The role of input in second language (L2) speech learning

VIth International Conference on Native and Non-native Accents of English
Łódź, 6-8 December 2012

James Emil Flege, PhD
Professor Emeritus
Speech & Hearing Sciences
University of Alabama at Birmingham
Aim

Explore a basic question regarding the acquisition of L2 speech:

*How important is phonetic input for second-language (L2) speech acquisition?*

For example: If I pronounce my L2 badly (i.e., with non-native characteristics) is it because I didn’t get *enough* input? Or because I didn’t get the *right* input?
Background

Various factors have been proposed over the past 6 decades to explain differences in learners’ success in L2 acquisition:

- Interference between the L1 and L2
- Individual differences (e.g., auditory acuity, working memory)
- Motivational differences
- Maturational state (critical period hypothesis)
- Differences in amount/kind of input
Background

Input has generally not been considered very important

*Historical trends in research* - the perceived importance of various factors, as judged by researchers in the field, has changed over the years

Why? Real scientific progress; or fads?
Based on my experience with the literature:

1950-1970 interference is the dominant “explanation”

Generally thought that the maintenance of L1 structures in the production/perception of L2 could explain most aspects of non-nativeness (foreign accent) in the L2
Background

1970-2000

Beginning with Lenneberg’s book (1969) emphasis shifted from a focus on abstract language structures to general properties of human development and neural/cognitive maturation.

These aspects follow a fairly predictable time course, and thus were thought to explain “age” effects.
Maturation as dominant explanation

“Input differences are not a good explanation for age effects, because it is precisely in the linguistic domains where input varies least—phonology—that the age effects are most readily apparent.”

“...the severe decline of the ability to induce abstract patterns implicitly is an inevitable consequence of fairly general aspects of neurological maturation

DeKeyser (2000) p. 518-519
Background

The relative importance of various factors has never been established empirically. Here, however, is my impression based on 30 years of research.

Without denying the potential importance of other factors…

In my view, the kind and/or quality of input is the single most important factor affecting nativeness of L2 segmental production and perception.
L2 input?

The perceived unimportance of input by many/most L2 researchers is to me, at least, surprising. After all…

- Input is crucial in L1 acquisition
- Children generally more successful than adults in L2 speech learning
- Children receive more and/or better L2 input than adults (although differences not yet quantified)
L2 input?

I have frequently wondered why (some) researchers seem to have concluded that:

• Input is more important for learning the L1 than for learning an L2
• The enormous variation in ultimate success that we typically see among Late learners is due to unspecified “individual differences” in learning ability, not differences in the kind/amount of input received
After all, it is widely accepted that

• Children establish long-term memory representations for phonetic categories based on input distributions (the range of tokens that have been experienced)

• During L1 acquisition, children translate information stored and structured in categories into articulatory motor programs that reproduce what has been heard (from others + self)
Children learning English receive two input distributions for the voiced “phonemes” /b d g/.

They are exposed to just one broad distribution for the voiceless phonemes /p t k/ (long-lag VOT).
Native speakers of English show rate-dependent processing of stops differing in VOT.

Later in life, when asked to rate the perceived “goodness” of instances of voiceless stop phonemes, monolingual native English adults prefer tokens having long-lag VOT.
Native speakers of English show rate-dependent processing of stops differing in VOT.

Goodness rating task

- Stimuli having the VOT values of stops heard most frequently in English are judged to sound best.
- A clear link between input and perception which, in turn, is linked to speech production patterns.

input distribution: L1
Interestingly:

- In production, VOT values in stops depend on speaking rate (slow rate – longer).
- Goodness ratings also changed appropriately in “fast” vs. “slow-rate” stimuli.

(Flege et al. 1996)
Input distribution: L1s = Spanish, English

Crucially:

Children exposed to different languages are exposed to different input distributions for the contrastive categories their L1 uses to distinguish meaning (words)
Input distribution: L1s = Spanish, English

Here are hypothetical input distributions for /p t k/ in Spanish and English (the patterns are surely correct, the exact details have never been established)

Spanish monolinguals hear short-lag stops (green)

English monolinguals hear long-lag stops (black)
What is heard eventually determines how monolinguals speak, leading to cross-language differences in production.

Here we see mean VOT measured in productions of the voiceless stops /p t k/.

English monolingual adults and 9-year-old children produced longer VOT than same-aged Spanish monolinguals.

Flege & Eefting (1987)
In both languages adults produced stops with somewhat longer VOT values than do 9-year-old children.

Implies that L1 phonetic development is not yet complete at age 9.

Flege & Eefting (1987)
Input distribution: L1 perception

- Children gradually become “language specific speaker-hearers” of their L1
- Nativeness depends on developments in both production (obvious) and perception (less obvious to the casual observer)
- Part of L1 development is the eventual “alignment” of perception and production patterns
"phoneme boundary" (perceptual crossover) between sounds heard as /d/ vs. /t/ (in msec of VOT in synthetic stimuli)

longer for English than Spanish

longer for adults than 9-year-olds (in both languages)

perception of /da/-/ta/ VOT continuum
Flege & Eefting (1986)
Input distribution: L1 perception

- Phoneme boundaries (crossovers) occurred at significantly longer VOT values for monolingual English adults than monolingual English 9-year olds (p < .05)

- Implies that L1 perceptual development not yet complete at age 9 years

perception of /da/-/ta/ VOT continuum
Flege & Eefting (1986)
Input distributions: L1

- It takes years for monolinguals to become mature “speaker-hearers” of their L1
- Logically: the same should apply to adults learning an L2.
- **Therefore:** We need to give adult L2 learners enough time/input to learn before concluding that they have failed!
Bilingual input distributions

Here is the “best case” input scenario for Spanish learners of English /p t k/

• These learners, who are to become bilingual, will hear only “authentic” long-lag VOT values in English

• In Spanish, they will hear only authentic short-lag VOT values
Bilingual input distributions

The “best case” scenario probably never occurs

Spanish learners of English are probably exposed to something like this (hypothetical data)

They get “accurate” VOT from native English speakers, and “inaccurate” (red) values from other native speakers of Spanish who speak English with a foreign accent (including “compromise” values for English /p t k/
Bilingual input distributions

- We can be sure that most (all?) individuals who learn an L2 receive different input distributions than do monolingual native speakers of the target L2.

- Can these (presumed) input differences explain non-nativeness in the production and perception of the L2?
L2 input distributions

• Let’s consider two studies that examined the acquisition of English /p t k/ by native Spanish speakers

• Both studies focused on individuals who began learning English as an L2 in childhood (Early learners)
Input distributions in L2

- **Study 1** carried out in Puerto Rico – where most persons who speak English learned it as an L2 (virtually no monolingual native speakers of English in Puerto Rico)

- **Study 2** carried out in Austin, Texas – where there are many monolingual native speakers of English.

- Assumption: the early learners in Puerto Rico received more “Spanish-accented” English input than those who learned English L2 in Texas
Puerto Rico study: Flege & Eefting (1987)

examined VOT in /p t k/ as produced by

1. Monolingual native speakers of Spanish (adults and children) living in Puerto Rico

2. Control groups
   Monolingual native speakers of English (adults and children) living in the US
Puerto Rico study: Flege & Eefting (1987)

also examined early Spanish-English bilinguals

3. Native Spanish **adults** who had learned English in a bilingual school in Puerto Rico

4. Native Spanish **children** who had learned English in a bilingual school in Puerto Rico
Early bilinguals (children and adults) produced English stops (/p t k/) with values that were too short compared to Spanish & English monolinguals producing /p t k/.

Puerto Rico study: Flege & Eefting (1987)
The Early bilinguals’ English stops had VOT values that were about midway between the values observed for Spanish and English monolinguals.
Puerto Rico study: Flege & Eefting (1987)

- This 1987 finding was a big surprise
- It was widely believed (then, as now) that individuals who learn an L2 before the critical period - Early learners – will produce and perceive their L2 perfectly

In fact, the Texas study showed the expected results
Texas study: Flege (1991)

The Texas study examined VOT in stops produced by

- Adult monolingual speakers of English and Spanish;
- Native Spanish adults who learned English as children (Early learners);
- Native Spanish adults who learned English in adulthood (Late learners).
Participants in Texas study  
(Flege 1991)

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Age</th>
<th>AOA</th>
<th>Length of Residence</th>
<th>English Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish monolingual</td>
<td>10</td>
<td>30</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>English monolinguals</td>
<td>10</td>
<td>26</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Early S-E bilinguals</td>
<td>10</td>
<td>23</td>
<td>2</td>
<td>21</td>
<td>82%</td>
</tr>
<tr>
<td>Late S-E bilinguals</td>
<td>10</td>
<td>34</td>
<td>20</td>
<td>14</td>
<td>66%</td>
</tr>
</tbody>
</table>

Early vs. Late groups selected on the basis of age of arrival (AOA) in the United States
Participants in Texas (Flege 1991) study

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Age</th>
<th>AOA</th>
<th>Length of Residence</th>
<th>English Use</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spanish monolingual</strong></td>
<td>10</td>
<td>30</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td><strong>English monolinguals</strong></td>
<td>10</td>
<td>26</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td><strong>Early S-E bilinguals</strong></td>
<td>10</td>
<td>23</td>
<td>2</td>
<td>21</td>
<td>82%</td>
</tr>
<tr>
<td><strong>Late S-E bilinguals</strong></td>
<td>10</td>
<td>34</td>
<td>20</td>
<td>14</td>
<td>66%</td>
</tr>
</tbody>
</table>

As is typical, the Early learners reported using English more than the Late learners.
Early learners produced VOT values accurately.

Values for Late learners were intermediate to values observed for Spanish, English monolinguals.

**Graph:**

- **Word-initial /t/ (utterance initial)**
- **Early**
- **Late**
- **Mean VOT (ms)**
- **English mono**
- **Spanish mono**
In fact, all of the individual Early learners produced native-like VOT values.
There was greater variability among the Late learners. Only a few of them produced English stops accurately.
What accounts for variability among the Late learners?

Differences in amount of input?

in quality of input?

Motivation?

Individual differences?
The filtering hypothesis

One potential explanation for why input is considered unimportant is the hypothesis that cross-language differences get filtered out. Several variants of this hypothesis exist:

• Having learned to perceive in a “phonemic” way, all cross-language phonetic differences are filtered out, just as if the phonetic differences didn’t exist;
• L2 properties/features not needed for phonemic contrasts in L1 are ignored/not stored/not detected auditorily;
• Some/all features/properties in L2 sounds get warped/distorted if they are used differently in L2 than in L1
The filtering hypothesis

• If L1/L2 differences are not auditorily accessible to L2 learners, then input can’t be important.

• However, according to my Speech Learning Model (SLM):

  Learners of all ages can eventually gain perceptual access to properties needed to define L2 speech sounds, even if such properties are not needed to differentiate L1 speech sounds (see Flege, 1992, 1995, 1999, 2002, 2003, 2007)
L2 input filtered out?

- One way to evaluate the role of “filtering” is to consider how native speakers of Japanese produce and perceive English /r/ and /l/.
- These two English liquids differ phonetically from any sound found in Japanese.
- If Japanese adults are able to gain access to the phonetic properties that differentiate English /r/ from /l/, and that differentiate these English liquids from the closest Japanese sound, then English input should matter.
L2 input filtered out?

Specifically

Japanese learners with a relatively long length of residence (LOR) in the US should be more native-like than those with a shorter LOR

(Assumption; amount of L2 input increases systematically with LOR)

Flege et al. (1995, 1996) compared Japanese adults differing in length of residence (LOR) in the US
Two groups of Japanese Late learners

<table>
<thead>
<tr>
<th></th>
<th>Native English</th>
<th>Experienced Japanese</th>
<th>Inexperienced Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chronological age, years</strong></td>
<td>36</td>
<td>44</td>
<td>35</td>
</tr>
<tr>
<td><strong>Age of arrival in USA, years</strong></td>
<td>--</td>
<td>23</td>
<td>34</td>
</tr>
<tr>
<td><strong>Length of residence in USA, years</strong></td>
<td>--</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td><strong>Use of English</strong> 1=never 7=frequently</td>
<td>--</td>
<td>6.0</td>
<td>2.4</td>
</tr>
</tbody>
</table>

- “Experienced” had a longer residence (LOR) than the “Inexperienced” group
- and so, presumably, more English input
Two groups of Japanese late learners

<table>
<thead>
<tr>
<th></th>
<th>Native English</th>
<th>Experienced Japanese</th>
<th>Inexperienced Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronological age, years</td>
<td>36</td>
<td>44</td>
<td>35</td>
</tr>
<tr>
<td>Age of arrival in USA, years</td>
<td>--</td>
<td>23</td>
<td>34</td>
</tr>
<tr>
<td>Length of residence in USA, years</td>
<td>--</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td>Use of English, 1=never 7=frequently</td>
<td>--</td>
<td>6.0</td>
<td>2.4</td>
</tr>
</tbody>
</table>

- “Experienced” group also reported using English more frequently than “Inexperienced” (on 7-point scale)
Flege et al. (1995) examined /r/, /l/ production
  – Elicited 3 speech samples, each containing words beginning in /r/ and /l/
  – Productions evaluated by native English listeners (dependent variable = % correct identification as intended by talker)
L2 input filtered out?

Definition task: respond “light” to “What we get from the sun”

Reading task

Spontaneous task

Much better production of /r/ and /l/ by experienced than inexperienced late learners
L2 input filtered out?

Most Experienced Japanese in native English range
L2 input filtered out?

- Flege et al (1996) examined the identification of four English consonants (/r/ /l/ /w/ /d/) in naturally produced English words.
- English /w/ and /d/ expected to be easy – comparable sounds exist in Japanese.
L2 input filtered out?

- As expected: /w/ and /d/ were easy to identify for both J groups
- However, both J groups had difficulty identifying /r/ and /l/
- Experienced better than Inexperienced … but certainly not perfect
L2 input filtered out?

- Why did both J groups identify /r/ better than /l/?
- English /r/ is perceived to be more distant phonetically from the single liquid of Japanese than is English /l/
- Thus according to the SLM, English /r/ is more likely to be treated as a “new” (non-L1) sound
L2 input filtered out?

Problem for SLM

- If /r/ is treated as new, shouldn’t at least some of the Japanese Late learners have identified it as accurately as do native English speakers?
- Yes… and in fact they did, but this requires some explanation
L2 input filtered out?

- The /r/ and /l/ stimuli occurred in naturally produced English words.
- The words used as stimuli made up “minimal pairs” (e.g., “road”, “load”).
- The words were presented one at a time, of course... but listeners have lexical knowledge (they know that there are multiple “_oad” words in English).
L2 input filtered out?

• The native English listeners correctly identified /r/ and /l/ in all words, regardless of the relative degree of familiarity of the two members of a minimal pair.
L2 input filtered out?

- However, the Japanese Ss’ identifications were influenced by subjective word familiarity (evaluated by ratings obtained after the perception experiment).
- For example: they identified /r/ better in “road” (more familiar than “load”) than in “rook” (less familiar than “look”)
L2 input filtered out?

Necessary to analyze identification of /r/ and /l/ in a specific subset of the stimuli, that is, the stimuli that were part of balanced minimal pairs (both words equally familiar to the J participants)

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>examples</th>
<th>ave. ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>r &gt; l</td>
<td>8</td>
<td>road-load</td>
<td>r=5,7 l=2,4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ride-lied</td>
<td></td>
</tr>
<tr>
<td>r = l</td>
<td>7</td>
<td>rate-late</td>
<td>r=5,1 l=5,5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rack-lack</td>
<td></td>
</tr>
<tr>
<td>r &lt; l</td>
<td>8</td>
<td>rook-look</td>
<td>r=1,9 l=6,0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rake-lake</td>
<td></td>
</tr>
</tbody>
</table>
Perception of /r/

For /r/ tokens in balanced minimal pairs, some J Late learners showed the same perfect performance as native English monolinguals.

5/12 Experienced
1/12 Inexperienced
L2 input filtered out?

Perhaps differences in input can explain some of the inter-subject variability in perception – and ultimately production

<table>
<thead>
<tr>
<th></th>
<th>simple correlation</th>
<th>AOA partialled out</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of Arrival (AOA)</td>
<td>-.50*</td>
<td>--</td>
</tr>
<tr>
<td>Length of residence (LOR)</td>
<td>.61*</td>
<td>.41*</td>
</tr>
<tr>
<td>% English at home</td>
<td>.42*</td>
<td>.03</td>
</tr>
<tr>
<td>% English at work/school</td>
<td>.30</td>
<td>.16</td>
</tr>
<tr>
<td>% English at parties/with friends</td>
<td>.62**</td>
<td>.52*</td>
</tr>
<tr>
<td>% English with children/younger relatives</td>
<td>.51*</td>
<td>.19</td>
</tr>
<tr>
<td>% English overall in last 5 years</td>
<td>.60**</td>
<td>.39</td>
</tr>
</tbody>
</table>

correlations with % correct /r/ scores (in balanced minimal pairs)
Length of residence (LOR)

- Those who posit that input is (relatively) unimportant have tended to focus on LOR.
- LOR in crude measure of amount of L2 input; and it says absolutely nothing about the quality of input received.
- Problem: possible to live for years in a predominantly L2-speaking environment without using the L2.
- So why is LOR often used as a predictor variable in research? (Answer: because it is easy)
LOR: length of residence

• Generally, amount of L2 input increases with each year of residence
• Problem: the relation between amount of input and years of residence is non-linear
• Example: LOR = 10 years does not necessarily indicate twice as much input as LOR = 5 years
LOR: length of residence

- Generally speaking, LOR has **not** proven to be an important predictor of L2 speech learning
- Example: Flege et al. (1995) – to be reviewed shortly
- Lack of correlation between LOR and dependent variables offered as evidence that input is relatively unimportant for L2 speech learning (DeKeyser & Larson-Hall, 2005)
Flege et al. (1995) examined 240 native Italian immigrants to Canada selected on the basis of AOA. There was a strong correlation between the degree of foreign accent and AOA ($r = -0.84$, $p < 0.01$). The earlier the arrival, the better the L2 pronunciation.
However, only a modest correlation existed between foreign accent and LOR ($r = .28$, $p < 0.01$)

The longer the residence in a predominantly L2-speaking environment, the better the L2 pronunciation
• The AOA-foreign accent correlation remained significant when the effect of LOR was partialled out.

• However, the LOR-foreign accent correlation became non-significant when the effect of AOA was partialled out.

• Conclusion: LOR is not an independent predictor of foreign accent.

<table>
<thead>
<tr>
<th></th>
<th>simple</th>
<th>partial</th>
</tr>
</thead>
<tbody>
<tr>
<td>FA-AOA</td>
<td>-.84*</td>
<td>-.84*</td>
</tr>
<tr>
<td>FA-LOR</td>
<td>.28*</td>
<td>-.19</td>
</tr>
</tbody>
</table>
LOR

• Does this kind of evidence – replicated several times – prove that amount of L2 input is unimportant?

• Maybe not. Perhaps we need to obtain better measures of L2 input, or else restrict hypotheses regarding LOR.

• Perhaps variations in LOR (and input, more generally) is important only in constrained circumstances, for example:
  – For Early but not Late learners
  – For beginning but not later stages of L2 learning
  – for “naturalistic” but not “classroom” learning
Flege et al. (2006) assessed foreign accent in Koreans differing in LOR in the US (3 vs. 5 years)

- Tested both children and adults
- No difference in foreign accent for Koreans differing in LOR (3 vs. 5 years)
- Same finding for adult and child immigrants
Participants tested at two times (T1, T2) separated by 1.2 years.

Difference between T1 and T2 barely significant for children; non-significant for adults.

A similar study by Aoyama et al. (2003) yielded a similar pattern of results for LOR.

Note: averaged over both LOR groups.
LOR

• Aoyama et al. (2003) examined Japanese immigrants to the US
• Overall, the Japanese participants had lived for a shorter time in the US than had the Koreans
• Like the Koreans, the Japanese Ss were tested twice
  – LOR at Time 1 = 0.5 years
  – LOR at Time 2 = 1.6 years
• Aoyama et al. (2003) results

• improvement from T1 to T2 for children

• but not adults
Questions about LOR

1. Why did foreign accent improve over 1 year for children but not adults?
2. Why was the LOR effect stronger in the Japanese than Korean study?

My hypotheses: effects of LOR (and “input”, more generally) are more likely
– In beginning than later stages of L2 learning
– For Early than Late learners
A speculative account

**Hypothesis:** a greater effect of input on performance in beginning than later stages of L2 learning

- True just for Early learners?
- Is no further learning possible after the rapid initial phase?

Growth curve: early learners

![Growth curve graph]

- Degree of Foreign Accent
- LOR (months)

Aoyama et al., Flege et al.

Japanese

Koreans
LOR

- Another reason to suppose that LOR is important, at least for children, is a case study by Winitz et al. (1995)
- Results supports the view that, even for children, the kind of input received is important
Winitz et al. (1995)

- Recorded a Polish boy ("AO") for 7 years after his arrival in the US
- This boy was chosen for study because of the special circumstances in which he was learning English as an L2
- Lived in a small town in Missouri where there were few/no other immigrants
- Assured that most input received by AO was "authentic" native-speaker input
LOR in early learners

Winitz et al. (1995) also recorded in a single session:

• other immigrant boys who lived in a large city and so were exposed to foreign-accented input
• Age-matched monolingual English boys

English sentences recorded by all participants were rated for foreign accent (5-point scale)
AO “passed” as a native speaker after about 1 year in the U.S.

The other non-native boys produced sentences that were clearly foreign-accented.

Winitz et al. (1995)
What about Late learners?

- Does their pronunciation of the L2 never improve, no matter how much input they receive?
- Some evidence seems to support of this idea
- However, careful consideration suggests otherwise
LOR

• Flege et al. (unpublished) carried out a longitudinal study of native Spanish adults who had immigrated to the US

• All were living in Birmingham, AL
  – 15 native Spanish late learners
  – 4 native English controls

• Recorded every 6 months for 5 years (11 data samples)
n = 15 native Spanish late learners in Flege et al. (unpublished) study

<table>
<thead>
<tr>
<th></th>
<th>$M$</th>
<th>$SD$</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronological age (years)</td>
<td>34</td>
<td>6</td>
<td>24</td>
<td>41</td>
</tr>
<tr>
<td>AOA (years)</td>
<td>29</td>
<td>6</td>
<td>19</td>
<td>39</td>
</tr>
<tr>
<td>LOR (years) Session 1</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Overall En use Session 1</td>
<td>60%</td>
<td>21</td>
<td>20%</td>
<td>100%</td>
</tr>
<tr>
<td>Overall Sp use Session 1</td>
<td>40%</td>
<td>21</td>
<td>0%</td>
<td>80%</td>
</tr>
</tbody>
</table>

Mean values for 15 participants at Time 1
Longitudinal study (Flege et al, unpubl.)

**Finding**: no significant change in foreign accent over 5 years
LOR

- Does this demonstrate that, in Late learners, input has no affect on L2 pronunciation?
- Maybe not.

- The LOR of the native Spanish Ss at the first time of testing was 4 years
- Perhaps they were already too far along the “learning curve” when tested at Time 1
Flege & Fletcher (1992) compared Spanish late learners differing in LOR. Shorter LOR = 0.7 years, Longer LOR = 14.3 years. **Finding:** less foreign accent by longer than shorter LOR group (for 2 of 3 sentences).
LOR

• A closer look at the 5-year longitudinal study (unpublished) provides some hints regarding the role of input

• In a post-hoc analysis, we compared the three Ss each having the “best” and “worst” pronunciation of English
15 native Spanish participants

Native English controls (n = 4)
LOR

• What explains the enormous and consistent difference between the “best” and the “worst” late learners?
• Possible answer: more English use with native speakers by the the 3 “best” than the 3 “worst”
The 3 “best” Late learners reported using English more than the 3 “worst” Late learners, especially with “friends” (optional context)
Self-reports of English use

However, use of English “with friends” -- very high at the beginning of the study – decreased over the course of the study. Why? Once settled in the US, did these Ss “retreat” to their L1?
• Is there a comparable diminishing effect of input for Early and Late learners?
• Maybe not.
• Assuming that adults and children get equal L2 input (unlikely!), we might speculate on slightly different growth functions
On this speculative account:

- Children learn more rapidly than adults in the initial stage of learning.
- And, although their learning slows, children continue learning gradually thereafter.
- But not adults?
Caveats

- Unfortunately, we don’t have enough data in hand to make any specific claims.
- Moreover: possible changes over time in the ameliorative effects of additional input are likely to confuse amount and kind of input received.
- As for adults: existing data indicates they can benefit from input – at least if they receive the right kind of input in the right circumstances.
LOR

- Flege & Liu (2001) tested 60 Chinese adults living in the US
- All were affiliated in some way with a large, urban medical center, but differed in LOR
  - “shorter LOR” range = 0.5 - 3.8 years (n=30)
  - “longer LOR” range = 3.9 - 15.5 years (n=30)
LOR

• Half of the participants in both LOR groups were **full-time students**. assumed to receive much native-speaker input because most of their teachers and fellow students were English monolinguals

• The other half were **non-students** (laboratory workers, research scientists, stay-at-home spouses) assumed to receive relatively little native-speaker input
Flege & Liu (2001) administered 3 tests

- **Comprehension of English**
  45-items Univ. of Michigan Listening Comprehension Test

- **Grammaticality Judgment Test**
  144-item, derived from Johnson & Newport (1989)

- **Phonetic perception test (identification of English consonants)**
  word-final tokens of /b d g p t k/ in noise
For all 3 tests

- **Students** with a longer LOR ($M = 7.3$ years) obtained significantly higher scores than students with a shorter LOR ($M = 2.5$ years)
- No significant differences between **non-students** with a relatively longer LOR ($M = 6.6$ years) and those with a shorter LOR ($M = 1.7$ years)
LOR

Conclusion from Flege & Liu (2001)

• LOR mattered for late learners, but only those who really needed to use English on a daily basis (students)

“Fossilization”

• It is widely believed that beyond a certain point, learning in an L2 stops.
• Indeed, I have raised the possibility that additional input may have comparatively little ameliorative effect for Late learners.
• However, we have seen evidence that in certain circumstances, learning does take place for Late learners.
• Should we reject the construct of “fossilization” – which implies no further learning beyond a certain point of development?
“Fossilization”

• When does fossilization occur?
• Birdsong & Molis (2001. p. 239) posited that L2 learning asymptotes after 10 years
• Does empirical evidence exist in support of fossilization?
  (showing the something doesn’t happen is difficult)

Participants

Our sample consisted of 61 native speakers of Spanish. Of these, 29 were Early Arrivals (AoA ≤ 16) and 32 were Late Arrivals (AoA ≥ 17). We attempted to ensure that learners were at asymptote by setting the mean length of residence in the United States at 10 years. Between early- and late-arriving groups, the length of residence is approximately equal. Thus, for Early Arrivals, the mean length of residence is 12.2 years compared to 10.5 years for Late Arrivals. According
“Fossilization”

• The best way to test for fossilization is to carry out a longitudinal study

• Let’s consider the well-known foreign accent study of Flege et al. (1995)

• Examined 240 Italian immigrants (+ native English controls)

• Ss repeated English sentences following a model; later rated for foreign accent by native English listeners
Fossilization

Flege et al. (1995)

Native Italian immigrants to Canada

foreign accents in English depends importantly on age of arrival (AOA) in Canada

240 Native Italian Ss' Production of English Sentences

Mean Rating

Strongest Accent

No Accent

NE control subjects

Age of Arrival (years)

r = -0.845

df = 238
“Fossilization”

• Flege & MacKay (unpublished) re-recorded 160 of the original 240 Italian immigrants a second time 10.5 years later
  – Mean LOR at Time 1 = 32.5 years
  – Mean LOR at Time 2 = 43.0 years
• Identical equipment and procedures, but just 3 of the 5 original stimulus sentences were evaluated for degree of foreign accent

• **Aim:** Determine if these highly experienced late learners were capable of additional speech learning
Finding: a strong correlation between foreign accent obtained from the same Ss at Time 1 and Time 2

• No apparent change in L2 pronunciation over 10.5 years
• Evidence for that the Ss L2 system had “fossilized”?
• Maybe not, if data considered in detail

\[ r(158) = .97 \]
“Fossilization”

Summary statistics for foreign ratings obtained for all 181 Ss (160 native Italian)

<table>
<thead>
<tr>
<th></th>
<th>Mean rating</th>
<th>SD</th>
<th>SE</th>
<th>2 SE’s</th>
</tr>
</thead>
<tbody>
<tr>
<td>1992</td>
<td>6.15</td>
<td>2.07</td>
<td>.154</td>
<td>.308</td>
</tr>
<tr>
<td>2003</td>
<td>6.03</td>
<td>2.14</td>
<td>.159</td>
<td>.318</td>
</tr>
</tbody>
</table>
The pronunciation of 61 native Italian participants got slightly worse over time.

The pronunciation of 22 native Italian participants got slightly better.
## “Fossilization”

<table>
<thead>
<tr>
<th></th>
<th>1992</th>
<th>2003</th>
<th>difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>worsened</td>
<td>5.4</td>
<td>4.8</td>
<td>-0.6 (p &lt; .001)</td>
</tr>
<tr>
<td>(n = 61)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>improved</td>
<td>5.3</td>
<td>6.0</td>
<td>+0.7 (p &lt; .001)</td>
</tr>
<tr>
<td>(n = 22)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n.s.</td>
<td>0.1</td>
<td>1.2</td>
<td>+1.1 (p &lt; .05)</td>
</tr>
</tbody>
</table>

Why did the English pronunciation of some native Italian participants improve, whereas that of others worsen?
“Fossilization”

Test for differences in self-reported use of English between “worsened” and “improved” Ss. No significant differences between these two subgroups, either in 1992 or 2003.

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, 2003</td>
<td>1.13</td>
<td>0.29</td>
</tr>
<tr>
<td>Age, 1992</td>
<td>1.16</td>
<td>0.28</td>
</tr>
<tr>
<td>AOA</td>
<td>2.50</td>
<td>0.12</td>
</tr>
<tr>
<td>% Italian, 1992</td>
<td>0.30</td>
<td>0.59</td>
</tr>
<tr>
<td>% English, 1992</td>
<td>0.46</td>
<td>0.50</td>
</tr>
<tr>
<td>% English, 2003</td>
<td>0.01</td>
<td>0.93</td>
</tr>
<tr>
<td>% Italian, 2003</td>
<td>0.01</td>
<td>0.94</td>
</tr>
<tr>
<td>M use of Italian, 1992*</td>
<td>0.02</td>
<td>0.88</td>
</tr>
<tr>
<td>M use of English, 1992*</td>
<td>0.06</td>
<td>0.81</td>
</tr>
<tr>
<td>M use of Italian, 2003</td>
<td>0.83</td>
<td>0.36</td>
</tr>
<tr>
<td>M use of English, 2003</td>
<td>0.03</td>
<td>0.86</td>
</tr>
</tbody>
</table>

*average of three 7-point rating scale questions regarding language use at home, in social contexts, and overall in past 5 years
“Fossilization”

However:

• An analysis of VOT in word-initial stop production suggests that some of the native Italian participants changed their pronunciation of English.

• Even though they had already been using English as an L2 for decades.
“Fossilization”

- Flege & MacKay (unpublished) re-recorded words beginning with /p t k/
- Some stimuli, recording procedures, measurement techniques used at Time 1 (1992) and Time 2 (2003)
“Fossilization”

- VOT in word initial /p t k/

All AOA-defined subgroups produced longer (more English-like) VOT values in 2003 than in 1992.

However, the differences were small and only one (in red) was significant.
“Fossilization”

- A single significant 9-ms increase for one subgroup does not provide impressive counterevidence to “fossilization”
- But who would we expect to show an increase in English VOT?
- My answer: those whose English use had increased
“Fossilization”

We defined 3 groups (n = 24 each) based on changes in self-reported English use in 1992 (Time 1) and 2003 (Time 2)

- **Group 1**: less English use in 2003
- **Group 2**: no change
- **Group 3**: more English use in 2003
“Fossilization”

Subgroups defined on the basis of change in % English use from 1992 to 2003

<table>
<thead>
<tr>
<th></th>
<th>less English</th>
<th>No Change</th>
<th>more English</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Age of arrival</em></td>
<td>14.2</td>
<td>11.9</td>
<td>11.3</td>
</tr>
<tr>
<td><em>Chron. age 1992</em></td>
<td>45.9</td>
<td>46.4</td>
<td>41.4</td>
</tr>
<tr>
<td><em>LOR in 1992</em></td>
<td>31.0</td>
<td>33.8</td>
<td>29.5</td>
</tr>
<tr>
<td><em>LOR in 2003</em></td>
<td>41.5</td>
<td>44.3</td>
<td>40.0</td>
</tr>
<tr>
<td>% <em>Italian, 1992</em></td>
<td>26.9</td>
<td>24.5</td>
<td>43.9</td>
</tr>
<tr>
<td>% <em>Italian, 2003</em></td>
<td>54.7</td>
<td>24.5</td>
<td>24.0</td>
</tr>
<tr>
<td>% <em>English, 1992</em></td>
<td>72.8</td>
<td>74.9</td>
<td>49.4</td>
</tr>
<tr>
<td>% <em>English, 2003</em></td>
<td>44.4</td>
<td>74.9</td>
<td>73.5</td>
</tr>
</tbody>
</table>

A 24% increase in English use over the 10-year study period
“Fossilization”

Subgroups defined on the basis of change in % English use from 1992 to 2003

<table>
<thead>
<tr>
<th></th>
<th>less English</th>
<th>No Change</th>
<th>more English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of arrival</td>
<td>14.2</td>
<td>11.9</td>
<td>11.3</td>
</tr>
<tr>
<td>Chron. age 1992</td>
<td>45.9</td>
<td>46.4</td>
<td>41.4</td>
</tr>
<tr>
<td>LOR in 1992</td>
<td>31.0</td>
<td>33.8</td>
<td>29.5</td>
</tr>
<tr>
<td>LOR in 2003</td>
<td>41.5</td>
<td>44.3</td>
<td>40.0</td>
</tr>
<tr>
<td>% Italian, 1992</td>
<td>26.9</td>
<td>24.5</td>
<td>43.9</td>
</tr>
<tr>
<td>% Italian, 2003</td>
<td>54.7</td>
<td>24.5</td>
<td>24.0</td>
</tr>
<tr>
<td>% English, 1992</td>
<td>72.8</td>
<td>74.9</td>
<td>49.4</td>
</tr>
<tr>
<td>% English, 2003</td>
<td>44.4</td>
<td>74.9</td>
<td>73.5</td>
</tr>
</tbody>
</table>

A 28% decrease in English use over the 10-year study period
“Fossilization”

- Examined VOT values in the speech of the 3 subgroups that had been defined on the basis of self-reported English use patterns.

- A positive value here indicates an increase in VOT – therefore more native-like productions in 2003 than 1992.
“Fossilization”

- Members of the subgroup that used English more at Time 2 than at Time 1 showed the largest increase in VOT values over the same time period.
- The subgroup that used English less at Time 2 than at Time 1 showed no change in VOT production.

Note: the VOT difference between subgroups was significant (p < .01)
“Fossilization”

• Is this finding of theoretic importance?
• Perhaps, at least if one considers input from the standpoint of the distributions of VOT values experienced over the course of the life span
More English

<table>
<thead>
<tr>
<th></th>
<th>less English</th>
<th>No Change</th>
<th>more English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of arrival</td>
<td>14.2</td>
<td>11.9</td>
<td>11.3</td>
</tr>
<tr>
<td>Chron. age 1992</td>
<td>45.9</td>
<td>46.4</td>
<td>41.4</td>
</tr>
<tr>
<td>LOR in 1992</td>
<td>31.0</td>
<td>33.8</td>
<td>29.5</td>
</tr>
<tr>
<td>LOR in 2003</td>
<td>41.5</td>
<td>44.3</td>
<td>40.0</td>
</tr>
<tr>
<td>% Italian, 1992</td>
<td>26.9</td>
<td>24.5</td>
<td>43.9</td>
</tr>
<tr>
<td>% Italian, 2003</td>
<td>54.7</td>
<td>24.5</td>
<td>24.0</td>
</tr>
<tr>
<td>% English, 1992</td>
<td>72.8</td>
<td>74.9</td>
<td>49.4</td>
</tr>
<tr>
<td>% English, 2003</td>
<td>44.4</td>
<td>74.9</td>
<td>73.5</td>
</tr>
</tbody>
</table>

For computational Purposes:
Assume that someone using English 100% of the time hears 1000 tokens per year (1000 = unity)

Period 1: arrival in Canada until 1993 = 29.5 yrs @ 49.4% = 14,573 tokens

Period 2: 1993-2003 = 10.5 yrs @ 73.5% = 7,717 tokens

From this exercize, we conclude that the “more English” group experienced a 53% increase in the total number of number of English /p t k/ tokens ever heard!
Less English

<table>
<thead>
<tr>
<th></th>
<th>less English</th>
<th>No Change</th>
<th>more English</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age of arrival</strong></td>
<td>14.2</td>
<td>11.9</td>
<td>11.3</td>
</tr>
<tr>
<td><strong>Chron. age 1992</strong></td>
<td>45.9</td>
<td>46.4</td>
<td>41.4</td>
</tr>
<tr>
<td><strong>LOR in 1992</strong></td>
<td>31.0</td>
<td>33.8</td>
<td>29.5</td>
</tr>
<tr>
<td><strong>LOR in 2003</strong></td>
<td>41.5</td>
<td>44.3</td>
<td>40.0</td>
</tr>
<tr>
<td><strong>% Italian, 1992</strong></td>
<td>26.9</td>
<td>24.5</td>
<td>43.9</td>
</tr>
<tr>
<td><strong>% Italian, 2003</strong></td>
<td>54.7</td>
<td>24.5</td>
<td>24.0</td>
</tr>
<tr>
<td><strong>% English, 1992</strong></td>
<td>72.8</td>
<td>74.9</td>
<td>49.4</td>
</tr>
<tr>
<td><strong>% English, 2003</strong></td>
<td>44.4</td>
<td>74.9</td>
<td>73.5</td>
</tr>
</tbody>
</table>

For computational Purposes:

Assume that someone using English 100% hears 1000 tokens per year

Period 1: arrival until 1993 = 31.0 yrs @ 72.8% = 22,569 tokens
Period 2: 1993-2003 = 10.5 yrs @ 44.5% = 4,672 tokens

In comparison, the “less English group have heard just 21% more tokens between Time 1 and Time 2.
Towards a quantitative model

In order to understand the effect of input during speech learning, we have to know:

– what kind of tokens have been heard
– to what distribution (phonetic category) they have been added
– how/if changes in the “auditory image” of the category changes
– how/if these perceptual changes result in modifications in phonetic implementation (production) of the phonetic category
Towards a quantitative model

Let’s imagine that when our native Italian participants arrived in Canada they had heard this array of VOT values in voiceless Italian stop consonants.
Towards a quantitative model

Then they began hearing English stops
Towards a quantitative model

Distribution of VOT in heard tokens

frequency
Towards a quantitative model

Distribution of VOT in heard tokens
Towards a quantitative model

Distribution of VOT in heard tokens

frequency
Towards a quantitative model

• Why did the native Italian participants who began producing English /p t k/ with longer VOT values begin using English more?

• Who were they speaking to?
  – Native English monolinguals?
  – Italians who spoke English with a foreign accent? (If so, we need to add English stops produced with “compromise” values into the input mix.)
Towards a quantitative model

- Perhaps members of the “more English” group used English more in 2003 than 1992 because the contexts in which they used English had expanded.
- Perhaps they came into contact with more English monolinguals with whom they had to speak English.

If so, then a 52% increase in the number of heard English tokens might have shifted the perceptual representation of voiceless English stops … and how they are produced … even after decades of prior L2 use.
Some conclusions

Input is generally considered unimportant for L2 speech learning. Why? Primarily because …

• We don’t know very much at present concerning what kind of input – and how much input – learners receive in their L2
• Even if we did know, we don’t know to what extent new input modifies long-term memory representations … and ultimately how these long-term memory representations guide production
My “hunch” is that input is an important determinant of success in L2 speech learning.

My motto: “You are what you eat, phonetically”

In fact, I think that input will eventually be shown to be more important than other potential determinants of “ultimate attainment” in L2 speech learning.
Some conclusions

Unfortunately, progress will be slow until we develop methods with which to actually measure (not just estimate!) L2 speech input
This concludes my remarks
Thanks for your attention!
Discussion?

Tuscania (VT) Italy
view from our street in Tuscania


