Rapid phonetic drift in second language acquisition

Charles B. Chang

Department of Linguistics, University of California, Berkeley
1203 Dwinelle Hall, Berkeley, CA 94720-2650, USA
cbchang@cal.berkeley.edu

December 19, 2009

Abstract

This paper is concerned with the nature and time course of phonetic drift in one's first language during acquisition of a second language. In a longitudinal study of native English speakers' acquisition of Korean, 20 late learners produced the same English and Korean items at weekly intervals over the course of intensive beginning Korean language classes. Results of acoustic analyses indicate that learning Korean stops influences the production of English stops (in terms of voice onset time and/or fundamental frequency onset) in as little as one week, with the English sounds approximating the characteristics of the Korean sounds to which they are most phonetically similar. These results indicate that first-language phonological categories are affected by second language learning on a very short time scale, suggesting that the equivalence classification between first- and second-language categories that gives rise to this phonetic drift may be rather low-level in nature.

Keywords: English, Korean, second language acquisition, phonetic drift, voice onset time, fundamental frequency

1 Introduction

When we learn a second language in our adult years, what happens to our native language, the language we learned first in childhood? Early observations of interaction between first- (L1) and second-language (L2) phonologies focused more on the influence of L1 on L2 and the phenomenon of "foreign accent" than on the influence of L2 on L1, as a result of two related assumptions: the existence of a so-called "critical period" for language acquisition and a unidirectional kind of cross-language influence. The classic view of the critical period (cf. Penfield & Roberts 1959; 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 because they have not yet passed this critical period of neural plasticity. In contrast, adults become neurally rigid, implying that the linguistic structures of L1 are fossilized in adulthood. It follows that while the

L1 may cause some interference in adult acquisition of an L2, the L1 itself should not be affected.

Recent work in phonetics and second language acquisition has challenged both of these assumptions. Some researchers (e.g. Flege 1987a) have pointed out numerous problems with the basic enterprise of proving that a maturationally defined critical period exists. Furthermore, there is mounting evidence that L1 can, in fact, be affected by the learning of an L2. Recognition of L2 influence on L1 goes back as early as Selishchev (1925), but is first discussed extensively in the work of Flege and colleagues (e.g. Flege 1987b, 1995, 2002; Flege *et al.* 2003; Flege 2007). In particular, the Speech Learning Model (SLM) introduced by Flege (1995:239) proposes that "[p]honetic categories established in childhood for L1 sounds evolve over the life span to reflect the properties of all L1 or L2 phones identified as a realization of each category." Thus, the notion of a static, fossilized L1 is replaced with that of a dynamic and ever-changing L1.

In a seminal study examining cross-language influence in adult L2 acquirers, Flege (1987b) provides evidence that the phonetic space of L1 categories changes when speakers are immersed in an L2 environment for an extended period of time. This study examined the speech production of L1 French-L2 English speakers and L1 English-L2 French speakers—both groups 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-lag voice onset time, or VOT, in native French) was produced with VOTs that were longer than native by both groups, and that English /t/ (produced with long-lag 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 in native French) was produced with formants that were higher than native by both groups; similarly, English /u/ (produced with a relatively high second formant in native English) was produced with formants 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. In his studies of VOT in L1 English-L2 Portuguese speakers, Major (1992, 1996) finds similar evidence of phonetic drift in L1 as a consequence of L2 acquisition, although he observes that "the influence of L2 is most prevalent in casual styles of L1 and may or may not be present in formal varieties" (Major 1992:204).

While Flege (1987b) and Major (1992) examined cross-sectional samples of several speakers at one time point, Sancier & Fowler (1997) followed one L1 Portuguese-L2 English speaker over time as she traveled between the U.S. and her native Brazil. 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, conversely, longer VOTs in both languages following months of immersion in English, although the magnitude of the difference between the two conditions was very small (on the order of 5 ms). The phonetic drift of both languages in the direction of the ambient language is accounted for in terms of "[humans'] disposition to imitate, phonological correspondence, and preeminence of recency" (Sancier & Fowler 1997:432). First, that the ambient language has an effect on any aspect of production 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 was already a teenager when she began learning English in earnest) 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.

In short, 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 a long time. As alluded to by Sancier & Fowler (1997), the framework of the SLM 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. But this begs the question: when and how is equivalence between L1 and L2 categories established? Previous work on adult L2 learners has not been able to address this issue because it investigates the pronunciation of L2 learners who are at or close to the "end state" of L2 acquisition, i.e. fluent or highly proficient bilinguals who have spent years in an L2 environment. Moreover, this work focuses on languages which share the same alphabet and consequently represent many similar sounds with identical graphemes (e.g. English, French, Portuguese, Italian, Spanish).

Consequently, our theory of multilingual phonetic competence is still very incomplete. Because the literature skips over the period of adult 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 longitudinal 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*). As schematized in Figure 1, the Korean categories differ from each other with

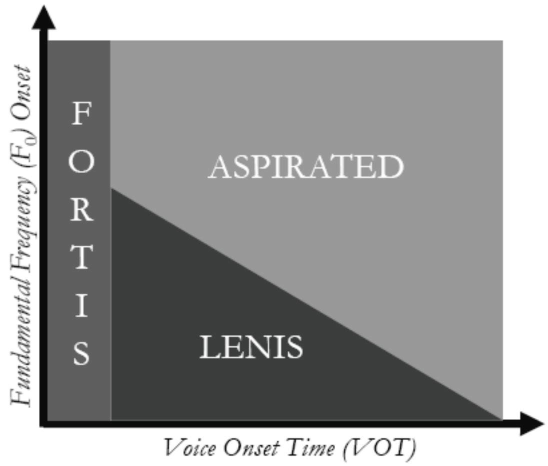

Figure 1: Schematic of f_0 onset vs. VOT in the three-way Korean laryngeal contrast among lenis, fortis, and aspirated plosives and affricates (based on Kim 2004).

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 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 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. All were paid for their participation. On average these learners received four hours of Korean instruction a day, for a total of approximately 82 hours of instruction by the end of the program (roughly equivalent to one semester of college-level Korean). In exit questionnaires, participants reported that class time constituted the majority of their experience with Korean, both in terms of listening and speaking.

2.2 Procedure

The 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

Korean		English		Korean		English	
바	/p͉a/	*'bot*	/bɑt/	히	/hi/	*heed*	/hid/
빠	/p͈a/					*hid*	/hɪd/
파	/pʰa/	*pot*	/pɑt/	해	/hɛ/	*hate*	/heᶦt/
다	/p͉a/	*dot*	/dɑt/			*head*	/hɛd/
따	/p͈a/					*had*	/hæd/
타	/tʰa/	*tot*	/tɑt/	호	/ho/	*hoed*	/hoᵘd/
가	/k͉a/	*got*	/gɑt/	후	/hu/	*who'd*	/hud/
까	/k͈a/			흐	/hɯ/	*hood*	/hʊd/
카	/kʰa/	*cot*	/kɑt/	허	/hʌ/	*hut*	/hɐt/
사	/s͉a/	*sod*	/sɑd/	외	/ɥɛ/	*wait*	/weᶦt/
싸	/s͈a/	*shot*	/ʃɑt/			*wet*	/wɛt/
시	/s͉i/	*seed*	/sid/	위	/ɥi/	*wee*	/wi/
씨	/s͈i/	*sheet*	/ʃit/	알	/al̢/	*all*	/ɒl/

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

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 (Forster 2008) on a laptop computer.

2.3 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. The same set of stimuli was used in every week of the study (see Table 1 for a full list[1]).

[1] IPA transcriptions use the extended IPA symbols for weaker and stronger articulations (International Phonetic Association 1999:189) to transcribe lenis and fortis obstruents.

2.4 Acoustic analysis

Participants' recordings were acoustically analyzed in Praat (Boersma & 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. This was done by subtracting the learner's mean f_0 over the duration of the study and dividing by the square root of the learner's variance in f_0 over the duration of the study.

As four tokens were collected of each word, the data presented in Section 3 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 Change in English stops

VOT in English shows a trimodal distribution reflecting three phonetic categories of voicing lag with respect to stop release (cf. Figure 2): 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. Because of this trimodal distribution of VOT, speakers' productions (in particular, of English voiced stops) are divided into subsets for the purposes of analysis according to which VOT range they fall in: "prevoiced" (<0 ms), "short-lag" (0–30 ms), or "long-lag" (>30 ms).

An examination of short-lag productions of English voiced stops over time reveals that while VOT in these productions does not change much, f_0 onset rises significantly (cf. Figure 3). A repeated-measures analysis of variance (ANOVA) with within-subjects factors Place (of articulation of the stop), Token, and Time (point in the language program) shows a highly significant main effect on VOT of Place [$F(2,4)=125.86$, $p<0.001$], as expected, but

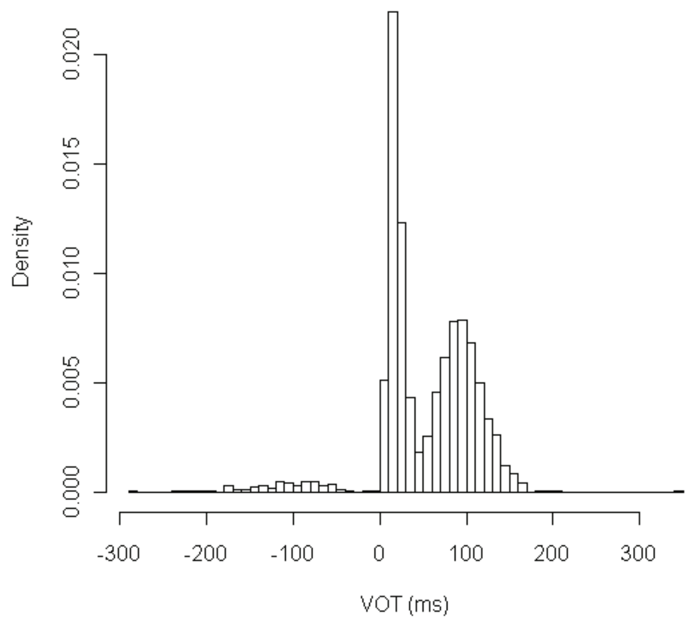

Figure 2: Histogram of VOT in word-initial productions of English plosives.

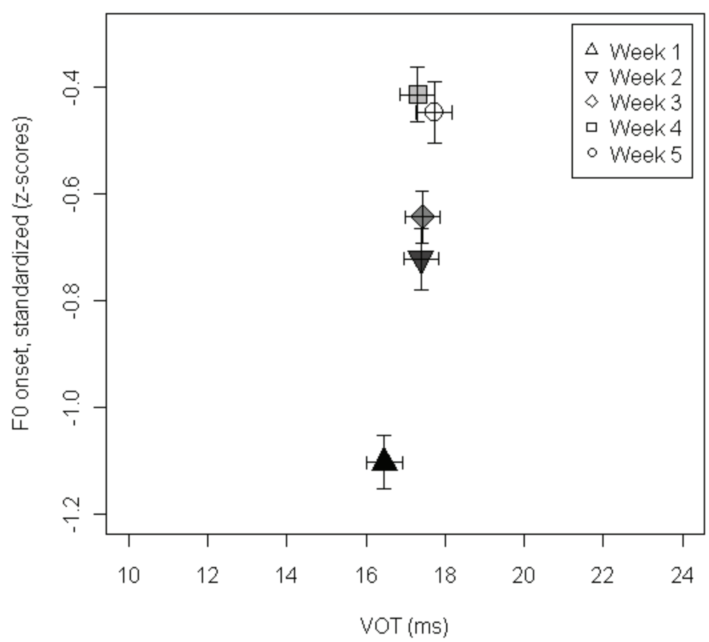

Figure 3: Mean f_0 onset vs. mean VOT in short-lag productions of English voiced plosives over time. The progression of weeks corresponds to different plot shapes with increasingly lighter shading. Error bars indicate ± 1 standard error about the mean.

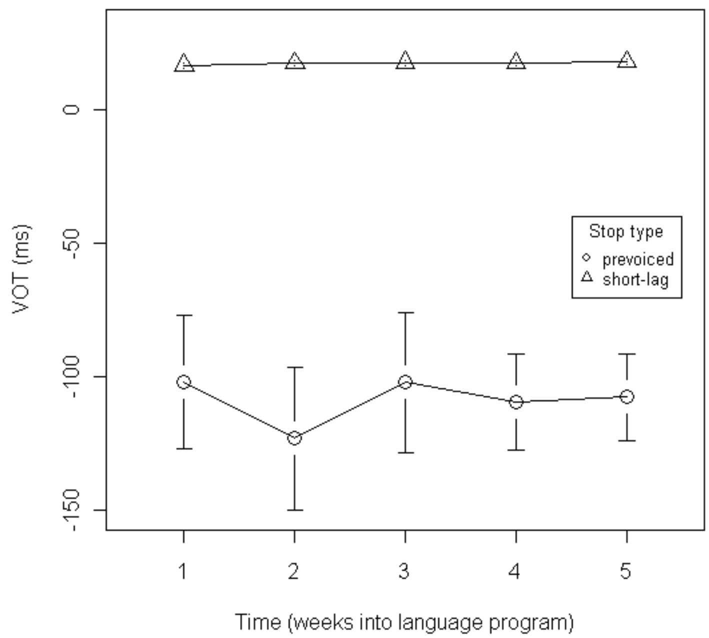

Figure 4: Mean VOT in prevoiced (circles) and short-lag (triangles) productions of English voiced plosives over time. Error bars indicate 95% confidence intervals.

no effect of Token [$F(3,3)=1.98$, $p=0.3$] or Time [$F(4,25)=1.56$, $p=0.2$] and no interactions among these factors. In contrast, a repeated-measures ANOVA reveals a highly significant main effect of Time on f_0 onset [$F(4,25)=5.05$, $p=0.004$]. There is also a main effect of Place [$F(2,4)=7.08$, $p=0.05$], but no effect of Token [$F(3,3)=1.79$, $p=0.3$] and no interactions.

Note that the patterns found in short-lag productions of voiced stops hold true of prevoiced productions of voiced stops as well. In Figure 4, it can be seen that, with the exception of a dip in VOT in prevoiced stops in Week 2, both subsets of voiced stop productions show little change in VOT over time. In Figure 5, it can be seen that both subsets of voiced stop productions steadily increase in f_0 onset over time, the f_0 of the prevoiced subset remaining below that of the short-lag subset in all weeks as expected.

An examination of long-lag productions of English voiceless stops over time reveals that both VOT and f_0 onset increase significantly (cf. Figure 6). A repeated-measures ANOVA with the same factors as above shows highly significant main effects on VOT of Place [$F(2,22)=12.70$, $p<0.001$] and Time [$F(4,61)=9.50$, $p<0.001$], but no effect of Token [$F(3,38)=0.37$, $p=0.8$]. With respect to f_0, a repeated-measures ANOVA reveals no effect of Place [$F(2,22)=1.44$, $p=0.3$], but significant effects of Time [$F(4,61)=5.15$, $p=0.001$] as well as Token [$F(3,38)=4.08$, $p=0.01$]. The effect of Token here is attributable to a general decrease in f_0 over the course of the experiment, likely due to subject fatigue. As with the ANOVAs of VOT and f_0 in voiced stops, there are no interactions among factors in the ANOVAs of VOT and f_0 in voiceless stops.

In short, over the course of Korean classes learners' English voiced stops do not increase significantly in VOT, but do increase significantly in f_0 onset, while their English voiceless

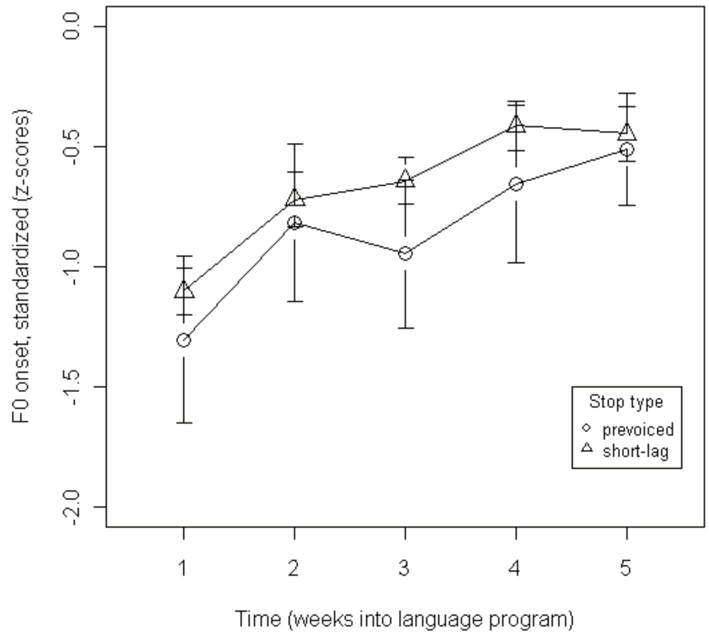

Figure 5: Mean f_0 onset in prevoiced (circles) and short-lag (triangles) productions of English voiced plosives over time. Error bars indicate 95% confidence intervals.

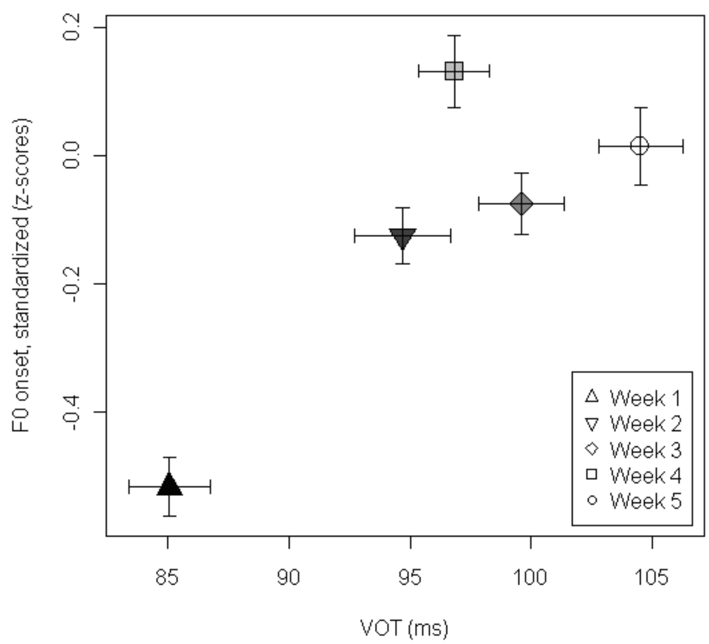

Figure 6: Mean f_0 onset vs. mean VOT in long-lag productions of English voiceless plosives over time. The progression of weeks corresponds to different plot shapes with increasingly lighter shading. Error bars indicate ± 1 standard error about the mean.

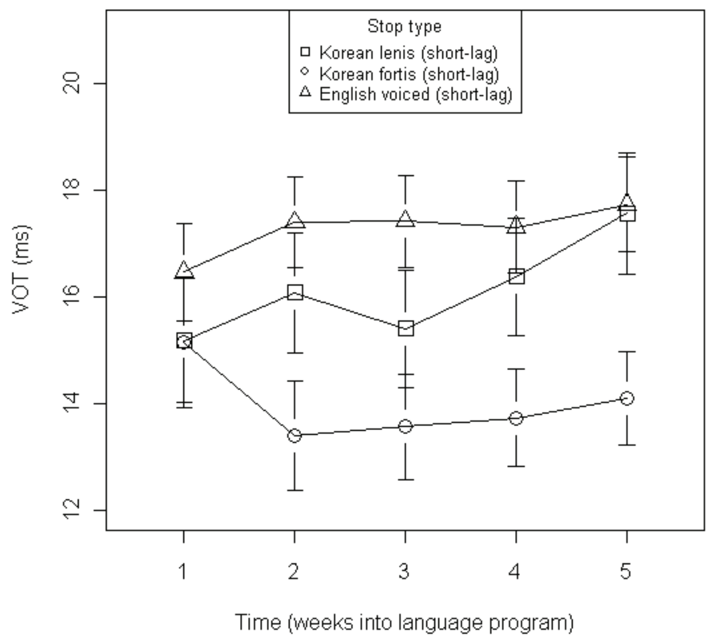

Figure 7: Mean VOT over time in short-lag productions of English voiced plosives (triangles), Korean lenis plosives (squares), and Korean fortis plosives (circles). Error bars indicate 95% confidence intervals.

stops increase significantly in both VOT and f_0 onset. Explanations for these changes are examined in the next section.

3.2 Change in English vs. change in Korean

As is to be expected, learners' Korean shows developments over the course of the Korean language program at the same time their English is changing. Comparisons of longitudinal changes in L1 and L2 reveal several noteworthy patterns, described below.

3.2.1 Short-lag stops

With regard to the short-lag stop series, the VOT of short-lag English voiced stop productions drifts slightly upward while the VOT of short-lag Korean fortis stop productions drifts slightly downward, such that these two stop types go from being not significantly different in VOT in Week 1 to being significantly different in VOT by Week 5 (cf. Figure 7). The magnitude of the difference between the two categories is small (approximately 4 ms), but significant [$t(352)$=5.74, p<0.001 with Bonferroni correction]. Interestingly, the VOT of short-lag productions of Korean lenis stops shows a developmental pattern that is somewhat intermediate between that of voiced and fortis stops. Lenis stops start off with the same VOT as fortis stops in Week 1, but then steadily increase in VOT over time to the VOT level of voiced stops in Week 5. As discussed in Section 3.1, there is no main effect of Time

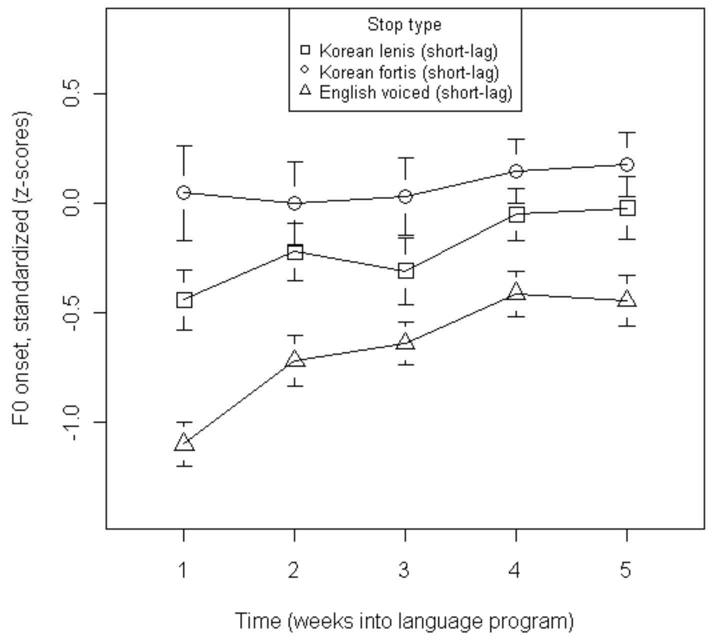

Figure 8: Mean f_0 onset over time in short-lag productions of English voiced plosives (triangles), Korean lenis plosives (squares), and Korean fortis plosives (circles). Error bars indicate 95% confidence intervals.

on VOT in short-lag voiced stop productions. Time does not have a main effect on VOT in short-lag fortis stop productions, either [$F(4,22)=1.10$, $p=0.4$]. However, Time does have a main effect on VOT in short-lag lenis stop productions [$F(4,19)=3.84$, $p=0.02$], attributable to the upward trend seen in Figure 7.

As for f_0, the f_0 onset of short-lag fortis stop productions stays steady, while that of short-lag voiced stop productions and that of short-lag lenis stop productions both drift upward (cf. Figure 8). In addition to the main effect of Time on f_0 of short-lag voiced stop productions, there is also a marginally significant main effect of Time on f_0 of short-lag lenis stop productions [$F(4,19)=2.47$, $p=0.08$]. As was the case with VOT, the f_0 of lenis stops patterns in between that of voiced and fortis stops, staying at an intermediate level in every week. The result of these f_0 developments is that voiced stops (as well as lenis stops) become more similar to fortis stops in f_0 onset. The standard f_0 distance between voiced and fortis stops starts off at 1.2 standard deviations in Week 1, but shrinks to 0.6 standard deviations by Week 5.

3.2.2 Long-lag stops

While there are only minute changes in the VOTs of short-lag voiced, fortis, and lenis stop productions, there are substantial increases in the VOTs of long-lag voiceless and aspirated stop productions (cf. Figure 9). The VOT of long-lag lenis stop productions remains relatively steady over time, but the VOTs of long-lag voiceless and aspirated stop productions

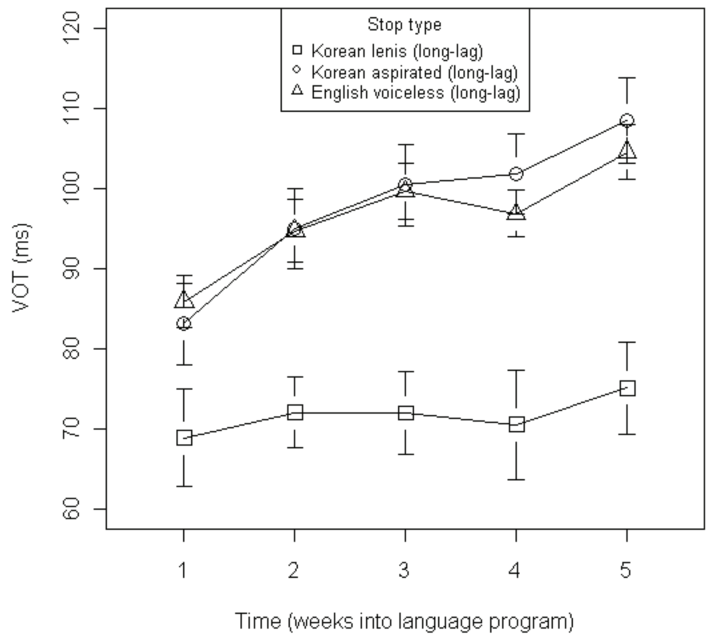

Figure 9: Mean VOT over time in long-lag productions of English voiceless plosives (triangles), Korean lenis plosives (squares), and Korean aspirated plosives (circles). Error bars indicate 95% confidence intervals.

rise significantly. Over five weeks, voiceless VOT increases by 19 ms over its initial level in Week 1, while aspirated VOT increases by approximately 25 ms. These two sets of stops are not significantly different from each other in VOT at any time point, although they begin to pull apart in Week 4, at which point they are marginally different from each other [$t(346)$=-1.70, p=0.09].

Time has a main effect not only on VOT in long-lag voiceless stop productions (cf. Section 3.1), but also on VOT in long-lag aspirated stop productions [$F(4,25)$=6.00, p=0.002]. This result follows from the pattern seen in Figure 9, where it is apparent that the VOTs of voiceless and aspirated stops closely follow each other. The increase in VOT of aspirated stops is unsurprising, as it is consistent with learners' approximation of the longer VOT norm for Korean aspirated stops, which are approximately 35 ms more aspirated than English voiceless stops on average (cf. Lisker & Abramson 1964:394–397). The increase in VOT of voiceless stops, however, cannot be explained in the same manner, since—at a VOT of 86 ms in Week 1—they start well above the VOT norm for English voiceless stops (cited as approximately 70 ms, cf. Lisker & Abramson 1964).

The way in which voiceless and aspirated stops move in lockstep suggests in fact that they have undergone equivalence classification, and that what is happening in terms of VOT development is that as the aspirated stops come to be produced with increasingly native-like VOT, the voiceless stops "go along for the ride" and increase in VOT as well, even though this increase results in their becoming less native-like vis-à-vis the phonetic norms of English.

With respect to f_0, the f_0 onset of long-lag aspirated stop productions—with the excep-

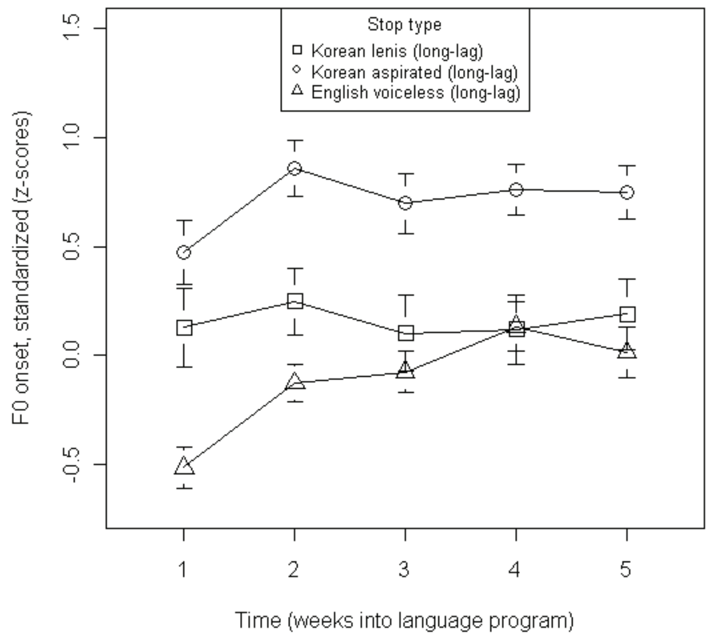

Figure 10: Mean f_0 onset over time in long-lag productions of English voiceless plosives (triangles), Korean lenis plosives (squares), and Korean aspirated plosives (circles). Error bars indicate 95% confidence intervals.

tion of a spike upwards in Week 2—stays relatively steady, as does the f_0 onset of long-lag lenis stop productions, which stay at a lower level overall (cf. Figure 10). There is no main effect of Time on long-lag aspirated f_0 onset [$F(4,25)=0.84$, $p=0.5$] or on long-lag lenis f_0 onset [$F(4,10)=0.19$, $p=0.9$]. In contrast, the f_0 onset of long-lag voiceless stop productions drifts steadily upward, as discussed in Section 3.1. The result of this f_0 increase is again that English stops—the voiceless stops in this case—become more similar to Korean stops in f_0 onset. The standard f_0 distance between voiceless and aspirated stops starts off at 1.0 standard deviations in Week 1, but shrinks to 0.7 standard deviations by Week 5.

In short, comparisons of the patterns of VOT and f_0 change in English and Korean stops suggests that English voiced stops are linked to Korean fortis stops, while English voiceless stops are linked to Korean aspirated stops. Neither voiced stops nor fortis stops show a substantial change in VOT, although they move slightly away from each other in VOT over time. However, voiced stops slowly approximate the elevated f_0 onset of fortis stops. As for voiceless stops and aspirated stops, they both 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 & Fowler (1997). The VOTs of these two stop types show no significant differences from each other over five weeks, although they push apart a bit in the last two weeks of the study. Like the voiced stops, the voiceless stops approximate the higher f_0 onset of the parallel Korean stop series (the aspirated stops, in this case), with the result that by the end of the study, the standard f_0 distance between parallel English and Korean stop categories shrinks by 0.3–0.6 standard deviations ($\approx$ 5–10 Hz).

4 Discussion and Conclusions

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 (1987b). 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 1987b: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. As alluded to in Section 1, in the context of a multi-tiered model of speech production (cf. Keating 1984; Flege & Eefting 1988; Laeufer 1996) 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. The data reviewed in Sections 3.2.1 and 3.2.2 are consistent with this schema of equivalence classification. These equivalence relationships are further supported by evidence from learners' production of fortis stops, which they occasionally produce as prevoiced—often robustly so—even though such prevoiced productions are virtually 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 transliterated with graphemes for voiced stops (in fact, they are transliterated as <pp, tt, kk> in the Revised Romanization system now widespread throughout South Korea). 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 change in tandem as seen in Section 3.2.2, 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, relative to those of voiced and voiceless stops, are somewhat intermediate. 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, the relatively low f_0 onset of initial lenis stops makes them similar to voiced stops; furthermore, in medial position lenis stops are usually more similar in VOT to voiced stops, with voicing during closure or a short-lag VOT. Interestingly, by showing some modest changes in VOT and f_0 limited to short-lag productions, 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 significant, as well as large, changes in both 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 (1987b).

The cross-language links proposed here 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. Despite these patterns of approximation, however, contrast between categories is always maintained (cf. Figure 11). Returning to the original research questions, then, 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.

Thus, the current proposal is that the phonetic drift in L1 seen here is ultimately caused by the early establishment of equivalences between similar L1 and L2 categories, the sort of equivalence relationships that have been proposed to exist for proficient L1-L2 bilinguals. However, two alternative explanations for these results should be addressed.

The first alternative account is that the changes in L1 arise not from relations to the developing L2, but from the experience of having to communicate with Korean locals, who are almost invariably L2 English speakers.[2] In other words, could the changes be attributable to the increased usage, and thus generally higher activation level, of foreigner-directed speech forms? While the changes in the voiceless stops can be explained this way, the changes in the voiced stops cannot, since if the changes merely arose from hyperarticulated pronunciations, one would expect the voiced stops to shift in ways opposite to the attested patterns: voiced stops should become more voiced (i.e. at least prevoiced stop productions should decrease in VOT), as well as lower in f_0 onset. However, what actually happens is that voiced stops change little in VOT (in fact, the VOT of short-lag productions slightly *increases*) and rise in f_0 onset. Thus, the present results cannot be accounted for by the influence of foreigner-directed speech.

The second alternative account is that increased familiarity with the experimental task and stimuli—combined with a higher level of comfort with the experimental environment—leads to the L1 changes observed by way of allowing participants to give more "confident" pronunciations.[3] Indeed, higher levels of speaker confidence might very reasonably result in higher levels of f_0 onset; one can even imagine that more confident speakers might aspirate their voiceless stops more. However, it is unclear why this increased confidence should affect only English stop productions and not Korean stop productions. As shown in Figures 8 and 10, with the exception of short-lag lenis stop productions Korean stops do not show a clear trend of increasing f_0 onset over time. If anything, though, one would expect Korean stops

[2]Thanks to Ann Bradlow for bringing up this possibility.

[3]Thanks to John Ohala for this suggestion.

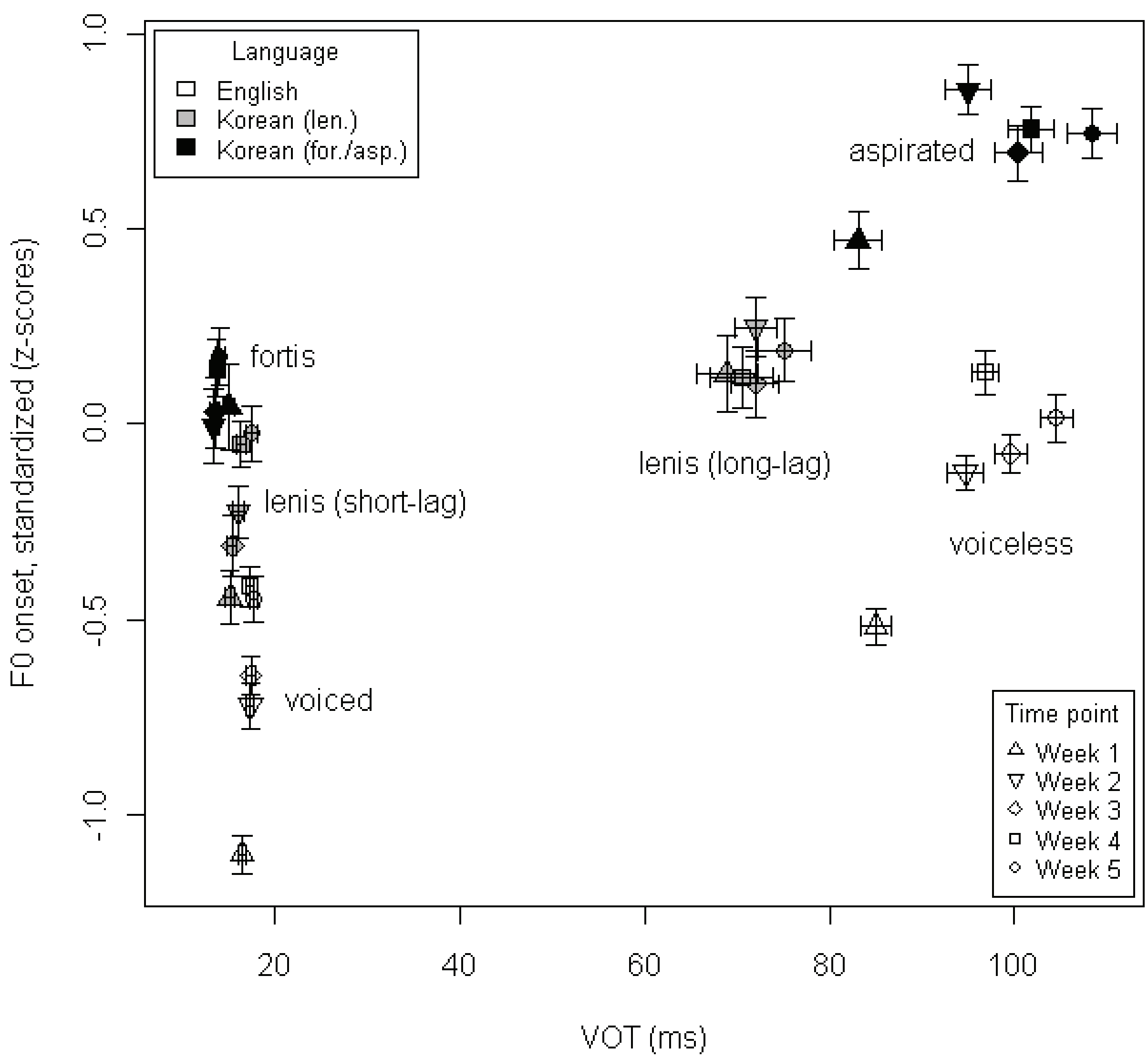

Figure 11: Mean f_0 onset vs. mean VOT in English voiced and voiceless plosives and Korean lenis, fortis, and aspirated plosives. The progression of weeks corresponds to different plot shapes, while different laryngeal categories are denoted by different degrees of shading. Error bars indicate ±1 standard error about the mean.

to most clearly show the effects of increased confidence, as the Korean forms are the ones which are actually unfamiliar to learners to begin with. The fact that they show such little change in comparison to English stops suggests that the present results are unlikely to have resulted from gradually increasing levels of speaker confidence.

Finally, the issue of task order should be addressed as well, since the Korean-English order of stimulus presentation (cf. Section 2.2) was purposefully kept the same across participants, rather than counterbalanced.[4] Could pronouncing Korean words first have affected the way participants pronounced the English words that followed? The short answer is yes, but this is a trivial question for two reasons. First, while it is quite likely that completion of the initial Korean production task affected performance on the subsequent English production task, this effect is controlled for by the fact that the task order was the same at all five time points in this study.[5] Second, the way in which task order would have affected performance is precisely the subject of this study: cross-language influence via L1-L2 category relationships. In many ways, if participants' production of English stops was influenced by their production of Korean stops shortly beforehand, the arguments advanced in this study are further supported, since this short-term influence should also occur via links of equivalence classification between similar L1 and L2 categories.

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 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] /p͈a/ 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

[4]In cross-sectional experiments, the control against effects of particular experimental conditions (e.g. task order) is to vary these conditions across participants so that patterns in the data cannot be attributed to the particular conditions that participants were subject to. On the other hand, in longitudinal experiments, the control against effects of particular experimental conditions is to keep the conditions the same across time points so that any effect of these particular conditions on the measured variable is automatically parceled out (i.e. if conditions are kept the same, then any change observed over time cannot be attributed to differences in conditions between time points).

[5]Thus, task order could only affect the results via an interaction with time (i.e. different effects of the task order on the measured variable at different time points), but there is no immediately apparent reason to posit such an interaction.

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.

Acknowledgments

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. The author is grateful to Ann Bradlow, Keith Johnson, John Ohala, Robert Port, members of the UC Berkeley Phonology Lab, and audiences at the 158th Meeting of the Acoustical Society of America and the 2009 Seoul International Conference on Linguistic Interfaces for helpful comments and feedback; to the Fulbright Korean-American Educational Commission—especially Jai-Ok Shim, Christine Arrozal, Luke Dolge, Jamal Grimes, Aimee Jachym, Jessica Lee, and Alexis Stratton—for invaluable assistance with study logistics; and to all study participants for their perseverance during a hot and humid summer. Portions of these data are also discussed in Chang (2009). Any errors are those of the author.

References

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

Chang, C. B. (2009). Short-term phonetic drift in an L2 immersion environment. In *Current Issues in Linguistic Interfaces (Volume 2)*, edited by Y.-S. Kang, S.-W. Tang, C. Kim, J.-Y. Yoon, Y.-S. Kang, K.-A. Kim, H. Yoo, Y. Jang, & H.-K. Kang, 31–46. Seoul, South Korea: Linguistic Society of Korea.

Flege, J. E. (1987a). A critical period for learning to pronounce foreign languages? *Applied Linguistics*, *8*, 162–177.

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

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange, 233–272. Baltimore, MD: York Press.

Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. In *An Integrated View of Language Development: Papers in Honor of Henning Wode*, edited by P. Burmeister, T. Piske, & A. Rohde, 217–244. Trier, Germany: Wissenschaftlicher Verlag.

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In *Laboratory Phonology 9*, edited by J. Cole & J. I. Hualde, 353–382. Berlin: Walter de Gruyter.

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*, 729–740.

Flege, J. E., Schirru, C., & MacKay, I. R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*, 467–491.

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 *Tone: A Linguistic Survey*, edited by V. A. Fromkin, 77–111. New York: Academic Press.

International Phonetic Association (1999). *Handbook of the International Phonetic Association*. Cambridge, UK: Cambridge University 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 *Proceedings of the 8th International Conference on Spoken Language Processing (INTERSPEECH-2004)*, 49–52. Jeju Island, Korea: International Speech Communication Association.

Laeufer, C. (1996). Towards a typology of bilingual phonological systems. In *Second-Language Speech: Structure and Process*, edited by A. James & J. Leather, 325–342. Berlin: Mouton de Gruyter.

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

Lisker, L. & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384–422.

Major, R. C. (1992). Losing English as a first language. *The Modern Language Journal*, *76*, 190–208.

Major, R. C. (1996). L2 acquisition, L1 loss, and the critical period hypothesis. In *Second-Language Speech: Structure and Process*, edited by A. James & J. Leather, 147–159. Berlin: Mouton de Gruyter.

Penfield, W. & Roberts, L. (1959). *Speech and Brain Mechanisms*. New York: Atheneum.

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

Selishchev, A. M. (1925). Des traits linguistiques communes aux langues balkaniques: Un balkanisme ancien en bulgare. *Revue des études slaves*, *5*, 38–57.