How the revised Speech Learning Model (SLM-r) works for stop consonants

James Emil Flege

Professor Emeritus
School of Health Professions
University of Alabama at Birmingham

a Research being reported was supported by grants from the National Institute for Deafness & Other Communicative Disorders (USA)

b Contact info: jeflege@uab.edu website: www.jimflege.com
Introduction

The purpose of the revised Speech Learning Model (SLM-r) is to account for how phonetic segments are learned across the lifespan.

The aim of this talk is to show how the SLM-r might be applied to the learning of word-initial stops by native speakers of Spanish and Italian.

Historical note: The SLM was proposed in Flege (1995). I introduced a revision of the model during a plenary talk at the ISCA Workshop on "Plasticity in Speech Perception" held June 15-17, 2005 at University College London. The revised model, the SLM-r, will be formally presented in 2018.
Introduction

The SLM-r focusses on the learning of position-sensitive allophones, not phonemes, because it is a phonetic, not phonological model.

Introduction

We administered feedback training on edited word-final English stops to Chinese adults

Closure voicing and release bursts had been removed from the stimuli. Such stimuli can be identified correctly by native English adults. However, Chinese learners of English L2 have difficulty

Training increased % correct identification significantly
• for both native Mandarin and Cantonese Ss
• However, Cantonese performed better than Mandarin before, during and after training
Introduction

Why the Cantonese vs Mandarin difference?

Cantonese has [p t k] in final position, Mandarin has none

If /p t k/ are realized in Cantonese and Mandarin with exactly the same VOT values in word-initial, the SLM-r would predict the same course of learning for speakers of both languages

I say IF because

• I haven’t seen a direct comparison
• the SLM-r generates predictions only when perceived cross-language phonetic dissimilarity is known
Introduction

Cross-language VOT differences

Originally thought that VOT values in the worlds’ language fell into one of three “universal” phonetic categories. Not so
Introduction

Cross-language VOT differences

We subsequently learned VOT can vary widely across languages
Introduction

Cross-language VOT differences

Children

• slowly learn to re-produce the language-specific VOT values of their native language
• they learn to re-produce what they have heard
• “attunement” to the phonetic properties of the sounds of a language takes many years, both in L1 and L2 learning
Introduction

VOT is popular in SLA research because

- data easy to obtain and to analyze
- VOT can be measured rapidly and accurately

Many studies have been published, but not all can be trusted because it is easy to make errors when examining VOT

It is crucial
to obtain VOT values that are representative of how L2 learners actually produce stops in L2 words

Not always done. Consider the error we made eliciting data in the Flege & Eefting (1987) study
Introduction

We (Flege & Eefting, 1987) tested 50 Dutch university students in The Netherlands

10 Engineering students had not studied English beyond the basic requirements required by the University; 40 other students were English majors

Nor surprisingly:
most of the “English majors” had a better overall pronunciation of English (less foreign accent) than the Engineering students
Introduction

We asked the Dutch students to read English words beginning with stops consonants from a list, saying them at the end of a carrier phrase (The word is__)

Some English majors produced VOT values that greatly exceed values produced by native English speakers

Some English major evidently understood the aim of the research

- they exaggerated VOT (under volitional control)
- their productions were not representative of how they normally produced English stops, and so not valid
Introduction

A second example: failing to control for speaking rate

Birdsong (2003) tested native English speakers who had learned French after the closure of a “critical period”

His aim: determine if Late learners who had lived in Paris for many years could produce French /p t k/ with accurate short-lag VOT

Of 21 Late learners tested, 14 (67%) met Birdsong’s statistical criterion for “native-like” production (< 1 SD from native French mean)

*note*: similar results for /p/ and /t/
Introduction

Unfortunately, the Late learners produced vowels that were 48% longer than those produced by native French speakers

A problem because

in English, VOT gets longer as vowel duration increases

Uncontrolled difference in speaking rate worked against Birdsong’s hypothesis that Late learners can produce L2 stops accurately if they get abundant L2 input
Introduction

I corrected Birdsong’s VOT data using the results of Theodore et al. (2009)

• The overall mean value for Late learners’ production of VOT in French decreased by 9 msec

• The remaining native vs non-native difference shrunk to just 2 msec, about the expected size of measurement error (see, e.g., Schmidt & Flege, 1995, p. 46)

• Had the Late learners spoke at the same rate as the native French speakers, 3 additional Late learners would have been credited with “native-like” VOT production
Introduction

The most serious and pervasive error in SLA research is to ignore the time dimension.

Children learning their L1 soon manage to recognize and produce recognizable differences between minimally paired words:

- tall vs doll,
- sought vs sod

However: developmental research indicates that children learning English need **at least 10 years** of native-speaker input to reach adult-like levels in producing and perceiving the VOT dimension in word-initial stops.
Introduction

When you measure VOT you get numbers. But what do those numbers mean?

It often happens in SLA that

a VOT difference between Late learners and native speakers is interpreted to mean that L2 learners were incapable of learning L2 stops. Perhaps due to having passed a “critical period”

This interpretation may be incorrect because it ignores the time dimension

A native vs non-native difference sometimes indicates “learning in progress”, not diminished capacity for learning

Results obtained by Flege & Eefting (1986) illustrate how long phonetic learning takes in the L1
Introduction

Flege & Eefting (1986) examined identification of stops differing in VOT. Listeners identified randomly presented VOT stimuli as /b/ or /p/

The phoneme boundaries (50% crossover) occurred at significantly longer value for native English than Spanish adults. This was the expected cross-language difference
Introduction

8 to 9 year-old monolingual Spanish and English children also showed the expected cross-language difference.

However, we obtained adult vs child differences in both languages. Adults needed to hear longer VOT values than children before crossing over from predominantly /b/ to /p/ responses. Neither group of 8-9 year-olds had completed phonetic learning of stops in their L1.
Introduction

Flege & Eefting (1986) tested additional groups of monolingual English children.

The phoneme boundaries of even 17-year-olds differed significantly from adults’
Introduction

Flege & Eefting (1986) also observed adult vs child differences in the production of VOT

Spanish and English adults produced /t/ with longer VOT values than Spanish and English-learning children

(The difference, which averaged 6 msec, narrowly missed reaching significance)
Introduction

According to the SLM-r

1. the same processes and mechanisms are used for learning speech across the lifespan

2. the most important aspect of speech learning is the establishment of language-specific phonetic categories

3. L2 learners of all ages possess all of the basic capacities used by children learning their L1, including the ability to establish new phonetic categories
Introduction

phonetic category formation

Children learning an L1

- discover the existence of contrastive phonetic categories through awareness of lexical contrasts (minimal and near-minimal pairs)
- begin to define phonetic categories through exposure to distributions of tokens
- L1 input, mostly from native speakers, “fills” developing L1 phonetic categories
Introduction

phonetic category formation

“category centers” emerge, gradually permitting children to

• identify an array of L1 phones as instances of a single category with increasing speed and accuracy

• recognize speech sounds in non-ideal listening conditions
Introduction

**phonetic category formation**

As more input is received, L1 phonetic categories continue to develop slowly via the cognitive mechanisms of

- **Acquired distinctiveness**: which augments sensitivity to differences between categories localized at the boundary between categories;

- **Acquired similarity**: which reduces sensitivity to differences between members of the same category within the perceptual space occupied by the category.
Introduction

**phonetic category formation**

SLM-r proposes that the *same processes and mechanisms* are used for L1 and L2 speech learning

However

- This does not guarantee identical results in L1 and L2 learning
- L1 and L2 learning can never be *identical* because the L1 already exists when L2 learning begins
Introduction

phonetic category formation for L2 sounds

Children and adults learning an L2

Must discover **differences** between L1 and L2 phonetic categories

- lexical definition of L1 vs L2 contrasts matters little because learners relate L1 to L2 sounds via the mechanism of equivalence classification
- a high degree of **perceived cross-language dissimilarity** promotes the discovery of new L2 categories
- Once L2 categories have been identified, they are gradually defined by input distributions

Note: How to measure perceived dissimilarity will be considered after I finish this talk, time permitting
Introduction

To illustrate the importance of *perceived cross-language dissimilarity* let’s consider five hypothetical languages ("a" to "e") having language-specific VOT values differing by 20 msec each.

![Hypothetical distributions for 5 languages](image)

- a
- b
- c
- d
- e

**VOT**

**Frequency**
Introduction

What does the SLM-r predict regarding the learning of Language “d” (like English) by native speakers of

<table>
<thead>
<tr>
<th>L1</th>
<th>differences from L2</th>
<th>what will happen?</th>
</tr>
</thead>
<tbody>
<tr>
<td>“a”</td>
<td>60 msec</td>
<td>very easy because of the large difference?</td>
</tr>
<tr>
<td>“b”</td>
<td>40 msec difference</td>
<td>greater difficulty than for speakers of “a”?</td>
</tr>
<tr>
<td>“c”, “d”</td>
<td>both differ by 20 msec</td>
<td>equal learning difficulty?</td>
</tr>
</tbody>
</table>

hypothetical distributions for 5 languages

![Frequency](image)
Introduction

The SLM-r makes no predictions

<table>
<thead>
<tr>
<th>L1</th>
<th>differences from L2</th>
<th>what will happen?</th>
</tr>
</thead>
<tbody>
<tr>
<td>“a”</td>
<td>60 msec</td>
<td>very easy because of the large difference?</td>
</tr>
<tr>
<td>“b”</td>
<td>40 msec difference</td>
<td>greater difficulty than for speakers of “a”</td>
</tr>
<tr>
<td>“c”, “d”</td>
<td>both differ by 20 msec</td>
<td>equal learning difficulty?</td>
</tr>
</tbody>
</table>
Introduction

Why not? Acoustically measured VOT (a reflection of unseen laryngeal timing differences) is not equivalent to perceived cross-language phonetic dissimilarity

- 20 msec differences: not necessarily the same if greater/less than target language norm
- 40 differences not necessarily twice as large as 20 msec difference
- 60 msec difference: does not guarantee success because sufficient L2 input is necessary

<table>
<thead>
<tr>
<th>L1</th>
<th>differences from L2</th>
<th>what will happen?</th>
</tr>
</thead>
<tbody>
<tr>
<td>“a”</td>
<td>60 msec</td>
<td>very easy because of the large difference?</td>
</tr>
<tr>
<td>“b”</td>
<td>40 msec difference</td>
<td>greater difficulty than for speakers of “a”</td>
</tr>
<tr>
<td>“c”, “d”</td>
<td>both differ by 20 msec</td>
<td>equal learning difficulty?</td>
</tr>
</tbody>
</table>
Introduction

Perceived cross-language dissimilarity

Consider, for example, a native Spanish-speaking woman who

- is first exposed to English at age 20 when she immigrates to the US
- uses English regularly for 10 years
- but only 50% of the time, so just 5 years of English input
Introduction

Perceived cross-language dissimilarity

This hypothetical learner of English L2 is not expected to have created native-like phonetic categories for English /p t k/ because

• children learning English as an L1 need far more than 5 years of input (roughly 10 years)

• much of the English input she receives is likely to be Spanish-accented, providing an incorrect model
Introduction

An obvious point: If you can’t hear it you can’t learn it

Are 20 cross-language VOT differences audible? Flege (1984) used a paired comparison test to evaluate the ability of English monolinguals to detect cross-language VOT differences

Stimuli: English /ti/ and /tu/ tokens produced by native English women and native French women who spoke English with foreign accent.

<table>
<thead>
<tr>
<th></th>
<th>/i/ context</th>
<th>/u/ context</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>native English (n=8)</td>
<td>79</td>
<td>78</td>
<td>79</td>
</tr>
<tr>
<td>native French (n=8)</td>
<td>63</td>
<td>46</td>
<td>55</td>
</tr>
<tr>
<td>Δ</td>
<td>16</td>
<td>32</td>
<td>24</td>
</tr>
</tbody>
</table>
Introduction

Stimuli:

Vowels cross-spliced so that the native and non-native CV stimuli differed only in VOT

The native- vs non-native VOT differences were twice as large in the /u/ than /i/ context (32 vs 16 msec)

Task of 10 native-English speaking listeners: decide which of two stimuli in each pair produced by a non-native

<table>
<thead>
<tr>
<th></th>
<th>/i/ context</th>
<th>/u/ context</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>native English (n=8)</td>
<td>79</td>
<td>78</td>
<td>79</td>
</tr>
<tr>
<td>native French (n=8)</td>
<td>63</td>
<td>46</td>
<td>55</td>
</tr>
<tr>
<td>Δ</td>
<td>16</td>
<td>32</td>
<td>24</td>
</tr>
</tbody>
</table>
Introduction

The monolingual native English listeners performed

• significantly above-chance (50% correct) in identifying the French-accented stimuli

• As expected, better performance for /u/ context (32-msec difference) than /i/ context (16-msec difference)

The SLM-r focuses on individual rather than group performance

Possible that: some English monolinguals better able to detect cross-language VOT differences

If so: may be better/faster in learning French stops
Introduction

According to the SLM-r

If perceived cross-language phonetic dissimilarity is insufficient for category formation, learners will continue using L1 categories for perceptually linked L2 categories.

However they will nevertheless show phonetic learning because, by hypothesis all audible input is used all of the time. L1 categories will be updated to reflect L2 input.
Introduction

The SLM-r predicts that in the absence of category formation

- “merged” categories reflect the properties of all L1 and L2 tokens identified as instances of the “merged” category

Perceptually linked L1 categories change slowly. In time, learners will produce L1 stops with VOT values slightly longer than L1 monolinguals, and L2 stops with VOT values slightly shorter than L2 monolinguals.
Organization

In the remainder of this talk I will

1. present data for English /p t k/

2. present data for English /b d g/ (the learning of which differs fundamentally from the learning of /p t k/)

3. Conclude
Results for /p t k/

Flege et al. (1998) tested 41 native Spanish adults living in Birmingham, Alabama (USA)

Produced 60 English words beginning with /t/. The 60 words differed according to text frequency and so were expected to differ in familiarity for the native Spanish learners of English

Part 1: lexical knowledge

• determine which English words the native Spanish learners knew
• words rated for subjective familiarity, estimated age of acquisition words

Part 2: production

Words elicited using delayed repetition task

heard: “X is the next word”
(constant moderate speaking rate modelled)
responded “Now I say X”
Results for /p t k/

Flege et al. (1998) obtained what is now a familiar pattern

- All Early learners produced native-like VOT values
- Great inter-subject variability for Late learners: some native-like, others produced English /t/ as if it were Spanish /t/
Results for /p t k/

Analyses of four subgroups attempted to explain inter-subject variability:

1. Early learners (all)
2. 12 randomly selected NE (of 20 original)
3. 12 Late learners with longest VOT (“good”)
4. 12 Late learners producing with shortest VOT (“bad”)

![Graph showing mean VOT (msec) for NE, Early, Late good, and Late bad groups. The Late bad group has a significantly lower mean VOT compared to the NE group.](image)
Results for /p t k/

Significant between groups differences in VOT

NE, Early, Late-good > Late-bad (p < .05)

Crucial questions for L2 speech learning

• Why did all Early learners resemble NE?
• Why did the two subgroups of Late learners differ (one native-like, the other not)?
Results for /p t k/

No difference between words known, not known
- Left: % of 60 words known: NE > Early > Late-good, Late-bad (p < .05)
- Right: VOT for words known, not known
Results for /p t k/

The two groups of Late learners did not differ in lexical factors

- subjective lexical familiarity
- estimated age of learning the English words

NE, Early > Late-good, Late-bad
NE, Early < Late-good, Late-bad
Results for /p t k/

The two groups of Late learners did not differ in:

- self-rated grammatical ability in English
- pronunciation of English
- estimated effort expended learning English
Results for /p t k/

The two groups of Late learners did not differ in

- chronological age
- age of arrival in the U.S.
- length of residence (LOR) in the United States

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of arrival</td>
<td>7.9</td>
<td>27.0</td>
<td>29.4</td>
</tr>
<tr>
<td>chronological age</td>
<td>28.2</td>
<td>31.2</td>
<td>37.3</td>
</tr>
<tr>
<td>length of residence</td>
<td>20.3</td>
<td>4.2</td>
<td>7.9</td>
</tr>
</tbody>
</table>

1 < 2.3
1>3
1 > 2.3
Results for /p t k/

Nor did the two groups of Late leaners differ in *quantity* of English input

- Early learners had received about 14 years of English input, presumably most of which came from NE speakers. May explain their success.
- The two groups of Late learners did not differ in years of English input (3 vs 4 years)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>% English home</td>
<td>48.7</td>
<td>31.3</td>
<td>7.3</td>
</tr>
<tr>
<td>% English work</td>
<td>71.9</td>
<td>84.7</td>
<td>85.4</td>
</tr>
<tr>
<td>% English w/ friends</td>
<td>70.4</td>
<td>63.9</td>
<td>30.2</td>
</tr>
<tr>
<td>Ave. % English</td>
<td>63.7</td>
<td>59.9</td>
<td>41.0</td>
</tr>
<tr>
<td>LOR * Ave. % English</td>
<td>13.7</td>
<td>2.7</td>
<td>3.9</td>
</tr>
</tbody>
</table>
Results for /p t k/

I hypothesize that the two groups of Late learners differed in quality of L2 input.

For the SLM-r, quality of input defined by the distributions of tokens identified as instances of the /p t k/ categories.

Distributions of VOT values (1,200 per group)
Results for /p t k/

He we see distributions of 1,200 VOT values obtained for
• native English
• native Spanish Ss who arrived before age 21
• Native Spanish Ss who arrived after age 21
Likely that native Spanish learners of English hear some mixture of all three input distributions. But what exactly?

Distributions of VOT values (1,200 per group)
Results for /p t k/

According to the SLM-r
phonetic categories developed for English /p t k/ will be defined by the
distribution of VOT heard in English words
The “good” Late learners may have been exposed to less Spanish-accented English than the “bad” Late learners
Results for /p t k/

The SLM’s proposal that L2 speech learning is input driven requires accurate measurements of both the quantity and quality of L2 input

• Such data do not now exist

• Fortunately, the technology needed to obtain such data does exist and can be applied to L2 speech learning in the near future

For now, the most convincing evidence that quality of input is an important determinant of VOT production and perception comes from a comparison of research carried out in two locations

  o Mayaguez, Puerto Rico
  o Austin, Texas
Results for /p t k/

Flege & Eefting (1987) tested two groups of university students in Puerto Rico. Both began learning English at school age (5-6 years)

Group 1 started school in New York, then returned to Puerto Rico for High School. Group 2 began learning English in a bilingual school, never leaving PR

Both groups of Early learners produced English /p t k/ with

- significantly longer VOT than Spanish monolinguals
- and significantly shorter VOT values than English monolinguals
Results for /p t k/

According to the SLM-r, development of language-specific implementation rules are guided by language-specific phonetic categories.

The production results just presented lead to the expectation of

• similar differences in perception between native English speakers and the Early learners tested in Puerto Rico.
Results for /p t k/

Here we see the mean phoneme boundaries - 50% cross-overs from primarily /d/ to /t/ judgments

As expected, the Early learners’ phoneme boundaries occurred at
- significantly longer VOT than Spanish monolinguals’
- significantly shorter VOT values than English monolinguals’

Why? The Early learners in Puerto Rico usually heard English produced with Spanish accent by other native speakers of Spanish.
Results for /p t k/ 

Early learners have been tested in places where they usually hear English spoken by native speakers. These Early learners closely resemble the native English speakers they have heard.

Example: data obtained in Texas by Schmidt & Flege (1996)
Results for /p t k/

The Early learners in Puerto Rico seem to have established new phonetic categories for English /p t k/.

However, their categories differed from those of English monolinguals because they based on different input distributions.

What about Early and Late learners who get most of their input from native speakers of English?
Results for /p t k/

Flege & Schmidt (1996) obtained perception data from Early and Late learners in Austin Texas

Ss identified members of /bi/ to /pi/ continua, then rated the same stimuli for “goodness” as instances of the English /p/ category

Here we see goodness ratings obtained from English monolinguals for one VOT continuum
Results for /p t k/

Here are the ratings obtained from the NE monolinguals for both VOT continua.

The shift in ratings across continua, the “speaking rate effect” on perception (yellow), reflects changes in VOT that occur in speech production as a function of changes in speaking rate.
Results for /p t k/

For example: English monolinguals gave higher “goodness” ratings to a stimulus having VOT = 80 msec if it seemed to have been produced at a slow compared to fast rate

By hypothesis, this perceptual adjustment is mediated by a multidimensional phonetic category for /p/
Results for /p t k/

Early leaners: all showed a speaking rate effect like native English monolinguals

**Conclusion**: they had established new phonetic categories for English /p t k/
Results for /p t k/

Two patterns obtained for Late learners.

• those who produced English /p t k/ with accurate VOT values showed a speaking rate effect
• those who produced English stops with Spanish-like short-lag VOT values showed no speaking rate effect

**Conclusion**: some but not all Late learners established new phonetic categories for English stops
Data for /p t k/

The SLM-r predicts that speakers of Romance language who receive sufficient native speaker input will establish English-like phonetic categories for long-lag stops.

Supported by findings obtained by Flege et al. (1995a) for highly experienced Italian speakers of English.

Italian Ss immigrated to Canada in the 1950s and 1960s. Tested at this Catholic church in Ottawa, ON (Canada).
Data for /p t k/

The 240 Italian immigrants assigned to 10 groups of 24 each. The Italians

- had lived in Canada (Ottawa) for decades
- spoke English mostly with native English speakers
- usually spoke English more than Italian, but the difference in language use narrowed as AOA increased
Data for /p t k/

In a later study we determined where the Italians immigrants spoke Italian.

We asked 190 Italians whether they would normally use English or Italian in 32 specific social contexts.

In many contexts (e.g. pharmacy) English was nearly always used.

In other contexts (with family members, at church) English was seldom used.

### Examples of social contexts

<table>
<thead>
<tr>
<th>(A)</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>at a pharmacy</td>
<td>187</td>
</tr>
<tr>
<td>buying a car</td>
<td>178</td>
</tr>
<tr>
<td>At doctors office</td>
<td>183</td>
</tr>
<tr>
<td>attending school play</td>
<td>177</td>
</tr>
<tr>
<td>(B)</td>
<td>n</td>
</tr>
<tr>
<td>talking to your oldest child</td>
<td>113</td>
</tr>
<tr>
<td>at barber/hairdresser</td>
<td>102</td>
</tr>
<tr>
<td>talking to house painter</td>
<td>91</td>
</tr>
<tr>
<td>visiting someone in hospital</td>
<td>77</td>
</tr>
<tr>
<td>(C)</td>
<td>n</td>
</tr>
<tr>
<td>talking to younger sibling</td>
<td>52</td>
</tr>
<tr>
<td>visiting a close friend</td>
<td>55</td>
</tr>
<tr>
<td>talking to someone after mass</td>
<td>43</td>
</tr>
<tr>
<td>Talking to your father</td>
<td>3</td>
</tr>
</tbody>
</table>
Data for /p t k/

The 240 Italian immigrants asked which was the “better” of their two languages, English or Italian

Those who arrived in Canada before age 13 usually said English was their better language, those who arrived later usually said it was Italian

Importantly: nearly all said they could keep only one language, it would be English. Why? Most people they knew spoke only English
Data for /p t k/

Flege et al. (1995a) elicited English sentences using a delayed repetition technique. Example

Voice 1 (question)  What did Paul eat?
Voice 2 (answer)  Paul ate carrots and peas.
Voice 1 (question)  What did Paul eat?
... Pause
... Beep
participants repeat the answer (target sentence)
Data for /p t k/

We measured VOT in the three stops in the sentence. No significant between group differences obtained.

*One-way ANOVAs carried out for the three stops all yielded non-significant effects of Group [F(10,253) = 0.738 – 1.498, p > 0.10]
Data for /p t k/

To increase statistical power, I re-assigned the Ss to groups of 80 each based on age of arrival. The Late learners who arrived after the age of 16 never differed from the native English speaker

Supports the SLM-r prediction.
Data for \(/p\ t\ k/\)

Flege, Imai & MacKay (in prep.) examined 160 Italian immigrants in Canada. Assigned to groups of 40 each according to years of English input (LOR * % English use)

Ss repeated English words they heard modelled via a loudspeaker \((X\ is\ the\ next\ word)\)

The repeated the word at the end of a carrier phrase \((\text{Now I say } X)\)
Data for /p t k/

The only group of Italian immigrants differing significantly from NE speakers had received 16 years of English input and arrived in Canada at an average age of 17 years.

Was their inaccurate production of English /p t k/ due to passing a "critical period"? No

![Bar chart showing mean VOT in English /p t k/ (msec) with data points for different years of English input and age of arrival in Canada.](chart.png)
Data for /p t k/

I selected two subgroups of 13 each from the first group (average input = 16 years, average AOA = 17)

There were 40 Ss to chose from, so possible to select subgroups of 13 each who were matched for AOA in Canada (17 for both) and length of residence (40 years for both)

The “Much Input” group used English far more often than the “Little Input” group (60% vs 23%) and so had received far more years of English input (24 vs 9 years)

<table>
<thead>
<tr>
<th></th>
<th>Little input</th>
<th>Much input</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>Age of arrival</td>
<td>16.6</td>
<td>16.6</td>
</tr>
<tr>
<td>Length of residence</td>
<td>39.4</td>
<td>40.1</td>
</tr>
<tr>
<td>% English use</td>
<td>23.3</td>
<td>60.3</td>
</tr>
<tr>
<td>Years English input</td>
<td>9.0</td>
<td>23.7</td>
</tr>
<tr>
<td>Years native speaker input</td>
<td>7.2</td>
<td>19.0</td>
</tr>
</tbody>
</table>
Data for /p t k/

I compared the “Much Input” and “Little Input” subgroups to 13 randomly selected native English speakers.

As predicted by the SLM-r, the Late learners who received adequate L2 input did not differ from NE. Those who had not yet received adequate input did differ significantly (p < .05).

Recall: it takes children learning English as an L1 at least 10 years of full-time native speaker input to “master” English /p t k/.

If we assume that 20% of the input received by “Little Input” group was Italian accented, their years of native speaker input is reduced from 9 to just 7 years.
Data for /p t k/

According to the SLM-r

Phonetic categories are updated **constantly** across the lifespan based on input being received

This proposal clearly contradicts the notion well accepted by SLM researchers, **fossilization**

According to this notion

learners of an L2, especially Late learners, make progress up to a certain point, then stop

Above: a recent survey showed that 9 out of 10 Late learners under the age of 36 live in constant fear of fossilization
Data for /p t k/

A longitudinal study (Flege & MacKay, in prep) sheds light on the notion of fossilization.

The study focused on VOT in word-initial English stops.

We tested 150 Italian immigrants in Canada at two intervals separated by 10 years.

Crucially: the two tests were identical (same stimuli, procedures, equipment, testing location).

Details: Ss were recorded using a head-mounted Shure (Model SM10A) microphone and a portable cassette tape recorder (Sony Model TC-D5ProII). Ss first heard test words to be produced at the beginning of a fixed carrier phrase (“__ is the next word to say”), then repeated it at the end of another carrier phrase (“Now I say __”). The VOT values presented here were an average of just six (pick, peak, tack, tag, cap, cab) of the 25 words elicited.
Data for /p t k/

As expected, the VOT values obtained at the two sessions separated by 10.5 years were strongly correlated both for the 20 native English speakers (left) and the 150 native Italian speakers (right).
Data for /p t k/

Despite the overall high correlation:

Some Italian immigrants to Canada increased VOT from Time 1 (1992) to Time 2 (2003) whereas VOT for others decreased over time.

To learn more, we examined in detail the:

- **20 Italians** who showed the largest VOT increases over time.
- **20 Italians** who showed the largest VOT decreases over time.
Data for /p t k/

Our analysis of these two subgroups of Italians and 20 native English speakers yielded a Group x Time interaction ($p < .0001$) because the “Increase” group differed significantly from the NE speakers in 1992 but not 2003 whereas the opposite held true for the “Decrease” group.

What accounts for these VOT changes over time?
Data for /p t k/

We can rule out a number of possible explanations for the between-group difference

<table>
<thead>
<tr>
<th></th>
<th>Increase subgroup</th>
<th>Decrease subgroup</th>
<th>prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of arrival in Canada, in years</td>
<td>10.1</td>
<td>13.8</td>
<td>.08</td>
</tr>
<tr>
<td>Chronological age at Time 2, in years</td>
<td>53.0</td>
<td>56.6</td>
<td>.13</td>
</tr>
<tr>
<td>Length of residence in Canada at Time 2</td>
<td>42.2</td>
<td>43</td>
<td>.99</td>
</tr>
<tr>
<td>ability to speak, understand English (7=good)</td>
<td>6.1</td>
<td>6.2</td>
<td>.93</td>
</tr>
<tr>
<td>ability to speak &amp; understand Italian 7=good</td>
<td>5.6</td>
<td>6.2</td>
<td>.20</td>
</tr>
</tbody>
</table>
Data for /p t k/

The most likely explanation: significantly more English input over time for the “Increase” group (p < .05) but a non-significant trend in the opposite direction for the “Decrease” VOT.

Why a change in English use? Probably due to some major life change (new job, spouse, neighborhood). I don’t know, because we didn’t identify this important change when we collected data in 2003.
Data for /p t k/

Did the VOT change over time result just from quantitative change in L2 input?

Unlikely. Was probably accompanied by a change in quality of input (more or less Italian-accented English heard)

More research needed
/p t k/ summary

The studies just reviewed (and others not mentioned) suggest that

- Early learners outperform Late learners because they obtain more and better L2 input
- Even Late learners can establish new phonetic categories for English /p t k/ and produce these stops accurately if they get enough native speaker input

We do not yet know how native speaker input is needed because we don’t have accurate measures of quantity and quality of L2 input

Fortunately: the technology needed to obtain such measures now exists, and will soon be put to use in L2 speech research

Time will tell
We need accurate input data to interpret data like that obtained for this Late learner who produced English /t/ with VOT values intermediate to the phonetic norms of English and Spanish (arrow)

Did she create a phonetic category that accurately reflected what she heard? Or did she form a merged L1/L2 category because she was unable to establish a new phonetic category?
Goal of phonetic learning: English /b d g/

We will now consider English /b d g/

Not possible for speakers of a Romance language such as Italian to establish new short-lag categories for English /b d g/

Why not? The short-lag VOT range is already occupied by the /p t k/ categories
Goal of phonetic learning: English /b d g/

Speakers of Romance language could do absolutely nothing and it would be just fine

They could use their pre-voiced /b d g/ in English. After all, English /b d g/ can be produced with full pre-voicing
Goal of phonetic learning: English /b d g/

However, they do something, confirming the SLM-r prediction that L2 learners use all audible input all of the time.

The SLM-r predicts learning, just not the kind seen for /p t k/

Specifically: Italian will continue using Italian /b d g/ to produce and perceive English /b d g/ but their L1 phonetic categories will gradually be restructured to reflect phonetic input from English.
Goals of phonetic learning: /b d g/

English monolinguals produce /b d g/
- with fully pre-voicing (continues until release)
- partial pre-voicing (dies out before release)
- short-lag VOT

Italian monolinguals nearly always (99.5%) produce /b d g/ with full pre-voicing (MacKay et al. 2001, Fig. 1)

VOT values in word-initial /b/ tokens produced by 20 Italian monolinguals

adapted from MacKay et al. (2001), Fig. 1
Data for /b d g/

MacKay et al. (2001) tested four Italians immigrants differing orthogonally in Age of arrival in Canada and language use

Stimuli

Word-initial English tokens of /b d g/ and /p t k/ presented in noise

Task

6-alternative forced-choice test (possible responses < b d g p t k>)

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Age of arrival</th>
<th>Length of residence</th>
<th>% use Italian</th>
<th>Years education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native English</td>
<td>18</td>
<td>0</td>
<td>50</td>
<td>-</td>
<td>18</td>
</tr>
<tr>
<td>Early-low</td>
<td>18</td>
<td>7</td>
<td>42</td>
<td>7%</td>
<td>14</td>
</tr>
<tr>
<td>Early-high</td>
<td>18</td>
<td>8</td>
<td>40</td>
<td>43%</td>
<td>11</td>
</tr>
<tr>
<td>Late-low</td>
<td>18</td>
<td>20</td>
<td>31</td>
<td>10%</td>
<td>2</td>
</tr>
<tr>
<td>Late-high</td>
<td>18</td>
<td>20</td>
<td>29</td>
<td>53%</td>
<td>2</td>
</tr>
</tbody>
</table>
Data for /b d g/

The four Italian groups substantially in Years of English and Italian input received over the course of their lives.

Expectation: the more often the Italians heard /b d g/ produced with partial pre-voicing or short-lag VOT in English, the more they would produce and perceive stops like native English speakers.

*Years of English input was calculated by multiplying years of residence in Canada (LOR) by percent use of English. Years of Italian input was calculated age at the time of immigration * 100% + LOR in Canada multiplied by percent Italian in Canada.*
Data for /b d g/

All four groups of Italians misidentified short-lag tokens of English /b d g/ as /p t k/ more often than the NE speakers.

However: Just one group -- Late learners who continued to use Italian often -- differed significantly from NE.

Importantly: this was the only group that had received far more Italian than English input.
Data for /b d g/

In the same study (MacKay et al., 2001) we also examined production. We elicited production 20 /b/ tokens in words like bad, bade and bed. The words were produced

• in isolation, or
• as the middle of three words (e.g., heed … bat … hid) separated by pauses

Details

• Five test words were spoken by a male with pre-voicing and also by a female with short-lag VOT. Repetitions of the two models, following a filled delay, did not affect how the words were produced

• The stimuli to be repeated occurred a carrier phrase “X is the next word to say” in the single word condition, or “X Y Z are the next words to say” in the 3-word condition. There were no significant difference across the two repetition conditions
Data for /b d g/

All four groups, including Early learners who frequently used English, produced English /b/ significantly less often with short-lag VOT than English monolinguals.
Data for /b d g/

However, both groups of Early learners often produced /b/ with partial prevoicing which died out before release, something not seen in Italian.

As a result, only the two groups of Late learners produced /b/ with full prevoicing significantly more often than the NE speakers.
Data for /b d g/ 

/b/ tokens produced with full pre-voicing had voicing (glottal pulsing) just prior to release.

Important because in Italian the presence of voicing just before release is the primary cue to the distinction between /b d g/ and /p t k/
Data for /b d g/

We see a correlation between how often the four groups of Italians produced /b/ with full pre-voicing and how often they misidentified short-lag tokens of English /b d g/ as /p t k/

The more often the Italian immigrants produced full pre-voicing, which assured voicing just before release, the more often they misidentified English stops without this cue (short lag realizations of English /b d g/) as /p t k/
Data for /b d g/

Flege & MacKay (in prep.) tested 190 Italian immigrants in Canada

• assigned to ten groups of 19 based on AOA

• examined the /b/ in “Barbara” produced utterance-initial position in both English and Italian sentences
Data for /b d g/

Elicitation procedures used by Flege & MacKay (in prep.)

The Italians produced alternating blocks of question-answer-question sequences in English and Italian materials (English monolinguals just English materials). The Italian sentences were translation equivalents of the English sentences.

Ss heard a question, heard the answer to give, heard the question again and they produced the modelled “answer”. In English sentences “Barbara” was produced with two syllables and short-lag VOT. In Italian sentences the target word “Barbara” was produced with three syllables and full pre-voicing

<table>
<thead>
<tr>
<th>#</th>
<th>Question</th>
<th>Modelled response</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Does she have a sister?</td>
<td>Barbara is Patty’s sister</td>
</tr>
<tr>
<td>7</td>
<td>Who’s taller?</td>
<td>Barbara is a little taller than Patty</td>
</tr>
<tr>
<td>8</td>
<td>Who did you invite to dinner?</td>
<td>Barbara and Patty will be coming</td>
</tr>
</tbody>
</table>

Above: three of 10 pairs used to elicit production of /b/ in English
Data for /b d g/

The native English group (AOA = 0) produced English /b/ tokens nearly twice as often with short-lag VOT than with pre-voicing ($M = 65.4\%$ vs. $34.6\%$).

All 10 Italian groups produced English /b/ less often with short-lag VOT than pre-voicing
Data for /b d g/

The magnitude of difference in the frequency of pre-voicing vs short-lag realizations of English /b/ reflected between group difference in Italian vs English input.

All participants who arrived in Canada after age 13 had heard Italian more than English.

The Italian vs English input differences

- increased as AOA increased
- was related to the frequency of pre-voicing in English /b/

Years of English input was calculated as LOR * % English. Years of Italian input, on the other hand, was calculated as LOR * % Italian [Italian input in Canada] + age at the time of emigration to Canada * 100% Italian.
Data for /b d g/

For the 10 Italian groups, we see a correlation between how often the immigrants had heard /b/ produced with pre-voicing over the course of their lives and how often they produced pre-voicing in English /b/

![Graph showing correlation between pre-voiced /b/ tokens heard and produced, with R² = 0.8248]

Note: this analysis assumed that 34.6% of English tokens and 99.5% of Italian tokens heard by participants were pre-voiced
Ottawa: L2-on-L1 effects

Flege, Imai & MacKay (in preparation) recruited four groups of 40 Italians each who had lived in Canada for $M = 33$ years.

The two groups of Early learners had received more English than Italian input over the course of their lives, the Late learners less English than Italian.

<table>
<thead>
<tr>
<th>N (m/f)</th>
<th>AOA</th>
<th>LOR</th>
<th>% EN</th>
</tr>
</thead>
<tbody>
<tr>
<td>15/25</td>
<td>4.6</td>
<td>46</td>
<td>81</td>
</tr>
<tr>
<td>18/22</td>
<td>9.7</td>
<td>44</td>
<td>77</td>
</tr>
<tr>
<td>20/20</td>
<td>14.8</td>
<td>42</td>
<td>65</td>
</tr>
<tr>
<td>17/23</td>
<td>19.8</td>
<td>39</td>
<td>53</td>
</tr>
</tbody>
</table>

Note: AOA, Age of arrival in Canada from Italy; LOR, length of residence in Canada, years; %EN, self-estimated percentage use of English.
Ottawa: L2-on-L1 effects

Flege, Imai & MacKay (in preparation) licited English words beginning with /p t k/ and 22 /b/-initial words (beat, bit, boat, etc.)

Participants heard an English model such as “bit is the next word” and responded by saying, for example, “Now I say bit”.

This context favors the production of pre-voicing by native speakers of English. Voicing present in the preceding vowel continues into the closure interval, but not necessarily the entire closure interval.
Data for /b d g/

Here we see /b/ data for the four Italian groups (n = 40 each) differing in years of English input.

As expected for this phonetic context, there were no significant between group differences in the frequency of pre-voicing. Not even in the frequency of partial pre-voicing in which voicing died out before stop release.
Ottawa: L2-on-L1 effects

After completing the English protocol, the 160 Italian immigrants and 20 Italian monolinguals in Rome repeated **Italian** words beginning with /p t k/ and /b d g/ that they had heard over a loudspeaker.

Eight Italian words each beginning with /b d g/ (e.g., *bocca, doccia, gonna*) were measured acoustically.

An important difference: the Italian words were produced in isolation whereas the English words considered earlier were produced following a vowel.
Data for /b d g/

Here we see data for /b d g/ in Italian words.

We see a high frequency of pre-voicing even though the context (absolute utterance initial position) does not favor it.
Data for /b d g/

An effect of learning English on Italian production is unmistakable. Compared to Italian monolinguals

- All four groups of immigrants produce pre-voicing significantly less often (p < .05)
- The two groups of immigrants having the most exposure to English produced significantly more Italian stops with partial pre-voicing (p > .05)
- Those with the most English exposure produce short-lag stops significantly more often
Data for /b d g/

According the SLM-r account

The Italians restructured their Italian /b d g/ categories for use in both English and Italian. The more English input they received, the more their /b d g/ categories resembled those of English monolinguals.

Learning English changed how the immigrants produced /b d g/ in their native language, Italian.

If the influence of English on Italian /b d g/ was sufficient, it in turn changed how the bilinguals produced Italian /p t k/.
Data for /b d g/

I identified two groups of Early learners who differed according to how often they produced Italian /b d g/ with English like short-lag VOT

<table>
<thead>
<tr>
<th></th>
<th>Frequent SL stops in Italian</th>
<th>Infrequent SL stops in Italian</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>18</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>age of arrival in Canada</td>
<td>7.8</td>
<td>7.8</td>
<td>n.s.</td>
</tr>
<tr>
<td>length of residence</td>
<td>44.3</td>
<td>45.7</td>
<td>n.s.</td>
</tr>
<tr>
<td>% English use</td>
<td>81.1</td>
<td>76.3</td>
<td>n.s.</td>
</tr>
<tr>
<td>% short lag realization of Italian /b d g/</td>
<td>70.1%</td>
<td>9.2%</td>
<td>p &lt; .001</td>
</tr>
</tbody>
</table>
Data for /b d g/

One group produced Italian /b d g/ with short-lag VOT values 70% of the time. The other group just 9% of the time

<table>
<thead>
<tr>
<th></th>
<th>Frequent SL stops in Italian</th>
<th>Infrequent SL stops in Italian</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>age of arrival in Canada</td>
<td>7.8</td>
<td>7.8</td>
</tr>
<tr>
<td>length of residence</td>
<td>44.3</td>
<td>45.7</td>
</tr>
<tr>
<td>% English use</td>
<td>81.1</td>
<td>76.3</td>
</tr>
<tr>
<td>% short lag realization of Italian /b d g/</td>
<td>70.1%</td>
<td>9.2%</td>
</tr>
</tbody>
</table>
Data for /b d g/

The Early learners who frequently produced Italian /b d g/ with short-lag VOT values produced Italian /p t k/ with significantly longer VOT values than Italian monolinguals and the Early learners who showed little effect of English on production of /b d g/ (p < .05).

The modified /b d g/ categories «pushed» the VOT of /p t k/ upwards to maintain phonetic contrast. Just as in historical sound change.
Conclusions

The evidence presented here supports the SLM-r view that

1. L2 speech learning is data driven (as is the case for L1 speech learning)

2. L2 learners make effective use of all audible input

3. Learners of all ages can learn to produce and perceive L2 phonetic segments accurately if they have the necessary input

4. Learning proceeds differently when new phonetic categories can be established for L2 sounds and when they can not

5. Specifically, for native speakers of Romance language, the learning of /p t k/ and /b d g/ proceeds differently because category formation is possible in the first but not second case
The End

Thanks for your kind attention

Above: a public park in Tuscania (VT)
Evidence of phonetic system pressure in the L1 production of Italian-English bilinguals is reminiscent of a finding reported for monolingual speakers of two varieties of French.

Caramazza & Yeni-Komshian (1974) found that – presumably due to “contact” with English – monolingual French Canadians were more likely to produce French /b d g/ with short-lag VOT values than monolingual French speakers in France.

<table>
<thead>
<tr>
<th>Dialect</th>
<th>% Lead VOT</th>
<th>% Short-lag VOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quebec (Canada)</td>
<td>41%</td>
<td>59%</td>
</tr>
<tr>
<td>Nantes (France)</td>
<td>94%</td>
<td>6%</td>
</tr>
</tbody>
</table>
L2-on-L1 effects

The French Canadians, in turn, produced French /p t k/ with longer VOT values than the French monolinguals living in France.
What is VOT?

VOT in word-initial stops may co-vary or co-exist with other acoustic properties having perceptual cue value:

1. height of the following vowel (e.g., Mortensen & Tøndering, 2013);
2. degree of stress/emphasis (Lisker & Abramson 1967)
3. speaking rate, including duration of the following vowel (e.g., Theodore et al. 2009) which may vary according to the number of syllables in the word and position in utterance;
4. speech clarity (e.g., Kessinger & Blumstein, 1997; Smiljanić & Bradlow, 2005)
5. socio-phonetic factors (e.g., Docherty et al., 2011)
6. burst and aspiration intensity (e.g., Repp, 1979)
7. F0 onset frequency and movement pattern (e.g., Hombert, Ohala & Ewan, 1979; Dmitrieva et al. 2015)
8. F1 onset frequency and movement pattern (e.g, Hillenbrand, 1984)
9. spectral tilt, H1-H2 (e.g., Kong et al., 2012)