# Short-Term Phonetic Drift in an L2 Immersion Environment [1]

**Charles B. Chang**
University of California, Berkeley
cbchang@post.harvard.edu

**Abstract:**
This paper investigates the nature and time course of phonetic drift in L1 by examining the very first weeks of 20 L1 English speakers' acquisition of Korean as L2. Acoustic analyses of these learners' L1 and L2 production over time indicate that learning L2 stops affects the production of L1 stops (in terms of VOT and/or $f_0$ onset) in as little as one week, with the L1 sounds approximating the characteristics of the L2 sounds to which they are most phonetically similar. These results indicate that L1 phonological categories are affected by L2 learning on a very short timescale, suggesting that the equivalence classification that gives rise to this phonetic drift may be rather low-level in nature.

## 1. Introduction

When we learn a second language, what happens to our native language – the language we learned first? For a long time, nothing was thought to happen, since the earliest research on the interaction of first (L1) and second language (L2) phonologies was based on two related assumptions: the assumption of a so-called "critical period" for language acquisition and the assumption of unidirectionality of cross-language influence. The classic view of the critical period (cf. Lenneberg 1967) holds that biological changes in brain development are responsible for the general decline in ability to learn another language with increasing age; children are better at acquiring language than adults, then, because they have not yet passed this critical period of neural plasticity. This idea of neural plasticity in children and, by implication, neural rigidity in adults suggests that the linguistic structures of L1 are fossilized by the time one reaches adulthood. Therefore, while the L1 may cause some interference in the acquisition of an L2, the L1 itself should not be affected.

More recent work in phonetics and second language acquisition has challenged both of these assumptions. Some researchers (e.g. Flege 1987b) have pointed out numerous problems with the basic enterprise of proving that a critical period exists. Furthermore, there is mounting evidence that L1 can, in fact, be affected by the learning of an L2. In two influential studies, Flege (1987a) and Sancier and Fowler (1997) provide evidence that the phonetic space of L1 categories changes when speakers are immersed in an L2 environment for an extended period of time. Flege (1987a) examined the speech production of L1 French-L2 English speakers and L1 English-L2 French speakers – both groups being highly experienced in their L2 after having lived in an L2 environment for a number of years – in three case studies focusing on the realization of /t/, /u/, and /y/. With respect to the voiceless stops, Flege found that French /t/ – produced with short VOT in native French – was produced with VOTs that were longer than native by both groups, and that English /t/ – produced with long VOT in native English – was produced with VOTs that were shorter than native by both groups. With respect to the back vowels, French /u/ – produced with a low second formant (F2) in native French – was produced with F2s that were higher than native by both groups; similarly, English /u/ – produced with a relatively high F2 in native English – was produced with F2s that were lower than native by the L1 French speakers. On the other hand, French /y/, the only phoneme under investigation with no phonological counterpart in English, was produced in a native-like fashion by both groups. While Flege (1987a) examined a large sample of speakers at one time point, Sancier and Fowler (1997) followed one L1 Portuguese-L2 English speaker over time as she traveled between the U.S. and her native Brazil. In a study concentrating on this speaker's production of voiceless stops, they found that she produced shorter VOTs in both Portuguese and English stops immediately following months of immersion in Portuguese and longer VOTs in both languages following months of immersion in English, although the magnitude of the difference between

[1] The research presented in this paper was supported by a National Science Foundation Graduate Research Fellowship, a UC Berkeley Center for Korean Studies Graduate Student Fellowship, and travel grants from the UC Berkeley Department of Linguistics. I am grateful to Keith Johnson, John Ohala, Robert Port, and members of the UC Berkeley Phonology Lab for their helpful comments and feedback; to Jai-Ok Shim, Christine Arrozal, Luke Dolge, Jamal Grimes, Aimee Jachym, Jessica Lee, and Alexis Stratton for their invaluable assistance with study logistics; and to all study participants for their perseverance during a hot and humid summer. Any errors are my own.

the two conditions was very small (on the order of 5 ms). Sancier and Fowler explain the phonetic drift of both languages in the direction of the ambient language in terms of "[humans'] disposition to imitate, phonological correspondence, and preeminence of recency" (1997: 432). First, that the ambient language has an effect on anything is due to humans' tendency to imitate what they hear. Second, that the ambient language has an effect on production in another language is due to a connection between phonologically corresponding categories in the two languages; in other words, hearing Portuguese /t/ can affect the production of English /t/ because at some level they are the "same thing": voiceless coronal plosives. Finally, that L2 categories can have an effect on L1 categories even when they are acquired so late (as was the case with the speaker in Sancier and Fowler's study, who did not learn English until she was well into her teenage years) is due to the heavy weighting of recently experienced exemplars in memory. In this way, recent L2 experience can affect L1 representations even though an individual may have much more cumulative experience with L1.

Thus, these studies show that phonetic characteristics of L1 categories shift towards the phonetic norms of similar L2 categories when speakers have been living in an L2 environment for years or even months. As alluded to by Sancier and Fowler (1997), the framework of Flege's (1995) Speech Learning Model analyzes this sort of change in L1 as arising from an "equivalence classification" of similar L1 and L2 sounds that ties them to the same higher-level category, thereby allowing both sounds to be affected by input in L1 or L2. One important question remains unanswered, however: if equivalence classification of L1 and L2 categories is responsible for the mutual influence of L1 on L2 and L2 on L1, how and when does this occur? Previous work in this area has not been able to address this question because it investigates the pronunciation of fluent or highly proficient bilinguals who have spent a long time in an L2 environment and, moreover, focuses on languages that share the same alphabet (e.g. English, French, Portuguese, Italian, Spanish). These facts limit the generalizability of current knowledge in two ways. First, only a small percentage of the population of L2 learners ever reaches the level of proficiency attained by the fluent L2 speakers who are the object of most previous studies. Second, the focus on instances of L1-L2 contact in which sounds that are phonetically similar are also connected via identical orthographic representations is not representative of the many other cases of language contact in which L1 and L2 do not share the same writing system.

Consequently, our theory of multilingual phonetic competence is still very incomplete. Because the literature skips over the period of L2 acquisition during which cross-language connections are likely established, it has not been able to address the question of whether equivalence classification is an automatic linguistic phenomenon that occurs early in L2 acquisition or a more considered, metalinguistic phenomenon that occurs later in L2 development. In addition, conclusions about phonetically based equivalences between L1 and L2 categories have been confounded by the orthographic relationship between the languages examined previously, especially in light of the prominent role that written representations typically play in formal L2 education. This confound makes it unclear whether equivalence classification of similar sounds is actually based upon the phonetic relationship between the sounds, or simply based on orthographic identity between the sounds.

In this paper, I delve deeper into the nature and time course of L1 phonetic drift by examining the very first weeks of native English speakers' immersion in a Korean language environment. In doing this, I broaden the scope of previous research on L1 phonetic drift by (i) investigating novice L2 learners, and (ii) concentrating on a pair of languages that do not share the same writing system. Given that L1 phonological categories can be affected by similar L2 categories, are they affected from the very first stages of L2 acquisition, or is it only later in the process that bidirectional connections between L1 and L2 categories may be observed? The null hypothesis in this case is that L1 categories remain unchanged in the short term, but this remains to be demonstrated empirically. A production experiment was therefore carried out to investigate change in novice learners' L1 and L2 production over time. Here I concentrate on learners' production of laryngeal categories, focusing on the two acoustic dimensions that serve as the primary cues to the Korean laryngeal contrast among lenis, fortis, and aspirated stops in word-initial position: voice onset time (VOT) and fundamental frequency ($f_0$) onset (cf. Kim 2004, *inter alia*). The Korean categories differ from each other with respect to these dimensions in the following way: lenis stops have medium/long-lag VOT and relatively low $f_0$ onset; fortis stops have short-lag VOT and relatively high $f_0$ onset; and aspirated stops have long-lag VOT and relatively high $f_0$ onset. Since in each of these dimensions there are two categories with considerable overlap (lenis and aspirated on VOT, fortis and aspirated on $f_0$), both VOT and $f_0$ are necessary cues for making a three-way contrast. In contrast, VOT alone largely suffices to make the two-way distinction between "voiced" and "voiceless" stops in English (VOT being longer in voiceless stops than in voiced stops), although the English categories also differ in terms of the $f_0$ onset that follows ($f_0$ being lower following voiced stops and higher following voiceless stops, cf. Hombert 1978).

## 2. Methods

### 2.1. Procedure

The longitudinal production experiment was conducted weekly starting from one week into the language course participants were taking. The task involved was a reading task in which participants read aloud a set of Korean and English stimuli. The experiment took place in a quiet dormitory room and was divided into two parts by language. In the first part, all Korean stimuli were presented, and in the second part, all English stimuli were presented, with a break in between the two parts. In both parts of the experiment, stimuli were presented a total of four times, once each in four randomized blocks following a practice session of five items. Each item was presented on screen for 1.5 seconds and then replaced by a picture of a green traffic light to cue the participant to produce the item. Audio was recorded via a head-mounted condenser microphone for a period of 2 seconds starting at the time point at which the green light appeared on screen, and the inter-stimulus interval from the end of this recording to the presentation of the following item was 1 second. All stimuli presentation and audio recording was done in DMDX 3.2.6.3 (Forster 2008) on a laptop computer.

### 2.2. Stimuli

The set of stimuli consisted of 22 Korean and 23 English monosyllables representing most of the phonemic contrasts in the two languages, with members of a subgroup of stimuli being maximally similar in segmental makeup (e.g. Korean /hu/ vs. English /hud/). English monosyllables were of the form CVC to allow for lax vowels, while Korean monosyllables were generally of the form CV to make them easier for novice learners to read (cf. Table 1). The same set of stimuli was used in every week of the study.

**Table 1. Korean and English stimuli used in the production experiment**

| *Korean* | | *English* | | *Korean* | | *English* | |
|---|---|---|---|---|---|---|---|
| 바 | /pa/ | *'bot* | /bɑt/ | 히 | /hi/ | *heed* | /hid/ |
| 빠 | /p*a/ | | | | | *hid* | /hɪd/ |
| 파 | /p$^h$a/ | *pot* | /pɑt/ | | | *hate* | /he$^ɪ$t/ |
| 다 | /ta/ | *dot* | /dɑt/ | 해 | /hɛ/ | *head* | /hɛd/ |
| 따 | /t*a/ | | | | | *had* | /hæd/ |
| 타 | /t$^h$a/ | *tot* | /tɑt/ | 호 | /ho/ | *hoed* | /ho$^ʊ$d/ |
| 가 | /ka/ | *got* | /gɑt/ | 후 | /hu/ | *who'd* | /hud/ |
| 까 | /k*a/ | | | 흐 | /hɯ/ | *hood* | /hʊd/ |
| 카 | /k$^h$a/ | *cot* | /kɑt/ | 허 | /hʌ/ | *hut* | /hɐt/ |
| 사 | /sa/ | *sod* | /sɑd/ | 외 | /ɥɛ/ | *wait* | /we$^ɪ$t/ |
| 싸 | /s*a/ | *shot* | /ʃɑt/ | | | *wet* | /wɛt/ |
| 시 | /si/ | *seed* | /sid/ | 위 | /ɥi/ | *wee* | /wi/ |
| 씨 | /s*i/ | *sheet* | /ʃit/ | 알 | /aɭ/ | *all* | /ɒl/ |

### 2.3. Participants

Participants were 20 late learners of Korean (3 males, 17 females; 21-26 years old), functionally monolingual native speakers of American English with no prior exposure to Korean undergoing a six-week course of intensive Korean immersion instruction at the time of the study. On average these learners received four hours of instruction a day, for a total of approximately 82 hours of instruction by the end of the program (roughly the equivalent of one semester of college-level Korean). In exit questionnaires, they reported that class time constituted the majority of their experience with Korean, both in terms of listening and speaking.

### 2.4. Acoustic Analysis

Participants' recordings were acoustically analyzed in Praat 5.0.26 (Boersma and Weenink 2008). Manual measurements of VOT and $f_0$ onset were taken on learners' productions of the 15 words beginning with plosives (voiced and voiceless stops in the case of English; lenis, fortis, and aspirated stops in the case of Korean). VOT was measured off a wide-band Fourier spectrogram with a Gaussian window shape (window length: 5 ms; dynamic range: 50 dB; pre-emphasis: 6.0 dB/oct) as the time at voicing onset minus the time at the stop burst; thus, VOT was positive when the voicing onset followed the stop burst ("lag-voiced" stops) and negative when the voicing onset preceded the stop burst ("prevoiced" stops). To get stable measurements of $f_0$ onset, the combined wavelength of the first three regular glottal periods in the vowel was marked off on the waveform and converted into a frequency value (by inverting and then multiplying by 3). Initial periods were skipped if they were irregular (e.g. more than 33% longer or shorter than the following period); however, tokens requiring more than five periods of the vowel onset to be skipped were discarded.

In order to put male and female learners on the same $f_0$ scale, raw $f_0$ measurements were standardized to *z*-scores by learner and language. This was done by subtracting the learner's mean $f_0$ for the given language over the duration of the study and dividing by the square root of the learner's variance in $f_0$ for that language over the duration of the study.

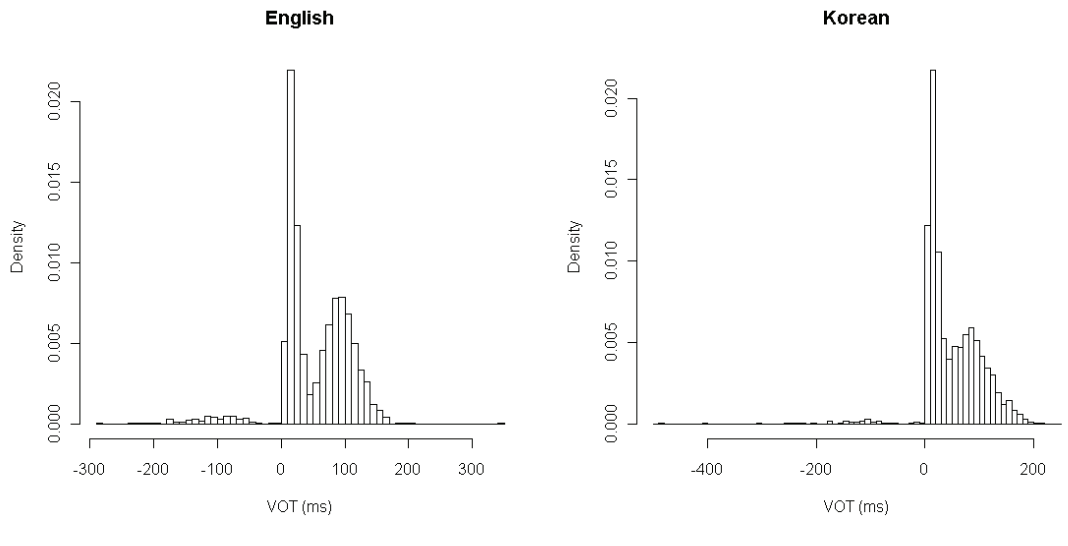
As four tokens were collected of each word, the data presented below are based on a total of approximately 60 tokens per learner per week (24 of the English words, 36 of the Korean words). Tokens with yawning, coughing, sighing, etc. were discarded.

## 3. Results

### 3.1. VOT over time

VOT in both languages shows a trimodal distribution reflecting three phonetic categories of voicing lag with respect to stop release (cf. Figure 1): voicing that begins prior to release (prevoicing), voicing that begins shortly after release (short-lag), and voicing that begins after a long delay following release (long-lag). Prevoiced stop tokens, while by far the least common of the three types (occurring at rates between 1% and 13% depending on the laryngeal category and time point in the study), are produced by learners in both languages. Prevoiced tokens, as well as long-lag tokens, have a wide VOT range – well over 100 ms in both cases. In contrast, short-lag tokens have a much narrower VOT range of approximately 40 ms that is centered around 10-20 ms in both languages.

**Figure 1. Histograms of overall VOT distributions in English and Korean**



An examination of VOT over time reveals no clear trends in the VOT of Korean fortis stops or English voiced stops (cf. Figures 2-3). Productions of these stops span all three voicing types shown in Figure 1 and are subdivided accordingly in the figures below ("prevoiced": less than 0 ms; "lag-voiced": 0 ms or greater; "short-lag": 0-30 ms; "long-lag": greater than 30 ms). Stars indicate the significance level of a *t*-test on VOT in consecutive weeks (1 star: $p < 0.05$; 2 stars: $p < 0.01$; 3 stars: $p < 0.001$).

**Figure 2. Mean VOT in short- and long-lag productions of fortis stops over time**

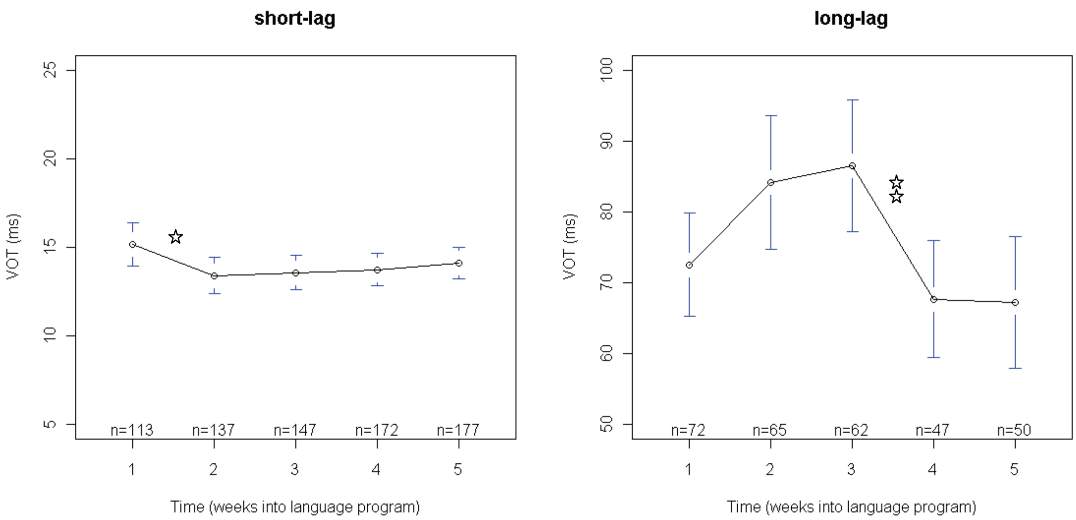
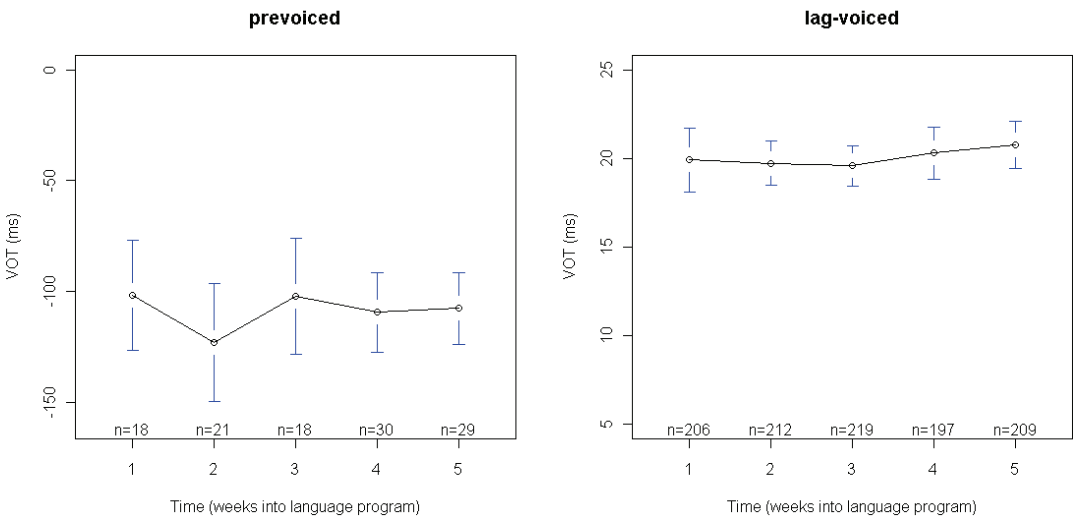


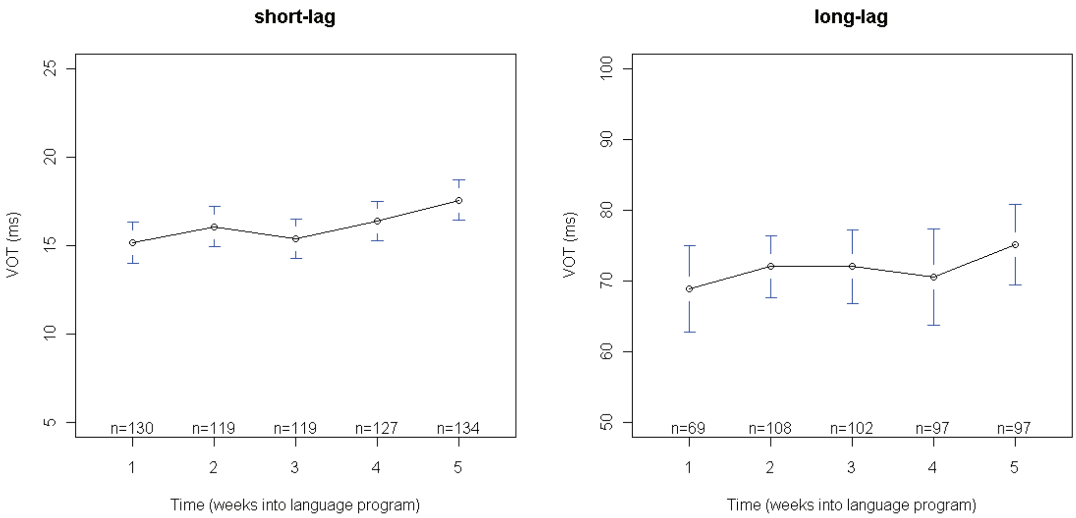

**Figure 3. Mean VOT in prevoiced and lag-voiced productions of voiced stops over time**




Although the differences in VOT between Weeks 1 and 2 in short-lag fortis and between Weeks 3 and 4 in long-lag fortis are significant, a repeated-measures analysis of variance (ANOVA) shows no main effect of time on VOT in fortis stops and no main effect of time on VOT in voiced stops – overall or in any of the VOT subsets shown above.

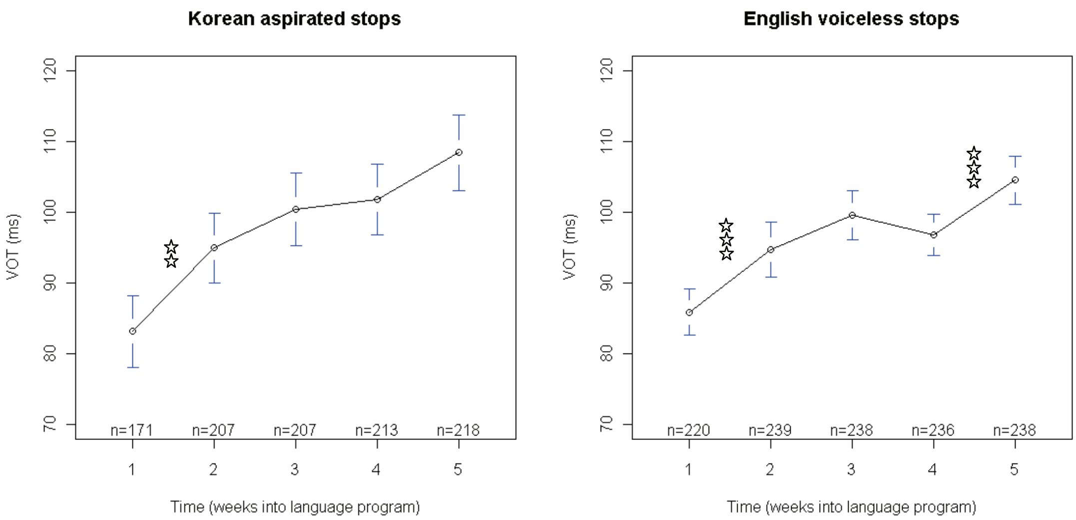
In the case of Korean lenis stops (cf. Figure 4), a repeated-measures ANOVA shows no main effect of time on VOT in long-lag lenis, but does show a main effect of time on VOT in short-lag lenis ($F[4, 19] = 3.841$, $p < 0.05$), where VOT gradually increases such that by Week 5, VOT is about 2.4 ms longer than in Week 1. The difference in VOT between these two endpoints is significant ($t[262] = -2.907$, $p < 0.01$); however, planned comparisons show no significant differences in VOT between consecutive weeks in either the short-lag or the long-lag VOT range and, unlike the short-lag VOT range, no significant difference in VOT between Weeks 1 and 5 in the long-lag VOT range.

**Figure 4. Mean VOT in short- and long-lag productions of lenis stops over time**




While there is no clear change in the VOT of fortis or voiced stops and only a slight increase in the VOT of short-lag productions of lenis stops, there is a clear and considerable increase in the VOT of Korean aspirated stops and English voiceless stops (cf. Figure 5).

**Figure 5. Mean VOT in long-lag productions of aspirated stops and voiceless stops over time**

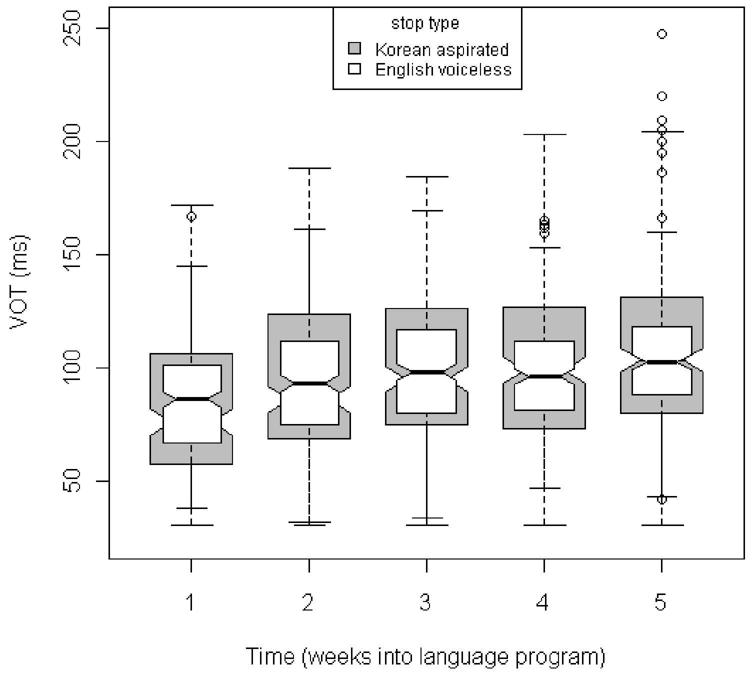



A repeated-measures ANOVA shows a main effect of time on VOT in aspirated stops ($F[4, 25] = 5.997$, $p < 0.01$), as well as a main effect of time on VOT in voiceless stops ($F[4, 61] = 9.503$, $p < 0.001$). Planned comparisons of consecutive weeks show that the increases in VOT of aspirated stops between

Weeks 1 and 2 ($t[262] = -2.907$, $p < 0.01$), of voiceless stops between Weeks 1 and 2 ($t[447] = -3.431$, $p < 0.001$), and of voiceless stops between Weeks 4 and 5 ($t[462] = -3.393$, $p < 0.001$) are all highly significant. The magnitude of the overall increase from Week 1 to Week 5 is also quite large for both stop categories: 25.3 ms in the case of aspirated stops, and 18.7 ms in the case of voiceless stops.

Comparing the VOT of English voiceless stops to that of Korean aspirated stops reveals that the two closely follow each other. In Week 1, mean VOT of long-lag productions of voiceless stops starts off higher than mean VOT of long-lag productions of aspirated stops, but then the two means get closer together, with aspirated VOT surpassing voiceless VOT in Week 4 (cf. Figure 6). However, planned comparisons of time-aligned aspirated VOT and voiceless VOT show that there is no reliable difference between the two means in any week. In other words, over the duration of the study Korean aspirated stops and English voiceless stops are produced with VOTs that are not statistically different from each other.

**Figure 6. Box plot of VOT in long-lag productions of aspirated stops (gray boxes) and voiceless stops (white boxes) over time**




In short, of the five laryngeal categories examined only three (Korean lenis in short-lag productions, Korean aspirated, and English voiceless) show a significant change in VOT over time. Short-lag lenis productions increase only slightly in VOT, whereas the latter two categories show an increase in VOT on the order of 20 ms by the final week of the study – a net change much larger in magnitude than the 5-ms effect found in Sancier and Fowler (1997).

### 3.2. $F_0$ onset over time

An examination of (standardized) $f_0$ onset over time reveals strong trends only for the English laryngeal categories. With regard to the Korean categories, first, there is no clear trend in the $f_0$ onset of Korean fortis stops (cf. Figure 7). Although the difference in $f_0$ onset between Weeks 1 and 2 in long-lag fortis productions is significant, a repeated-measures ANOVA shows no main effect of time on $f_0$ onset in fortis stops – overall or in any of the VOT subsets shown below. In contrast, English voiced stops show a steady increase in $f_0$ onset over time (cf. Figure 8), and a repeated-measures ANOVA shows the effect of time on $f_0$ onset in voiced stops to be highly significant ($F[4, 58] = 5.716$, $p < 0.001$). In particular, there are significant increases in $f_0$ onset of lag-voiced productions between Weeks 1 and 2 ($t$[414] = -4.846, $p < 0.001$) and between Weeks 3 and 4 ($t$[408] = -3.539, $p < 0.001$). From Week 1 to Week 5, $f_0$ onset goes up by about 0.8 standard deviations (equivalent to 15 Hz on average) – an increase of considerable magnitude.

**Figure 7. Mean standard $f_0$ onset in short- and long-lag productions of fortis stops over time**

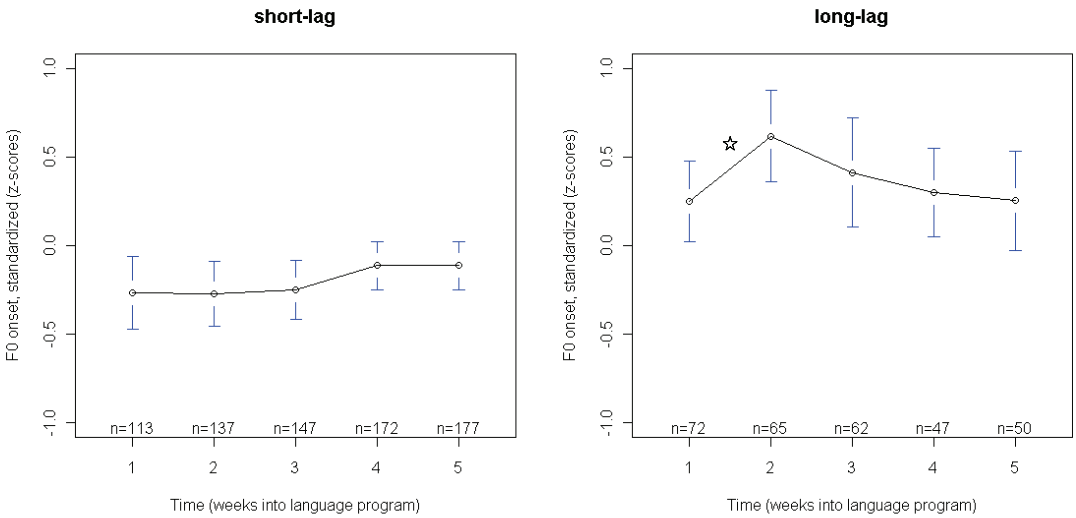
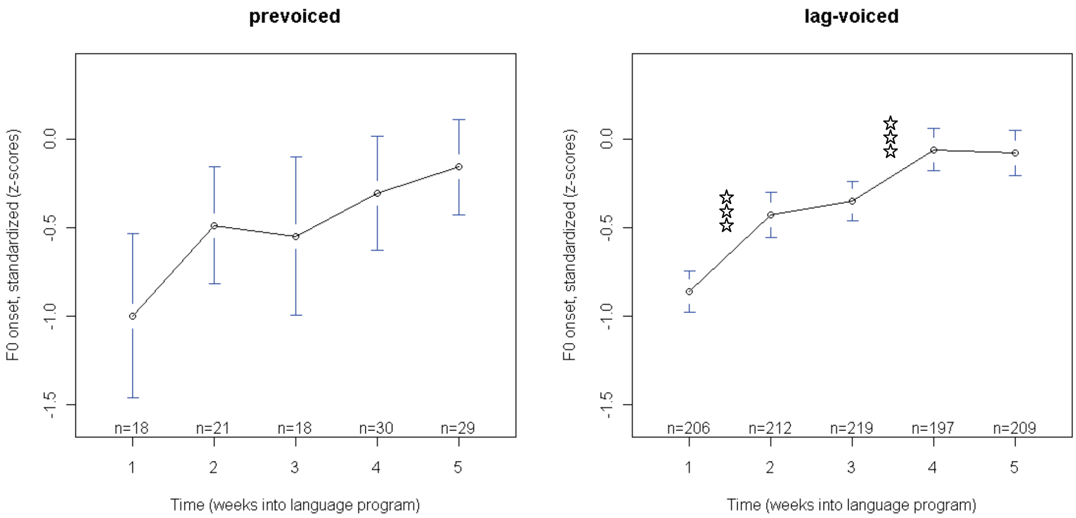
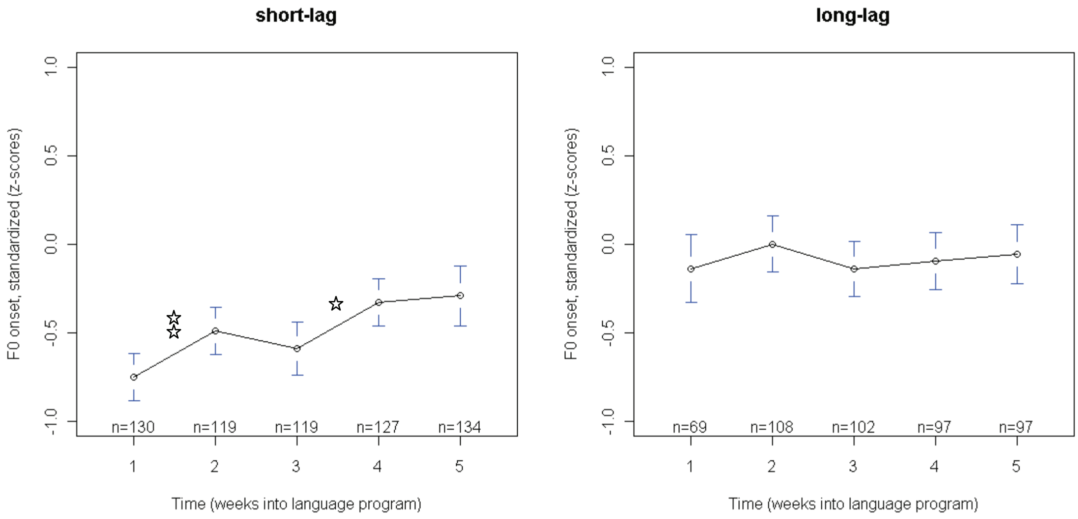



**Figure 8. Mean standard $f_0$ onset in prevoiced and lag-voiced productions of voiced stops over time**



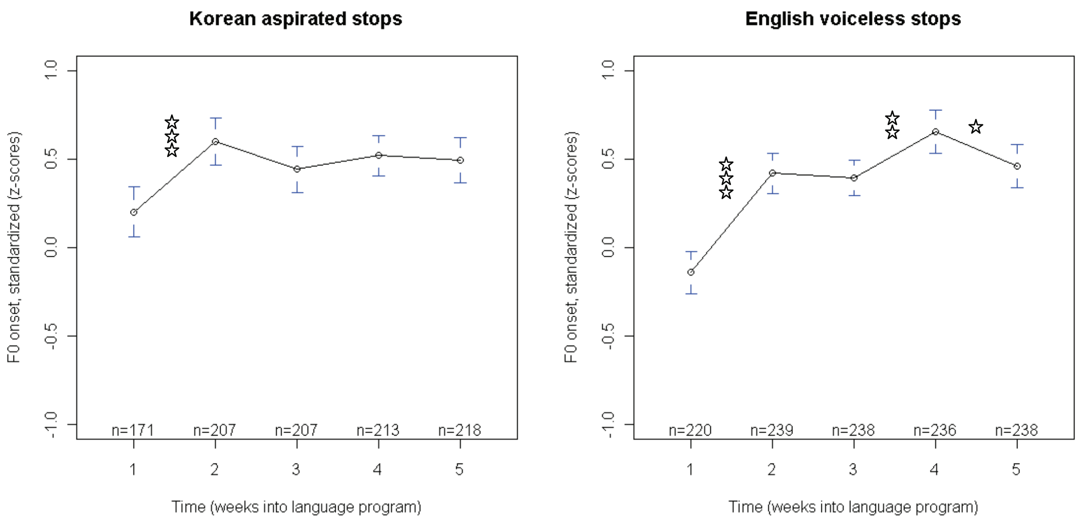

As for Korean lenis stops (cf. Figure 9), there is a marginally significant main effect of time on $f_0$ onset in short-lag productions ($F$[4, 19] = 2.530, $p = 0.074$), but no main effect of time on $f_0$ onset in long-lag productions. The differences in $f_0$ onset in short-lag productions between Weeks 1 and 2 ($t$[247] = -2.741, $p < 0.01$) and between Weeks 3 and 4 ($t$[239] = -2.543, $p < 0.05$) are both significant. The net increase in $f_0$ onset in short-lag lenis productions from Week 1 to Week 5 is approximately 0.46 standard deviations (a change equivalent to 8 Hz on average).

**Figure 9. Mean standard $f_0$ onset in short- and long-lag productions of lenis stops over time**




Finally, analysis of $f_0$ onset in Korean aspirated stops and English voiceless stops reveals that, similar to the pattern in VOT, $f_0$ onset generally increases in the long-lag productions of both stop types (cf. Figure 10). However, in the case of aspirated stops, there is no main effect of time on $f_0$ onset, although the increase in $f_0$ between Weeks 1 and 2 is significant ($t[365] = -4.014$, $p < 0.001$). On the other hand, there is a main effect of time on $f_0$ onset in voiceless stops ($F[4, 61] = 4.802$, $p < 0.01$), where $f_0$ increases by about 0.6 standard deviations (equivalent to 11 Hz on average) from Week 1 to Week 5.

**Figure 10. Mean standard $f_0$ onset in long-lag productions of aspirated stops and voiceless stops over time**




A comparison of $f_0$ onset in voiceless stops and aspirated stops at the same time points shows that these two stop types change in $f_0$ in a similar way, as was the case for VOT. Recall from Figure 10 that the $f_0$ increase in the aspirated stops happens primarily between Weeks 1 and 2, while the $f_0$ increase in the voiceless stops is a more general pattern. The effect of this difference in pattern of $f_0$ change is that here, too, the means of the two stop types get closer together (cf. Figure 11).

**Figure 11. Box plot of $f_0$ onset in long-lag productions of aspirated stops (gray boxes) and voiceless stops (white boxes) over time**

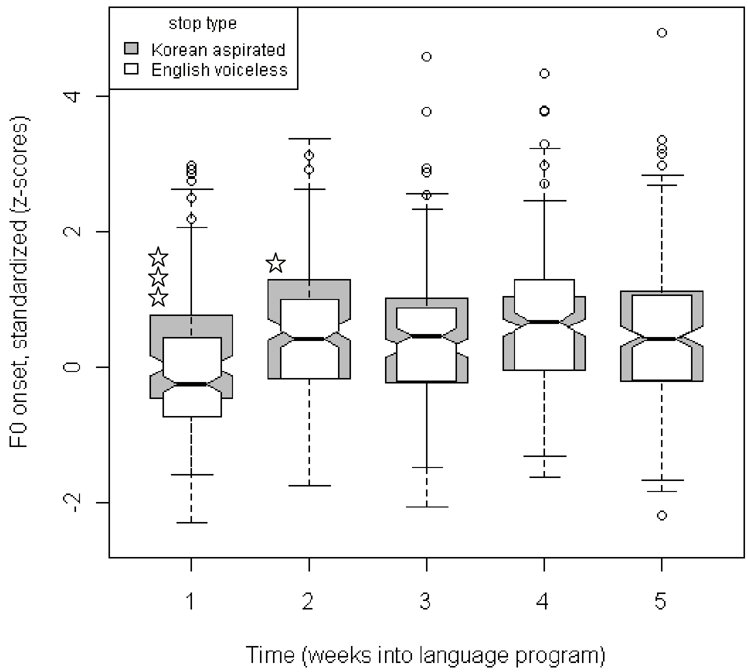
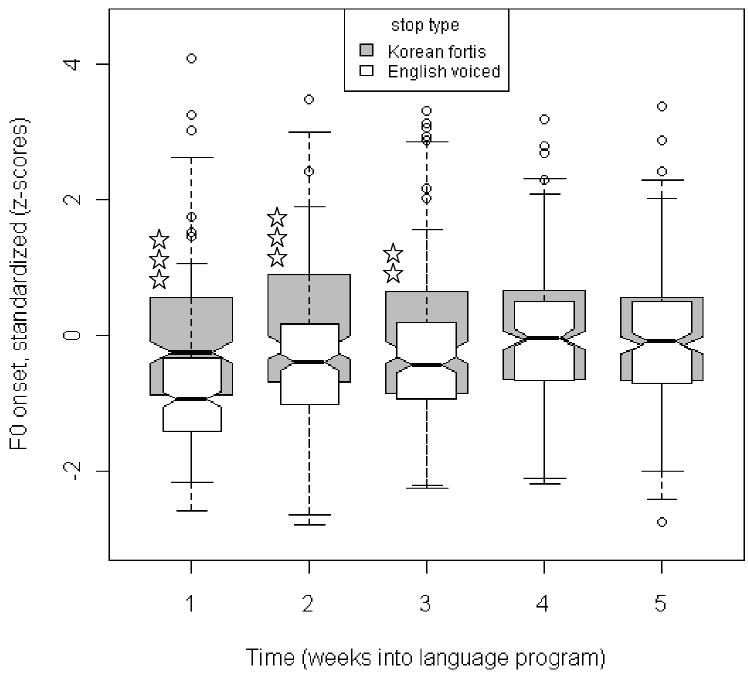



**Figure 12. Box plot of $f_0$ onset in lag-voiced productions of fortis stops (gray boxes) and voiced stops (white boxes) over time**



In Week 1, mean standard $f_0$ onset of long-lag productions of English voiceless stops starts off at 0.34 standard deviations below that of long-lag productions of Korean aspirated stops – a difference that is highly significant ($t[356] = -3.640$, $p < 0.001$) – but then the difference between the two means diminishes. In Week 2, mean voiceless $f_0$ is 0.18 standard deviations lower than mean aspirated $f_0$, a difference that is still significant ($t[419] = -2.045$, $p < 0.05$). However, from Week 3 on there is no longer a significant difference between the two means.

English voiced stops and Korean fortis stops display a similar pattern. As shown in Figures 7-8, $f_0$ onset increases in voiced stops, but not in fortis stops. A comparison of these two stop types at the same time points indicates that their means move closer together over time as well (cf. Figure 12). In Week 1, mean standard $f_0$ onset of lag-voiced productions of voiced stops starts off at 0.79 standard deviations below that of lag-voiced productions of fortis stops, a highly significant difference ($t[353] = -8.001$, $p < 0.001$), but as with the voiceless stops and aspirated stops, the difference between the two means diminishes over time. In Week 2, mean voiced $f_0$ moves to within 0.44 standard deviations of mean fortis $f_0$, a difference that is still highly significant ($t[393] = -4.269$, $p < 0.001$). A week later, mean voiced $f_0$ moves to within 0.30 standard deviations of mean fortis $f_0$, still a significant difference ($t[383] = -3.125$, $p < 0.01$). However, starting from Week 4 there is no longer a significant difference between the two means.

To summarize, none of the Korean stop categories shows a significant main effect of time on $f_0$ onset, although there are significant differences in $f_0$ onset between particular weeks. Only the English voiced and voiceless stops show a significant main effect of time on $f_0$ onset, which increases by 0.6-0.8 standard deviations (≈ 11-15 Hz) over the duration of the study.

## 4. Discussion and Conclusion

Why is the phonetic space of the English stops changing in this manner? This drift in L1 seems to be based in the sort of equivalence classification between L1 and L2 categories described by Flege (1987a). In his words, equivalence classification in general is "a basic cognitive mechanism which permits humans to perceive constant categories in the face of the inherent sensory variability found in the many physical exemplars which may instantiate a category" (Flege 1987a: 49). Applied to the learning of L2 phones, this process targets L2 phones that are phonetically similar to L1 phones and classifies them as the same category, thereby preventing a new category from being established for the L2 phone. In the context of a multi-tiered model of speech production (cf. Keating 1984, Flege and Eefting 1988, Laeufer 1997), classifying a new L2 sound as equivalent to an old L1 sound means essentially linking them at the highest level of representation, creating a cross-language connection between the two sounds that allows input in L1 or L2 to affect the sound in the other language as well. However, because levels of representation below the linked level of representation can remain separate, L1 and L2 sounds that have undergone equivalence classification may nevertheless be realized distinctly. For example, L1 English-L2 Spanish speakers who produce Spanish voiceless stops with VOTs that are longer than native Spanish VOTs (due to influence from long-lag English voiceless stops that have been linked to them) could still maintain a contrast between the two sets of voiceless stops by producing Spanish ones with VOTs that are reliably shorter than those of their English ones.

Once the phonetic properties of the Korean and English laryngeal categories are compared, these categories are predicted to be linked in the following way on the basis of their similarity in VOT: fortis to voiced, and aspirated to voiceless. This schema of equivalence classification is supported by evidence from learners' production of fortis stops, which they occasionally produce as prevoiced – often robustly so – even though such prevoiced productions are almost certainly absent from their Korean input. Fortis stops are realized by native Korean speakers as voiceless both word-initially and word-medially, with a relatively long duration of (voiceless) closure apparent in word-medial productions. Moreover, fortis stops are not generally Romanized with graphemes for voiced stops. Thus, the pattern of voicing fortis stops (which occurs 4-11% of the time depending on the time point in the study) appears to arise from a phonetically-based equivalence classification with English voiced stops, which are prevoiced at similar rates (8-13%) over the course of the study. This suggests that aspirated stops and voiceless stops, which seem to change in tandem as seen above, are linked to each other on a similar basis.

Whether lenis stops are linked to a particular L1 category is less clear, as their phonetic properties are somewhat intermediate between those of voiced and voiceless stops and, moreover, lenis stops do not pattern like either L1 category in this study. On the one hand, lenis stops are more similar in VOT to voiceless stops than to voiced stops in initial position, where they are typically realized with a substantial amount of aspiration. On the other hand, in medial position they are usually more similar in VOT to voiced stops, with voicing during closure or a short-lag VOT; the relatively low $f_0$ onset of initial lenis stops makes them similar to voiced stops as well. Interestingly, by showing some modest changes in VOT and $f_0$ limited to the short-lag range, lenis stops in this study seem to pattern neither like the voiced stops (which show a significant change in $f_0$, but not VOT) nor like the voiceless stops (which show large changes in VOT and $f_0$). It may be that as the "odd category out", lenis stops constitute the sort of "new" category for L2 Korean learners that French /y/ is argued to be for L2 French speakers in Flege (1987a).

The cross-language links proposed above are consistent with the patterns of change observed in this study: change in the English categories approximates the characteristics of the Korean categories to which they are most phonetically similar. Voiced VOT does not change significantly, since voiced stops are already similar to fortis stops in VOT; however, voiced $f_0$ onset rises in approximation to the elevated $f_0$ onset of fortis stops. Voiceless VOT increases along with aspirated VOT, and as happens with voiced $f_0$, voiceless $f_0$ rises in approximation to the elevated $f_0$ onset of aspirated stops. Thus, returning to the original research questions, it appears that equivalence classification of L1 and L2 sounds occurs in the very first weeks of L2 acquisition, and that it is based upon phonetic similarity, not orthographic identity.

As for the level at which equivalence classification occurs, evidence from a separate imitation task in which the vast majority of participants show command of only two English-like laryngeal categories (cf. Chang 2009) indicates that active command of L2 categories is not a prerequisite for equivalence classification with L1 categories. In other words, an L2 learner may equate L2 categories with L1 categories without being able to reliably produce these L2 categories. This result suggests that cross-language equivalence classification may be a low-level process – perhaps based on comparison of the acoustic details of L1 and L2 sounds – since high-level L2 categorical knowledge does not need to be fully developed for the linkage to occur. The implication is that equivalence classification might occur on a token-to-token basis rather than a category-to-category basis. In this way, L1 sounds and L2 sounds could become linked even when the L2 sounds are not yet solidly connected to an L2 category. Another possibility is that exactly the opposite occurs: with only superficial knowledge of L2 categories to draw upon (in this case, knowledge that there are three separate L2 laryngeal categories), learners make an explicit connection between the unfamiliar L2 categories and familiar L1 categories. Over the course of this study, learners made passing remarks (e.g. "[fortis] /p*/ is like [English] /b/") suggesting that explicit linkage of L1 and L2 categories could indeed be the basis of the cross-language connections observed here. Further work is required to clarify the role of "bottom-up" acoustic comparison vs. "top-down" categorical knowledge in L1-L2 equivalence classification. However, we should keep in mind that these two methods of cross-language linkage are not mutually exclusive, and that in the end both may be needed to fully account for changes in L1 and L2 production that occur over the course of L2 acquisition.

**References:**


Boersma, P. & Weenink, D. 2008. Praat: Doing phonetics by computer. Version 5.0.26. http://www.praat.org.

Chang, C. B. 2009. Phonological categories in early L2 acquisition. Poster presented at the 83rd Annual Meeting of the Linguistic Society of America. San Francisco, CA: Hilton San Francisco, January 10.

Flege, J. E. 1987a. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15(1):47-65.

Flege, J. E. 1987b. A critical period for learning to pronounce foreign languages? *Applied Linguistics* 8: 162-177.

Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In Strange, W., ed., *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Perception*, 233-272. Baltimore: York.

Flege, J. E. & Eefting, W. 1988. Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *Journal of the Acoustical Society of America* 83(2):729-740.

Forster, J. C. 2008. DMDX. Version 3.2.6.3. http://www.u.arizona.edu/~kforster/dmdx/dmdx.htm.

Hombert, J.-M. 1978. Consonant types, vowel quality, and tone. In Fromkin, V. A., ed., *Tone: A Linguistic Survey*, 77-111. New York: Academic Press.

Keating, P. 1984. Phonetic and phonological representations of stop consonant voicing. *Language* 60:286-319.

Kim, M. 2004. Correlation between VOT and F0 in the perception of Korean stops and affricates. In *INTERSPEECH 2004*, 49-52.

Laeufer, C. 1997. Towards a typology of bilingual phonological systems. In James, A. R. & Leather, J., eds., *Second-Language Speech: Structure and Process*, 325-342. Berlin: Mouton de Gruyter.

Lenneberg, E. 1967. *Biological Foundations of Language*. New York: John Wiley & Sons, Inc.

Sancier, M. L. & Fowler, C. A. 1997. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 27(4):421-436.