of 12 years were «accent free» even after decades of immersion. This finding supports the CPH. However, less than half of the Ss who began learning L2 prior to the supposed end of a CP for L2 learning did not meet the lax criterion for «accent free». This finding diverges from expectations generated by the CPH.
Doubts regarding the CPH

• The two studies just cited made use of a “retrospective developmental design”, examining adult Ss who differed according to their chronological age at the time they had immigrated and began learning their L2.

• Would foreign accent (FA) be found for children who were currently in the process of learning an L2?

• Flege et al. (in press) did, in fact obtain such evidence
Doubts regarding the CPH

Flege et al. (in press).

The Ss were native Korean children who had been living in North America for either 3.5 or 5.5 years as well as age-matched native English speaking children who were born and raised in an English-speaking community. FA in English sentences was rated using a 9-point scale. All Ss were tested twice.
Doubts regarding the CPH

Flege et al. (in press) found that Korean children produce English sentences with a detectable foreign accent (FA)

Analysis of subgroups

• 10-year-old Koreans who had arrived in North America four years earlier, that is, at the chronological age of 6 years, received significantly lower ratings than did 10-year-old native English children
Conclusions regarding the CPH

• The presence of FA in children is not something that is likely to be observed casually. The FA measure used here was fairly fine-grained.

• This and similar findings convinced us that many, perhaps even most children who learn an L2 will speak it with a detectable foreign accent, even following years of immersion.

• This finding is not something one would expect if the presence of a FA were the result of having passed a “critical period”.

• The presence of FA in many adult L2 learners of L2 is of course consistent with CPH. However, work by Bongaerts et al. has demonstrated that some adult L2 learners manage to speak their L2 without FA.
Contrastive Analysis (CA) hypothesis

• In 1979-1984, most L2 research was framed in terms of the CAH which posited that
  - L2 phonemes that are similar to L1 phonemes will be “easy” to produce;
  - L2 phonemes that are different from L1 phonemes will be “hard” to produce.

• Interference was seen as the major cause of most learning problems: what you already know in L1 will sometimes help you but just as likely it will hurt you because, after all, the L2 differs from the L1.
Doubts regarding the CAH

• Flege (1987) examined native English speakers’ production of the French vowels /y/ and /u/
• Both French vowels are likely to be heard (classified) as English /u/ (work by Bernie Rochet) even though
  • English /u/ differs from its French counterpart, being fronted in vowels space (i.e., having higher F2 values)
  • French /y/ has been described as being “radically different” from any English vowel, and so might be treated as a “new” vowel (see Delattre, 1964, p. 83)
Comparison of acoustic values for two French vowels (red symbols) and six English vowels.

The mean F1 and F2 values (in Hz) of French vowels were drawn from a study by Delattre and by Debrock & Forrez.

The values for English vowels were drawn from a study by Hillenbrand et al. (1995)
Doubts regarding the CAH

Flege (1987) tested 3 groups of native English speakers, all women:

**Group B**: American college students who had just returned to Chicago after a 9-month academic program in Paris, France.

**Group C**: Somewhat older native speakers of American English. All had obtained advanced degrees in French, taught French at American university and had lived in France for at least a brief period.

**Group D**: Americans who lived in Paris for $M = 10$ years
Doubts regarding the CAH

The remaining two groups tested by Flege (1987) consisted of native French-speaking women

- Group B:
- Group C:
- Group D:
- Group E: All were living in Chicago when tested and had lived there for an average of 10 years
- Group F: French monolinguals living in Paris, France
Mean F2 values reported by Flege (1987) for the high vowels /y/ (white) and /u/ (black) as produced by five groups. The French monolinguals produced the largest F2 differences between /y/ (a front rounded vowel) and /u/ (a back rounded vowel).
Doubts regarding the CAH

• All three groups of native English-speaking women (B, C & D) in the Flege (1987) study produced French /u/ with significantly higher F2 values than did the French monolinguals. This means that they were producing “fronter” variants of the French vowel, presumably as the result of the influence of the more fronted American English vowel /u/.

• However, none of the three native English groups differed significantly from the French monolinguals when producing the “new” French vowel, /y/.

• Possible interpretation: they could produce French /y/ accurately because there was no interference from a vowel in their L1 phonetic inventory.
Conclusions regarding the CAH

• Comparisons of acoustic values suggested that French /y/ may be more dissimilar from the closest English vowel than is French /u/. This needs to be verified in a formal perceptual test.

• The Flege (1987) results suggest that adult learners of an L2 may be more successful at producing a “new” vowel in the L2 than an L2 vowel that resembles a vowel already found in the L1

• This is the opposite of what one might expect from the CAH which posits that “similar” is easy to learn whereas “new” will be difficult
Abstract linguistic analyses

- In the period 1979-1984 it was widely believed that the phonologies of an L1 and an L2 come into contact at an abstract phonemic level.
- On this view, learners perceive the sounds of an L2 through the grid of their existing L1 “phonology” (see work by Trubetzkoy).
- As a result, learners perceive (hear) and produce (articulate) L2 words as if they were concatenations of L1 phonemes. In other words: “new wine in old bottles” This is admittedly poetic, but is it true?
Abstract linguistic analyses

• In this analysis, phonemes are viewed as a set of freely commutable elements that can be arranged to construct large lexicons.

• The phonemes themselves are regarded as being formed by bundles of commutable distinctive features (only some of which are associated with specific acoustic and articulatory dimensions)

• The prevailing view was that learners of an L2 cannot use a feature exploited by an L2 if it were not already deployed in the L1 to contrast meaning
Doubts regarding abstract analyses

- Flege & Port (1981) evaluated these assumption of generative phonology by examining the production of English /p/ by native speakers of Saudi Arabian Arabic.
- All were young men who had come to the U.S. on scholarship to study at Indiana University.
Doubts regarding abstract analyses

- Flege & Port (1981) selected Arabic as the target L2 in this study because the Arabic phonemic inventory has /b/, /d/, /t/ and /k/, but no /p/ or /g/
- Arabic necessarily has a [voicing] feature (for the /d/-/t/ contrast) and [place] feature (for the /b/-/d/ contrast)
- If learning to produce L2 sounds occurs at a phonemic level then the Saudi Ss should have been able to learn to produce English /p/ by recombining the Arabic [voicing] and [place] features
Doubts regarding abstract analyses

• Acoustic phonetic measurements by Flege & Port (1981) indicated that the native Arabic participants produced with English /p/ with temporal properties appropriate for a bilabial stop, but with the glottal pulsing that is characteristic of a phonologically voiced stop.

• Not surprisingly, the Saudis’ English /p/ productions were often heard as /b/ by native English-speaking listeners.

• Conclusion: The Saudi Ss did not re-combine abstract features. Their difficulty is best described as learning to produce a new phonetic segment, not a new phoneme.
Doubts regarding abstract analyses

- McAllister, Flege & Piske (2002) evaluated L2 learners’ ability to learn to use a new distinctive feature: the use of a phonemic [length] feature that is needed to produce and perceive vowel distinctions in Swedish.

- Two non-native groups consisted of speakers of English and Spanish who had lived for more than 10 years in Stockholm, Sweden.

- A phonemic [length] feature is not used to distinguish vowels in either English or Spanish. These nonnative Ss (but not the Estonian controls) had to learn it in Swedish if they were to produce and perceive Swedish vowels adequately.
Doubts regarding abstract analyses

McAllister et al. (2002) examined four Swedish long-short vowel pairs that differed according to the [length] feature:

- Two pairs of mid-vowel contrasts. The vowels in these pairs are relatively similar in vowel quality (formant frequencies) but differ substantially in duration, which is measured in msec at a phonetic level of analysis.
- The other two pairs consisted of high or low vowels. These non-mid vowel contrasts are based on both duration and spectral quality.
Doubts regarding abstract analyses

McAllister et al. (2002) recruited four groups of 20 Ss who were native speakers of:

- Swedish,
- English
- Spanish
- Estonian

The Estonians were selected as a control group because Estonian has vowel contrasts based on [length]. If the Estonian Ss had difficulty producing or perceiving Swedish vowels distinguished by [length] it could not be attributed to the inability to learn a new abstract feature
Doubts regarding abstract analyses

The words used as stimuli in the perceptual experiment carried out by McAllister et al. (2002) were all highly frequent words known to the non-native Ss.

The words were recorded by an adult male native speaker of Swedish. Half of the stimulus words contained a phonologically long vowel, the other half of the words had a phonemically short vowel.
Doubts regarding abstract analyses

McAllister et al. (2002) made copies of all words. The copies were then altered digitally

• Long vowels in the copies were shortened, making possible but non-occurring Swedish words
• Short vowels in the copies were lengthened to make long vowel, again creating non-words
Doubts regarding abstract analyses

• The task of participants in the McAllister et al. (2002) perception experiment was to indicate if each stimulus had been produced “correctly”. In effect, the task was a 2-alternate non-word recognition task.

• The task was very easy for native speakers of Swedish.

• An analogous task in English would be to ask if /fiʃ/ is a correctly produced English word. In this illustrative example, the English vowel /i/ substitutes the vowel /ɪ/, making a non-word, and so the correct response would be “incorrect” (that is to say, not a word).
A Swedish long-short contrast: /ø/ vs /øː/

Mean % correct scores. The Swedish and Estonian Ss obtained high scores because they could determine whether familiar words contained a phonemically long or short vowel. The English Ss obtained lower scores. Some Spanish Ss performed at or below chance indicating that they didn’t know if words in their Swedish lexicon had a long or a short vowel; these Ss had clearly not learned to use a new L2 feature.
McAllister et al. (2002) carried out separate analyses for:

- Two mid-vowel pairs in which the contrast was based almost entirely on duration. The opposing members of both pairs of vowels differed little in terms of spectral quality.
- Two non mid-vowel pairs (1 high, 1 low). In these pairs the contrast was based on both duration and spectral quality differences.
McAllister et al. (2002) expected better performance on the non-mid (high- or low) vowel pairs than for the two mid-vowel pairs.

If someone learning Swedish as an L2 learner were oblivious to the duration differences distinguishing the non-mid vowel pairs they could always rely on the spectral differences that accompanied the duration differences.
Doubts regarding abstract analyses

The native English and Spanish Ss obtained higher scores for non-mid than for mid-vowel pairs because for non-mid pairs they could use of both temporal and spectral cues whereas only temporal cues were available for the mid-vowel pairs. No such difference was observed for the Ss whose L1 makes use of a [length] feature.
Doubts regarding abstract analyses

Some conclusions from McAllister et al. (2002)

1. Estonians re-used the [length] feature of their L2 when learning Swedish L2.

2. English & Spanish Ss continued to rely on spectral cues to perceptually distinguish long vs. short Swedish vowels. Some showed evidence of little or no sensitivity to the [length] feature.

3. It is difficult for many adult learners of an L2 to acquire sensitivity to an acoustic phonetic dimension – duration – that is not used, or else is used differently, in the L1.
Doubts regarding abstract analyses

However

• Notable differences existed between languages (Spanish vs. English) and between individual Ss within a language
• Most native English Ss and some native Spanish Ss showed significantly above-chance performance, which indicates at least some sensitivity to the new feature [length] that presumably did not exist before exposure to Swedish. (More research needed!)
• Some native English Ss showed native-like performance, indicating they had learned a new L2 feature. What is the source of these individual differences?
Doubts regarding abstract analyses

Now let’s consider the results of three studies that examined production of the rhotic English vowel /ə-/ (as in *bird, heard*) by native speakers of Italian:

Munro, Flege & MacKay (1996)
Doubts regarding abstract analyses

- English /ə-/ differs from any vowel found in the Italian vowel inventory
- When Italians mimic an American accent in Italy they rhotacize vowels in an exaggerated manner, indicating they are aware of this acoustic phonetic property and can, under the right circumstances, produce it
- The perceptual dissimilarity of English /ə-/ from any Italian vowel was demonstrated in a perceptual assimilation experiment carried out by Flege & MacKay (2004)
- Acoustic differences between /ə-/ and Italian vowels are shown in the next slide
Acoustically, the /ə/ of English differs substantially from any vowel in the Italian inventory, both in terms of (a) F1 and F2 formant frequencies and (b) the frequency of the third formant, F3. One might infer that the rhotic vowel will be treated as «new» but perceptual data is needed to confirm this impression. If so, it will be necessary to determine if, over time, Italian learners of English begin producing /ə/ accurately.
Doubts regarding abstract analyses

- Munro, Flege & McKay (1996) examined 240 native speakers of Italian who had immigrated to Canada. The participants differed, among according to
  - Their original age of arrival (AOA) in Canada (an independent variable)
  - Years of residence in Canada, abbreviated LOR (which was moderately correlated with AOA)
- Production samples were obtained by having the Ss repeat English words following a filled delay (delayed repetition task)
- Vowels production accuracy was asssed by having native English listeners rate the vowels (5-pt scale)
Munro et al. (1996). Mean ratings /ə/ production accuracy by native English-speaking listeners. The talkers were groups of native Italian Ss selected on the basis of age of arrival (AOA) in Canada as well as a native English control group.
Munro et al. (1996). Number of native Italian Ss per group (n = 24) whose /ə-/ productions received a rating within 2 SDs of the mean rating obtained for production of this vowel by native English speakers.

Most Italian Ss who arrived in Canada before the age of 12 years produced /ə-/ well, but few of those who arrived later in life did so. Why? A limitation on production? Or on perception?
Doubts regarding abstract analyses

Flege, Schirru & MacKay (2003) examined production of /ɚ/ by 5 groups of 18 participants each:

- Native speakers of English (NE)
- Two groups of “early learners” who arrived in Canada from Italy as children but differed according to average self-reported use of their L1, Italian use (means = 7% vs. 43%)
- Two groups of “late learners” who arrived in Canada later in life, subdivided according to self-reported use of Italian (means = 10% vs. 53%)
NE-listeners used 4 labels to classify English vowels, presented in separate blocks. Tokens were considered “accurate” if classified “good” or “acceptable”. Non-parametric tests evaluated the number of Ss in each group (max = 18) whose vowels were produced accurately. Results shown here only for /ə/
Flege et al. (2003) results

• The non-parametric analyses revealed that native speakers of English produced /ɚ/ more accurately than did both groups of late learners (Late-low Italian use of Italian, Late-high Italian use) but did not differ significantly from either group of early learners.

• Suggests a greater difficulty learning a new “feature” ([rhotic]) as age of first exposure to an L2 increases.
Production of English /ə/ by native speakers of Italian

Flege et al. (2003) results

• However, detailed acoustic analyses suggested that the Italian late learners of English do acquire sensitivity to the [rhotic] feature.

• The analysis examined Bark-transformed F3-F2 differences. This derived acoustic measures provides a perceptually relevant index of the [rhotic] dimension in speech production.
Flege et al. (2003). The early learners were found to have produced significantly smaller F3-F2 differences in the /ɚ/ tokens than did the late native Italian learners of English. In other words, the early learners were more successful than the late learners in producing the “rhotic” feature (property).
Production of English /ə/ by native speakers of Italian

- The Late bilinguals produced significantly larger F3-F2 differences than did the native English speakers.
- However, some late bilinguals seem to have acquired sensitivity to [rhotic] feature.
- In fact, as can be inferred from the box and whiskers graphic, a few produced /ə/ with native-like F3-F2 values. (What accounts for this inter-subject variability!?)
- Regression analyses indicated that, for all 11 English vowels examined, the more the native Italian Ss continued to use their L1, the less accurate produced English vowels. Does the inter-subject variability depend on amount of continued interference from the L1? On amount of L2 input?
Conclusions regarding abstract analyses

• Both early and late learners are able to gain access to features (acoustic phonetic dimensions not used to contrast L1 phonemes.

• As a group, Late learners are less likely to do so than early learners. However, there tends to be a lot of inter-subject variability in groups of late learners, and so success cannot be ruled out solely as a function of age of first exposure to an L2.

• Amount of continued L1 use was found to predict success in learning to produce L2 vowels. It will be important in future research to understand better the sources of variation in L2 learning, especially among late learners.
Categorical perception (CP)

• The CP paradigm was applied to most cross-language perception in studies carried out in the period 1979-1984

• The CP paradigm was motivated by the consistent finding that discrimination is more accurate for pairs of stimuli straddling a “phoneme boundary” (identified as belong to two categories) than for stimuli identified as instances of a single category

• For example; English voiceless stop tokens having VOT values of 25 and 65 msec will not be discriminated if both are identified as the phoneme /t/

• The audible acoustic differences between two such tokens were said to be “filtered out”
Categorical perception (CP)

• The CP paradigm was applied to most cross-language perception in studies in the period 1979-1984

• For example, in 1981 MacKain, Best & Strange (Appl. Psycholing. 2: 369-390) found evidence of the “categorical” perception of a synthetic /r/-/l/ continuum by five native Japanese adults had on average 2.3 years of “conversational” experience in English
The “doom” scenario

- My reading of the literature from this period, which had already made great advances compared to the impressionistic analyses of years past, suggested that there was little hope that adults could learn the fine-grained phonetic features of an L2, much less create new phonetic categories for sounds in the L2 that differed sufficiently from sounds in the L1 inventory.

- I call this the “doom” scenario
The “doom” scenario

1. The sound systems of two languages differ in terms of number and kinds of phonemic categories, whose physical phonetic realizations may also differ.

2. Children learn to realize (produce) speech sounds in a native-like way as they begin to note and organize the sensory properties of sounds they encounter in their linguistic environment.

3. Production and perception become “aligned” during the course of L1 acquisition. In normal L1 development, perceptual development normally “leads” the fine tuning of patterns of speech articulation.
The “doom” scenario

4. The CP paradigm suggests that within-category phonetic variation will be discarded.

5. During L2 acquisition, the learner may encounter L2 sounds that differ phonetically from the closest sound in the L1 inventory.

6. If the L2 sounds as classified implicitly by the learner as belonging to the most similar phoneme of the L1, then the CP paradigm leads us to expect that even audible L1-L2 phonetic differences will be discarded.
The “doom” scenario

7. If the L2 sounds as classified implicitly by the learner as belonging to the most similar phoneme of the L1, then the CP paradigm leads us to expect that even audible L1-L2 phonetic differences will be discarded.

8. If L1-L2 phonetic differences are discarded (“filtered out”), phonetic learning will not be possible.

9. It will also be impossible to create new phonetic categories.
The “doom” scenario

• Fortunately, there is no reason to be so “gloomy and doomy”

• Abundant evidence existed even in 1979-1984 for those who sought it that within-category phonetic information is available to listeners “although the retrieval of this information … will depend on the level of processing” (Pisoni & Tash, 1974)
No doom, no gloom

Let’s consider the results of two studies which suggest that applying the “phonological grid” of the L1 to sounds encountered on the phonetic surface of an L2 via categorical perception does not cause L2 learners to filter out audible (at a sensory level) cross-language phonetic differences

- Flege (1984)
- Flege & Hammond (1982)
Flege (1984) examined speech samples produced in English by adult native speakers of English and French. The stimuli consisted of:

- Unmodified tokens of the syllable /tu/ edited from phrases such as *two little girls*
- Hybrid /tu/ tokens created by splicing instances of /t/ or /u/ edited out of original syllables and then cross-spliced. One segment of the hybrid syllables was a single segment produced a NE speaker, the other segment varied
- The final set of stimuli consisted of the first 30-ms of /t/, essentially a release burst plus a bit of aspiration
No doom, no gloom

• The English stimuli presented in pairs, of which one member was produced by a native speaker of English and the other member consisted of at least a segment produced by a *native speaker of French*

• The listeners’ task on each trial was to decide which of the two stimuli was “foreign”

• The listeners were given no training or feedback on the task.

• The task was un-speeded; however, a response was required on every trial before listeners could move on to the next trial.
No doom, no gloom

Flege (1984). % identification of stimuli as the “foreign” member of a pair of stimuli.

The NE listeners usually identified nonnative /tu/ productions as “foreign” and seldom identified native-English produced tokens as such.

Accuracy decreased but remained well above change when just one segment, or part of a segment, was produced by a non-native speaker.
The foreign accent findings of Flege (1984) demonstrated that NE listeners could detect specific cross-language phonetic differences which represented, in English, “within-category” variation:

- Production of /u/ as a “backer” (non-front) vowel than is typical for English, thereby resembling French /u/
- Production of /t/ with VOT values that were too short (resembling values typical for French /t/)
- A tendency to realize /t/ with a dental rather than alveolar place of articulation
No gloom, no doom

• The differences between aspects of phonetic implementation in the English spoken by monolinguals and French-accented English are often smaller than the phonetic differences between French and English.
• Flege (1984) reasoned that if native English monolinguals could detect foreign accent on the basis of small phonetic differences, that such differences were audible.
• This indicates that applying the CP paradigm to L2 learning is inappropriate.
• Next we ask: is this kind of performance seen in Flege (1984) possible outside the laboratory where listeners must pay attention to meaning?
No gloom, no doom

• Flege & Hammond (1982) tested 50 native English-speaking students who were enrolled in first-year Spanish classes at the University of Florida (Gainesville)

• All participants were familiar with Spanish-accented English having been born in raised in Florida and being enrolled in a Spanish class that was being taught – mostly in English! – by native Spanish instructors who spoke English with fairly strong Spanish accents
The Ss tested by Flege & Hammond (1982) inserted a set of English words into the carrier phrase (The__is on the__). The Ss were asked to produce the sentences, recorded for later analysis, with a “Spanish accent”. No coaching, explanation or training was provided on the foreign accent imitation task.

<table>
<thead>
<tr>
<th>Lexical Items</th>
<th>Substitute</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>nose, cheese, hose</td>
<td>s/z</td>
<td>141 (47%)</td>
</tr>
<tr>
<td>vice, veil, vase</td>
<td>b/v</td>
<td>129 (43%)</td>
</tr>
<tr>
<td>fig, pig, wig</td>
<td>i/I</td>
<td>127 (42%)</td>
</tr>
<tr>
<td>book, hook, crook</td>
<td>u/U</td>
<td>61 (20%)</td>
</tr>
<tr>
<td>shell, sheet, sheep</td>
<td>č/š</td>
<td>49 (16%)</td>
</tr>
<tr>
<td>bean, phone, bone</td>
<td>η/n</td>
<td>1 (0.3%)</td>
</tr>
<tr>
<td>tape, tube, toad</td>
<td>d/t</td>
<td>0 (0.0%)</td>
</tr>
</tbody>
</table>
No gloom, no doom

• Flege & Hammon (1982) later noted all segmental substitution in the target words.
• The Ss were subdivided into groups according to how many substitutions typical for Spanish-accented English they produced (“most” vs. “least”)
• The authors reasoned that the “most” group (Ss who produced many expected substitution) had more experience with Spanish accented English than the “least” group (Ss who produced fewer expected substitutions)
No gloom, no doom

• Flege & Hammond (1982) also measured VOT in 6 /t/ tokens for each participant.

• This acoustic phonetic dimension was of interest because
  • Native speakers of Spanish often produce English /t/ with VOT values that are intermediate to the values produced by Spanish monolinguals (short-lag) and by English monolinguals (long-lag)
  • If VOT is shortened sufficiently in /t/ it may be heard as /d/
No gloom, no doom

• Flege & Hammond (1982) found that the Ss imitating a Spanish accent in English were never heard, when transcribed by 2 phoneticians, to have substituted an English /d/ for the [t] of Spanish-accented English (SAE)

• This could mean either that participants in the foreign accent imitation task (a) were unable to detect the shortening of VOT that is typical for SAE, or (b) were unable themselves to shorten VOT in their imitations

• The VOT measurements revealed that interpretation “b” was correct
No doom, no gloom

Flege & Hammond (1982). Mean VOT (msec) in English words produced by two experimental groups and one control group (n = 10 each)

Group A1 produced the largest number of expected segmental substitutions
Group A2 produced by fewest segmental substitution
U – Ss in the control group simply read the speech materials without trying to produce a Spanish foreign accent
No gloom, no doom

Flege & Hammond (1982). The distribution of VOT values (msec) in keywords produced by two groups of native English students asked to imitate a Spanish foreign accent in English and by members of a control group who produced the same speech materials normally, i.e., without trying to imitate a Spanish accent.

![Graph showing VOT values](image)

**Figure 1.** Frequency of VOT values measured in stops produced by three groups of speakers.
No gloom, no doom

- Examination of the frequency histogram suggests that at least the Ss in Group A1 had detected the shortened VOT values typical of SAE and were able to reproduce it in their imitations of SAE.
- The Ss’ knowledge of SAE developed through direct exposure to English spoken with a Spanish accent. The Ss stored within-category phonetic information in long term memory.
- The same capabilities are likely to be available to all young adults who set out to learn an L2
No gloom, no doom

In conclusion

• Applying the categorical perception (CP) paradigm or the “phonological grid filter” paradigm to L2 speech learning does not seem to be well founded.

• It appears that adults who are exposed to a foreign or second language can (eventually) detect within-category cross-language phonetic differences and store this information in long-term memory representations.
Unidirectional L1→L2 interference

• When I was writing my PhD dissertation at Indiana University in 1979, students from W. Africa told me something important.
• They reported that when they returned home, sometimes after several years of uninterrupted residence in the US, their family & friends made fun of them for “affecting” an American accent in their L1.
• I inferred that learning an L2 (in this case English) might have influence their production of the L1.
• In the period 1979-1984 “interference” of the L1 on the L2 was well documented but there was no interest or discussion of L2→L1 interference.
Unidirectional L1→L2 interference

This lack of attention to something later recognized as “obvious” was probably the result of a number of assumptions held by most investigators prior to 1979:

1. Bilinguals “switch” between separate self-contained L1 & L2 phonological systems. When the L1 is “on” the L2 must necessarily be “off” and so cannot influence the L1.
2. Errors in an L2 occur because it has not been properly or full learned, something not possible for an L1.
3. What is learned early on in the L1 “stays learned” (see work by Roman Jakobson)
Doubts regarding unidirectional interference

Now let’s review the results obtained in two studies that provide evidence of “L2 effects on L1”, what we might call “reverse interference”. These studies are:

- Yeni-Komshian et al. (2000)
- Yeni-Komshian & Flege (unpubl.)
Doubts regarding unidirectional interference

Yeni-Komshian et al. (2000) tested 240 Korean adults living in United States

• The native Korean Ss, all long-term residents of the U.S., were selected on the basis of their age of arrival in the U.S.

• English and Korean monolinguals produced sentences in those languages; the 240 bilinguals produced sentences in both languages

• The English and Korean sentences were rated for overall degree of foreign accent by English & Korean monolinguals, respectively
Doubts regarding unidirectional interference

Mean foreign accent ratings obtained for Korean and English sentences by Yeni-Komshian et al. (2000)

Isolated symbols indicate the mean ratings obtained for sentences produced by English & Korean monolinguals

The Korean adults who arrived in the US prior to the age of 8 years produced Korean sentences with what seems to have been an American foreign accent
Doubts regarding unidirectional interference

- Yeni-Komshian & Flege (unpubl.) elicited isolated Korean words beginning with the consonants /s/, /s’/, /th/, /t’/

- After being digitally prepared, the stimuli produced by 240 Koreans living in the U.S. were randomly presented in separate blocks to native Korean-speaking listeners.

- Production of the word-initial consonants was judged to be:
  - 4 very good
  - 3 okay
  - 2 distorted
  - 1 wrong consonant
Doubts regarding unidirectional interference

Results obtained by Yeni-Komshian & Flege (unpublished).

The brackets enclosed +/- 1 S.E.

Korean consonants produced by nearly all of the 240 Korean adults living in the U.S. received lower ratings than did consonants produced by Korean monolinguals in Korea.

Only consonants produced by Korean adults who had arrived in the U.S. prior the age of 8 years were judged, on average, to be less than adequate.
Doubts regarding unidirectional interference

• Flege & MacKay (unpublished) examined the production of Italian words spoken by 80 native speakers of Italian who were long-time residents of Canada and by monolingual native speakers of Italian recorded in Padova, Italy.

• The voiced stops /b d g/ can be realized with pre-voicing in English but are usually realized with short-lag VOT values (no voicing in the closure).

• In Italian, on the other hand, /b d g/ are realized with lead VOT values, that is, with pre-voicing (glottal pulsing) during the period of closure, before release.
Results obtained by Flege & McKay (unpublished). The brackets enclosed +/− 1 S.E.

Nearly all of the 80 Italians living in Canada realized Italian /b d g/ tokens as short-lag stops at least some of the time, whereas this was observed seldom in the speech of Italians living in Italy.

The earlier in life the Italian-English bilinguals had immigrated to Canada the more often they produced Italian /b d g/ in an English-like fashion, that is, as short-lag stops.
Conclusions so far

The picture of L2 speech learning that began to emerge differed substantially from what was generally assumed and/or believed in the period 1979-1984

• We found no evidence for a sharp drop in L2 speech learning ability at the age of 12 years, and evidence that “pre-critical period” learns speak the L2 with a FA;

• Cross-language phonetic differences are not filtered out, and so might trigger modifications and/or additions to the learner’s phonetic repertoire

• The L1 and L2 sub-systems do not exist in splendid isolation

• Learning an L2 may affect how the L1 is produced, especially for early leaners
Outline

1. Purpose of the Speech Learning Model (SLM)
2. Historical background
3. **Core aspects of the SLM**
4. Some predictions generated by the SLM
5. How to falsify the SLM
6. Future directions
The Speech Learning Model (SLM)

The SLM was developed to make sense of the empirical results we had begun to obtain. The model rests on several basic premises:

1. L2 learners can, given adequate and sufficient input, perceive the phonetic properties of L2 speech sounds accurately
2. As in L1 development, L2 speech learning (a) takes time, and (b) is influenced importantly by the nature of input received
3. As in L1 development, production is guided by perceptual representations stored in long-term memory
The Speech Learning Model (SLM)

Further, the SLM proposes that

4. The processes and mechanisms that guide successful L1 speech acquisition—including the ability to form new phonetic categories—remain intact and accessible across the life span

5. The phonetic elements that make up the L1 and L2 phonetic subsystems exist in a “common phonological space”, and so mutually influence one another
SLM hypotheses

- The greater the perceived dissimilarity of an L2 sound from the closest sound of the L1, the more likely a new category will be formed for the L2 sound.
- Category formation for an L2 sound becomes less likely through childhood as representations for neighboring L1 sounds develop.
- When a category is not formed for an L2 sound because it is too similar to an L1 counterpart, the L1 and L2 categories will assimilate, leading to a “merged” L1-L2.

(illustrations/examples to follow)
SLM hypotheses

• By way of illustrating the hypotheses just stated, let’s imagine the vowel spaces of an L1 and L2
• Let’s also make some simplifying assumptions to facilitate the discussion
Let’s imagine that there are 5 vowels in the L1, depicted here by ellipses in a 2 dimensional high-low vs. front-back vowel space.

Our imaginary language is similar to real languages such as Spanish.
Let’s suppose that the L2 has 7 vowels and that perception of vowels of the L2, like those of the L1, are based entirely on center formant frequency values (no role of either duration or formant movement patterns).

Here we see varying degrees of overlap in the acoustically defined vowel space between 5 L2 vowels and the 5 vowels of the L1. Two L2 vowels occupy space not exploited in the L1.
Many researchers would immediately conclude that the two non-overlapping L2 vowels will be treated as “new”

But wait! In the period 1984-1993 the SLM proposed that vowels in an L2 could be classified as identical, similar, or new. This tri-partite division was abandoned for several reasons in 1994, over a decade ago.

Whether L2 learners will treat a vowel in the L2 as “new” will emerge over time. This determination cannot be made by looking at plots of acoustic data.
The SLM regards perceived cross language phonetic dissimilarity as a continuum that must be measured in a perceptual experiment. (Basic technique: have listeners rate pairs of stimuli made up of one L1 vowel token and one L2 vowel token.)

The L2 /ɒ/ would probably be at the high end of the dissimilarity continuum, and the L2 /i/ and /e/ at the low end of the continuum. However, this must be established empirically.
The SLM generates several predictions.

First, L2 vowels rated a phonetically similar to an existing L1 vowel will be produced fairly well in early stages of L2 acquisition. They are said to have gotten a “free ride”

Second, L2 vowels rated as very dissimilar from the closest L1 vowel might be produced poorly in the earliest stages of L2 learning. Perhaps they will be substituted using one or more L1 vowels that are adjacent to the L2 vowel.
However, in the “long run” (e.g., decades of predominant L2 use) such vowels should be produced more accurate than vowels that are less dissimilar from the closest L1 vowel. This is the expected outcome when new phonetic categories are established for certain vowels in the L2.

Note that without a time dimension – which might be simulated through groups differing in L2 experience – the SLM cannot generate testable predictions.
SLM hypotheses

• According to the SLM, degree of perceived cross-language phonetic dissimilarity exerts an import role in determining how successfully vowels in an L2 will eventually be produced.
• This is because perceived L1-L2 dissimilarity is the key to understand if new categories will or will not be established.
• However, the SLM posits that interactions between vowels in the combined L1-L2 vowel space also play a role.
• Once again, let’s make some simplifying assumptions to facilitate the discussion
Illustrating assimilation, dissimilation

Let's imagine a 7-vowel L1 vowel system …
Illustrating assimilation, dissimilation

… and a 10-vowel L2 vowel system (unfilled ellipses)

The SLM proposes that when learners are unable to create a new category for an L2 vowel because it is too similar to an existing L1 vowel, the two vowels will eventually form a composite — coming to resemble one another.

When categories ARE created for an L2 vowel, it and the closest L1 vowel are predicted to dissimilate in order to minimize perceptual confusions in the combined L1-L2 vowel space.

(Recall that fluent bilinguals often insert L1 materials into the L2, and vice versa.)
Observing assimilation and dissimilation processes in L2 acquisition requires a lot of data and time. These processes are probably observed best over real rather than apparent time, i.e., in longitudinal research rather than research comparing groups differing in (presumed) L2 input and use.

The figure at the right illustrates what might happen when learners of the L1 (7 vowels) learn the hypothetical L2 (10 vowels): dissimilation of three L2 vowels from neighboring L1 vowels, and assimilation of the remaining seven L1 vowels.
Outline

1. Purpose of the Speech Learning Model (SLM)
2. Historical background
3. Core aspects of the SLM
4. Testing SLM predictions
5. How to falsify the SLM
6. What is most needed now?
Testing SLM predictions

Hypothesis: When category formation does not occur – because an L2 sound differs insufficiently from the closest L1 sound, the L2 sound and the closest L1 sound will assimilate

- Productions of the L2 sound will continue to resemble the L1 sound
- Productions of the L1 sound will shift in the direction of the L2 sound
Testing SLM predictions

Flege (1987) examined the production of French and English /t/ by members of two groups
• American women who had lived in Paris for $M = 10$ years
• French women who had lived in Chicago for $M = 10$ years
Flege (1987) measured the VOT of /t/ produced in the initial position of words in English and French by two groups of bilinguals as well as the productions of French and English monolinguals.

The dashed lines indicate the mean value of stops produced by the monolinguals in English, French.

Left: mean VOT of productions of the English-French bilinguals in English (L1) and French (L2).

Right: production of the French-English bilinguals in English (L2) and French (L1).
Flege (1987) The red arrows indicate productions in the L2

(a) For the American women in Paris, French /t/ is produced with shorter VOT (English-like) values than would be typical for French

(b) For the French women in Chicago, English /t/ produced with longer VOT values than would be typical for English

Both groups learned something in the L2, making modifications in the right directions, but neither group produced stops in the L2 accurately
Testing SLM predictions

Flege (1987) The red arrows here indicate productions in the L1

(c) For the American women in Paris, English /t/ was produced with shorter (French-like) VOT values than would be typical for English

(d) For the French women in Chicago, French /t/ was produced with longer (English-like) VOT values than would be typical for French

This supports the hypothesis that L1 and L2 phonetic elements exist in a common space and mutually influence one another.
Testing SLM predictions

**Prediction:** When a new category is formed for an L2 sound, it and/or the nearest L1 sound may dissimilate so that they are more distant from one another in phonetic space. This process may make production of one or both phonetic elements less accurate from the point of view of normative values obtained from monolinguals.

This renders concrete the adage (see work by F. Grosjean) that a bilingual cannot be “two monolinguals in one person”
Testing SLM predictions

• Flege & Eefting (1986, 1987) examined production of phonologically voiceless stops, /p t k/, in Spanish and English words.

• They recorded four groups of participants. Of these:

1. Spanish monolinguals: adults & children living in Puerto Rico
2. English monolinguals: adults & children in Alabama
Flege & Eefting (1986, 1987) found that phonologically voiceless stops are realized differently by monolingual speakers of two languages. Both Spanish adults and children produced /p t k/ as short-lag stops. Both English adults and children produced the “same” stops with substantially longer VOT values.
Testing SLM predictions

Flege & Eefting (1986, 1987). The between-language differences in production were closely related to a between-language difference in perception (location of the “phoneme boundary” between voiced and voiceless stops).

Perception and production are said to be “aligned”. The small difference in production between English adults vs. children seen here was also “aligned” to differences in perception.
Testing SLM predictions

Flege and Eefting (1986, 1987) also recorded two groups of early bilinguals

3. Native Spanish adults who had begun learning English when attending a bilingual school in an essentially monolingual Spanish community in Puerto Rico

4. Native Spanish children who were currently enrolled in the same bilingual school in Puerto Rico where English was used as the language of instruction in most classes
Flege & Eefting (1986, 1987) also tested in two languages native speakers of Spanish who had learned English as an L2.

Both child and adult bilinguals produced the phonologically voiceless stops /p t k/ with substantially longer VOT values in English L2 than in Spanish L1.
Testing SLM predictions

Flege & Eefting (1986, 1987). Productions of Spanish /p t k/ by two groups of Spanish monolinguals (children, adults) and by the two groups of early Spanish-English bilinguals (children, adults).

The monolinguals produced Spanish /p t k/ as short-lag stops (VOT < 30 msec). Both groups of bilinguals produced voiceless Spanish stops with even shorter VOT values. This provides evidence of dissimilation.
Testing SLM predictions

According to the SLM, children are more likely to form phonetic categories for L2 sounds than adults because their L1 categories are less fully developed and represent weaker “attractors” for sounds encountered on the phonetic surface of an L2.

The SLM maintains, however, that the processes and mechanisms subserving the completely successful acquisition of an L1 by monolinguals are used by persons who acquire an L2, even in adulthood. Thus, according to the SLM, even adults retain the capacity to form new categories for L2 sounds if given the right kind of input and the opportunity (time) needed to do so.
Testing SLM predictions

• Flege & Eefting (1988) provided evidence of category formation for /p t k/ by early Spanish-English bilinguals

• These authors tested (a) Spanish-English bilinguals and (b) Spanish & English monolinguals

• The participants’ task was to imitate members of a synthetic /d/ to /t/ continuum made up of stimuli differing in VOT, which ranged from lead (pre-voiced) values to long-lag values (an aspirated [\textipa{th}])
Flege & Eefting (1988). This figure shows the distribution of VOT values in the 900 imitations of members of a VOT continuum by Spanish monolingual children. The children did not accurately reproduce the VOT values present in the stimuli. They instead tended to produce stops having VOT values in the lead (pre-voiced) range or with VOT values in the short-lag range. Both are typical for Spanish.
Flege & Eefting (1988). Here are the results for the monolingual Spanish adults, who did not accurately imitate VOT values in the stimuli. We again see two distributions of VOT values in the imitation responses, one typical for the Spanish /d/, the other for the /t/ of Spanish. The authors concluded that Ss rapidly classified the initial stops in the perceptual stimuli in terms of phonetic categories established in L1 acquisition and then produced them according to the VOT value specified by the phonetic categories.
Dramatically different results can be seen in the distribution of the 900 imitations of the same VOT continuum by monolingual English children. These children produced few pre-voiced (lead VOT) stops. They showed two distributions in the lag VOT region. The short-lag values are typical for English /d/ and the long-lag values are typical for English /t/.
Flege & Eefting (1988). Much the same pattern of results was for the **monolingual English adults**.
Flege & Eefting (1988). Distribution of VOT values obtained during imitation of the VOT continuum by Spanish-English bilingual children. These children three distinct distributions of VOT values in their imitation responses. The authors interpreted this as evidence for the existence of three distinct phonetic categories.
The conclusion regarding the existence of three distinct phonetic categories was reinforced by the finding obtained for the bilingual Spanish-English adults. The authors concluded that the Spanish-English bilinguals retained two phonetic categories established during L1 acquisition (lead, short-lag) and added a third phonetic category needed for the long-lag stops of English when they acquired English as an L2.
Testing SLM predictions

• Flege, Schmidt & Wharton (1996) and Schmidt & Flege (1995) provided evidence of category formation for the long-lag /p/ of by a few late Spanish-English bilinguals

• Participants rated the randomly presented member of a VOT continuum for category “goodness”

• In English, VOT values in word-initial /p/ tokens shorten as speaking rate increases

• The authors created two VOT continua, one that stimulated speech produced at a slow rate, the other at a faster rate.
For both continua, ratings obtained from native English Ss increased (indicating a better perceived “goodness” as instances of the /p/ category) as VOT increased, then a systematic decrease in the ratings as VOT values in the stimuli went beyond values typical for English.

The pattern indicates a match between production and perception. The Ss accepted longer VOT values as “good” in the slow rate continuum accordance with the fact that VOT values are longer in speech produced at a slow rate.

Native speakers of English show rate-dependent processing of stops differing in VOT.
Native speakers of Spanish who had learned English as adults (late bilinguals) were also tested.

There is relatively effect of variations in speaking rate on VOT in Spanish short-lag stops.

Here we see the results obtained for 4 of 15 late bilinguals tested. They produced English /p/ with short-lag values that are typical for Spanish (range of mean production values = 13 – 18 msec)

The perception data shown here – no indication of a speaking rate effect on the goodness judgements coincides with the production data. These late bilinguals had not created new phonetic categories for English /p/
A very different picture emerged for the 4 (of 15) late bilinguals who managed to produce English /p/ with long-lag VOT values.

The four late bilinguals showed a clear speaking rate effect on the goodness judgements. This, taken together with the production results, suggests that these late bilinguals had created new phonetic categories for English /p/
Outline

1. Purpose of the Speech Learning Model (SLM)
2. Historical background
3. Core aspects of the SLM
4. Some predictions generated by the SLM
5. How to falsify the SLM
6. What is most needed now?
How to falsify the SLM

• A theoretical model is interesting and useful only to the extent that it can be falsified.
• Here I’ll suggest some ways the SLM might be falsified.
• But first, a few comments are necessary regarding the adequacy of testing methods.
Adequate measurement of L2 speech are needed

*For examinations of L2 segmental production we need to ask*

- Do native speakers of the target L2 hear segments produced by L2 learners as they were intended (categorical judgment)? If so, do the native listeners rate the L2 segments as “distorted” or “foreign accented” (qualitative)?
- When measured acoustically, do relevant dimensions in the target L2 segments differ significantly from native speakers’ productions?
Adequate measurement of L2 speech are needed

For examinations of L2 segmental perception we need to ask:
- Do L2 learners correctly identify L2 segments?
- If so, do they do so as rapidly as native speakers?
- Do they show a greater influence of semantic/lexical context than native speakers of the target L2?
Appropriate participants must be tested

- It is unreasonable to conclude that L2 learners are unable to L2 phonetic segments accurately if they have only (or mostly) heard inaccurate productions of the L2 phonetic segments
- It takes children years to learn to accurately produce phonetic segments in their L1. It is inappropriate to conclude that L2 learners are unable produce an L2 segment accurately until they have received at least as much input as is needed by monolingual L1 learning children
To establish upper limits on phonetic learning in an L2, it makes sense to find participants who

1. Have used the L2 frequently for many years;
2. Tehd, necessarily given #1, to use their L1 infrequently
3. Tend to use their two languages in different contexts and with different people

Some would regard #3 as too severe given that bilinguals often fluently mix their two languages, inserting materials from one language into the other. However, if such a condition is not imposed, then native vs. non-native differences observed in the L2 might result from a psycholinguistic difficulty in separating the L1 & L2 subsystems rather than difficulty in purely phonetic learning.
Perceived L1-L2 dissimilarity must be assessed

Perceived cross-language similarity phonetic dissimilarity can be “guessed” at by looking at acoustic data (e.g. formant values plotted in a 2- or 3-dimensional vowel space).

However, predictions generated by the SLM cannot be tested based on such guesses.
Assess perceived cross-language dissimilarity

Alas, currently used perceptual testing procedures are inadequate. In evaluations used now:

• Ss (listeners) hear a specific L2 phone
• They are asked classify it in terms of abstract (imagined) L1 category
• They are then asked to rate the L2 phone for goodness of fit to the (imagined) L1 category

Problems: How can we be sure that listeners (Ss) performing such a task are really “thinking about” the L2 phone when making subsequent judgments? How do Ss match imagined abstract representations to the auditory properties associated with a specific phone? How does the experimenter integrate ratings with classifications?
Assess perceived cross-language dissimilarity

Better: present pairs of L1 and L2 tokens to be rated for degree of cross-language phonetic distance

• Each L1 and L2 category is represented by multiple natural tokens, produced by monolinguals in multiple contexts

• The two stimuli in each pair are rated on an EAI scale ranging from “very similar” (1) to “very dissimilar” (5)

• To ensure use of the entire scale by all listeners (Ss) pairings should include L1 sounds that are “near” each L2 sound, those that are maximally distant, and those that are “in between”
Assess perceived cross-language dissimilarity

Once mean distances have been calculated on all pairings, the experimenter focuses on the ratings obtained for pairings that represent the closest L1 sound to each of the target L2 sounds of interest. (It might be valuable to consider the two closest L1 sounds if two L1 sounds appear to be roughly equidistant to a target L2 sound.)

The next slide presents some hypothetical data showing the perceived distances of five L2 vowels from the closest L1 vowel.
Assess perceived cross-language dissimilarity

Hypothetical data. Degree of perceived cross-language phonetic dissimilarity of 5 vowels in an L2 from the perceptually closest L1 vowel. Flege et al. (1994, JASA, 95: 3623ff) that the differential categorization of a pair of vowels augments perceived dissimilarity.
Hypothetical data. We expect that at the time of first exposure, the vowels of the L2 will differ in perceived distance from the closest L2 vowel. The SLM predicts that the greater the perceived distance, the greater the likelihood of category formation. An augmentation of perceived dissimilarity, especially if seen in L2 vowels that are relatively distant from the closest L1 vowel, would indicate the formation of a new category.
The absence of an augmentation of perceived cross-language phonetic dissimilarity over time, especially in multiple experiments, would disconfirm the SLM hypothesis. So, too, would augmentations in perceived distance for relatively similar L2 sounds rather than for relatively dissimilar L2 sounds.
How to falsify the SLM

At a young age children can accurately recognize and identify the phonemes of their L1. However, the phonetic categories used in language comprehension, especially in non-ideal conditions, required finely tuned phonetic categories that take years to fully develop.

According to the SLM, category formation (CF) becomes less likely as L1 categories develop

- **Hypothesis**: as L1 categories become more robust through childhood, they “more powerful attractors” for L2 speech sounds (see Baker, Trofimovich, Mack & Flege, 2002)
How to falsify the SLM

• Let’s imagine another hypothetical experiment that might falsify core tenets of the SLM.
• It is a repetition of the earlier experiment that compares Ss differing in age
• The groups consist of individuals aged 9 and 21 years who are learning the same L2 in similar circumstances
• The dissimilarity ratings will be obtained in 2 sessions separated by 9 years, so the Ss will be 18 and 30 years of age at Time 2
How to falsify the SLM

The SLM predicts that there will be a greater augmentation of perceived cross-language dissimilarity for at least the most distant vowel by the children compared to the adults. When individual data is examined, more children than adults should show evidence of augmented L1-L2 dissimilarity.
How to falsify the SLM

• The observation of a greater augmentation in perceived L1-L2 distances for early than late learners – or for more of the early than the late learners (considered individually) would support the hypothesis that category formation (CF) for L2 sounds becomes less likely as the L1 system develops.

• The conclusion regarding CF would be further confirmed if the vowels showing evidence of CF demonstrated greater improvements in production accuracy than did those showing no augmentation of perceived cross-language phonetic dissimilarity.
Outline

1. Purpose of the Speech Learning Model (SLM)
2. Historical background
3. Core aspects of the SLM
4. Some predictions generated by the SLM
5. How to falsify the SLM
6. What is most needed now?
What is needed most now?

1. More adequate methods of participant selection
2. Standardized measure of perceived L1-L2 phonetic distance
3. More precise measures of L2 input
4. Quantitative methods to model the relation between L2 input and L2 performance
5. Explanations for the large individual differences seen in many study of L2 learning, especially among late learners
What is needed most now?

6. The development of a standardized test of category formation using non-overt responses such as MMN or neural imaging)
7. Large scale studies examining multiple measures of production & perception specifically designed to falsify the SLM
8. Patience
9. Hard work
10. Imagination
11. Luck
Thank you

- Thanks are also due NIH for its continued financial support for this research
- To my many collaborators, including: K. Aoyama; W. Baker; O-S. Bohn; W. Eefting; S. Fletcher; R. Fox; E. Frieda; S. Guion; R. Hammond; J. Hillenbrand; S. Imai; S. Liu; M. Mack; V. Mann; R. McAllister; D. Meador; M. Munro; T. Nozawa; T. Piske; T. Riney; C. Schirru; A. Schmidt; H. Southwood; Y. Takagi; P. Trofimovich; K. Tsukada; A. Walley; C. Wang; R. Wayland.
- Special thanks are due to Ian R. A. MacKay, Robert F. Port, and Grace Yeni-Komshian
References


