Aspir(at)ing to Speak Like a Native: Tracking Voice Onset Time in the Acquisition of English Stops

Mike Klein
George Mason University

Abstract

Much attention has been paid in recent decades to the development of new phonetic categories in the process of second language acquisition. Empirical studies, many of them focusing on voice onset time (VOT) in the acquisition of stops, have abounded, and a number of theories have been advanced to account for the data uncovered in these studies. In this paper I present the results of a corpus study focusing on VOT in the production of English stops by native Mandarin, Spanish, and French speakers who acquired English at different ages and possess differing lengths of residency in an English-speaking area. Data for the native Mandarin and French speakers (although not the native Spanish speakers) strongly suggest that learners with a significant period of residency tend to produce more authentic English voiced stops. Three theories of phonological learning – the Speech Learning Model (Flege, 1995), Full Transfer/Full Access (Escudero & Boersma, 2004; Schwartz & Sprouse, 1996), and Statistical Learning (McMurray, Aslin, & Toscano, 2009) – are evaluated in light of these new data; the results strongly favor Full Transfer/Full Access and present a new challenge to the Speech Learning Model.

1. Introduction

When speakers use a language they learned after childhood, native-speaking interlocutors generally have no trouble detecting their "foreign accents." This is true even of advanced learners with flawless syntax and a robust lexicon: phonology, it seems, gives them away every time (Scovel, 1995). While there are undoubtedly numerous acoustic cues in a given speech signal that allow us to detect a foreign accent, it has long been recognized that voice onset time (VOT) in the production of stops is particularly salient – and particularly amenable to measurement and analysis. VOT is a measure of the relative timing of two articulatory gestures associated with stop consonants: the release of the stop and the start of vocal fold vibrations (in other words, voicing). Positive values of VOT are found when voicing begins after a stop is released and thus correspond to aspirated plosives; likewise, negative VOT values indicate that voicing begins before a stop is released and are therefore associated with (partially or fully) voiced stops.

Experimental studies have shown that an L2 (second language) speaker's realization of VOT is closely correlated with native speakers' perceptions of foreign accent. Major (1987) measured VOT values for items read from a word list by adults learning English as a foreign language in Brazil and also had native English speakers rate their accents using recorded speech samples. He found a tight correlation: "the higher the accent score [i.e. the more native-like a speaker is rated], the closer the VOT conforms to the American English norm" (p. 199). Using almost the same methodology, Flege and Eefting (1987) achieved similar results using a group of 50 adult learners of English from the Netherlands, again finding a correlation between VOT and perception of foreign accent. These findings were confirmed yet again by Riney and Takagi (1999), this time by way of a long-term study of Japanese speakers of English as a foreign language. While a full understanding of all the factors that lead to the perception of a foreign accent remains an ongoing research program, it is safe to say that VOT is a major contributor; as Riney and Takagi conclude, "in L2 pronunciation there is a basic correlation between GFA [global foreign accent] and VOT" (Riney & Takagi, 1999, p. 298). Correlation, of course, does not prove causation; other factors could be just as important in the detection of a foreign accent. However, these studies suggest that VOT can stand as a convenient standard-bearer for foreign accent in general.

The central role of VOT in studies of the acquisition of L2 phonology is the result of at least three factors. First, as mentioned above, it is a relatively straightforward matter to measure VOT using a laptop computer armed with acoustic analysis software like Praat (Boersma & Weenink, 2012). Second, as we have seen, the study of VOT in L2 production can serve as a convenient, more concrete proxy for the study of foreign accent in general, a much trickier thing to evaluate or even define. Finally, the phonological systems of some very commonly studied languages differ significantly in the arrangement of their stops with regard to VOT, making it easy to find subjects whose L1 and L2 feature phonemes are not likely to match up in VOT. In the United States, for instance, it is typically not difficult to find native Mandarin and Spanish speakers who are learning English, and all three of these languages feature systems of stop phonemes with very different VOT values (see Figure 1). This is fortunate for researchers who want to study the role of VOT in phonological acquisition: without such variety, it would be difficult to pose and answer interesting questions about what language learners do when faced with discrepancies between L1 and L2 phonemes.

Figure 1 shows the arrangement of stop contrasts with regard to VOT in four languages. Since the actual realization of stop phonemes depends on many factors (including, but probably not limited to, place of articulation, register, phonetic environment, and individual variation), Figure 1 is intended to give an impression of the kinds of contrasts that are possible, rather than a guide to actual VOT values in these languages. Each language, for a given place of articulation, contrasts two or three phonemes, but the contrasts are different in each case. Despite their traditional label, English voiced stops /b/, /d/, and /g/ are typically only partially voiced, with small negative VOT values (light blue box); unvoiced stops, at least in word-initial position, are aspirated, with moderately large positive VOT (dark blue box). In Mandarin, all stops (indeed all obstruents) are unvoiced, and the contrast is instead between unaspirated /p/, /t/, and /k/, with small positive VOT (orange box), and aspirated /p^h/, /t^h/, and /k^h/, with large positive VOT (brown box), slightly higher than English voiceless stops. The Spanish system contrasts fully voiced /b/, /d/, and /g/, with large negative VOT (light green box), with unvoiced and unaspirated /p/,

/t/, and /k/ (dark green box), similar to those in Mandarin. (The French system, also relevant for this study, is essentially the same as Spanish.) Finally, Thai boasts a three-way contrast between fully voiced /b/, /d/, and /g/ (pink box), unvoiced and unaspirated /p/, /t/, and /k/ (lavender box), and heavily aspirated /p^h/, /t^h/, and /k^h/ (purple box).

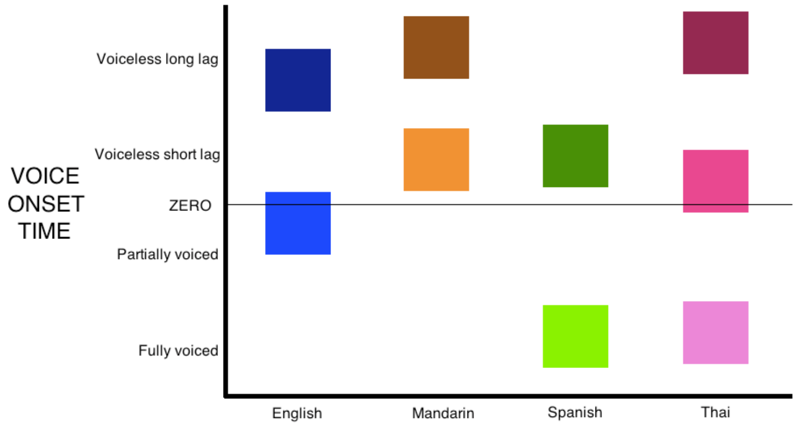


FIGURE 1. Stop contrasts in English (Lisker & Abramson, 1964), Mandarin (Chao & Chen, 2008), Spanish (Williams, 1977), and Thai (Gandour, Petty, Dardarananda, Dechongkit, & Mukngoen, 1986).

The general research question addressed in this paper is this: what happens when adults learn a language with a phonological system featuring stop phonemes with VOT values not found in their native language? More specifically, can adult learners improve their L2 pronunciation with experience, producing L2 stops with VOT values more closely approximating native norms? This paper, of course, is not the first to address VOT in second language acquisition; on the contrary, there has been a thriving literature on the subject for several decades. To the best of my knowledge, however, no study has focused on the possibility that production of L2 stops might change systematically over time as adults learn a second language. As we will see, the answer to this question has important ramifications for second language phonological theory, since different theories of phonological acquisition make different predictions with regard to the possibility that VOT values change over time for language learners.

In the next section of this paper, I will review several experimental studies that have been published on VOT and on second language phonology in general. I will present these studies in the context of three theories of phonological acquisition: Flege's Speech Learning Model (Flege, 1995), probably the gold standard in the field; Full Transfer/Full Access, introduced by Schwartz and Sprouse (1996) in the domain of syntactic acquisition but since extended to phonology as well (Escudero & Boersma, 2004); and the Statistical Learning Model as formalized by McMurray et al. (2009), who provide a model of L1 phonological acquisition that is readily extendible to L2 acquisition. In Section 3, I present the results of an extensive corpus study of English speech samples recorded by native speakers of Mandarin, Spanish, and French. The corpus used here is the George Mason University Speech Accent Archive (Weinberger, 2012). With more than 1,700 speech samples accompanied by demographic information on each contributor (including age of English onset and length of residency in an English-speaking country), the Archive is ideally suited to address, by way of a semi-longitudinal study, the possibility that learners can improve their pronunciation by producing VOT values closer to native English norms as they gain experience. In Section 4, I present a general discussion of the experimental results, evaluating each of the previously discussed models of L2 acquisition in light of these new data. Finally, Section 5 provides a brief conclusion to the paper.

2. Three Models of Phonological Acquisition

In this section, I will introduce three prominent theoretical models of phonological acquisition. In the course of discussing these models, I will review a number of experimental studies that have been published regarding second language phonological acquisition, with special emphasis on studies that have shaped our thinking on VOT in the acquisition of L2 stop contrasts. In a few cases I will highlight the advantages of my experimental design compared to previous studies that did not attempt to systematically deal with the possibility that learners' realization of L2 stops can change over time. I will end this section with a discussion of the differing behavioral predictions of each model with regard to the corpus study presented later in this paper.

2.1. Speech Learning Model

The Speech Learning Model (SLM), discussed and developed throughout the decades-long research program of James Emil Flege and fleshed out most fully in Flege (1995), attempts to explain why L2 learners are not as successful as L1 learners at acquiring the phonetic details of a particular language and why certain sounds are harder to acquire than others. The SLM ultimately attributes these differences to the auditory processing limitations of late L2 learners, rendering them less able than early learners to perceive and internalize certain phonetic contrasts. Young learners are able to perceive all of the phonetic contrasts that are relevant to the phonological systems of the world's languages. Accordingly, these learners have no problem acquiring the phonetic categories of any language they learn, no matter how similar or different these categories are to those of a previously learned language. On the other hand, later learners are increasingly unable to perceive (without explicit training; see Tees & Werker (1984)) some of the finer contrasts between L1 and L2 sounds that they encounter, by hypothesis making it difficult for them to produce native-like phonetic categories in their second language. Where L2 sounds differ significantly from sounds found in their L1, no learners are predicted by the SLM to have difficulty in forming new categories. It is precisely those L2 sounds that are similar but not identical to L1 sounds that present problems for adult learners, who are likely to assimilate them into their existing L1 categories rather than form new L2 categories. According to the SLM, fossilization is a good possibility here: due to these auditory processing limitations, learners will never be able to break free from the phonetic details of their first language phonetic category, resulting in a persistent "foreign accent."

Flege (1980) was the first study within the budding SLM model to tackle the question of VOT in the acquisition of L2 stops. In this study, Flege examined various aspects of the production of voiceless stops (/p/, /t/, and /k/) among Saudi Arabic speakers learning English in the United States. He divided them into two groups of six participants each; every member of the first group (Ar1) had spent less than one year in the U.S., while members of the second group (Ar2) had all spent more than a year in the U.S., with an average of 39 months of residency. He asked them, as well as a native English-speaking control group (Am), to produce CVC words in a carrier sentence and measured VOT for the initial stops in these words. His results are summarized in Table 1.

	Am	Ar2	Ar1
Voice-onset time			
/p/ in *pat*	46	21	14
	(4)	(11)	(10)
/t/ in *tab*	62	29	32
	(11)	(14)	(10)
/k/ in *cab*	67	47	41
	(12)	(11)	(7)

TABLE 1. Summary of VOT results from Flege (1980).

As can be seen from Table 1, which shows the average VOT and the standard deviation (in parentheses) for each group and each phoneme, the two native Arabic groups differed markedly from the American control group, but did not differ significantly from each other. Native Arabic norms for VOT are smaller than English values, around 37 ms for /t/ and 52 ms for /k/ (Flege, 1979); the Arabic speakers in this study were therefore producing sounds much more like L1 Arabic than L2 English. Note that the very low VOT values for /p/ in Table 1 are likely a result of the fact that learners are actually associating English /p/ with Arabic /b/, since most varieties of Arabic have no voiceless bilabial stop /p/.

Although Flege's study does attempt to deal with the possibility of change over time by contrasting the two Saudi groups, his null result – that there is no significant difference between those who had been in the U.S. less than one year and those who had a longer length of residency – is not particularly persuasive. First, there were only six subjects per group, arguably not a large enough participant base to draw any reliable conclusions. Second, even those in group Ar2, with a longer length of residency, averaged only 39 months in the U.S.; consequently, if learners require more than a few years of exposure to shift their VOT towards native English norms, Flege would have been unable to detect those changes. Finally, acknowledging that the results were not statistically significant, the differences in the mean VOT for /p/ and /k/ were moving in the right direction with more experienced speakers producing longer (more English-like) VOTs; although this was not true for /t/, the difference between the two groups was very small in this case. In the corpus study to be presented in this paper, I hope to improve on Flege (1980) by tracking

VOT over a longer period of time (in some cases several decades) among a larger participant pool (about 50 subjects for each L1 group).

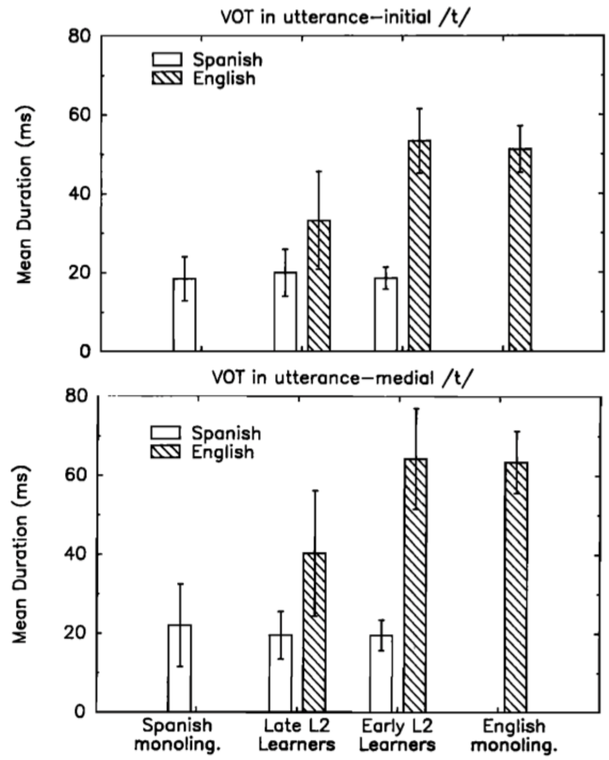


FIGURE 2. Summary of VOT results from Flege (1991).

Flege (1991) revisited the question of VOT in L2 acquisition, this time comparing a group of L1 Spanish early learners of English (age of onset less than seven years) with a group of late learners, also native Spanish speakers. Monolingual Spanish- and English-speaking control groups were also included in his experimental model. He focused only on the phoneme /t/, having participants read sentences like *take a textbook* (or *tengo un tigre*) and measuring VOT for /t/ in utterance-initial and utterance-medial position. His results are summarized in Figure 2.

As can be seen, monolingual Spanish speakers produced tokens of /t/ with an average VOT of around 20 ms, while monolingual English speakers averaged 50 to 60 ms. Early L2 English learners, in accord with the predictions of the Speech Learning Model, were able to fully acquire English /t/, with average VOT values indistinguishable from monolingual English speakers. Interestingly, however, late L2 learners produced VOT values that were, on average, intermediate between Spanish and English norms, hovering around 30 to 40 ms.

To explain this result within the SLM, Flege appeals to “realization rules” (Flege, 1991, p. 407), which allow an individual speaker to pronounce a given phoneme differently in different situations. These rules are independently needed to account for the apparently universal fact that speakers modify their pronunciation as a function of things like rate of speech and social context (or in experimental contexts; see Goldrick (2004); Nielsen (2011); Kirov & Wilson (2012)). Flege believes that these same kinds of rules allow L1 Spanish learners of English to supply longer (but not quite native-like) VOT values for /t/ when speaking English than they do when speaking Spanish, despite the fact that (by hypothesis within the SLM) there is only one phonetic category to cover both languages. While this explanation is not implausible, it is not the only possible explanation for the intermediate VOT values produced by late English learners in this study. If, in fact, VOT values steadily improved with experience so that inexperienced learners produced very Spanish-like stops while more seasoned learners produced more English-like tokens, we would expect precisely the kind of result shown in Figure 2. The average VOT for late learners would naturally fall between that of native English and native Spanish values, since it would reflect the behavior of speakers all along the continuum from L1-like to L2-like production of /t/. It is precisely this possibility that will be explored later in this paper.

Most recently, still working within the Speech Learning Model, Gonzalez Lopez (2012) examined a group of sixteen native English-speaking college students who were intermediate learners of L2 Spanish. These subjects were asked to read sentences in English, Spanish, and a mixture of the two languages. This last sentence type, illustrated in Figure 3, was included because part of Gonzalez Lopez’ research question (not directly related to the subject of the present paper) involved the interaction between L1 and L2 sounds when subjects were forced to switch quickly between them.

a. Spanish → English CS
Pre-switch /t/ Switch /t/ Post-switch /k/
/**t**/odos mis amigos | /**t**/alked Spanish as /**k**/ids.
'All my friends talked Spanish as kids.'

b. English → Spanish CS
Pre-switch /t/ Switch /t/ Post-switch /p/
The /**t**/yphoon damaged | /**t**/echos y /**p**/aredes.
'The typhoon damaged ceilings and walls.'

FIGURE 3. Bilingual sentences in Gonzalez Lopez (2012).

Gonzalez Lopez measured VOT for tokens of Spanish and English voiceless stops (/p/, /t/, and /k/) contained in these sentences. Her results for the monolingual Spanish and English sentences are shown in Figure 4; Figure 5 shows the results for the mixed sentences.

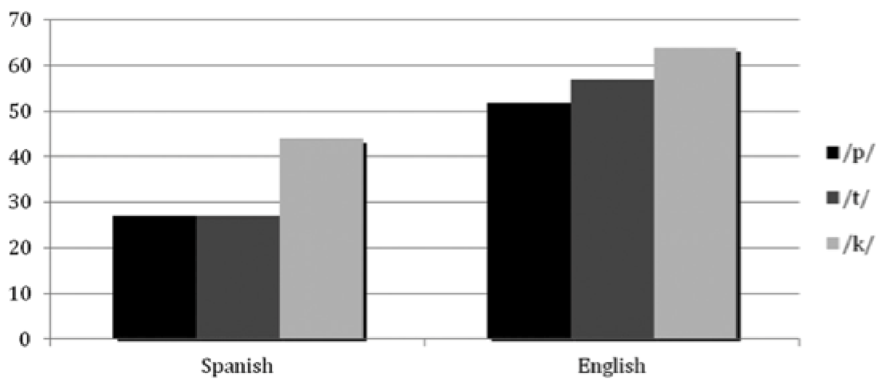


FIGURE 4. Results for monolingual sentences in Gonzalez Lopez (2012).

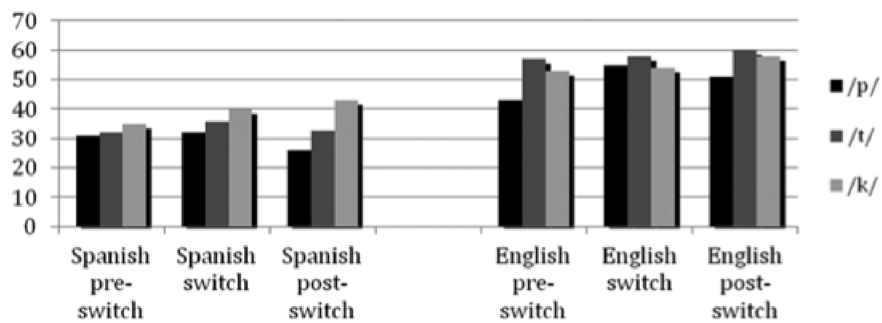


FIGURE 5. Results for mixed sentences in Gonzalez Lopez (2012).

As can be seen from both figures, but especially Figure 4, subjects produced notably lower values for voiceless stops in L2 Spanish than they did in L1 English. Gonzalez Lopez takes this as evidence that subjects are, in fact, able to form new phonological categories for Spanish voiceless stops, although – as can clearly be seen by comparing the configuration of bars in Figure 5 to those in Figure 4 – there was some interaction between L1 and L2 when subjects had to switch mid-sentence. While speakers are clearly treating Spanish and English sounds differently, they are not consistently producing native Spanish-like stops: average VOT for Spanish sounds in Figure 4 are still noticeably above the 20 ms found for native Spanish speakers in Flege (1991) (see Figure 2). Thus, neither Flege's explanation (one phonological category with different realization rules) nor the alternative hypothesized in this paper (a continuum of speakers with more or less native-like VOT values) can be entirely discounted.

2.2. Full Transfer/Full Access

The idea that second language learners initially transfer aspects of their L1 into their L2, but nonetheless are able to slowly approach L2 norms by tapping into the same learning mechanisms available in first language acquisition, was first proposed in the domain of syntax by Schwartz and Sprouse (1996). This theory, known as Full Transfer/Full Access, extends fairly naturally to phonological acquisition, where we know impressionistically that beginning L2 learners typically have very noticeable accents that reflect the phonological structure of their native language, but that pronunciation usually improves over time. Several studies have provided experimental confirmation of this qualitative impression about L2 phonological learning.

Escudero and Boersma (2004) examined the contrast between lax /ɪ/ and tense /i/ in words like *ship* and *sheep* among L1 Spanish learners of two varieties of English, those spoken in Scotland and Southern Britain. While members of both learner groups were ultimately able to acquire the contrast, those whose target language was Scottish English had a much easier time than those who were trying to learn Southern British English, acquiring the contrast more quickly. Escudero and Boersma showed that phonological differences between the two dialects could account for this result; starting with L1 Spanish, learners of Scottish English simply had less distance to cover in learning the contrast, with the direct result that they required less exposure to successfully master it. According to the authors, all learners should be able to acquire the contrast, but some phonological features require more time to acquire due to greater differences between L1 and L2. (Although

Escudero and Boersma do not discuss this possibility, it seems possible – and entirely consistent with Full Transfer/Full Access – that some phonological details might require so much time to acquire that typical adult learners never reach target-like norms.)

Halicki (2010) used a well-formedness judgment task to probe the intuitions of native English speakers with varying levels of experience in L2 French. She found that beginners did not possess native-like phonotactic knowledge, but that more experienced learners increasingly resembled native French speakers in their judgments of novel French-like and non-French-like vocabulary items. Halicki took this to indicate that L2 learners are able to access the same mechanisms as native learners in constructing the phonology of a language; if this is true, then lingering pronunciation problems among even advanced L2 learners may be more about insufficient executive control (Bialystok, Craik, & Ryan, 2006; Rodriguez-Fornells, Balaguer, & Münte, 2006) than incomplete acquisition of the phonological system itself.

As far as I am aware, no previous study has attempted to assess the predictions of the Full Transfer/Full Access model in the realm of VOT in the production of L2 stops. These predictions, however, should be amenable to evaluation using the methodology presented in this paper. We will return to this issue very shortly, in section 2.4.

2.3. Statistical Learning Model

Turning for the moment to the realm of first language acquisition, Maye, Werker, and Gerken (2002) demonstrated that the sound perception of infants was significantly affected by the statistical distribution of the sounds to which they were exposed (See also Maye & Gerken (2000) for an extension to adult participants).

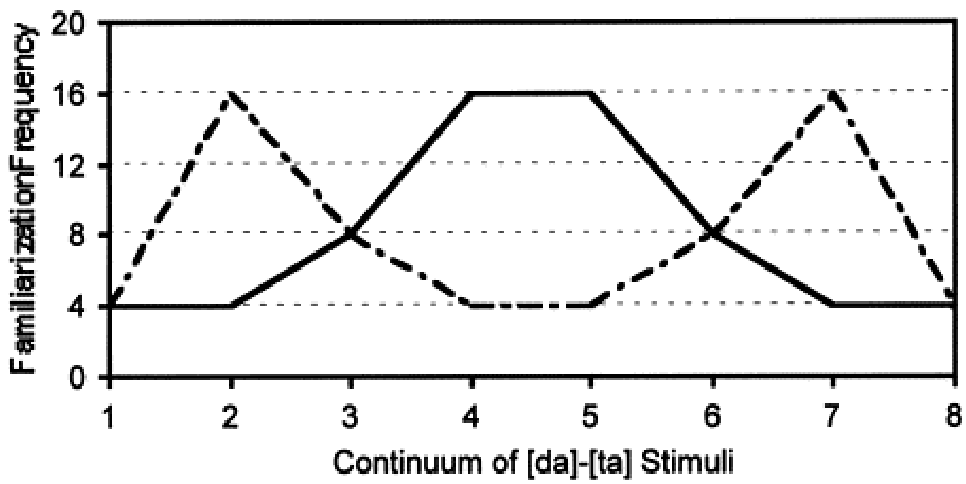


FIGURE 6. Unimodal and bimodal frequency distributions used in May et al. (2002).

Their experiment involved two phases: a familiarization phase, where 6- to 8-month-olds were exposed to either a bimodal or unimodal frequency distribution of tokens on the [da]-[ta] continuum (see Figure 6), and a testing mode, where looking times were used to find out if these young subjects were able to discriminate pairs of words at opposite ends of the continuum. The researchers found that looking time was significantly correlated with familiarization condition, indicating that infants who had been exposed to the bimodal distribution had a much easier time discerning between [d]-like sounds and [t]-like sounds. From a very young age, language learners are apparently highly sensitive to distributional patterns in their language input.

McMurray, Aslin, and Toscano (2009) built on the work of Maye et al. by developing a computational model that was successfully able to acquire the contrast between /t/ and /d/ in English. Their computational learner was supplied with data corresponding to the frequency distribution of VOT for these two sounds in English (Figure 7) and was able to discern two distinct categories, just as L1 learners of English can. This result lends credence to the notion that human language learners are also statistical learners of a sort.

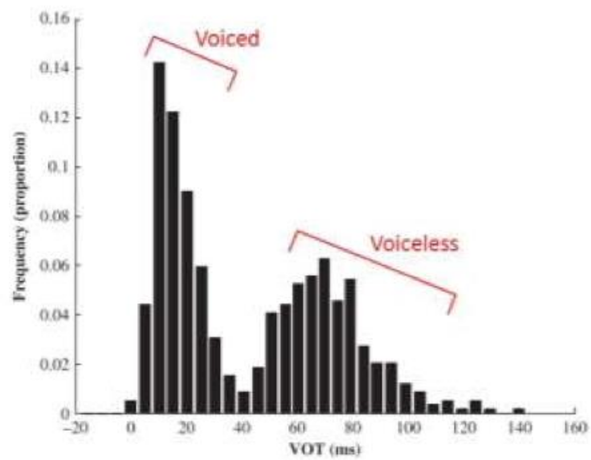


FIGURE 7. Frequency distribution of VOT for English /d/ and /t/, as employed in McMurray et al. (2009).

The Statistical Learning Model has not been explicitly applied to the formation of phonological categories in second language acquisition, but there is no reason it could not be extended to cover L2 as well as L1 acquisition. In the next section, we will do just that and explore the predictions of the Statistical Learning Model in terms of the corpus study to be presented in section 3 of this paper.

2.4. Experimental Predictions

Each of the three models considered in this section makes a different prediction with regard to the results of the corpus study to be presented in the next section of this paper. Recall that this study will focus on the question of whether adult L2 learners can improve their realization of English stops by producing more native-like VOTs with increased L2 experience, specifically with longer length of residency (LOR). Since each of the three models assumes that learners are endowed with different mechanisms for acquiring new phonetic categories, it should be possible to evaluate these models based on the behavior of a large group of adult learners with differing levels of English experience.

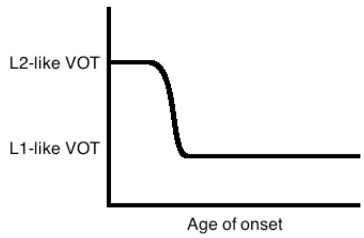
According to Flege's Speech Learning Model, late L2 learners are incapable of developing new phonetic categories for sounds similar to those found in their native language. Since this is exactly the task presented to the language learners to be considered in the corpus study, the Speech Learning Model predicts that age of English onset should be the major factor determining whether speakers are able to produce native-like values of

VOT when speaking L2 English. Early learners should do a very good job producing native English-like stops; later learners are expected to fossilize at L1-like values. This prediction is depicted in Figure 8(a).

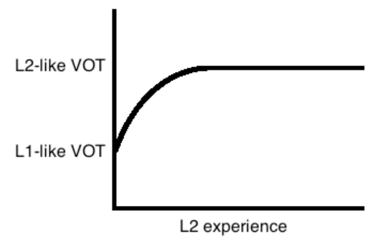
The Full Transfer/Full Access model offers a rosier picture for language learners. Although beginners are expected to supply VOT values more in line with their L1 than with L2 English norms (full transfer), they should be able to approach authentic native English behavior with increased experience (full access). This is shown in in Figure 8(b).

Finally, if we extend the Statistical Learning Model to second language acquisition and assume that adult learners as well as young learners have the capacity to form categories based on the statistical distribution of sounds, then it is inevitable that speakers should eventually be able to discern distinct categories for L1 and L2 sounds. This may, however, take quite a bit of time, especially if the peaks in the distribution are fairly close together, as they will be in the cases to be studied here. Before learners are able to establish a new category for L2 sounds, we might expect them to attempt to assimilate them into their L1 category, essentially expanding the old category to accommodate the new sounds until a new category is formed.

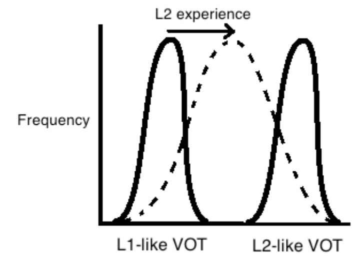
The predictions of the Statistical Learning Model, depicted in Figure 8(c), may prove to be very difficult to tease apart from the steady improvement predicted by Full Transfer/Full Access, but there are at least two differences. First, while the Statistical Learning Model predicts that average VOT for L2 stops will, at first, progress steadily from L1-like towards target-like values as the L1 category expands to accommodate new L2 data, this will not be the uninterrupted march predicted by Full Transfer/Full Access. It cannot be, since the category still has to accommodate L1 data as well. Assuming an (obviously idealized) even split between L1 and L2 data, we would never expect average VOT to progress beyond the halfway point between native L1 and L2 values during this intermediate phase, before a new L2 category can be formed. Second, the Statistical Learning Model predicts that, during this initial phase with a single category for L1 and L2 sounds, the production of L1 sounds will be affected as well. This second prediction will not be addressed by the current study, but some previous research has indicated that L1 pronunciation may indeed be affected by the learning of a second language (Pavlenko, 2000).



(a) SPEECH LEARNING MODEL: Only early learners can achieve native L2-like VOT values, so age of onset is the main determiner of VOT in L2 stops.



(b) FULL TRANSFER/FULL ACCESS: VOT values improve steadily from L1-like to authentic L2-like values, given sufficient linguistic experience.



(c) STATISTICAL LEARNING MODEL: L1 category initially expands to accommodate L2 data (dotted line); eventually separate categories are acquired (solid lines).

FIGURE 8. Predictions of the phonological learning models discussed in Section 2.

3. Corpus Study

In order to test my research question – Can adult learners pronounce L2 stops with more authentic VOT values if they spend more time in an English-speaking region? – I collected acoustic and demographic data from the George Mason University Speech Accent Archive (http://accent.gmu.edu; Weinberger, 2012). The Archive contains speech samples from more than 1,700 individuals of varying geographical and linguistic backgrounds, all of whom have submitted an audio recording of the same English paragraph. Demographic information for each subject includes birthplace, native language, other language(s) spoken, age, sex, age of English onset, learning method (academic or naturalistic), country of English-language residence, and length of residency in that country. Many samples feature phonetic transcriptions and an explanation of the prominent features of each subject's pronunciation, although these will not be used in the present study.

My focus will be on L1 Mandarin, Spanish, and French learners of English. For the native Mandarin speakers, I will be interested in the pronunciation of English voiced stop phonemes /b/, /d/, and /g/, which Mandarin speakers are likely to identify with their native voiceless and unaspirated phonemes /p/, /t/, and /k/. The English phonemes feature average VOT values lower than those for the Mandarin phonemes, so my hypothesis (adopting for the sake of simplicity the assumptions of the Full Transfer/Full Access model) is that subjects with more L2 experience (i.e., longer length of residency) will produce lower, and therefore more native-English like, VOT values. For the native

Spanish and French speakers, I will focus on the realization of English voiceless stops /p/, /t/, and /k/ in word-initial position, where they are aspirated and therefore have longer VOT values than the corresponding unaspirated voiceless Spanish and French phonemes. My hypothesis will then be that VOT will increase with L2 experience for the native Spanish and French speakers pronouncing these phonemes. The elicitation paragraph used for each sample on the Speech Accent Archive is given below; the words that will be used in the analysis of L1 Mandarin speakers are shown in red, while those that will be used in the analysis of the Spanish and French speakers are shown in green:

> Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.

The general procedure for each L1 group was to collect relevant acoustic information from the Speech Accent Archive samples using Praat (Boersma & Weenink, 2012), and then construct a series of linear mixed effects models to determine whether the available measure of L2 experience, length of residency, makes a unique contribution to average VOT values. I used the statistical package R (R Development Core Team, 2012) to perform statistical analysis and generate relevant graphs and charts.

In the remainder of this section, I will first describe and analyze the data from Mandarin learners of English, then the data from Spanish learners, and finally the data from French learners.

3.1. L1 Mandarin, L2 English

3.1.1 Participants and Materials

A total of 48 native speakers of Mandarin have submitted samples on the Speech Accent Archive; all of these samples, and the attached demographic information, will be used in the analysis presented in this section. In addition, samples from a randomly selected group of twenty native English speakers will serve as a control group to verify that L1 Mandarin speakers indeed behave differently in the pronunciation of voiced English stops and to allow us to test the hypothesis that these Mandarin speakers behave more like native English speakers with increased L2 experience.

3.1.2 Procedure

Each of the relevant sound files (48 L1 Mandarin, 20 L1 English) from the Speech Accent Archive was captured, as well as each subject's demographic information, which was collected in a spreadsheet. Praat (Boersma & Weenink, 2012) was then used to measure the VOT for the initial stop in each of the four relevant words (*Bob*, *big*, *bags*, and *go*) in each sample; these values were entered into the same spreadsheet. Finally, the data was imported into R in preparation for the statistical analysis to be described in the next section.

When measuring VOT, there were three distinct possibilities for the realization of each stop consonant: negative VOT (i.e., voicing), positive VOT (i.e., aspiration), and zero VOT. Figure 9, a screen shot from Praat, shows an example of negative VOT.

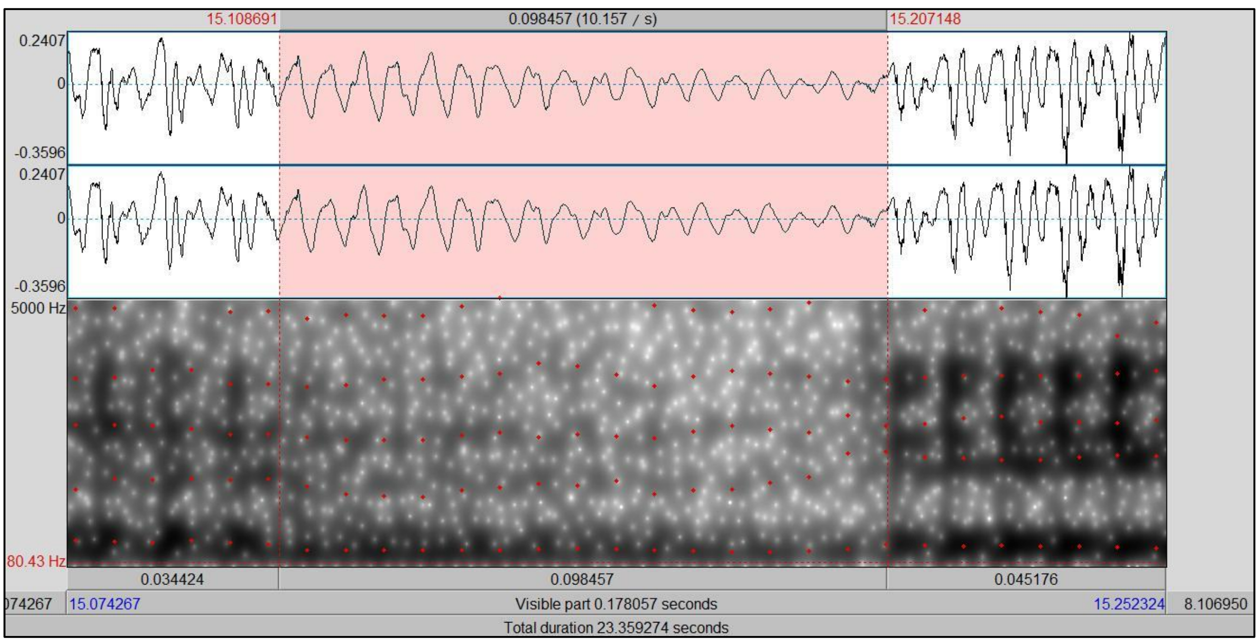


FIGURE 9. Waveform and spectrogram for *big* from subject mandarin33, illustrating negative VOT.

As can be seen by inspecting the waveform (top half of Praat display) in the pink shaded area, periodic vibrations continue throughout the closure between vowel sounds. In addition, the "voicing bar" – the region of low-frequency sound in the spectrogram (bottom half of Praat display) – is also visible throughout closure and is another sign of vocal fold vibration. In this case, the token of /b/ (in *big*) is fully voiced, and VOT is measured from the end of the previous vowel to the beginning of the next vowel (vowels are most easily identified from the more complicated waveform to the left and right of the pink shaded area). This VOT is assigned a negative value, since the start of voicing precedes the release of the stop.

Figure 10 shows an example of positive VOT.

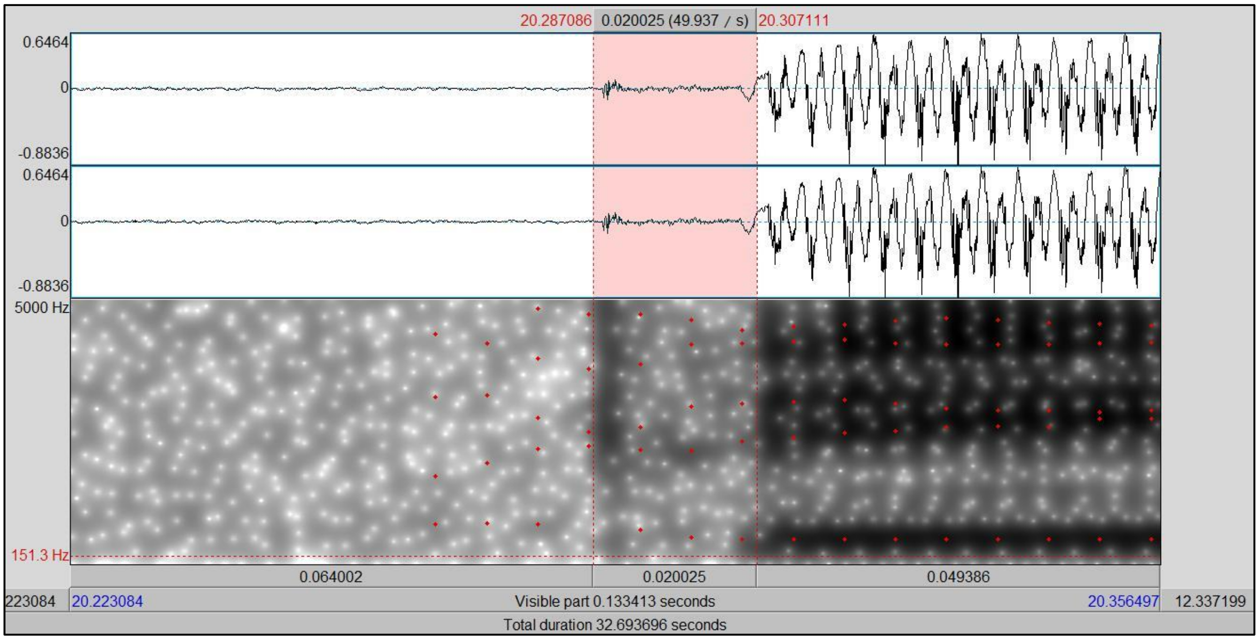


FIGURE 10. Waveform and spectrogram for *big* from subject mandarin18, illustrating positive VOT.

The release burst for /b/ is visible most clearly in the spectrogram as a brief period of "white noise" extending throughout the frequency range (but especially prominent in the higher frequencies); it is also accompanied by a slight perturbation in the waveform. The important matter here is that there is then a discernable lag between the release burst and the beginning of voicing for the vowel, clearly visible here as the mostly empty space after the release burst in the pink shaded region of the waveform and spectrogram. This lag was measured and assigned a positive VOT value, since voicing begins after the stop is released.

Finally, Figure 11 is an example of zero VOT.

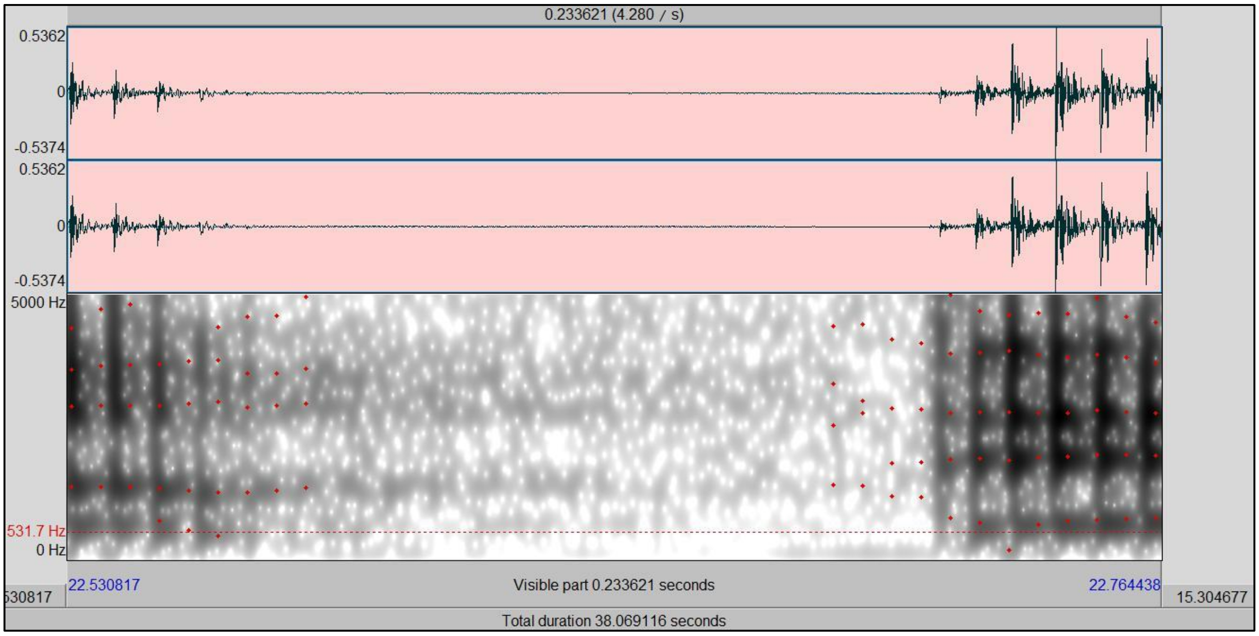


FIGURE 11. Waveform and spectrogram for *big* from subject mandarin30, illustrating zero VOT.

In this case, there is no indication of voicing leading up to the release of the stop: the waveform is completely flat between the two vowels, and no voicing bar is visible in the spectrogram. In addition, there is no evidence of a release burst; the acoustic calm of the unvoiced closure is disturbed only with the beginning of the vowel itself. For this reason, this sample (along with just a few other samples in the data set) was assigned zero VOT.

VOT measurement proceeded in this manner for all of the L1 Mandarin and English samples considered in this section, as well as for the L1 Spanish and French samples to be considered in Sections 3.2 and 3.3.

3.1.3 Results and Discussion

Figure 12 displays the mean VOT values produced by native English and native Mandarin speakers for the initial voiced stop in each word. Table 2 summarizes the descriptive and comparative statistics. (Appendix A contains all of the raw data used in this study.)

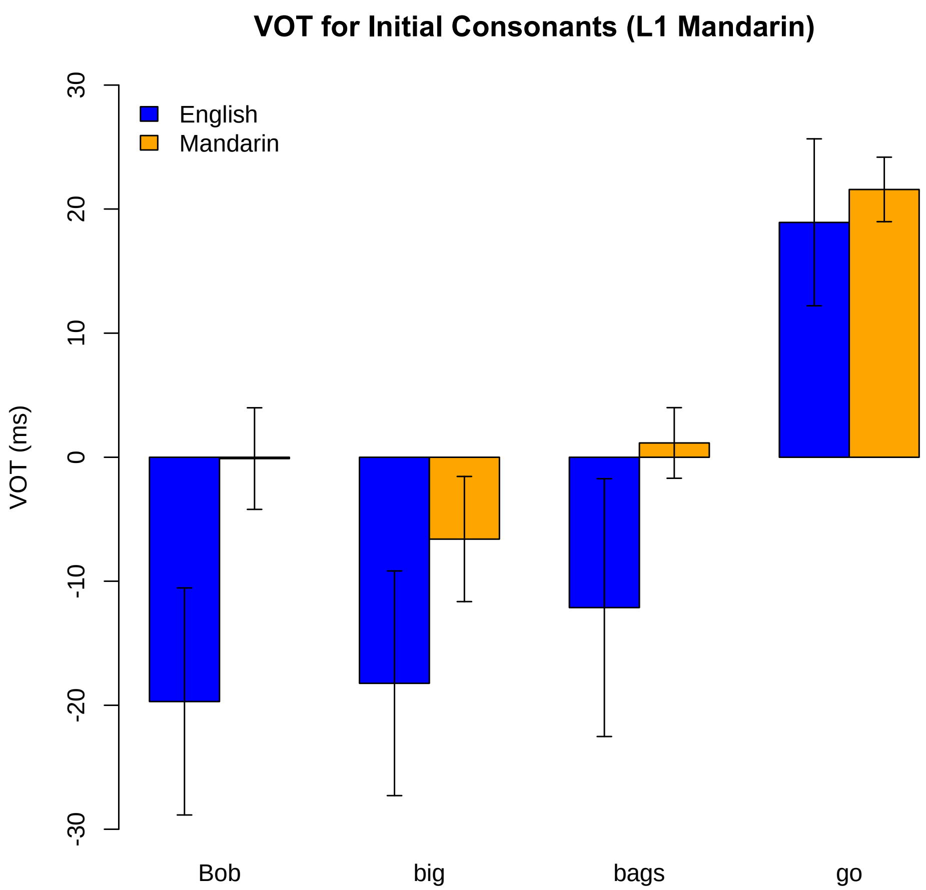


FIGURE 12. Average VOT for initial voiced consonants produced by L1 Mandarin and L1 English speakers.

	Mandarin		English		Welch's t-test		
Word	Mean VOT (ms)	Std. Dev.	Mean VOT (ms)	Std. Dev.	t	d.f.	p
Bob	-0.12	28.38	-19.70	40.93	-1.95	26.94	0.06 (n.s.)
big	-6.61	34.94	-18.23	40.50	-1.12	31.39	0.27 (n.s.)
bags	-1.14	19.75	-12.13	46.48	-1.23	21.92	0.23 (n.s.)
go	21.58	18.01	18.93	28.55	-0.37	22.27	0.72 (n.s.)
ALL WORDS	**4.00**	**28.05**	**-8.47**	**42.04**	**-2.41**	**105.97**	**< 0.02**

TABLE 2. Descriptive and comparative statistics for initial voiced stop data.

In each case, as expected under the assumption that L1 Mandarin speakers associate English voiced stops with their native unvoiced and unaspirated stops, these speakers produce longer average VOT values than native English speakers. Also as expected, there is a great deal of variation among the four words used here; this is most noticeable with *go*, which differs in place of articulation from the other three tokens, but there are differences among the words with initial /b/ as well, most likely due to phonological environment. As can be seen in the right-hand section of Table 2, although there is not sufficient data to find a statistically significant difference between L1 English and L1 Mandarin VOT values for each individual word, we do find a rather significant difference ($p < 0.02$) between the aggregate data sets. Based on this result and the fact that the difference was in the expected direction for each individual token, the analysis in the rest of this section will collapse data for all four words.

Figure 13 shows the relationship between age of onset and average VOT values for each L1 Mandarin speaker. From the figure it is clear that those who began to acquire English at an early age (earlier than about eight years) tend to produce more native English-like voiced stops (recall from Table 2 that the overall mean for L1 English speakers was -8.47 ms). Those who began to learn English later than age eight tend to produce more Mandarin-like stops, with higher average VOT values. To test the predictive power of age of onset, I performed a likelihood ratio test to compare two linear mixed effects models using the *lme4* package in R (Bates, Maechler, & Bolker, 2012). The first model was a "null" model, with just speaker and word (i.e., *Bob*, *big*, *bags*, or *go*) as random effects and no fixed effects at all; the second model added age of onset as a fixed effect. The difference between these two models was significant ($\chi^2 = 6.26$, $p < 0.02$), confirming that age of onset is a significant predictor of performance. (Detailed results for these linear mixed effects models, and those to be discussed below, are contained in Appendix B.)

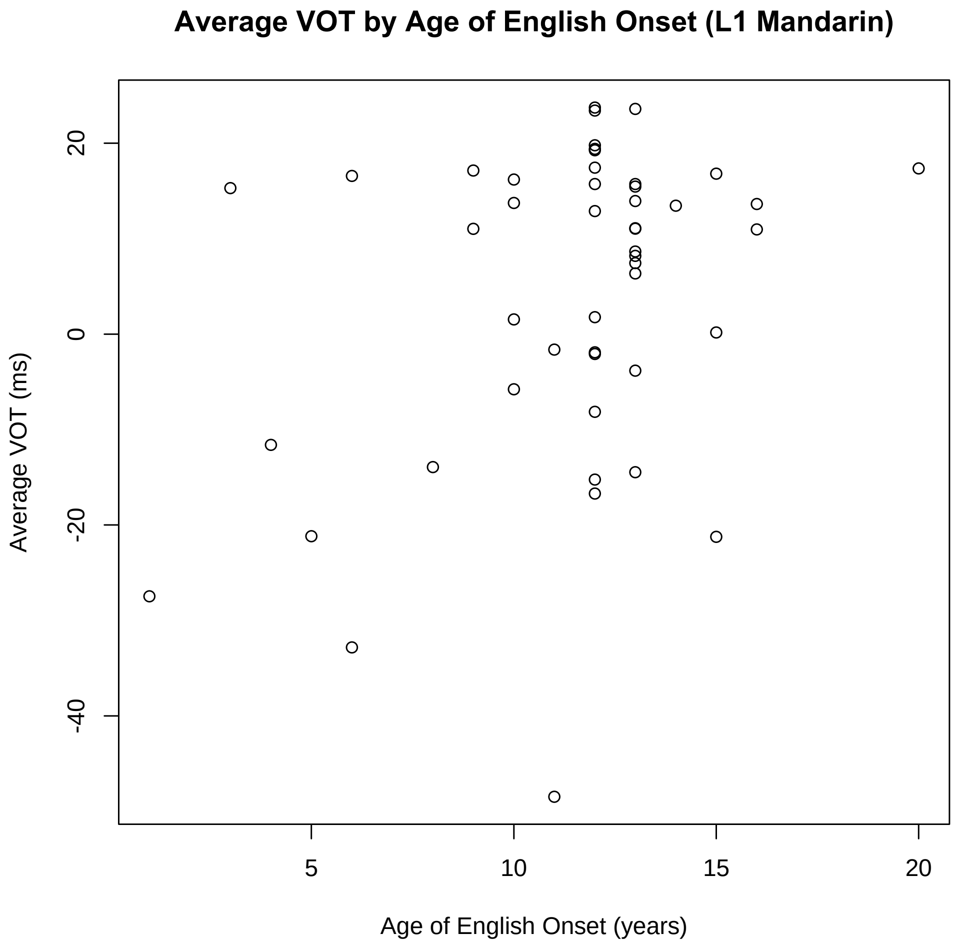


FIGURE 13. Age of onset and production of VOT by L1 Mandarin speakers.

Figure 14 shows the relationship between length of residency (LOR) and production of VOT for voiced English stops. Subjects with very long length of residency – more than about twenty years – tended to produce more English-like voiced stops with negative VOT, while those with shorter LOR produced more Mandarin-like stops. To test the predictive power of length of residency, I added LOR to the linear mixed effects model (along with age of onset) and compared it to the previous model, with only age of onset as a fixed effect. The likelihood ratio test confirms that LOR is a highly significant predictor above and beyond age of onset ($\chi^2 = 12.45$, $p < 0.001$).

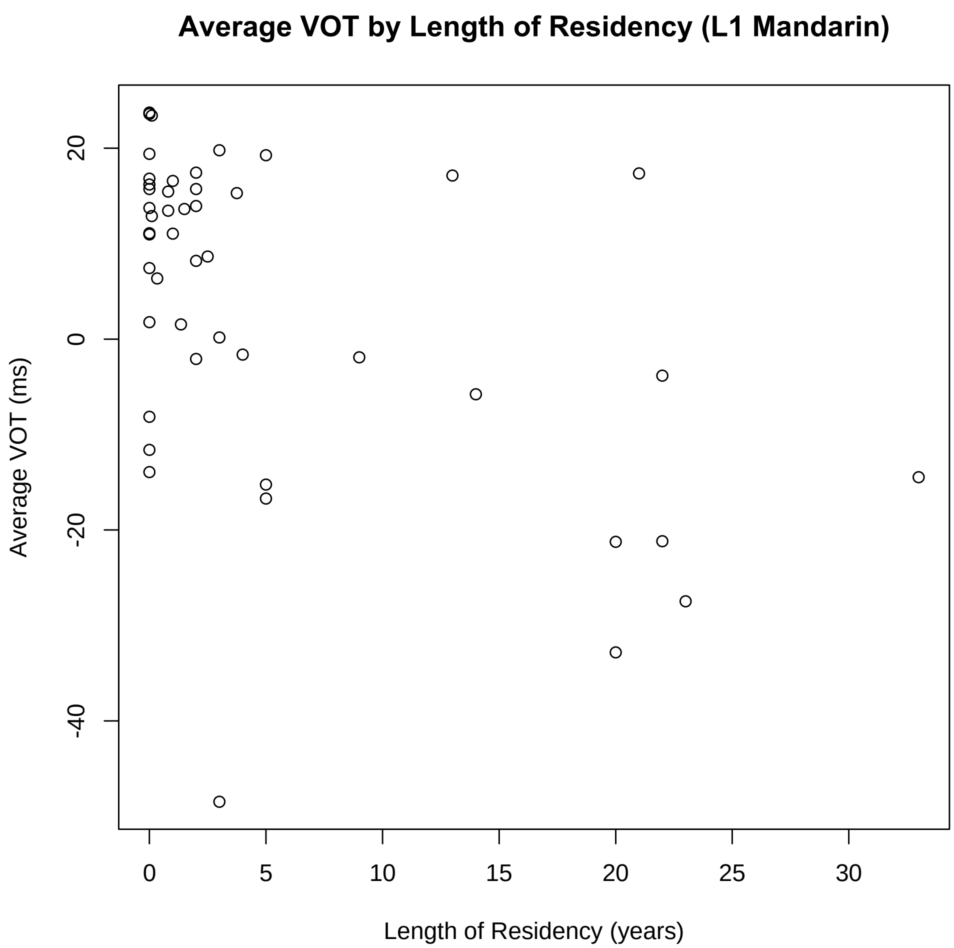


FIGURE 14. Length of residency and production of VOT by L1 Mandarin speakers.

3.2. L1 Spanish, L2 English

3.2.1 Participants and Materials

114 native Spanish speakers have submitted samples to the Speech Accent Archive, but the data set used here was limited to the most recently submitted 50 samples to make data collection more manageable and to keep the Mandarin, Spanish, and French data sets relatively even. In addition, samples from the same twenty L1 English speakers used in Section 3.1 were analyzed as a benchmark for comparison with the L1 Spanish samples.

3.2.2 Procedure

The procedure used here is identical to that described in Section 3.1.2, except that the phonemes of interest were the initial stops in the words *call*, *peas*, *toy*, and *kids*. VOT for the 50 L1 Spanish samples and the twenty L1 English control samples were measured using Praat, collected in a spreadsheet along with demographic information for each subject, and then imported into R for statistical analysis. Measurement of VOT using Praat proceeded in the same manner described in Section 3.1.2.

3.2.3 Results and Discussion

Presentation of data and analysis for the L1 Spanish speakers will proceed in the same manner as for the Mandarin speakers. Figure 15 and Table 3 give the basic results. VOT values for the initial voiceless English stop in each word differ in exactly the way we would expect if Spanish speakers associate aspirated (high-VOT) English stops with their native unaspirated (low-VOT) voiceless stops: the values, in each case, are larger for L1 English speakers than L1 Spanish speakers. Unlike the L1 Mandarin data, this difference is significant for each individual word as well as for the aggregate data. Data for all four words will again be collapsed in the analysis that follows.

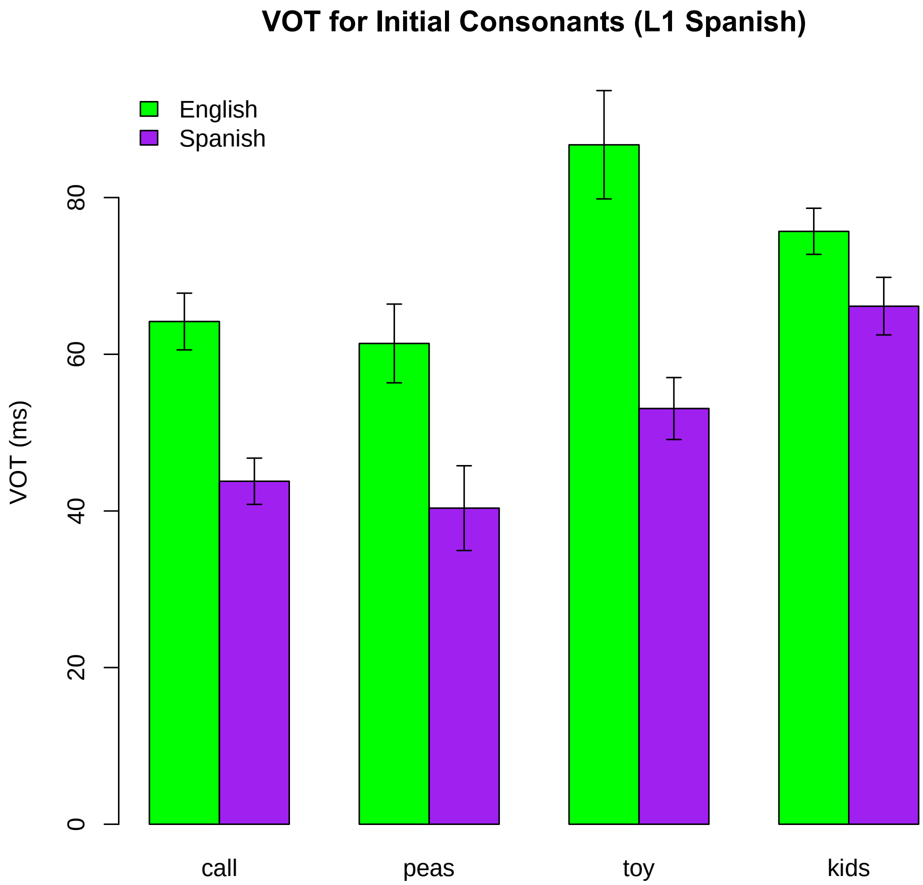


FIGURE 15. Average VOT for initial voiceless consonants produced by L1 Spanish and L1 English speakers.

	Spanish		English		Welch's t-test		
Word	Mean VOT (ms)	Std. Dev.	Mean VOT (ms)	Std. Dev.	t	d.f.	p
call	43.79	20.87	64.18	16.20	4.36	44.91	< 0.001
peas	40.36	38.24	61.37	22.46	2.85	58.24	< 0.01
toy	53.07	27.97	86.74	30.93	4.22	32.13	< 0.001
kids	66.14	25.97	75.69	13.14	2.03	64.09	< 0.05
ALL WORDS	**50.84**	**30.43**	**72.00**	**23.64**	**6.21**	**186.05**	**< 0.001**

TABLE 3. Descriptive and comparative statistics for initial voiceless stop data.

The effect of age of onset on production of voiced stops is shown in Figure 16. As might be expected, earlier learners produce stops with more English-like VOT values (recall from Table 4 that the native English mean VOT is 72 ms). A likelihood ratio comparison between a “null” model (with only random effects of speaker and word) and a model with an added fixed effect of age of onset confirms that this is a significant predictive factor ($\chi^2 = 4.35$, $p < 0.05$).

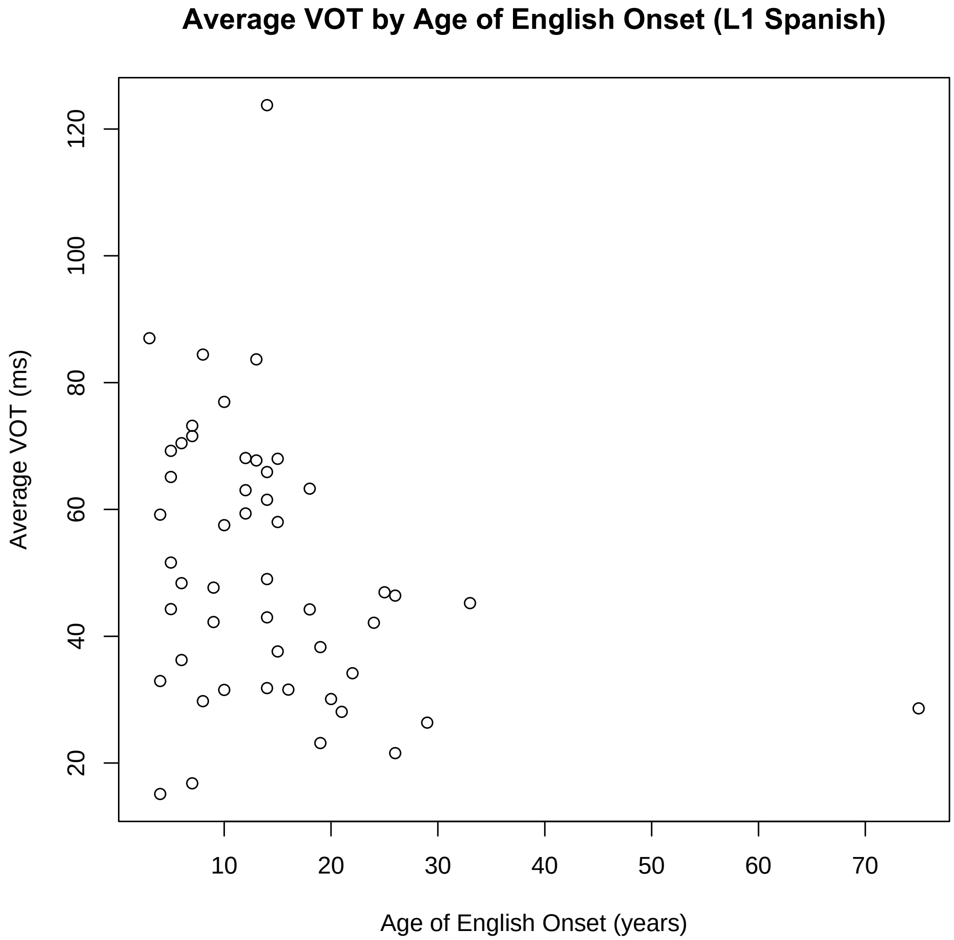


FIGURE 16. Age of onset and production of VOT by L1 Spanish speakers.

Figure 17 shows the relationship between mean VOT and length of residency. Impressionistically, it is difficult to discern a pattern from the graph; while there are fewer participants with longer LOR, their average VOT values do not seem to trend in any particular direction. The likelihood ratio comparison here confirms this impression: adding length of residency to the linear mixed effects model has no predictive value for the L1 Spanish speakers ($\chi^2 = 0.09$, $p = 0.76$ (n.s.)).

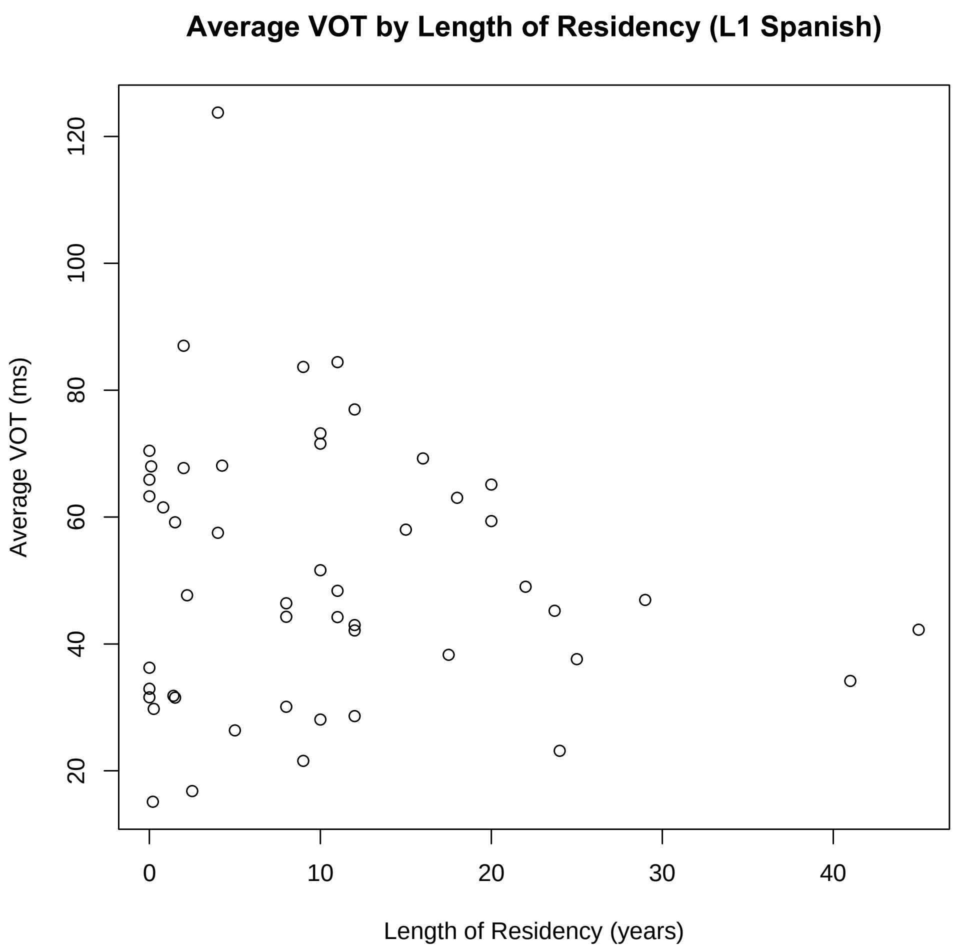


FIGURE 17. Length of residency and production of VOT by L1 Spanish speakers.

3.3. L1 French, L2 English

3.3.1 Participants and Materials

Data from all 51 native French speakers on the Speech Accent Archive were collected and processed using the methodology described in Sections 3.1.1 and 3.2.1. The same L1 English data used in previous sections was also used for comparison.

3.3.2 Procedure

The procedure used here is identical to that described in Section 3.2.2, targeting the initial stops in the words *call*, *peas*, *toy*, and *kids*.

3.3.3 Results and Discussion

Figure 18 and Table 4 give the basic results for L1 French speakers. VOT values for the initial voiceless English stop in each word differ in exactly the way we would expect if French speakers associate aspirated (high-VOT) English stops with their native unaspirated (low-VOT) voiceless stops: the values, in each case, are larger for L1 English speakers than L1 French speakers. This difference is highly significant for all individual words except *kids* and for the data set as a whole. Data for all four words will be collapsed in the analysis that follows.

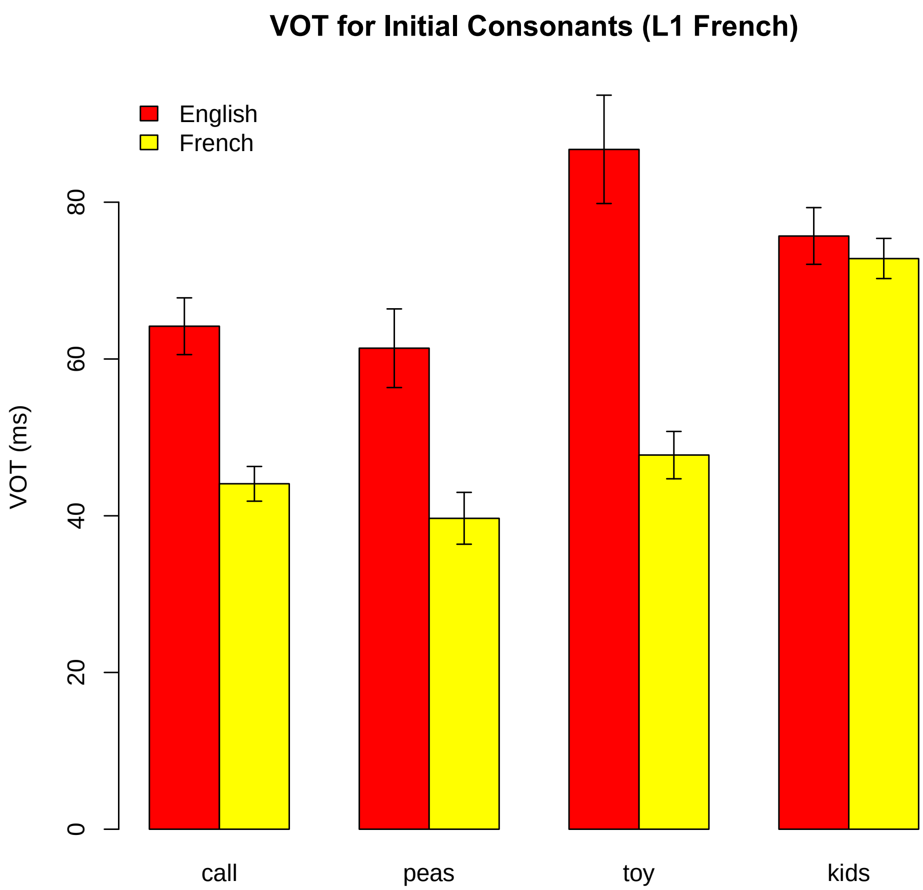


FIGURE 18. Average VOT for initial voiceless consonants produced by L1 French and L1 English speakers.

Word	French		English		Welch's t-test		
	Mean VOT (ms)	Std. Dev.	Mean VOT (ms)	Std. Dev.	t	d.f.	p
call	44.07	15.84	64.18	16.20	4.73	34.09	< 0.001
peas	39.67	23.61	61.37	22.46	3.61	36.42	< 0.001
toy	47.74	21.58	86.74	30.93	5.17	26.58	< 0.001
kids	72.81	18.29	75.69	13.14	0.74	48.25	0.46 (n.s.)
ALL WORDS	**51.07**	**23.72**	**72.00**	**23.64**	**-4.17**	**206.22**	**< 0.001**

TABLE 4. Descriptive and comparative statistics for initial voiceless stop data.

The effect of age of onset on production of voiced stops is shown in Figure 19. Perhaps surprisingly, no strong tendency is apparent: participants who began to acquire English earlier do not seem to produce more native-like (i.e. high-VOT) initial voiceless stops. A likelihood ratio comparison between a "null" model (with only random effects of speaker and word) and a model with an added fixed effect of age of onset confirms the impression that age of onset is not a strong predictor of VOT for L1 French speakers; the effect is only marginally significant ($\chi^2 = 3.52$, $p = 0.06$).

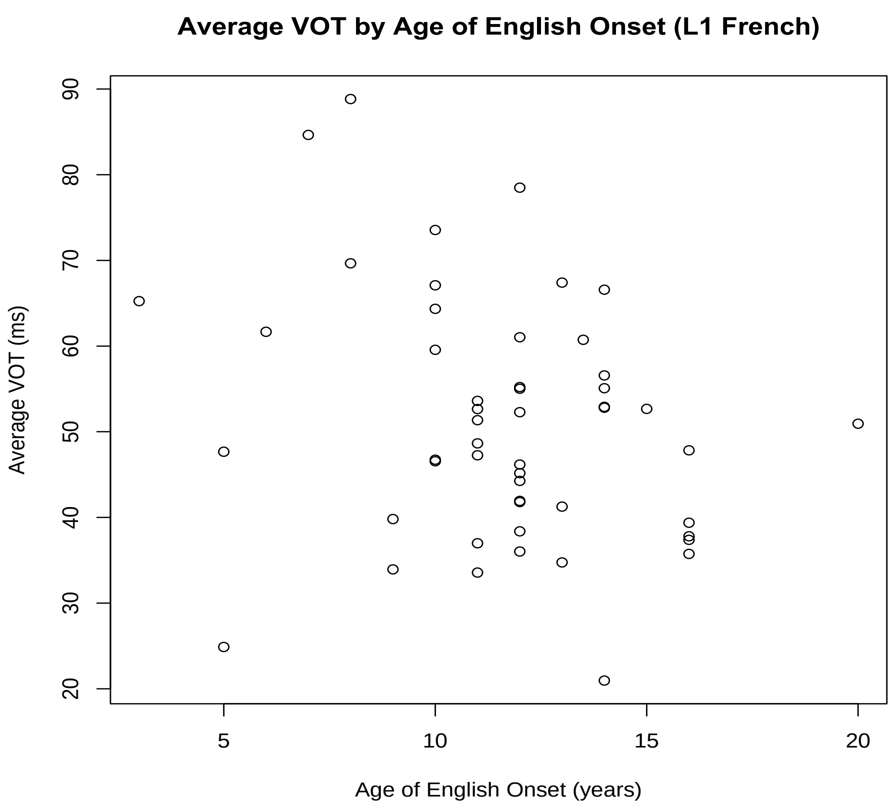


FIGURE 19. Age of onset and production of VOT by L1 French speakers.

Figure 20 shows the relationship between mean VOT and length of residency. A slight upward trend seems discernible here, indicating that speakers with longer LOR produce more English-like stops. Since the effect of age of onset proved to be marginal, I employ two additional mixed effects models to confirm the contribution of length of residency. The first model includes both age of onset and LOR as fixed effects; the likelihood ratio test confirms that this model significantly improves on the previous model that included only age of onset ($\chi^2 = 5.61$, $p < 0.05$). Furthermore, a model with only LOR as a fixed effect significantly improves on a model with only random effects ($\chi^2 = 6.41$, $p < 0.05$).

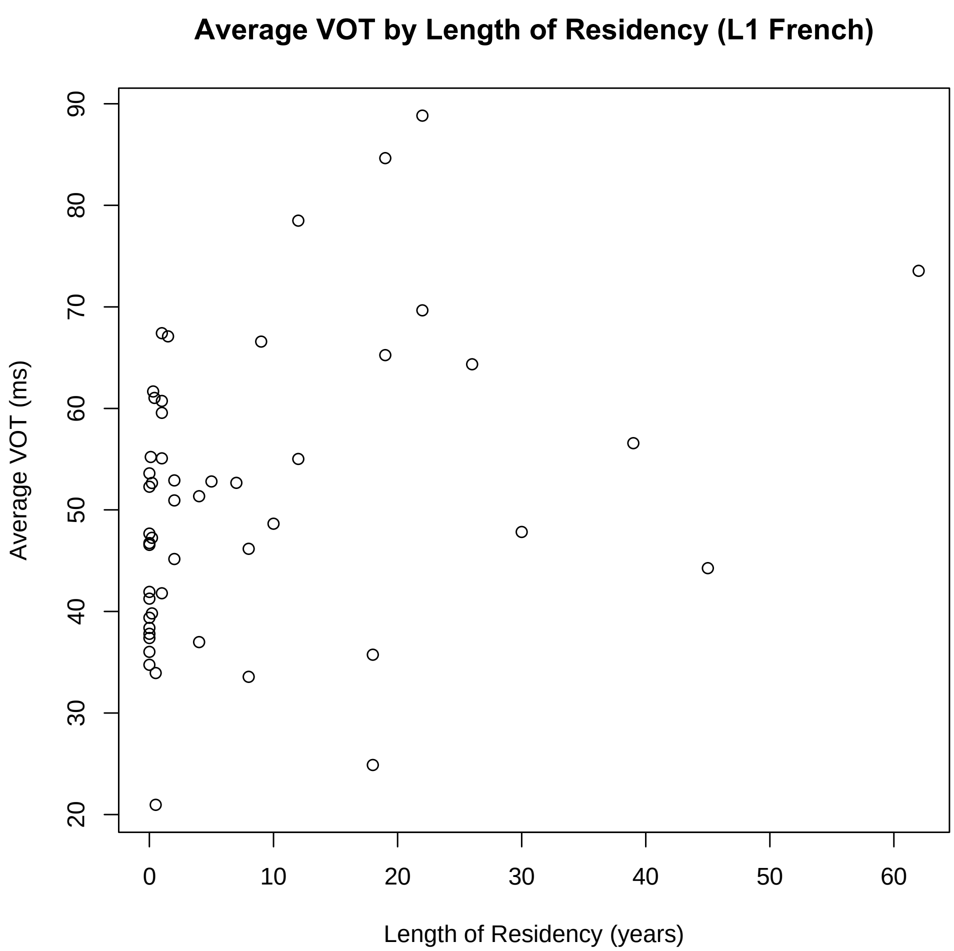


FIGURE 20. Length of residency and production of VOT by L1 French speakers.

4. General Discussion

It is not surprising that age of English onset emerged as a strong predictive factor for both Mandarin and Spanish speakers. Of all the linguistic domains, phonology is the most obviously susceptible to critical period effects (Lenneberg, 1967), so it is very much expected that early learners will generally manage to achieve native English-like VOT values while later learners will largely fail to do so. That is exactly what we found here for these two languages (see Figures 13 and 16). There is, however, one surprise in the results regarding age of onset: the null (or, at best, marginal) result obtained for the L1 French speakers. Linear mixed effects models led us to conclude that age of onset was not a significant predictor of VOT for these speakers. It may be the case that this factor has too narrow a range to reveal any significant correlation with VOT in this speaker group, since all participants had begun to learn English by the age of 20. A larger data set might reveal the expected result that younger learners should produce more English-like stops.

Turning to the main research question of this paper – whether L2 English learners can produce stops with more English-like VOT values with increased length of residency – the results are mixed, but point in a distinctly positive direction. LOR emerged as a very strong independent predictor of performance for the L1 Mandarin and French speakers, but had no effect on the performance of the L1 Spanish group. Whatever factor that accounts for this discrepancy is not likely to be found within phonological theory itself, since the relevant properties of Spanish and French stops are very similar. Both languages feature a contrast between fully voiced stops and voiceless, unaspirated stops, yet the pattern of acquisition of English stops differed markedly for the two groups. Since the effect of length of residency was observed for two very different languages, Mandarin and French, the preponderance of the evidence certainly indicates that this effect is real – that adult learners indeed can improve their pronunciation with experience, producing more native like values for VOT when producing English stops. The challenge is then to explain what is blocking this effect for the Spanish speakers. Sociolinguistic factors might plausibly come into play here. Perhaps Spanish speakers living in an English-speaking country tend to use their native language more often than similarly situated Mandarin or French speakers, causing them to retain L1 Spanish-like stops even with significant exposure to English. Or perhaps Spanish speakers, relative to Mandarin or French speakers, are more interested in maintaining their own linguistic identity than they are in assimilating to the dominant language community and therefore value the retention of certain aspects of their accent when speaking English. More detailed study is needed to provide a definitive answer to this problem.

To bring this discussion to a close, I would like to consider once again each of the models of phonological learning discussed in Section 2 of this paper, evaluating them in light of the results of the corpus study. The Speech Learning Model (Flege, 1995) predicts that late learners of a second language should not be able to form new categories for L2 sounds very similar to those found in their L1 due to deficiencies in linguistic perceptual ability. Only the results for the L1 Spanish speakers are consistent with this theory: early learners produced English-like stops, while later learners produced more Spanish-like stops, and length of residency had no effect on this pattern. The Speech Learning Model would have much more difficulty accounting for the Mandarin and French results, however. This model does not have any mechanism that could account for late learners improving over time, producing more native-like VOT values with longer LOR, as was robustly found in the present study. On the contrary, the SLM predicts fossilization at L1-like values when L1 and L2 sounds are as close together as they are in this study. Overall, the results discussed in this paper present a difficult challenge to proponents of the Speech Learning Model.

Precisely the opposite is true for the Full Transfer/Full Access Model (Escudero & Boersma, 2004; Schwartz & Sprouse, 1996), which predicts that, while beginning learners might produce sounds closer to their L1 than to their L2 target, they should be able to converge on L2 norms with the right kind and the right quantity of experience. The Mandarin and French data conform closely to this prediction; the right kind of experience, in this case, is residency in an L2-speaking country, and the right quantity appears to be twenty years or more of this kind of immersion. Since the Spanish data do not conform to this pattern, we cannot declare unequivocal victory for Full Transfer/Full Access; as discussed above, however, the evidence certainly seems to tip the scale firmly in that direction.

The Statistical Learning Model of McMurray et al. (2009) makes predictions similar to, but slightly different from those of Full Transfer/Full Access. On the analysis considered in Section 2, the Statistical Learning Model predicts that learners should initially produce L1-like stops, gradually expanding their existing category so that average VOT slowly approaches L2 norms, until finally forming an authentic L2 category with target-like production. The main observable difference should be a gap in the progress of VOT values: at some point, when a new category is formed by the learner, average values should jump more or less directly from some intermediate value to a target-like distribution. For reasons already amply discussed, the Spanish data will not provide support for any model that, like Statistical Learning, predicts progressive development, but it is worth taking a closer look at what the Mandarin and French data have to say. Figures 14 and 20 showed the mean VOT produced by each of the L1 Mandarin and French speakers, respectively, as a function of their length of residency. There is no clearly discernible gap in average VOT values of the kind that would be predicted by the Statistical Learning Model. It must be admitted that the data become somewhat sparse for larger values of LOR, so these data may well be insufficient to discern between the very similar predictions of Full Transfer/Full Access and the Statistical Learning Model.

5. Conclusion

In this paper, I presented the results of a corpus study using data from the George Mason University Speech Accent Archive. I looked at voice onset time (VOT) in the production of word-initial stops among three groups, native Mandarin, Spanish, and French speakers, all of whom were speakers of English as a second language. These three L2 English groups differed in expected ways from a control group of native English speakers, often producing stops with VOT closer to their L1 values or intermediate between L1 and target-like English values. Age of English onset emerged as a strong predictor of behavior among Mandarin and Spanish groups, in accord with the existence of a critical period effect for phonological acquisition. No such effect was observed for the French group, but this may have been due to the narrower range for age of onset in this group. More interestingly, length of residency in an English-speaking region emerged as a strong predictor of VOT production for the L1 Mandarin and French groups, providing strong support for the Full Transfer/Full Access model of acquisition, and a strong challenge for Flege's Speech Learning Model. This was not the case for the L1 Spanish speakers, who showed no indication of producing more authentic English stops as they gained linguistic experience. Since Spanish and French are very similar in the relevant phonological respects, it was speculated that sociolinguistic factors were at play in this null result for Spanish speakers. Further research is needed to confirm this speculation, and to more carefully evaluate these two models of phonological learning, as well as the Statistical Learning Model.

Appendix A: Corpus Study Data

VOT and demographic data for native English-speaking controls, L1 Mandarin speakers (Section 3.1), L1 Spanish speakers (Section 3.2), and L1 French speakers (Section 3.3) are presented in the tables below. Demographic data not considered in this paper are not included here, but are available from the George Mason University Speech Accent Archive (http://accent.gmu.edu) or from the author.

Speaker	Age	Sex	Word	VOT /b/ags (ms)	VOT /b/ig (ms)	VOT /b/ob (ms)	VOT /g/o (ms)	VOT /k/all (ms)	VOT /p/eas (ms)	VOT /t/oy (ms)	VOT /k/ids (ms)
english021	37	F	bags	16.236	16.719	-18.564	28.536	52.028	54.517	186.745	81.157
english023	43	M	bags	-119.096	-111.67	-100.662	-88.422	53.788	61.671	64.517	70.928
english044	63	F	bags	0	10.59	13.674	23.711	86.435	71.639	121.92	80.495
english068	52	M	bags	9.852	-57.656	-47.859	32.015	33.531	9.145	69.175	70.429
english075	32	M	bags	9.295	0	17.612	23.223	61.961	42.608	131.353	74.303
english133	36	F	bags	12.085	-41.47	0	10.98	52.855	69.256	85.302	69.581
english134	22	F	bags	0	9.913	-98.243	22.365	76.317	102.948	99.87	98.939
english135	22	M	bags	0	10.265	0	7.633	75.589	54.911	86.919	86.315
english137	26	M	bags	-87.744	-58.976	-50.565	14.189	45.894	83.066	57.234	69.804
english158	18	F	bags	10.792	7.853	-13.83	27.94	36.996	56.18	58.906	75.483
english159	18	M	bags	0	-12.113	15.995	24.581	81.758	73.422	100.253	81.335
english221	45	M	bags	21.834	9.282	12.804	31.794	64.042	26.068	78.713	78.256
english242	77	F	bags	0	10.399	17.984	46.095	90.917	42.027	84.82	91.703
english256	23	M	bags	-4.088	-19.271	0		85.38	54.32	78.845	93.326
english292	38	M	bags	-143.407	-104.351	-80.508	21.72	52.277	77.736	66.603	77.628
english325	32	M	bags	0	18.464	0	21.741	71.241	35.793	69.536	53.745
english332	20	F	bags	0	-9.808	-6.019		65.17	89.018	81.525	41.039
english369	27	M	bags	7.849	11.713	13.824	16.03	56.213	72.606	86.427	82.514
english412	31	F	bags	13.052	-62.617	-79.79	40.999	69.663	72.644	72.012	66.799
english441	19	F	bags	10.732	8.088	10.17	35.6	71.563	77.802	54.04	70.03

TABLE A1. Data for L1 English controls.

Speaker	Age	Sex	Onset	Method	LOR	VOT /b/ags (ms)	VOT /b/ig (ms)	VOT /b/ob (ms)	VOT /g/o (ms)
mandarin01	26	F	13	Academic	2	9.742	6.558	5.706	10.732
mandarin02	38	F	14	Academic	0.8	12.209	15.759	13.439	12.38
mandarin03	43	M	10	Academic	14	-24.069	-23.01	0	23.957
mandarin04	24	F	6	Academic	1	4.046	12.673	13.472	36.082
mandarin05	31	F	12	Academic	2	11.335	12.258	15.406	30.715
mandarin06	28	F	12	Academic	5	16.799	11.775	11.627	36.801
mandarin07	22	M	5	Naturalistic	22	-22.293	-19.251	-31.368	-11.79
mandarin08	29	M	12	Academic	5	-14.184	-34.379	-40.182	27.772
mandarin09	38	M	12	Academic	2	0	16.07	15.835	30.947
mandarin10	19	M	3	Academic	3.75	13.163	13.551	12.147	22.275
mandarin11	53	F	13	Academic	33	-32.249	-33.874	-10.768	18.996
mandarin12	23	M	1	Naturalistic	23	-19.799	-62.959	-41.224	14.114
mandarin13	29	M	13	Academic	0	11.17	7.146	7.926	18.123
mandarin14	49	M	20	Academic	21	20.333	13.963	11.172	23.905
mandarin15	28	F	11	Academic	4	-11.267	0	-19.984	24.737
mandarin16	32	M	10	Academic	1.35	10.521	-15.043	-12.838	23.499
mandarin17	26	M	13	Academic	2	-17.009	19.553	26.611	26.559
mandarin18	29	F	13	Academic	0	18.961	20.025	16.871	38.49
mandarin19	27	M	13	Academic	0.33	2.724	9.289	7.443	5.956
mandarin20	40	F	12	Academic	5	13.636	-83.527	-17.58	20.678
mandarin21	38	F	9	Academic	13	16.155	12.305	16.126	23.948
mandarin22	39	F	11	Academic	3	-57.214	-80.416	-85.928	29.729
mandarin23	46	F	13	Academic	22	-11.176	7.799	-33.972	21.994
mandarin24	21	F	10	Academic	0	12.044	8.142	13.664	21.07
mandarin25	28	F	13	Academic	0.8	11.763	12.44	8.606	28.962
mandarin26	31	F	12	Academic	3	17.622	17.237	25.629	18.583
mandarin27	18	M	15	Academic	3	21.515	14.73	-60.357	24.752
mandarin28	45	M	15	Academic	20	11.358	-130.403	13.723	20.395
mandarin29	24	M	4	Academic	0	-15.182	-11.19	-37.475	17.416
mandarin30	27	M	13	Academic	1	11.289	0	10.527	22.336
mandarin31	26	F	13	Academic	2.5	-4.698	12.246	-12.76	39.815
mandarin32	23	F	12	Academic	2	-14.242	-9.334	-11.677	26.895
mandarin33	25	M	6	Naturalistic	20	26.221	-98.457	22.872	-81.896
mandarin34	31	F	13	Academic	0	0	11.302	0	18.411
mandarin35	27	F	12	Academic	0	11.375	10.83	13.402	41.979
mandarin36	32	F	10	Academic	0	11.237	12.361	15.27	25.844
mandarin37	32	F	12	Academic	0	11.579	11.952	22.963	48.386
mandarin38	33	F	12	Academic	0.1	-11.267	25.431	49.484	29.997
mandarin39	24	F	12	Academic	0	11.171	-47.925	20.387	23.429
mandarin40	28	M	12	Academic	9	9.377	-63.113	16.322	29.776
mandarin41	34	F	13	Academic	0	10.765	13.639	14.385	24.07
mandarin42	47	F	16	Academic	0	9.42	9.157	9.021	16.239
mandarin43	24	F	15	Academic	0	12.215	12.769	19.604	22.621
mandarin44	21	M	16	Academic	1.5	8.602	9.362	11.03	25.508
mandarin45	42	F	12	Academic	0.1	7.827	7.552	16.074	20.075
mandarin46	43	F	9	Academic	0	0	10.82	16.969	16.327
mandarin47	28	F	8	Academic	0	9.568	7.097	-93.633	21.24
mandarin48	37	M	12	Academic	0	-76.162	10.019	10.44	23.143

TABLE A2. Data for L1 Mandarin speakers.

Speaker	Age	Sex	Onset	Method	LOR	VOT /k/all (ms)	VOT /p/eas (ms)	VOT /t/oy (ms)	VOT /k/ids (ms)
spanish065	32	M	12	Naturalistic	20	54.415	32.246	70.724	80.081
spanish066	30	M	12	Naturalistic	18	38.769	39.942	80.206	93.189
spanish067	48	F	14	Academic	4	70.222	184.319	73.62	166.856
spanish068	27	M	14	Academic	0.8	59.217	95.655	27.786	63.428
spanish069	20	F	7	Naturalistic	10	83.315	59.695	82.835	66.984
spanish070	30	F	10	Academic	1.5	34.04	0	25.631	66.475
spanish071	24	M	6	Academic	0	51.194	67.333	81.368	81.927
spanish072	20	F	5	Naturalistic	20	86.193	49.351	81.13	43.798
spanish073	19	M	15	Academic	0.1	38.752	76.528	51.241	105.443
spanish074	44	M	14	Academic	0	31.925	84.632	57.125	89.927
spanish075	28	F	5	Academic	8	31.359	25.066	62.302	58.411
spanish076	27	M	10	Academic	4	47.032	47.492	56.92	78.644
spanish077	45	M	13	Academic	2	66.845	41.32	70.234	92.552
spanish078	55	F	5	Academic	16	17.943	52.399	112.98	93.654
spanish079	70	F	22	Academic	41	38.907	21.359	21.512	54.883
spanish080	77	F	75	Naturalistic	12	52.26	12.985	15.363	33.889
spanish081	63	F	7	Academic	2.5	24.985	0	14.362	27.872
spanish082	49	M	14	Academic	12	42.398	30.012	38.303	61.175
spanish083	48	M	14	Academic	1.4	43.448	16.44	20.686	46.737
spanish084	19	F	4	Academic	0.2	29.931	-73.212	56.258	47.556
spanish085	44	F	6	Academic	0	34.265	-9.091	79.471	40.35
spanish086	23	M	4	Academic	0	27.881	14.974	39.523	49.335
spanish087	55	M	9	Academic	2.2	54.18	51.567	23.119	61.82
spanish088	53	M	25	Naturalistic	29	62.481	33.031	35.747	56.514
spanish089	22	M	13	Academic	9	41.061	84.4	107.741	101.442
spanish090	23	F	3	Naturalistic	2	51.092	124.663	73.645	98.638
spanish091	20	F	7	Naturalistic	10	80.414	55.705	86.815	63.357
spanish092	54	M	9	Naturalistic	45	46.796	32.513	46.431	43.237
spanish093	29	F	18	Naturalistic	11	36.902	47.135	39.364	53.506
spanish094	19	M	6	Academic	11	20.642	42.738	56.516	73.618
spanish095	31	F	21	Naturalistic	10	44.719	14.933	12.608	40.12
spanish096	29	F	12	Academic	4.25	32.441	47.725	78.028	114.255
spanish097	52	F	19	Academic	17.5	35.555	28.809	44.15	44.658
spanish098	34	M	29	Naturalistic	5	27.943	21.913	18.715	36.935
spanish099	20	F	4	Academic	1.5	74.937	21.734	91.542	48.497
spanish100	30	F	26	Academic	8	38.965	0	70.809	75.858
spanish101	39	F	19	Naturalistic	24	0	0	13.16	79.468
spanish102	31	M	26	Academic	9	26.933	12.942	13.181	33.198
spanish103	46	M	15	Naturalistic	25	33.829	29.789	19.498	67.33
spanish104	28	F	20	Naturalistic	8	35.148	8.086	32.67	44.436
spanish105	41	F	10	Academic	12	30.802	88.857	83.079	105.084
spanish106	21	M	5	Academic	10	42.12	50.727	55.44	58.189
spanish107	36	M	14	Naturalistic	22	24.823	69.382	52.792	49.101
spanish108	36	M	24	Academic	12	42.173	29.514	33.842	62.944
spanish109	21	M	16	Academic	0	31.474	17.199	24.756	52.923
spanish110	20	M	18	Academic	0	104.685	36.381	45.529	66.54
spanish111	55	M	33	Naturalistic	23.7	47.31	46.347	23.687	63.54
spanish112	25	M	15	Academic	15	46.973	38.122	94.403	52.53
spanish113	19	M	8	Academic	11	75.983	80.291	99.736	81.685
spanish114	22	F	8	Naturalistic	0.25	-6.355	34.094	57.048	34.273

TABLE A3. Data for L1 Spanish speakers.

Speaker	Age	Sex	Onset	Method	LOR	VOT /k/all (ms)	VOT /p/eas (ms)	VOT /t/oy (ms)	VOT /k/ids (ms)
french1	20	F	12	Academic	0.4	50.908	72.180	38.869	82.184
french2	19	M	14	Academic	0.5	29.588	-13.459	27.951	39.751
french3	22	F	11	Academic	0.2	43.053	29.583	58.736	79.217
french4	31	F	14	Academic	9	46.257	24.514	64.880	130.654
french5	36	F	11	Academic	4	30.860	41.743	26.847	48.497
french6	26	F	13	Academic	1	38.239	108.743	40.671	81.987
french7	18	M	5	Academic	18	24.347	14.812	19.118	41.237
french8	66	M	16	Academic	0	48.863	19.030	41.335	41.951
french9	21	M	11	Academic	0	67.608	38.246	58.987	49.542
french10	31	M	10	Academic	1	45.911	36.406	71.447	84.468
french11	31	M	11	Academic	8	22.362	27.510	19.977	64.378
french12	19	F	9	Academic	0.2	41.944	27.348	23.946	65.981
french13	19	M	12	Academic	0	31.455	17.067	31.389	64.183
french14	23	F	14	Academic	1	41.120	65.087	44.307	69.817
french15	32	M	12	Academic	0	27.841	36.574	28.223	60.838
french16	19	F	7	Academic	19	54.405	91.314	89.150	103.689
french17	39	M	14	Academic	5	48.865	37.200	40.721	84.420
french18	22	M	6	Academic	0.3	37.618	69.648	60.185	79.219
french19	39	M	12	Academic	12	64.505	33.742	59.691	62.167
french20	23	M	12	Academic	2	39.911	49.004	29.985	61.777
french21	20	M	9	Academic	0.5	31.218	18.685	27.966	57.891
french22	78	F	16	Academic	18	35.386	20.377	27.528	59.678
french23	76	F	16	Academic	0	30.734	45.619	27.745	53.442
french24	47	M	13	Academic	0	29.838	11.306	15.955	81.879
french25	20	M	12	Academic	0.1	25.423	67.046	61.941	66.468
french26	27	F	11	Academic	4	38.084	19.996	59.145	88.149
french27	38	F	12	Academic	8	46.696	46.031	36.098	55.847
french28	35	F	11	Naturalistic	10	47.760	25.425	52.347	69.042
french29	54	F	10	Academic	26	52.305	34.410	61.211	109.435
french30	37	M	12	Academic	12	117.219	56.513	48.400	91.850
french31	28	M	15	Academic	7	55.118	21.453	55.190	78.932
french32	60	M	16	Academic	0	50.787	18.056	19.164	61.496
french33	62	M	10	Academic	62	38.937	53.477	116.716	85.035
french34	56	M	16	Academic	30	72.404	13.036	20.144	85.709
french35	27	M	10	Academic	0	37.526	37.232	49.534	62.664
french36	32	F	10	Academic	1.5	47.575	61.665	75.937	83.196
french37	42	M	14	Naturalistic	2	30.641	60.586	67.291	53.120
french38	22	M	8	Academic	22	75.585	95.181	96.581	87.982
french39	28	M	13	Academic	0	35.424	27.228	24.003	78.350
french40	44	M	13.5	Academic	1	39.791	66.918	65.730	70.515
french41	24	M	12	Academic	0	38.011	22.842	47.815	59.028
french42	22	F	20	Academic	2	32.112	36.364	53.086	82.166
french43	22	M	11	Academic	0.2	53.512	29.783	41.035	64.658
french44	20	F	12	Academic	0	44.978	32.881	62.011	69.254
french45	22	M	8	Naturalistic	22	54.433	63.446	77.695	83.076
french46	22	M	10	Academic	0	40.328	19.824	45.129	80.962
french47	66	F	12	Academic	45	47.424	35.897	31.928	61.775
french48	19	F	3	Naturalistic	19	56.267	30.833	64.222	109.668
french49	39	M	12	Academic	1	41.735	30.169	36.038	59.251
french50	67	F	14	Academic	39	31.292	69.818	30.725	94.460
french51	18	M	5	Academic	0	33.459	24.772	59.840	72.580

TABLE A4. Data for L1 French speakers.

Appendix B: Linear Mixed Effects Model Results

Detailed results for each linear mixed effects model discussed in Section 3 are given below.

```
Linear mixed model fit by REML
Formula: M$VOT ~ (1 | M$Speaker) + (1 | M$Word)
  AIC  BIC logLik deviance REMLdev
 1801 1814 -896.5     1799    1793
Random effects:
 Groups    Name        Variance Std.Dev.
 M$Speaker (Intercept) 129.49   11.379
 M$Word    (Intercept) 137.33   11.719
 Residual              555.76   23.575
Number of obs: 192, groups: M$Speaker, 48; M$Word, 4

Fixed effects:
            Estimate Std. Error t value
(Intercept)    4.001      6.317   0.633
```

TABLE B1. L1 Mandarin, model with only random effects of speaker and word.

```
Linear mixed model fit by REML
Formula: M$VOT ~ M$EngOnset + (1 | M$Speaker) + (1 | M$Word)
  AIC  BIC logLik deviance REMLdev
 1796 1812 -892.9     1792    1786
Random effects:
 Groups    Name        Variance Std.Dev.
 M$Speaker (Intercept) 101.17   10.058
 M$Word    (Intercept) 137.33   11.719
 Residual              555.76   23.575
Number of obs: 192, groups: M$Speaker, 48; M$Word, 4

Fixed effects:
            Estimate Std. Error t value
(Intercept) -15.0947     9.7487  -1.548
M$EngOnset    1.6726     0.6538   2.558

Correlation of Fixed Effects:
           (Intr)
M$EngOnset -0.766
```

TABLE B2. L1 Mandarin, model with added fixed effect of age of onset.

```
Linear mixed model fit by REML
Formula: M$VOT ~ M$EngOnset + M$LOR + (1 | M$Speaker) + (1 | M$Word)
  AIC  BIC logLik deviance REMLdev
 1787 1806 -887.3     1780    1775
Random effects:
 Groups    Name        Variance Std.Dev.
 M$Speaker (Intercept)  49.409   7.0291
 M$Word    (Intercept) 137.332  11.7189
 Residual              555.757  23.5745
Number of obs: 192, groups: M$Speaker, 48; M$Word, 4

Fixed effects:
            Estimate Std. Error t value
(Intercept)  -6.8505     9.3236  -0.735
M$EngOnset    1.3635     0.5851   2.330
M$LOR        -0.9118     0.2469  -3.693

Correlation of Fixed Effects:
           (Intr) M$EngO
M$EngOnset -0.736
M$LOR      -0.239  0.143
```

TABLE B3. L1 Mandarin, model with added fixed effects of age of onset and length of residency.

```
Linear mixed model fit by REML
Formula: S$VOT ~ (1 | S$Speaker) + (1 | S$Word)
  AIC  BIC logLik deviance REMLdev
 1889 1902 -940.6     1887    1881
Random effects:
 Groups    Name        Variance Std.Dev.
 S$Speaker (Intercept) 327.34   18.092
 S$Word    (Intercept) 122.62   11.073
 Residual              511.45   22.615
Number of obs: 200, groups: S$Speaker, 50; S$Word, 4

Fixed effects:
            Estimate Std. Error t value
(Intercept)   50.839      6.304   8.064
```

TABLE B4. L1 Spanish, model with only random effects of speaker and word.

```
Linear mixed model fit by REML
Formula: S$VOT ~ S$EngOnset + (1 | S$Speaker) + (1 | S$Word)
  AIC  BIC logLik deviance REMLdev
 1888 1904 -938.8     1882    1878
Random effects:
 Groups    Name        Variance Std.Dev.
 S$Speaker (Intercept) 297.48   17.248
 S$Word    (Intercept) 122.62   11.074
 Residual              511.45   22.615
Number of obs: 200, groups: S$Speaker, 50; S$Word, 4

Fixed effects:
            Estimate Std. Error t value
(Intercept)  58.7141     7.2882   8.056
S$EngOnset   -0.5469     0.2595  -2.107

Correlation of Fixed Effects:
           (Intr)
S$EngOnset -0.513
```

TABLE B5. L1 Spanish, model with added fixed effect of age of onset.

```
Linear mixed model fit by REML
Formula: S$VOT ~ S$EngOnset + S$LOR + (1 | S$Speaker) + (1 | S$Word)
  AIC  BIC logLik deviance REMLdev
 1890 1910 -939.1     1882     1878
Random effects:
 Groups    Name        Variance Std.Dev.
 S$Speaker (Intercept) 305.50   17.479
 S$Word    (Intercept) 122.62   11.074
 Residual              511.45   22.615
Number of obs: 200, groups: S$Speaker, 50; S$Word, 4

Fixed effects:
            Estimate Std. Error t value
(Intercept) 59.44911    7.63905   7.782
S$EngOnset  -0.52832    0.26777  -1.973
S$LOR       -0.09807    0.29285  -0.335

Correlation of Fixed Effects:
           (Intr) S$EngO
S$EngOnset -0.424
S$LOR      -0.287 -0.207
```

TABLE B6. L1 Spanish, model with added fixed effects of age of onset of length of residency.

```
Linear mixed model fit by REML
Formula: VOT ~ (1 | Subject) + (1 | Word)
   AIC   BIC logLik deviance REMLdev
 724.8 734.5 -358.4    722.2   716.8
Random effects:
 Groups   Name        Variance Std.Dev.
 Subject  (Intercept) 100.49   10.024
 Word     (Intercept) 107.69   10.377
 Residual             253.77   15.930
Number of obs: 83, groups: Subject, 21; Word, 4

Fixed effects:
            Estimate Std. Error t value
(Intercept)    9.270      5.896   1.572
```

TABLE B7. L1 French, model with only random effects of speaker and word.

```
Linear mixed model fit by REML
Formula: VOT ~ Onset + (1 | Subject) + (1 | Word)
   AIC   BIC logLik deviance REMLdev
 725.3 737.4 -357.6    721.8   715.3
Random effects:
 Groups   Name        Variance Std.Dev.
 Subject  (Intercept) 105.82   10.287
 Word     (Intercept) 107.24   10.356
 Residual             253.83   15.932
Number of obs: 83, groups: Subject, 21; Word, 4

Fixed effects:
            Estimate Std. Error t value
(Intercept)   6.6472     7.3160   0.909
Onset         0.4320     0.7108   0.608

Correlation of Fixed Effects:
      (Intr)
Onset -0.590
```

TABLE B8. L1 French, model with added fixed effect of age of onset.

```
Linear mixed model fit by REML
Formula: VOT ~ LOR + (1 | Subject) + (1 | Word)
  AIC  BIC logLik deviance REMLdev
 1787 1804 -888.7     1781    1777
Random effects:
 Groups   Name        Variance Std.Dev.
 Subject  (Intercept) 115.44   10.744
 Word     (Intercept) 215.74   14.688
 Residual             265.94   16.308
Number of obs: 204, groups: Subject, 51; Word, 4

Fixed effects:
            Estimate Std. Error t value
(Intercept)  48.0296     7.6718   6.261
LOR           0.3750     0.1449   2.589

Correlation of Fixed Effects:
    (Intr)
LOR -0.153
```

TABLE B9. L1 French, model with added fixed effect of length of residency.

```
Linear mixed model fit by REML
Formula: VOT ~ Onset + LOR + (1 | Subject) + (1 | Word)
  AIC  BIC logLik deviance REMLdev
 1786 1806   -887     1779    1774
Random effects:
 Groups   Name        Variance Std.Dev.
 Subject  (Intercept) 109.33   10.456
 Word     (Intercept) 215.74   14.688
 Residual             265.94   16.308
Number of obs: 204, groups: Subject, 51; Word, 4

Fixed effects:
            Estimate Std. Error t value
(Intercept)  59.7657    10.4697   5.708
Onset        -0.9882     0.6009  -1.645
LOR           0.3435     0.1437   2.390

Correlation of Fixed Effects:
      (Intr) Onset
Onset -0.682
LOR   -0.200  0.133
```

TABLE B10. L1 French, model with added fixed effects of age of onset of length of residency.

Bibliography

Bates, D., Maechler, M., & Bolker, B. (2012). Linear mixed-effects models using S4 classes (Version 0.999999-0). Retrieved from http://lme4.r-forge.r-project.org/

Bialystok, E., Craik, F., & Ryan, J. (2006). Executive control in a modified antisaccade task: Effects of aging and bilingualism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(6).

Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer (Version 5.3.21). Retrieved from http://www.praat.org/

Chao, K.-Y., & Chen, L.-M. (2008). A Cross-Linguistic Study of Voice Onset Time in Stop Consonant Productions. *Computational Linguistics and Chinese Language Processing*, *13*(2), 215–232.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*(4), 551–585.

Flege, J. E. (1979). *Phonetic interference in second language acquisition* (Unpublished Ph.D. Thesis). Indiana University.

Flege, J. E. (1980). Phonetic approximation in second language acquisition. *Language Learning*, *30*, 117–134.

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustic Socity of America*, *89*(1), 395–411.

Flege, J. E. (1995). Second-language speech learning: Theory, Findings, and Problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–278). Baltimore, MD: York Press.

Flege, J. E., & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication*, *6*, 185–202.

Gandour, J., Petty, S. H., Dardarananda, R., Dechongkit, S., & Mukngoen, S. (1986). The acquisition of the voicing contrast in Thai: a study of voice onset time in word-initial stop consonants*. *Journal of Child Language*, *13*(3), 561–572.

Goldrick, M. (2004). Phonological features and phonotactic constraints in speech production. *Journal of Memory and Language*, *51*, 586–603.

Gonzalez Lopez, V. (2012). Spanish and English word-initial voiceless stop production in code-switched vs. monolingual structures. *Second Language Research*, *28*(2), 243–263.

Halicki, S. (2010). *Learner knowledge of target phonotactics: Judgments of French word transformations* (Ph.D. dissertation). Indiana University.

Kirov, C., & Wilson, C. (2012). The Specificity of Online Variation in Speech Production. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 3 4th Annual Conference of the Cognitive Science Society* (pp. 587–592). Austin TX: Cognitive Science Society.

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

Lisker, L., & Abramson, A. (1964). Voice Onset Time in the Production and Perception of English Stops. Presented at the LSA General Meeting, New York.

Major, R. C. (1987). English voiceless stop production by speakers of Brazilian Portuguese. *Journal of Phonetics*, *15*, 197–202.

Maye, J., & Gerken, L. (2000). Learning phonemes without minimal pairs. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development* (pp. 522–533). Somerville MA: Cascadilla Proceedings Project.

Maye, J., Werker, J., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), 101–111.

McMurray, B., Aslin, R., & Toscano, J. (2009). (2009). Statistical learning of phonetic categories: Insights from a computational approach. *Developmental Science*, *12*(3), 369–378.

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, *39*(2), 132–142.

Pavlenko, A. (2000). L2 influence on L1 in late bilingualism. *Issues in Applied Linguistics*, *11*(2), 175–205.

R Development Core Team. (2011). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Riney, T., & Takagi, N. (1999). Global Foreign Accent and Voice Onset Time Among Japanese EFL Speakers. *Language Learning*, *49*(2), 275–302.

Rodriguez-Fornells, A., Balaguer, R., & Münte, T. (2006). Executive control in bilingual language processin. *Language Learning*, *56*(1), 133–190.

Schwartz, B., & Sprouse, R. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, *12*, 40–72.

Scovel, T. (1995). Differentiation, Recognition, and Identification in the Discrimination of Foreign Accents. In J. Archibald (Ed.), *Phonological Acquisition and Phonological Theory*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Tees, R., & Werker, J. (1984). Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology*, *38*(4), 579–590.

Weinberger, S. (2012). Speech Accent Archive. Retrieved from http://accent.gmu.edu

Williams, L. (1977). The voicing contrast in Spanish. *Journal of Phonetics*, *5*, 169–184.