The role of input in the acquisition of L2 stops

James Emil Flege
Professor Emeritus
School of Health Professions
University of Alabama at Birmingham

SICSS 2017: 2017 Seoul International Conference on Speech Sciences

10–11 November 2017
Seoul National University, Korea

aSupported by the National Institute for Deafness & Other Communicative Disorders (USA)

bContact: jeflege@uab.edu  website: www.jimflege.com
The aim of this talk is to examine the role of input in the perception and production of L2 stop consonants.

I will focus exclusively on the Voice onset time (VOT) dimension, reviewing results we obtained in research carried out with native Italian learners of English in Canada.

The results we obtained have convinced me that input is a far more important determinant of success in L2 speech learning than the age at which L2 learning begins.
This conclusion is somewhat controversial

Many believe that native vs. non-native differences are often due to differences in age of L2 learning (AOL)

AOL is considered important because it is thought to index L2 learners’ state of neurological or cognitive development when learning begins

We selected Italian immigrants for our research in Canada based on their age of arrival (AOA) in Canada from Italy

AOA is important in our research because it conditioned the kind of experience our participants had with English, defining both the quantity and quality of input they received over the course of their lives
Introduction

The notion of a “Critical Period” has been applied to L2 speech learning since 1967, when Eric Lenneberg observed that

• after about the age of 12 years, foreign accents can not be “overcome easily” and the learning process itself requires “conscious and labored effort” (Lenneberg 1967, p. 176)
Lenneberg proposed that

- for language learning to be fully effective, it must occur before the closure of a **critical period** at about the age of 12 years;

- after the stage of development evident at that age, brain functions needed for speech acquisition may retain “some flexibility” but the “ability for self-organization and adjustment to the physiological demands of verbal behavior quickly declines” (1967, p. 158)
With respect to the role of input, Lenneberg suggested that

- after the closure of the critical period at about the age of 12 years, L2 learners are no longer able to make “automatic” use of input “from mere exposure” to the input (Lenneberg 1967, p. 176)

With respect to the use of the VOT dimension in the perception of L2 stop consonants, Stolten et al. (2014) recently noted that speech learning may be

“especially vulnerable to age-of-onset effects which, in turn, may be related to the early maturation of motor and auditory cortices of the brain” (p. 444)
Twenty years after Lenneberg published his famous monograph, I wrote an article questioning the relevance of the Critical Period Hypothesis (CPH) for L2 speech learning (Flege, 1987)

I noted that, as formulated, the CPH could not be disconfirmed and so was not an hypothesis is the true sense of the word

Even more importantly, I suggested that enthusiasm for the CPH tended to inhibit the search for valid explanations of age-related effects in L2 speech learning

Such as input
In fact, relatively little attention has been paid to the role of input in L2 learning.

Investigators who accept the CPH have tended either to deny the existence of input differences (DeKeyser, 2000) or, more often, to simply ignore them.

As I see it, the assumption that input differences are unlikely to explain differences between individuals and groups of L2 learners has limited progress in the field.

As an example:

The technology needed to accurately assess both the quantity and quality of L2 input already exists (see Flege, 2009, pp. 188-189). However, as far as I know, no one has ever exploited this technology in L2 speech research.
The Ottawa Italian-English project

Our 12-year program of research in Ottawa examined native speakers of Italian ("Italians" for short) who had immigrated to Canada in the 1950s and 1960s for economic reasons and had lived there for decades.
The Ottawa Italian-English project

The first study to be presented today was carried out in collaboration with Murray Munro and Ian MacKay.

We selected normal-hearing individuals living in Ottawa in a quasi-random manner on the basis of:

- gender
- age
- length of residence (LOR) in Canada
- age of arrival (AOA) in Canada

Above: Murray Munro and Ian MacKay
The Ottawa Italian-English project

We tested our Italian participants in an annex of Catholic church located in central Ottawa. This church has long been a center of the Italian-speaking community in Ottawa since it was constructed by the first Italian immigrants in 1913.

Our first study was carried out in 1992 and published as Flege et al. (1995a, b).

It examined ten groups of 24 Italians each who differed according to age of arrival (AOA) in Canada.

Above: St. Anthony of Padua, Ottawa ONT
The Ottawa Italian-English project

Here we see how often Italians in the ten AOA groups (max = 24) reported that English was the “better” of their two languages (blue). Early learners indicated that English was their better language, whereas Late learners usually indicated that it was Italian.

However, when we asked our participants which of their two languages they would “prefer” to keep if they could keep only one, nearly all responded “English” (orange) regardless of their AOA.
The Ottawa Italian-English project

The disparity between the “best” and “preferred” language responses reflects linguistic reality. Flege and MacKay (2011) found that Italians living in Ottawa often used Italian with family members and at church events.

In most other contexts, such as going to the barber or hairdresser (commercial transactions) or talking to a neighbor (personal interactions), they usually reported being more likely to use English than Italian.
The Ottawa Italian-English project

We eventually realized that our participants’ age of arrival (AOA) in Canada conditioned their future experiences with English. For example, most Italians who arrived in Canada before the age of 14 began school immediately and developed life-long relationships with native English classmates.

Most Italians who arrived slightly later in life, on the other hand, went to work immediately.

Rather than learning English from native English speaking classmates, these later arrivals tended to learn English from fellow Italian immigrants who spoke English with an Italian accent.
The Ottawa Italian-English project

The importance of **quality of input** can be readily understood by considering the effect of gender on degree of foreign accent.

Women who arrived in Canada before the age of 12 years soon began attending school.

They had a significantly better pronunciation of English than men having the same AOA and the same opportunity to learn English from native English classmates.

Note: *in this reanalysis I reassigned member of the original 10 groups of 24 each to four groups of 60 each.*
The Ottawa Italian-English project

However, women with a mean AOA of 19 years had significantly stronger foreign accents than men.

When women in this AOA group arrived in Canada, they normally worked at home.

Their primary model of English was provided by male relatives who worked outside the home.

Alas, these men usually spoke English with strong foreign accents.
The Ottawa Italian-English project

In addition to affecting quality of English input, we discovered that AOA also exerts an important long-term effect on the quantity of English input received.

Here we see the Italians’ self-estimates of % English and Italian use.

Participants who arrived as young children reported using English nearly four times as much as Italian.

Those who arrived as young adults used English only slightly more than Italian.
I’ll begin my review of results obtained in Ottawa by presenting a perception experiment (Flege & MacKay, in prep.) for which we recruited 18 native speakers of English and 190 Italians.

The Italians were highly experienced in English, having lived in Ottawa for an average of 40 years (range: 11 to 53).

Despite this, virtually all of them spoke English with detectable foreign accents.

Note: We evaluated degree of FA for 138 of the 190 NI participants who were assigned to AOA-defined subgroups of 23 each. All six NI groups received significantly lower ratings than the NE group.
Perceiving L2 stops

We assigned the 190 Italians to ten groups of 19 each based on their AOA in Canada.

The task of the Italians and members of the native English comparison group was to identify members of four 9-member natural-edited VOT continua. The continua ranged from:

- /b/ to /p/ (bi/-/pi/, /bu/-/pu/)
- /d/ to /t/ (/di/-/ti/, /du/-/tu/)

As expected (Bohn & Flege, 1993; Hazan & Boulakia, 1993) the Italians consistently labelled the lead VOT endpoints as voiced ("b" or "d") and the long-lag endpoints of the four continua as voiceless ("p" or "t")

Details: the CV stimuli in each continuum were randomly presented in isolation four times each. The continua were presented in counterbalanced order. Following a familiarization phase, participants identified stimuli as voiced ("b", "d") or voiceless ("p", "t")
Perceiving L2 stops

We focused instead on how participants identified the two short-lag stimuli found in each continuum:

- We reasoned that the Italians could **not** establish new phonetic categories for short-lag variants of English /b d g/ because their Italian /p t k/ categories already “occupied” this VOT range;

- Given the SLM hypothesis that phonetic learning occurs when adequate phonetic input is available, we predicted that our Italian participants would “re-structure” their Italian /b d g/ categories (see MacKay et al., 2001)

Detail: *One was an unedited token of stops produced by a monolingual native speaker of English. The other was copy of these tokens from which a 5-msec interval between located between stop release and vowel onset was removed*
Perceiving L2 stops

This led to the expectation that the Italians, just like English monolinguals, would identify English stops having both lead and short-lag VOT as phonologically voiced (“b” or “d”)

Labelling the short-lag English stops as “p” or “t”, on the other had, would indicate a lack of speech learning, perhaps due to a diminished capacity for perceptual learning
Perceiving L2 stops

As expected, the native English speakers identified stimuli having both lead and short-lag VOT (*blue arrows*) as voiced ("b" or “d”)

Details: *the CV stimuli in each of the four continua were randomly presented in isolation in counterbalanced order. Following a familiarization phase, participants identified four randomly presented tokens of each stimulus as voiced or voiceless*
Perceiving L2 stops

The Italians closely resembled the native English speakers when identifying endpoint stimuli in the bi-to-pi continuum.

As predicted, they also closely resembled the native English speakers in labelling, as “b”, the two stimuli having short-lag VOT.

This finding supported our “category restructuring” hypothesis.

Note: for clarity I omit the ID functions of five of the ten native Italian groups.
Perceiving L2 stops

Responses to the *bu-pu* continuum also supported the category re-structuring hypothesis. The Italians labelled the short-lag stimuli as “b” only slightly less often than the native English speakers did ($M = 90.8\%$ vs $97.2\%$)
Perceiving L2 stops

For the /d/-to-/t/ (*di-ti, du-tu*) continua, on the other hand, fewer Italians than native English speakers identified short-lag stimuli in an English-like way, that is, as “d” (*M* = 65.7% vs 95.1%)

A question we need to answer is: why did some Italian Late learners label the short-lag English stops in an Italian-like way (i.e. as “t”)

![Graph showing % "t" judgments for di-to-ti and du-to-tu](chart.png)
Perceiving L2 stops

Here we see the frequency of Italian-like ("t" or "p") responses for the native English speakers and the AOA-defined groups of Italians.

Six native Italian groups labelled short-lag English /d/ tokens as "t" significantly more often than the native English speakers ($p < 0.05$);

One Italian group labelled short-lag English /b/ tokens as "p" significantly more often than the native English speakers ($p < 0.05$).
Perceiving L2 stops

Did differences between the native English speakers and the Italians who arrived in Canada after the age of 12 years derive from differing states of neurological development when the native speakers and Italians first began to learn English?

Above: Nearly all of our native Italian participants embarked at Pier 21 in Halifax, Nova Scotia between 1948 and the mid-1970s.
Perceiving L2 stops

I don’t think so. First, consider individual differences in performance. We identified pairs of Italians who had the same AOA in Canada but differed in their labelling of short-lag English stops.

Despite having the same AOA, one member of each pair of participants identified the short-lag English stops in an English-like way (“d”), the other in an Italian-like way (“t”).
Perceiving L2 stops

Other problems for a Critical Period-type explanation are that:

1. Nearly all of our Italian participants labelled short-lag tokens of English /b/ in an English-like way (i.e., as “b”)
2. Most Italians labelled the short-lag tokens of /d/ in an English-like way (i.e., as “d”)

How could maturational state at the time of immigration affect just some native speakers of Italian for just one place of articulation?
Perceiving L2 stops

Also, results of the study enabled us to reject Lenneberg’s 1976 hypothesis that, following the closure of a critical period, L2 learners lose the ability to make effective use of input.

Each of the four VOT continuum had an unedited short-lag English stop and a copy from which 5 msec of VOT was removed.

If “post-Critical Period” learners of an L2 are unable to make “automatic” use of input “from mere exposure” (1967, p. 176) we would not expect Late learners to distinguish pairs of short-lag stops differing by 5 just msec of VOT.
Perceiving L2 stops

However, the Italians were quite sensitive to these small VOT differences. Increasing VOT by 5 msec increased the frequency of their “p” judgments by 5% and the frequency of their “t” judgments by 18%

The VOT differences were highly significant in a two-way ANOVA ($p < .0001$) but did not interact with the Group factor.

This means that the Italians’ ability to make “automatic” use of VOT input was unrelated to their AOA in Canada, and so unrelated to age of L2 learning.
Perceiving L2 stops

As seen here, the later our native Italian participants had arrived in Canada, the less English input they had received. Thus an alternative hypothesis is that the observed between-group differences were due to differences in input.

Before presenting an analysis of groups defined on the basis of L2 input, however, I first want to describe the technique we used to validate our Italian participants’ self-estimates of % English use.
Perceiving L2 stops

We asked the 190 native speakers of Italian in our study to indicate if they would be more likely to use English or Italian in 32 social contexts.

In some contexts most participants indicated being more likely to use English than Italian, presumably because few opportunities existed for using Italian.

<table>
<thead>
<tr>
<th>Item</th>
<th>Context</th>
<th>N of “English” responses (max=190)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>buying a refrigerator</td>
<td>188</td>
</tr>
<tr>
<td>7</td>
<td>while at a pharmacy</td>
<td>187</td>
</tr>
<tr>
<td>33</td>
<td>while buying a car</td>
<td>178</td>
</tr>
<tr>
<td>10</td>
<td>at a doctors office</td>
<td>183</td>
</tr>
<tr>
<td>5</td>
<td>attending a school play</td>
<td>177</td>
</tr>
</tbody>
</table>

Note: one of the original 33 items was excluded from analysis because too many participants indicated that the social context did not apply to them.
Perceiving L2 stops

Fewer Italians indicated being more likely to use English in other social contexts, however:

<table>
<thead>
<tr>
<th>Item</th>
<th>Context</th>
<th>N English</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>talking to your oldest child</td>
<td>113</td>
</tr>
<tr>
<td>19</td>
<td>while at the barber/hairdresser</td>
<td>102</td>
</tr>
<tr>
<td>27</td>
<td>talking to someone painting your house</td>
<td>91</td>
</tr>
<tr>
<td>24</td>
<td>visiting someone in the hospital</td>
<td>77</td>
</tr>
</tbody>
</table>

And in still other contexts, most Italians said they would be more likely to use **Italian** than English:

<table>
<thead>
<tr>
<th>Item</th>
<th>Context</th>
<th>N English</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>talking to your younger sister or brother</td>
<td>52</td>
</tr>
<tr>
<td>30</td>
<td>visiting a close friend</td>
<td>55</td>
</tr>
<tr>
<td>28</td>
<td>talking to someone after mass</td>
<td>43</td>
</tr>
<tr>
<td>2</td>
<td>talking to your father</td>
<td>3</td>
</tr>
</tbody>
</table>
Perceiving L2 stops

For each participant we determined the percentage of the 32 contexts in which English was the more likely of two languages to be used

The “English use in context” scores were highly correlated with the Italians’ self-estimates of % English use, $r(8)= 0.965$ for groups, $r(188) = 0.52$ for individuals

Also, our Italian participants’ estimates of overall % use of English correlated with the average of estimated use at English at home, at work and in various social contexts, $r(188) = 0.71$
Perceiving L2 stops

Here, now, is the analysis for groups defined on the basis of years of English input (LOR * % English use). The effect of Group in this analysis was non-significant for the b-to-p continua. For the t-to-d continua just 5 (rather than 6) Italian groups differed significantly from the native English speakers (asterisks)

We get a slightly better fit to the data for groups defined by L2 input rather than AOA

However, an examination of individual data leads me to doubt that the frequency of Italian-like responses to short-lag stops was determined by input
Perceiving L2 stops

We compared the performance of pairs of Italian participants who were matched for years of English input but differed in how often they labelled short-lag stops in an English-like way (“d”).

Regardless of years of English input received, one member of each pair of participants identified the short-lag stops in an English-like way (“d”), the other in an Italian-like way (i.e., as “t”).
Bilingual dominance

Having rejected both age of L2 learning and amount of L2 input as explanations for the labelling of short-lag stops in this study, the question remains:

What accounts for the between-group differences?

The answer seems to be: differences in bilingual dominance
Bilingual dominance

Previous research suggested that our Italian participants’ labelling of short-lag stops might be influenced by dominance

Elman et al. (1977) examined the identification of short-lag bilabial stops as “b” or “p”. English monolinguals labelled these stimuli as “b” far more often than Spanish monolinguals. Bilinguals were tested in two language “sets”. The overall decrease in English-like identifications (i.e. labelling short-lag stops as “b”) when switching from the English to the Spanish set averaged just 9%. However, much larger shifts were seen for “strong” than “weak” bilinguals.

According to Hazan and Boulakia (1993, 19) the labelling of short-lag stops by French-English bilinguals may depend both on the test procedures that are used and on what I refer to as “relative language dominance”
Bilingual dominance

Relative language dominance

We (Flege & MacKay, in prep.) asked our native Italian participants to rate their ability to speak, understand, pronounce and read/write both English and Italian using 7-point EAI scales.

Our measure of dominance was the **ratio** of self-estimated ability in English compared to Italian. On this scale:

- a value of 1.0 might be called “mixed” dominance.
- values >> than 1.0 would indicate dominance in English.
- values << than 1.0 would indicate dominance in Italian.
Bilingual dominance

What determines relative language dominance? Patterns of language use over long periods of time

The earlier in life the Italians had arrived in Canada, the more English input they had received

This is because the earlier arrivals had usually lived longer in Canada and tended to use English more frequently on a daily basis than the later arrivals
Bilingual dominance

Differences in language use, in turn, led to differences in relative language dominance.

Here we see groups of 19 native Italian participants each defined on the basis of years of English input.

The more English input the Italians had received, the higher were their relative dominance scores.
Bilingual dominance

Here we see the frequency of English-like ("d") categorizations of short-lag English stops by groups of 19 Italians each defined on the basis of relative dominance rather than AOA.

The Italians who had become English-dominant did not differ significantly from the native English speakers.

However, groups consisting of individuals who were not English dominant did differ from the native English speakers.

Note: the asterisks indicate significant differences from the NE group in one-sample $t$-tests (Bonferroni-corrected $p < .05$).
Bilingual dominance

The unusually strong effect of dominance seen in this study was due to two factors:

1. Our native Italian participants, when tested, were in a bilingual language mode with both their English and Italian phonetic systems activated

2. The identification task we used tapped a phonemic rather than phonetic processing level

Prior to identifying stops the Italians had responded to a detailed questionnaire focusing on bilingualism and then heard and repeated English and Italian sentences in alternating blocks
Importantly, there was no “correct” way for the Italians to label the short-lag stops. The identification task offered a binary choice between phonemes, not a choice between three phonetic categories.

The CV stimuli were presented in isolation, without pre-cursor material that would define the stimuli as being “English” or “Italian”.

A better testing format (see Appendix 1) would have been a lexical error detection task which yields responses at a phonetic processing level (see Flege & MacKay, 2004, Exp. 4).

Had this procedure been used to test participants in an English language mode, I think the Italian-like responses we saw earlier would disappear.

More research is needed, of course.
Perceiving L2 stops

To summarize

1. The Italians usually labelled both lead and short-lag VOT tokens of English /b/ and /d/ as phonologically voiced (/b/, /d/)

This demonstrated phonetic learning

About 1/3 of the responses obtained for short-lag stimuli in the d-to-t continua were Italian-like, but this seems to have been due to uncertainty as to what the “correct” label should be

The results support the hypothesis that our native Italian participants re-structured Italian phonetic categories to accommodate both lead and short-lag phonetic variants of phonologically voiced English stops
Perceiving L2 stops

To summarize

2. The Italians consistently identified long-lag tokens of English /p/ and /b/, even though these stops are realized as short-lag stops in the L1

This does not, however, demonstrate phonetic learning

Why not? Because we would expect even Italian monolinguals to do the same simply as the result of inter-lingual identification (see Bohn & Flege,

Given this uncertainty, I’ll now review a study examining the identification of English stops in a different way
Perceiving L2 stops

MacKay et al. (2001) examined the identification of English stops by 72 Italians living in Ottawa. As seen here, Early and Late learners were subdivided according to amount of Italian use (“high” vs “low”)

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Average % use of Italian$^a$</th>
<th>use of Italian in social contexts$^b$</th>
<th>N of interlocutors in Italian$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early-low</td>
<td>18</td>
<td>7%</td>
<td>12%</td>
<td>2.8</td>
</tr>
<tr>
<td>Early-high</td>
<td>18</td>
<td>43%</td>
<td>39%</td>
<td>5.0</td>
</tr>
<tr>
<td>Late-low</td>
<td>18</td>
<td>10%</td>
<td>18%</td>
<td>2.6</td>
</tr>
<tr>
<td>Late-high</td>
<td>18</td>
<td>53%</td>
<td>51%</td>
<td>4.4</td>
</tr>
</tbody>
</table>

$^A$ These values were the average of estimated use of Italian in the past 5 years, months, and weeks preceding the test;
$^B$ Participants estimated Italian use in 10 social contexts including home use, use with family members, at church, with friends and while shopping
$^C$ Total number of people with whom Italian was used “always” or “sometimes”
**Perceiving L2 stops**

**MacKay et al. (2001) procedures**

The stimuli consisted of non-words spoken by two monolingual native speakers of English. The non-words were formed by inserting tokens of /b d g p t k/ into a /'Cama/ frame.

The /b d g/ tokens were all produced with short-lag VOT ($M = 15$ msec). The response alternatives offered on each trial were

<table>
<thead>
<tr>
<th>p</th>
<th>t</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>d</td>
<td>g</td>
</tr>
</tbody>
</table>

To avoid ceiling effects, the stimuli were embedded in noise.
Perceiving L2 stops

None of the four Italian groups differed from the native speakers of English in identifying the long-lag English /p t k/ tokens (blue bars)

This is consistent with the hypothesis that the Italians had established new phonetic categories for English /p t k/
Perceiving L2 stops

For the short-lag tokens of /b d g/, however, the Late learners who continued to use Italian often (“Late-High”) made significantly more identification errors than the native English speakers (35% vs 11%, p < .05)
Perceiving L2 stops

Here we see identification errors that were due exclusively to the voicing feature. Once again, just Late-High differed significantly from the native English speakers ($p < 0.05$).

These results support the hypothesis that the Italians re-structured their Italian /b d g/ categories for use in English, but that success in doing depended on input.

Specifically, the unusually high error rate for Late-High seems to have resulted from insufficient English input.

$^{a}$Errors in the last side combined errors involving both the voicing feature and place of articulation
Perceiving L2 stops

It makes no sense to attribute the difference between the Late-High group and the native English speakers to age of L2 learning inasmuch as the Late learners who used Italian seldom (Late-low) did not differ significantly from the native English speakers.

Input seems to be the key here.

Members of the Late-High group had received significantly less English input than members of the other three Italian groups ($p < .05$).
Perceiving L2 stops

In fact, the input differences were dramatic. Here we see the mean years of English and Italian input that participants in the four Italian groups had received over the course of their lives.

Members of the Late-High group had received significantly more Italian than English input ($p < .05$).

There was no difference, or a significant difference in the opposite direction, for the other three Italian groups.

*Calculation of total language input: English = %English * LOR; Italian = %Italian * LOR + AOA*
Perceiving L2 stops

What should be considered “sufficient” input?

I proposed yesterday in my tutorial lecture that L2 learners may need as much as 10 years of native-speaker input to fully establish new phonetic categories for /p t k/

However, category formation is not possible for /b d g/

As we will see later, the question of interest for /b d g/ is not how much input the Italians have received but, rather, what proportion of lead and short-lag realizations of /b d g/ they have heard
So far I have been speaking of “L2 learning”. However, that term is a bit misleading for our population of Italian immigrants in Ottawa.

It is impossible to understand the production and perception of phonetic segments in the L2 of bilinguals without first understanding how the phonetic systems of bilinguals are organized.

According to the Speech Learning Model (Flege, 1995 ff.), the elements making up a bilinguals’ two phonetic subsystems interact with one another because they exist in a “common phonological space”.

This leads to us to expect an influence of L1 segmental production and perception of L2 segments and vice versa.
Bilingual dominance

With respect to labial stops, I hypothesize that our bilingual participants in Ottawa had three distinct phonetic categories:

1. a re-structured Italian “b” category used in both English and Italian
   
   Like the phonetic categories developed by English monolinguals in childhood, the bilinguals’ re-structured category encompassed both lead and short-lag VOT variants of /b/

2. a short-lag [p] category for use in Italian

3. a long-lag [pʰ] category for use in English
Before immigrating to Canada, our participants’ Italian phonetic system was organized as shown here. In this system, phonologically voiced (/b d g/) and voiceless (/p t k/) stops are realized with lead and short-lag VOT values, respectively.
As discussed in my tutorial lecture, when the Italians began learning English they encountered a phonetic system in which /p t k/ are consistently realized with long-lag VOT whereas /b d g/ are realized with either lead or short-lag VOT.
As our native Italian participants became bilingual they developed a phonetic sub-system for use when processing speech in an English language mode.

As the result of extensive experience in English, the structure of the Italians’ English phonetic sub-system came to closely resemble that of English monolinguals, regardless of when they began learning English as an L2.

Note: Differences between the Italian and native English speakers existed, of course, in the relative weighting of the phonetic properties defining /b d g/ but a discussion of cue weighting goes beyond the scope of this talk.)
As the result of developing an English phonetic sub-system, their Italian phonetic sub-system, used for processing speech in an Italian language mode, underwent change.

By hypothesis

The Italian-English bilinguals maintained their short-lag phonetic categories for Italian /p t k/

They used their “re-structured” /b d g/ categories when processing both Italian and English /b d g/
If my reasoning is correct, two things should happen. First, we should observe the production of Italian /b/ with a trace of English, that is, with instances of partial pre-voicing and short-lag VOT.

Second, we should observe increases in the VOT of Italian /p t k/.

Why? If Italian /b d g/ are realized as short-lag stops, the VOT of Italian /p t k/ must increase to avoid the production of phonologically voiced and voiceless stops having the same VOT values.
Producing /p t k/

This section will examine production of /p t k/ in word-initial position by participants in the Ottawa project.

Background

Previous research has suggested that the perceived phonetic dissimilarity of English and Italian realizations of /p t k/ is sufficient to permit the formation of new phonetic categories in English (see Flege 1995 ff).

Nearly all of the Italians in our research had received more than 10 years of native-speaker input, and so we expected accurate production by nearly all of them, regardless of AOA.
Producing /p t k/

Flege et al. (1995b) used a delayed repetition task to elicit the production of English words, some beginning with /p t k/

Model presented via loudspeaker: X is the next word

Participant’s response: Now I say _X_

X: pick, tack, tag, cap, cab

Delayed repetition task

This technique was used to reduce the influence of orthography and unwanted variation in VOT due to changes in speaking rate

Native speaker models were presented at a fixed interval. The materials intervening between each model and its repetition prevented direct imitations from sensory memory
Producing /p t k/

Here we see the mean VOT values obtained for ten AOA-defined groups of 24 Italians each.

No native Italian group differed significantly from the native English speakers in producing English /k/. However, two groups of Italians produced /p/ with significantly shorter VOT values, and six did so for /t/ (asterisks).
Producing /p t k/

This is not what I was expecting. Moreover, input differences can **not** account for these unexpected between-group differences.

Here we see the mean VOT values produced by two subgroups of 22 Italians each who were matched for AOA.

One subgroup had received more than 10 years of native-speaker input, the other subgroup had not.
Producing /p t k/

The Italians who had received 25 years of English input as well as those who had received just 8 years produced English /p/ and /t/ with significantly shorter VOT values than the NE speakers ($p < 0.01$)

Some explanation other than input is needed

An explanation exists, but before presenting it we must first consider a second set of data obtained from the same 240 native Italian participants
Producing /p t k/

Flege et al. (1995a) obtained VOT measurements for stops found in one five English sentences that were originally elicited for the evaluation of foreign accent.

As seen here, the English sentences were elicited using a delayed repetition technique.

<table>
<thead>
<tr>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voice 1 (question)</td>
</tr>
<tr>
<td>Voice 2 (model)</td>
</tr>
<tr>
<td>Voice 1 (question)</td>
</tr>
<tr>
<td>... Pause</td>
</tr>
<tr>
<td>... Beep, and then repetition of the model</td>
</tr>
</tbody>
</table>
Here we see the foreign accent ratings obtained for the sentence of interest, *Paul ate carrots and peas*

Significantly lower foreign accent ratings were obtained for seven AOA-defined native Italian groups (asterisks) than for native speakers of English (red)

Note: Sentences produced by many *Early learners* also received significantly lower ratings than were obtained for the NE speakers
Producing /p t k/

We measured VOT in the three word-initial stops found in the “Paul” sentence. No native Italian group differed significantly from the native English speakers for VOT in the three stops that were measured.

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

Note: these data were never published. At the time we were looking for segmental differences that predicted degree of foreign accent and there was, of course, no relation between VOT and FA.
For a closer look, I re-assigned the 240 Italians to three groups of 80 each. Group “AOA 16-23” consisted entirely of Late learners. This group of 80 Italian Late learners did not differ from the native English speakers for any of the three stops examined.
Language mode and activation

Variation in speaking rate might, in part, explain the differing results obtained for stops found in English words and in the “Paul” sentence. This is because speaking rate was controlled better in the elicitation of the sentences than the words.

The three Italian groups and the native English speakers produced sentences with nearly identical durations.

However, the Italians with a mean AOA of 19 years produced significantly shorter vowels than the native English speakers did.
Language mode and activation

The Late learners’ shorter vowels indicate a slightly faster speaking rate, which is expected to shorten VOT (see Appendix 3).

However, differences in speaking rate (i.e., vowel duration) alone can probably not account for the observed differences between Late learners and native English speakers in the production of words.

I suspect that the difference in English word production between the native English speakers and some Italian Late learners was due to code switching by the Late learners.
Let’s return to the word production experiment

Flege et al. (1995b) examined the distribution of VOT values obtained for Early and Late learners. Unlike the native English speakers and Early learners, some Late learners produced English /p/ and /t/ with Italian-like short-lag VOT values in English words. Why?
I hypothesize that some Late learners code switched, producing English words with an Italian pronunciation.

To evaluate this hypothesis, I identified the 33 Italians who usually realized English /p t/ with short-lag VOT, and 33 others who had the same AOA but never produced English /p/ and /t/ with short-lag VOT.

As expected, the two subgroups of Late learners differed significantly for VOT in words but not sentences.
Language mode and activation

The two sub-groups of 33 each were matched for AOA ($M = 18$ years for both). Not surprisingly, they spoke English with equally strong foreign accents.
Language mode and activation

We can infer that the Italian phonetic sub-systems of all 240 Italians in our study were activated when they produced the English words. This is because

• they were tested in a place where they usually spoke Italian (i.e., their parish church)

• they had responded earlier to a detailed questionnaire regarding their bilingualism

• they had just heard and re-produced English and Italian sentences in alternation

I hypothesize, however, that some Late learners were more accustomed to code switching than others
Language mode and activation

The 33 Late learners who produced English words with Italian-like VOT reported using Italian more than English at home. The reverse held true for the 33 Late learners who never produced English /p t/ with short-lag VOT (p < .05 by Mann-Whitney U-tests)

I speculate that Late learners who often spoke Italian at home were more accustomed to producing English words with an Italian pronunciation (code switching) than those who seldom used Italian at home.
Language mode and activation

The need for speculation (ugh!) might have been avoided had our research, carried out in 1992, controlled for both language mode and activation.

It did not.

To control language mode and activation in the future, researchers should consider using either a repeated measures (see e.g., Magloire & Green, 1999) or randomized block design (e.g., Antoniou et al., 2010).

Both design types have practical advantages and disadvantages (see Appendix 2).
Producing /b d g/

In this section we will examine the production of /b d g/ in both English and Italian. Before doing so, we must consider the role of “communicative pressure” in phonetic sub-systems.

Romance language learners of English are under communicative pressure to create new phonetic categories for English /p t k/.

VOT values that are too short contribute to the perception of foreign accent in English (Flege & Munro, 1992) and may trigger lexical confusions.

No such pressure exists for the creation of new short-lag categories for English /b d g/. In fact, it is a priori implausible to think that Italian learners of English would create new short-lag L2 phonetic categories for /b d g/ “on top of” already existent L1 categories for /p t k/.

The SLM predicts that learning will nevertheless occur.
The Speech Learning Model

Background

As discussed in my tutorial lecture, monolingual English adults produce /b d g/ in various ways: with full pre-voicing, pre-voicing that ends prior to stop release, or with short-lag VOT.

I suggested that children who learn English as an L1 respond to differences in the input they receive, creating either “lead” or “short-lag” phonetic categories for English /b d g/.

By the time our Italian participants began learning English they had already established “lead” VOT categories for Italian /b d g/.

I hypothesize that as a result of exposure to short-lag variants of English /b d g/ they re-structured their existing categories to encompass both VOT variants.

Note: as discussed in the tutorial lecture, the PRRs developed by English monolinguals who establish “lead” categories for /b d g/ do not yield pre-voiced stops in which pre-voicing always continues until release, as for monolingual speakers of Romance languages.
MacKay et al. (2001) examined the production of English /b/ by four groups of Italians who differed orthogonally in AOA and frequency of Italian use (“high” vs “low”). They 72 Italians and 18 native English speakers produced English words such as *bed* and *bad* in isolation after hearing stimulus sentences of the form

\[
\text{CVC is the next word to say}
\]

We did not measure VOT in /b/ because, given the phonetic structure of English, it made no sense to do so. Instead, we classified the 20 /b/ tokens available for each participant as having been produced with

- **full pre-voicing** that continued until stop release;
- **partial pre-voicing** that ceased before stop release;
- **short-lag VOT**

The restructuring of L1 categories
Perceiving L2 stops

Earlier we saw that Late learners, especially Late-H (high use of Italian), misidentified short-lag /b d g/ tokens as “p t k” more often than Early learners or native English speakers.

According to the SLM, the phonetic category representations developed for L2 speech sounds guides the development of the phonetic realization rules (PRRs) needed to produce those sounds.

This leads us to expect production differences between Late-H and the other groups.
Category restructuring in L2 speech learning

All four Italian groups produced English /b/ with short-lag VOT values (orange), something not seen in Italian. This demonstrates phonetic learning.

Importantly, however, all four Italian groups produced /b/ with short-lag VOT values significantly less often than the native English speakers did.

(p < .05 by Bonferroni-corrected t-tests)
Category restructuring in L2 speech learning

Some members of all four Italian groups produced /b/ with *partial* pre-voicing that ended before stop release. This is also something not seen in Italian, and so demonstrates phonetic learning.

In fact, one group of Italians, Early-Low (use of Italian), produced English /b/ with partial pre-voicing even more frequently than the native English speakers did.

Note: half of the English /bVC/ models were produced with pre-voicing, half with short-lag VOT. A preliminary analysis revealed no effect of the model on participants’ productions.
Category restructuring in L2 speech learning

Italian /b/ is produced with full pre-voicing. None of the Italian bilinguals examined always produced English /b/ with full pre-voicing. However, both groups of Late learners produced English /b/ with full pre-voicing significantly more often than the native English speakers.

Although the Early learners resembled the native English speakers to a greater extent than the Late learners, this does not necessary mean that they had learned English more or better than the Late learners.
Category restructuring in L2 speech learning

The pattern just seen in production is similar to the pattern we saw earlier for perception. Here we see the relation between the two

Italian-like identifications of short-lag tokens of English /b d g/ (as “p t k”) implies a reliance on closure voicing as a perceptual cue

The more our Italian participants relied on pre-voicing as a perceptual cue, the more often they produced /b/ with full pre-voicing
The restructuring of L1 categories

I hypothesize that between-group differences in both the perception and production of /b/ arose from differences in the relative frequency with which the Italians had heard /b/ produced with pre-voicing and short-lag VOT.

To evaluate this hypothesis, I estimated the percentage of pre-voiced tokens of /b/ the 72 native Italian and 18 NE participants were likely to have heard over the course of their lives.

This is a novel approach, and so I will briefly explain how these lifetime language input estimates were obtained.
The restructuring of L1 categories

The estimates were based on values obtained experimentally and from a detailed Language Background Questionnaire.

### Lifetime language input estimates

1. The frequencies with which English and Italian monolinguals produce /b/ with pre-voicing and short-lag VOT were based on values obtained by MacKay et al. (2001), Experiment 1: English - 24.6 vs. 75.6%, Italian - 99.5 vs. 0.05%;

2. Bilingual participants were assumed to have heard only Italian before immigrating to Canada;

3. Years of English and Italian input in Canada were estimated by multiplying years of residence in Canada by participants’ self-estimated % use of English and Italian

I assumed that the NE speakers received only English input. Estimates for them were obtained by multiplying chronological age by the English values in #1 above.
The restructuring of L1 categories

Here we see estimates of how often the native English speakers and four groups of Italians had heard /b/ produced with pre-voicing over the course of their lives.

The estimated frequency with which members of the five groups had heard /b/ produced with pre-voicing captured 80% of the variance in the frequency with which they produced /b/ with pre-voicing.

Note: weaker but still significant correlations were obtained for groups and individual participants when we considered just the Italians.
The restructuring of L1 categories

Flege & MacKay (in prep.) further examined the influence of input on the production of /b/

Participants

As described earlier for stop identification, we recruited 190 Italians in Ottawa who were assigned to ten groups of 19 each according to AOA. The mean AOA values for the ten Italian groups ranged from $M = 8$ to 27 years

Speech materials

The Italians produced alternating blocks of English and Italian materials; the native English speakers produced only English materials

We examined the /b/ in Barbara, which was the first word in translation equivalent English and Italian sentences
Once again, we used a delayed repetition task to elicit production

The three English and two Italian blocks (presented in alternation) consisted of 10 question-answer pairs (see examples below), yielding 9 tokens of English /b/ and 6 tokens of Italian /b/ for analysis

<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><em>Barbara is Patty’s sister</em></td>
</tr>
<tr>
<td>7</td>
<td>Who’s taller?</td>
<td><em>Barbara is a little taller than Patty</em></td>
</tr>
<tr>
<td>8</td>
<td>Who did you invite to dinner?</td>
<td><em>Barbara and Patty will be coming</em></td>
</tr>
</tbody>
</table>

Above: three of 10 pairs used to elicit production of /b/ in English
The restructuring of L1 categories

Crucially, the native-produced “Barbara” stimuli used to elicit production in English and Italian differed phonetically.

The two-syllable English models began with a short-lag /b/ whereas the three-syllable Italian models were all produced with lead VOT.

<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><em>Barbara is Patty’s sister</em></td>
</tr>
<tr>
<td>7</td>
<td>Who’s taller?</td>
<td><em>Barbara is a little taller than Patty</em></td>
</tr>
<tr>
<td>8</td>
<td>Who did you invite to dinner?</td>
<td><em>Barbara and Patty will be coming</em></td>
</tr>
</tbody>
</table>
The restructuring of L1 categories

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

Most Italians showed the opposite pattern, producing significantly more pre-voiced than short-lag stops ($p < .05$)

The one exception to this pattern were the Italians having a mean AOA of 8 years.
The restructuring of L1 categories

The difference between AOA-8 and the other Italian groups was probably due to input.

Here we see mean years of English and Italian input received by members of the ten Italian groups. AOA-8 was the only group to have received significantly more English than Italian input.

Note: *Years of English input was calculated as LOR * % English. Years of Italian input was LOR * % Italian + AOA (age at the time of emigration to Canada, prior to which Italian was used 100% of the time.*
The restructuring of L1 categories

As predicted by the category re-structuring hypothesis, the frequency with which the Italians produced English /b/ and Italian /b/ with pre-voicing was correlated [$r(8) = 0.77$ for groups, $r(188) = 0.71$ for individuals]

The English-Italian correlations would have been even higher without group AOA-8 who, as we saw, differed substantially from the other groups in terms of input.
The restructuring of L1 categories

Using the procedure described earlier, we (Flege & MacKay, in prep.) estimated the percentage of times that members of the ten AOA-defined Italian groups had heard /b/ produced with pre-voicing over the course of their lives.

This analysis replicated the one presented earlier. Once again, the input estimates captured 82% of the variance in the frequency of pre-voicing in productions of English /b/.

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

The final study to be presented is that of Flege, Imai & MacKay (in prep.)

For this study we recruited 160 Italians living in Ottawa; they were assigned to four groups of 40 each based on AOA

<table>
<thead>
<tr>
<th>N (m/f)</th>
<th>Age of arrival</th>
<th>Length of residence</th>
<th>% English use</th>
</tr>
</thead>
<tbody>
<tr>
<td>15/25</td>
<td>4.6 1.2-7.1</td>
<td>46 37-53</td>
<td>81 25-99</td>
</tr>
<tr>
<td>18/22</td>
<td>9.7 7.2-12.2</td>
<td>44 30-55</td>
<td>77 30-98</td>
</tr>
<tr>
<td>20/20</td>
<td>14.8 12.6-17.1</td>
<td>42 32-51</td>
<td>65 20-98</td>
</tr>
<tr>
<td>17/23</td>
<td>19.8 17.3-23.3</td>
<td>39 28-47</td>
<td>53 9-95</td>
</tr>
</tbody>
</table>
L2-on-L1 effects

As expected from our previous research in Ottawa, the AOA-defined groups differed according to input.

The two groups of Early learners had received more English than Italian input over the course of their lives whereas the Late learners had received less English than Italian input.
L2-on-L1 effects

We used the same protocol as Flege et al. (1995b) to elicit English words beginning with /p t k/ and /b/ (22 tokens)

Data collection

We used the same protocol as Flege et al. (1995b) to elicit English words beginning with /p t k/ and /b/ (22 tokens)

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

After completing the English protocol, the NI speakers repeated Italian words beginning with /p t k/ and /b d g/ that they had heard over a loudspeaker. The production of Italian words by Italian monolinguals was elicited in Rome

Crucially: the native-produced English words used to elicit production of English /b/ were produced with short-lag VOT whereas the native-produced Italian words used to elicit production of /b d g/ in Italian were produced with full pre-voicing
L2-on-L1 effects

Here we see the production data for tokens of English /b/ produced in utterances of the form *Now I say boot*”

As expected for stops in this utterance position (see Flege & Brown, 1982) the native English speakers produced /b/ far more often with pre-voicing than short-lag VOT.
L2-on-L1 effects

Of most interest here is the frequency with which English /b/ was produced with partial pre-voicing (i.e., pre-voicing that ended before release), something that does not occur in Italian.

All four groups of Italians resembled the native English speakers in producing /b/ with partial pre-voicing (orange bars).
L2-on-L1 effects

Now let’s consider production in Italian. The results for Italian /b d g/ show the unmistakable influence of English. When compared to Italian monolinguals, all four groups of Italian-English bilinguals were found to have produced Italian /b d g/

- significantly less often with full pre-voicing;
- significantly more often with short-lag VOT

(p < .05 by Bonferroni adjusted t-tests)
L2-on-L1 effects

Here are the mean VOT values obtained for English /p t k/ by Flege et al. (in prep.). These data closely resemble the data obtained earlier using the same protocol (Flege et al., 1995b)

The two groups of Late learners (mean AOA = 15, 20) produced significantly shorter VOT values than the native English speakers did

(p < .05 by Bonferroni adjusted t-tests)
L2-on-L1 effects

Once again, a close look at the data for English /p/ and /t/ suggests the possibility of code-switching by some Late learners.

The two groups of Late learners (AOA-15, AOA-20) produced English /p/ and /t/ with short-lag VOT values significantly more often than either the Early learners (AOA-5, AOA-10) or the native English speakers.

Note: For this analysis “short-lag” was defined as the range of values observed for the production of Italian /p/ and /t/ by Italian monolinguals.
L2-on-L1 effects

Here we see the mean VOT values for Italian /p t k/. As predicted, the Early learners (AOA-5, AOA-10) produced the phonologically voiceless Italian stops /p t k/ with significantly longer VOT value than the Italian monolinguals.

In my view, this change in the production of Italian /p t k/ was due to phonetic system pressure, not to the “loss” of Italian /p t k/ (Major, 1992).

Note: the four groups of bilinguals were compared to Italian monolinguals in four t-tests. The probability values seen in the graph have undergone the Bonferroni adjustment.
L2-on-L1 effects

This interpretation was confirmed by a post-hoc analysis of the results obtained for sub-groups of 18 Italians each:

• 1: Early learners who frequently produced Italian /b d g/ with short-lag VOT;
• 2: Early learners matched for AOA who seldom or never produced short-lag VOT;
• 3: Italian monolinguals

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>AOA</th>
<th>/b/</th>
<th>/d/</th>
<th>/g/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequent production of</td>
<td>18</td>
<td>7.8</td>
<td>58%</td>
<td>83%</td>
<td>69%</td>
</tr>
<tr>
<td>short-lag stops</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infrequent production of</td>
<td>18</td>
<td>7.8</td>
<td>6%</td>
<td>5%</td>
<td>17%</td>
</tr>
<tr>
<td>short-lag stops</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italian monolinguals</td>
<td>18</td>
<td>--</td>
<td>1%</td>
<td>0%</td>
<td>6%</td>
</tr>
</tbody>
</table>
L2-on-L1 effects

As expected, Early learners who often produced Italian /b d g/ with short-lag VOT produced Italian /p t k/ with VOT values averaging 15 msec longer than both the Early learners who seldom/never produced short-lag VOT and the Italian monolinguals.

The lengthening of VOT in Italian /p t k/ is similar to an effect noted for monolinguals in a comparison of two French dialects, and is likely the result of the same phonetic mechanism (see Appendix 5).
Longitudinal effects

One last finding of Flege et al. (in prep.) is important for an understanding of the ability of Late learners to use the input that is available to them.

Some believe that L2 input may be important only in early stages of L2 learning and that, later on, learners make less or no use of input due to “fossilization” (see Appendix 5).

Our results suggest otherwise.
The VOT data I just presented were obtained by Flege et al. (in prep.) in 2003.

All 160 Italians in this study had been tested previously in 1992 using the same protocol (see Flege et al., 1995a,b).

The 1992 data for 150 Italians and 20 native English speakers were still available, permitting longitudinal analyses.
Longitudinal effects

The foreign accent ratings obtained for three sentences spoken by the 150 Italians in 1992 and 2003 were very similar, $r(148) = 0.98$

The same held true for ratings obtained for 20 native English speakers even though their ratings clustered at the high ("no foreign accent") end of the scale, $r(18) = 0.93$
Longitudinal effects

The VOT values obtained in 1992 and 2003 for English /p t k/ were also correlated, both for the Italians and the native English speakers.

We nevertheless wondered why somewhat weaker correlations were obtained for the two sets of VOT measurements than for the foreign accent ratings.

<table>
<thead>
<tr>
<th>Group</th>
<th>$n$</th>
<th>VOT correlation 1992 - 2003</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native Italian</td>
<td>150</td>
<td>$r(148) = 0.83^*$</td>
<td>0.001</td>
</tr>
<tr>
<td>Native English</td>
<td>20</td>
<td>$r(18) = 0.74^*$</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Note: VOT increased from 1992 to 2003 by a constant 4.15 msec for both NE and NI speakers due to a change in measurement procedures. Voicing onset was defined as beginning at the first upward-going zero crossing in 1992, but at the first positive peak in the following vowel in 2003.
We identified the Italians who showed the largest and the smallest VOT changes over time. For these two sub-groups of 20 each and the remaining 110 Italians the 1992 vs 2003, the VOT correlations approximated the near-perfect correlation obtained for the foreign accent ratings.

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>VOT change (msec)</th>
<th>Correlation 1992 - 2003</th>
</tr>
</thead>
<tbody>
<tr>
<td>NI showing largest VOT increases</td>
<td>20</td>
<td>+18.37</td>
<td>r(18)=0.96*</td>
</tr>
<tr>
<td>NI showing largest VOT decreases</td>
<td>20</td>
<td>-18.06</td>
<td>r(18)=0.97*</td>
</tr>
<tr>
<td>Remaining NI participants</td>
<td>110</td>
<td>-0.04</td>
<td>r(108)=0.95*</td>
</tr>
</tbody>
</table>

Note: We subtracted 4.15 msec from the 2003 values obtained for all participants; see note on the previous slide.
Longitudinal effects

In an ANOVA we examined the VOT values obtained for the two sub-groups at two times (1992 vs 2003)

The main effects of Group and Time were non-significant

However, the G x T interaction was highly significant ($p < 0.0001$)

This is because the VOT values obtained for both sub-groups changed significantly over time, but in opposite directions
Longitudinal effects

The Italians who produced VOT values averaging 18 msec longer in 2003 than 1992 reported using English significantly more in 2003 than in 1992 ($p < 0.05$).

Those who produced VOT values averaging 18 msec shorter in 2003 than 1992 reported using English somewhat less in 2003 than when asked to estimate their % English use in 1992.
The research reviewed here leads to a number of conclusions that are important for both phonetic theory and practice.

**Practice**

1. It is inappropriate to calculate mean VOT for /b d g/ by averaging over lead and short-lag tokens. Mixing short-lag and long-lag VOT values for /p t k/ may also be inappropriate in L2 research if doing so makes it impossible to understand what has been learned, and what has not.

Some learners of English may, of course, produce English /p t k/ with short-lag VOT values because they haven’t learned anything, but this is unlikely for highly experienced learners. That being the case, it is important to find ways to reduce or eliminate code switching when variation in language mode is unintended.
Theory

2. Learning to perceive and produce the VOT dimension in an L2 depends more importantly on the phonetic input received than on the age at which L2 learning has begun.

3. For native speakers of Italian, the learning of English /b d g/ and /p t k/ proceed differently due to differing patterns of cross-language differences for phonologically voiced and voiceless stops.

4. Italians who learn English do not establish new phonetic categories for English /b d g/. Instead, they re-structure native-language phonetic categories for use in both Italian and English based on the input they have received.
5. On the other hand, Italians who receive adequate phonetic input in English do establish new phonetic categories for English /p t k/.

6. Doing so does not cause them to “lose” their short-lag phonetic categories for Italian /p t k/. As a result, they have three categories for stops consonants, not just two as is the case for monolingual speakers of English and Italian.

7. Italian-English bilinguals who begin producing Italian /b d g/ with short-lag VOT values will increase VOT in Italian /p t k/ in order to avoid producing Italian /b d g/ and /p t k/ with the same short-lag VOT values (phonetic system pressure).
Thanks for your kind attention

This talk, as well as an extended version of the talk can be downloaded from www.jimflege.com/conferences


Flege, J. 1987. The production of “new” and “similar” phones in a foreign language. *Journal of Phonetics* 15, 47-65


References


References


References


Appendix 1: experimental design

Here’s how one might go about eliciting categorization at a phonetic category level

Stimuli:
CVs differing in VOT are randomly presented in alternating blocks (A, B) following two different precursors that constrain lexical identity. Precursors for use with our du-tu continuum might be:

A) *She knows what Sam will CV* (do = /d/)

B) *Ben will stay until CV* (two = /t/)

Procedure:
Participants decide if the utterance-final CVs have been produced “correctly” or “incorrectly”. Those possessing a “restructured” /d/ category that includes lead and short-lag VOT variants will accept both variants in “A” blocks but reject both variants in the “B” blocks
Repeated measures design

Different monolingual speakers administer the L1 and L2 portions of the protocol in different testing locations.

Participants are unaware that the two seemingly unrelated sessions are actually part of a single protocol.

To create two distinct (L1, L2) language modes and deactivate the language not currently in use, participants must be convinced that both experimenters are truly monolingual. This will require an extended conversation prior to administering the L1 and L2 portions of protocol.

Disadvantages: Cost and the potential for creating selection biases (not all people who meet the selection requirements will be willing to participate in two relatively lengthy experimental sessions).
Appendix 2: experimental design

**Randomized block design**

Different groups of participants are tested a single time, either in the L1 or the L2 (not both). There is no need to vary testing location.

Participants are tested by different monolingual native speakers of the L1 and the L2.

The two groups of participants need to be matched for all factors known to influence bilingual speech production and perception.

**Disadvantages**: Cost, and the difficulty (impossibility?) of establishing completely matched groups of bilinguals.
A re-examination of the VOT data presented by Birdsong (2003) suggested that uncontrolled differences in speaking rate can affect the outcome of L2 speech research.

Birdsong tested 21 native speakers of English who had lived in Paris for an average of 11 years (range: 5 to 22 years).

His criterion for “native-like” L2 production by individual late learners was the production of mean VOT values in French /p t k/ that differed by less than 1 SD from the mean values obtained from French monolinguals.
Appendix 3: controlling for speaking rate

Here is Birdsong’s VOT data for /t/ (arranged by me in ascending order). The “very strict” 1 SD criterion was attained by 14 of the 21 Late learners. When the data for /t/ was pooled with data for /p/ and /k/, the criterion was met two-thirds of the time (40/62 = 65%)*

*Note. The /p/ data for subject 22, which appears to be spurious (David Birdsong, pers. Communication) has been omitted in this calculation.
Appendix 3: controlling for speaking rate

What about the one-third of participants whose stops were not classified as “native like”? 

Birdsong (2003) did not report values that permit the calculation of Years of L2 Input, nor indicate with whom his participants spoke French

One Late learner had lived in Paris for just 5 years. He can be excluded from our discussion of highly experienced Late learners

Perhaps all of the remaining Late learners who had received 10 years of native speaker input – the benchmark I proposed in Part 1 – would have been classified “native like”
Appendix 3: controlling for speaking rate

Even apart from this conjecture, we can be fairly sure that the data just presented underestimated the Late learners’ performance in French.

As expected (Guion et al., 2000; MacKay & Flege, 2004), the Late learners tested by Birdsong (2003) spoke more slowly than did members of the monolingual French comparison group. We know this because Birdsong (2003) also measured the duration of vowels immediately following the /p t k/ he had measured for VOT.

The Late learners’ vowels were 48% longer on average than vowels produced by the French monolinguals (Birdsong, 2003, Table 2).
Appendix 3: controlling for speaking rate

In English, the VOT in word-initial singleton stops is correlated with the duration of a following vowel. Here, for example, are data obtained for an English monolingual by Theodore et al. (2009)
Appendix 3: controlling for speaking rate

I adjusted Birdsong’s (2003) published VOT values by applying the findings obtained by Theodore et al. (2009) regarding the relationship between VOT and vowel duration*

My post-hoc analysis suggested that had the Late learners spoken at the same rate as the French monolinguals, their French VOT values would have shortened – thus becoming more French-like – by an average of 8.7 msec

The remaining difference between French monolinguals and Late learners dropped to just 2.1 msec, less than measurement error (Schmidt & Flege, 1995, p. 46)

*See Tables I and III, Theodore et al. (2009)
Appendix 4: Fossilization

It is well accepted in SLA research that after a certain point in time progress in an L2 slows, or even ends entirely because learners “fossilize”

Nearly everyone believes this, and I suspect that it holds true for many L2 learners. *Not* because L2 learners lose the capacity for further progress in the L2 but because they normally settle into a comfortable “niche”

Their input becomes relatively fixed. They tend to use the L1 and L2 in well defined contexts with well defined sets of people (family members, friends, service providers, acquaintances, etc.)
The change over time noted here does no agree with the traditional notion of “fossilization”. First, it is important to note that the “Increase” and “Decrease” subgroups did not differ in chronological age, age of arrival in Canada, length of residence in Canada or according to bilingual dominance.*

<table>
<thead>
<tr>
<th></th>
<th>Increased</th>
<th>Decreased</th>
<th>F(1,38)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>9m/11F</td>
<td>10m/10f</td>
<td>-</td>
</tr>
<tr>
<td>Chronological age (years)</td>
<td>53</td>
<td>57</td>
<td>3.41 n.s.</td>
</tr>
<tr>
<td>Age of arrival (years)</td>
<td>10</td>
<td>14</td>
<td>3.34 n.s.</td>
</tr>
<tr>
<td>Length of residence (years)</td>
<td>42</td>
<td>42</td>
<td>0.0 n.s.</td>
</tr>
<tr>
<td>Dominance ratio*</td>
<td>1.42</td>
<td>1.03</td>
<td>2.49 n.s.</td>
</tr>
</tbody>
</table>

*The ratio of self-rated ability to speak and understand English vs Italian
The participants who increased VOT had an average age of 43 years when they were tested the first time, and had already been using English on a daily basis for decades.

The longitudinal finding supports the view (e.g., Flege, 2009; Flege & MacKay, 2011) that L2 learners remain able to exploit the input they receive across the life span.

One wonders: why did English use patterns change for just a subset of the native Italian participants?

I suspect that such changes arise only as the result of important life changes such as moving to a new neighborhood, getting a new job or re-marriage. Clearly, however, more research is needed.

Appendix 4: Fossilization
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 – French Canadians were more likely to produce French /b d g/ with short-lag VOT values than monolinguals in France.

<table>
<thead>
<tr>
<th>dialect</th>
<th>Production of French /b d g/</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% lead VOT</td>
</tr>
<tr>
<td>Quebec (Canada)</td>
<td>41%</td>
</tr>
<tr>
<td>Nantes (France)</td>
<td>94%</td>
</tr>
</tbody>
</table>
Appendix 5: cross-dialect differences

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

adapted from Caramazza & Yeni-Komshian (1974), Fig. 5