Voice Onset Times of Voiceless Stops by Bilingual Korean-English Children: Phonetic Acquisition and Sociolinguistic Preferences

Danielle Fahey

1. Introduction

The acquisition of native-like accent in a second language (L2) depends on the acquisition of a number of phonological features. In particular, the acquisition of voice onset time (VOT) may be particularly challenging, with phonemic distinctions being made in some languages by minimal changes in VOT length (Cho & Ladefoged, 1999). Numerous studies of adult bilinguals indicate that attainment of native-like L2 VOT may not be possible (Flege, 1984, 1986, 1995; Flege & Hillenbrand, 1986). Instead, an intermediary VOT is acquired, sometimes even expressed across both languages the bilingual speaks. On the other hand, child bilinguals, particularly younger children, can acquire native-like L2 VOTs, with age of acquisition affecting their likelihood of attaining native-like VOT production. VOTs are also of interest for their sociophonetic variation, which allows speakers to demonstrate cultural membership and affiliation. Following attainment of these features, language learners are able to employ culturally associated sociophonetic variation as well. This study considers both the phonetic acquisition and sociophonetic variation of VOTs by 11 y.o. Korean-English bilingual children who have had 2 years of immersive L2 English exposure.

1.1. VOT

VOT is a sub-segmental feature of stops, measuring the time between the burst and the onset of vocal fold pulsing (Lisker & Abramson, 1964). VOT can be a positive or negative value, dependent upon whether the pulsing of the glottis starts before the burst (negative) or after the burst (positive). Voiced stops typically have a negative VOT. Fully voiced stops have a VOT that occurs simultaneously with the onset of the stop. Partially voiced stops have a VOT that begins during the closure of the consonant. For example, English phonemes [b], [d], and [g] in onset position are partially voiced stops. Aspirated voiceless stops, when followed by sonorants, have a positive VOT. The length of VOT indicates the strength of aspiration. Unaspirated voiceless stops have a VOT near zero. These stops have inaudible offsets because the voicing of subsequent sonorants begins at the release of the stop.

Languages vary greatly in VOT. Lisker and Abramson (1964) found three common length distinctions across languages: voiced (VOT of approximately −90ms), voiceless unaspirated (VOT of approximately +10ms), and voiceless aspirated (VOT of approximately +75ms). Cho and Ladefoged (1999) also indicated that languages cluster around four levels of VOT, which fall into the following ranges of timing: fortis (also called tenuis) typically has a VOT of 20-30 milliseconds (ms), short-lag (also called lenis or weakly aspirated) has a VOT of 50-60ms, moderately aspirated [kʰ] has a VOT of 80–90ms, and strong aspiration has a VOT over 100ms. Either moderately or strongly aspirated VOT may also be called long-lag. Languages may make phonemic or allophonic distinctions using VOT, but rarely use multiple VOT contrasts as phonemic distinctions.

The participants in this study speak English and Korean, languages which differ by VOT for stops. English uses voicing as a phonemic distinction and aspiration as an allophonic variation. The English
The inventory includes the following voiced stops: /b/, /d/, and /g/. English VOT for voiceless stops differs allophonically by position, with aspiration always occurring in word-initial position. English /p/, /t/, and /k/ VOTs are always positive in word-initial positions, as these are always aspirated, starting well after the burst (Lisker & Abramson, 1964). Word-initial aspiration is always long-lag in English. Yang (1993) reports word-initial voiceless stops have an average of 77ms for [p] (range: 47ms-142ms), 95ms for [t] (range: 54ms-193ms) and 88ms for [k] (range: 45ms-131ms).

On the other hand, Korean uses aspiration as a phonemic distinction and voicing as an allophonic variation in medial positions with VOT variations. In addition to the unvoiced stops of /p/, /t/, and /k/, the Korean inventory sometimes includes the palatal stop /c/ with the allophone [ɟ], but disagreement exists about the manner of this consonant (Shin, Kiaer & Cha, 2012). These consonants were not further considered since they are not shared across languages. In Korean, VOT is a three-way phonemic distinction using aspiration. These may be represented as [p], [p’], and [pʰ] for /p/ in fortis, short-lag, and long-lag, respectively. Yang (1993) reported Korean speakers producing word-initial /p/ with an average VOT of 12ms for [p] (range: 7ms-17ms), 35ms for [p’] (range: 14ms-65ms), and 75ms for [pʰ] (range: 39ms-124ms). Korean participants produced word-initial /t/ with averages of 16ms for [t] (range: 8ms-29ms), 39ms for [t’] (range: 15ms-81ms), and 75ms for [tʰ] (range: 41ms-113ms). For word-initial /k/, Korean participants produced word-initial /t/ with averages of 24ms for [k] (range: 7ms-40ms), 53ms for [k’] (range: 31ms-66ms), and 95ms for [kʰ] (range: 64ms-182ms).

The VOTs of voiceless stop consonants in English and Korean overlap (Cho & Ladefoged, 1999). For a side-by-side comparison of VOT averages and ranges across languages, see Figure 1. As can be seen above, the average English VOT of stops corresponds most closely to Korean long-lag VOTs. The English [p] VOT range overlaps with both longer productions of Korean [p’] but can be longer than shorter productions of [pʰ]. The average and range of the English [t] VOT can be considerably longer than the Korean [tʰ], by 20ms and 80ms respectively, but can also be produced shorter than the longer
productions of Korean [t’]. The average and range of the English [t] timing is longer than other English voiceless stops, as well as all realizations of the Korean [t]. By contrast, the average and range of the English [k] VOT is shorter than the [kʰ], by 7ms and 51ms respectively. The English [k] VOT range also overlaps with the Korean [k’].

1.2. Effects of language contact on VOTs

As with many aspects of grammar, L2 learners are prone to transfer their L1 phonology, their VOT in this case, to their L2 during the acquisition process. L2 speech errors commonly occur during production of phonemes that either do not occur in the L2 or are realized differently in the L2 from the L1 (Flege, 1985). However, unlike many other aspects of grammar, neither higher proficiency nor greater exposure seem to result in more native-like VOTs for most learners. Flege (1995) suggests that the difficulty in acquiring different realizations of VOT may be the result of L1 effects on categorical perception from the L1. Flege and Hillenbrand (1984, 1986) found that L1 English speakers in their study could not generate weakly aspirated French /t/ that approximated L1 French VOTs, no matter how much learners increased in exposure or French proficiency. Although Flege (1986) implied that adult learners would never match native-VOTs, Flege and Hillenbrand (1987) found that learners in their study eventually approximated L2 phonetic norms by exhibiting intermediary VOTs. The researchers argued that these intermediate VOTs indicate language learners can perceive some differences between the two languages’ VOTs. To explain the mechanism that pushed learners to assimilate the VOTs across their two languages, Flege (1995) introduced the Speech Learning model, which proposed that adult learners extend their perception of natively acquired categories to include the L2 categories in an “equivalence classification.” Thus, L1 English L2 French learners in Flege and Hillenbrand (1984, 1986) were thought not to have established new phonetic categories for weakly aspirated French /t/, but instead were thought to have judged the L2 French /t/ to be a realization of their L1 English /t/. Even when considering adult onset bilinguals judged to be native-like, most individuals measured in research to date do not produce VOTs that approximate native-speakers’. For example, Stölten, Abrahamsson, and Hyltenstam (2015) measured the L2 Swedish VOTs of voiceless stops by 41 Spanish early and late near-native-speakers. Researchers categorized early acquisition as beginning between 1 and 11 y.o. and late acquisition as beginning between 13 and 19 y.o. All learners were judged by monolingual Swedish speakers to be native-speakers and had lived in Sweden for 10 or more years. Measuring their word-initial Swedish voiceless stop consonant VOTs when producing isolated, real words, researchers found that learners’ mean VOT measurements were within the range produced by native-speakers 74% of the time for early bilinguals and only 40% for late bilinguals. Considering that all participants were judged to be native-speakers of their L2, these late bilinguals’ results particularly highlight the improbability of later onset bilinguals from ultimately obtaining native-like L2 VOTs.

In addition to L1 transfer, assimilation of VOTs by groups of speakers in settings of language contact has been found to affect languages, suggesting that changes to VOT result from bilinguals merging their VOTs across both languages. For example, in New Zealand, Maori stop consonant VOTs showed a measurable increase in length due to contact with New Zealand English, which had longer timings (MacLagan et al., 2009). Conversely, some speakers have shown a pattern of condensation in the VOTs of their native-language. Heselwood & McChrystal (1999) concluded that Punjabi participants collapsed the distinctions of stop VOT from three to two realizations in their Punjabi as an effect of English. While these findings suggest that the acquisition of L2 VOTs by adult language learners may progress beyond wholesale transfer from the L1, ultimate attainment of native-like L2 VOTs has not been found for most bilingual adults.

1.3. Child development of L2 VOTs

Since it has long been established that the attainment of native-like L2 VOTs is less likely in L2 acquisition following puberty, many researchers have investigated whether child bilinguals can attain native-like L2 VOTs. Investigations have considered a number of ages of second language acquisition (AoA). As was shown by Stölten et al. (2015), earlier AoA corresponds to a higher probability of ultimate attainment. One case study examined two sequential bilingual Spanish–English-speaking siblings, aged 4;0 and 7;0, on VOT values of word-initial stops (Bond, Eddey & Bermejo, 1980). When
comparing the timing of stops in single words utterances across both English and Spanish, both children appeared to produce VOT values that were language specific. The authors concluded that the native-like attainment of Spanish and English stop VOT's proscribed interlanguage influence across this feature.

Nevertheless, many studies of VOT development (Deuchar & Clark, 1996; Kehoe, Lleó & Rakow, 2004; Fabiano-Smith & Bunta, 2012) have found differential acquisition patterns when comparing simultaneous and early sequential bilingual children to monolingual children. (AoA between 0 and 5 y.o., which includes simultaneous and early sequential bilingual children, is termed early bilinguals for the purposes of this study.) Such studies often attribute these differences to interlanguage influence or language dominance. For example, Deuchar and Clark (1996) determined that a simultaneous Spanish-English child in England showed differential development from monolingual English-speaking children. By age 1;11, the child had developed voicing distinctions, primarily in her language dominance. For example, Deuchar and Clar the purposes of this study.) Such studies often attribute these differences to interlanguage influence or slow development from monolingual Spanish children in their development of Spanish VOT distinctions, matching findings in monolingual English peers. These children also did not seem to differ markedly from monolingual Spanish children in their development of Spanish VOT distinctions, matching findings in Deuchar and Clark (1996).

In a related study, Kehoe and colleagues (2004) investigated the VOT of 4 German-Spanish bilingual children between 2 and 3 years of age living in Germany. At the end of the study, not all of the children had fully developed VOT contrasts in German. One child made no VOT distinction between German and Spanish. Another had not yet acquired target-like VOTs in either language, although they had acquired a distinction between their German and Spanish. The other children in the study did acquire target-like German VOTs by the end of the study. Similarly to Deuchar and Clark’s (1996) findings, the children studied by Kehoe et al. (2004) seemed to demonstrate a pattern of delayed stop consonant VOT in one language (German). Also similarly to Deuchar and Clark (1996), Kehoe et al. (2004) highlighted that the children’s patterns of bilingual development were not uniform across languages, but suggested that interlanguage influence as well as language dominance may have affected VOT acquisition. Such research demonstrates that while many early bilingual children develop native-like VOTs, interlanguage influence may affect early development. Whether it is more probable that an early bilingual child will develop native-like VOT in both languages or whether this development pattern will ultimately result in a delay or even lack of native-like VOT attainment in one or both languages is not completely certain.

The picture is more convoluted for bilingual children acquiring between the ages of 5 and 11. While bilingual children of this age range are often not considered late bilinguals, for the purposes of this study, sequential bilinguals between the ages of 5 and 11 are termed “later bilingual children” to distinguish from sequential bilinguals with earlier onsets of acquisition, such as those mentioned above, in whom a native-like VOT developmental trajectory is often observed and sequential bilingual children who begin second language acquisition after puberty, such as those in Stölten et al. (2015). In a study that included later bilingual children, Heselwood and McChrystal (2000) further explored English-Punjabi bilingualism. Not only were bilinguals’ Punjabi VOT affected, but 10 y.o. English-Punjabi children produced more voicing lead in English voiced stops than did monolingual English children, indicating that the children did not produce native-like VOTs in either language. In another study considering later bilingual children, Khattab (2000) measured VOT values for three Arabic-English children. The children, 5, 7, and 10 y.o., produced VOT stop values similar to monolingual controls in English, both in weaker and stronger aspiration (allophonic contrasts). In Arabic, the children failed to produce both contrasts, although the monolingual Arabic children only inconsistently produced these same contrasts, known to be present in adult speech. The author concluded that the bilingual children’s VOTs did not differ markedly from English-speaking monolingual children’s, but the 5 y.o. and 10 y.o. children’s Arabic VOTs showed differences from monolingual Arabic children’s. Khattab (2000) theorized that the 10 y.o. child’s differences from monolingual Arabic children was due to minimal exposure to Arabic at a young age. Similarly, Yavaş (2002) reported that 7 y.o. Spanish-English bilingual children
participants produced unstable VOTs in both languages. Two of the 10 children studied failed to differentiate the VOTs between their languages. However, the children had received only one to two years of consistent English exposure since entering school.

For later sequential bilingual children, it seems as though both AoA and exposure may affect development of native-like L2 VOTs. However, when considering vowels, many researchers (Flege, MacKay & Piske, 2002; Sebastián-Gallés, Echeverría & Bosch, 2005) have shown that children as old as 13 y.o. can develop native-like L2 VOTs. For AoA during later childhood, many adult bilinguals demonstrated native-like VOT measures. However, later bilingual children when measured less than 10 years after AoA often demonstrated intermediate VOTs. L1 interference should not necessarily prevent native-like VOTs in the L2, depending on such factors as L2 exposure and dominance.

1.4. Sociolinguistic markers in L2

Besides being influenced by L1 transfer and interlanguage influence, research has also demonstrated that VOT may be affected by individual social choices. Social choices are indexed through variation to phonological feature at all levels, including supersegmentals, although speakers need not necessarily be conscious of the change. Changes to VOT have been used by speakers to index different generations (Takada & Tomimori, 2006), genders (Robb, Gilbert & Lerman, 2005; Oh, 2011), sexualities (Podesva, Roberts & Campbell-Kibler, 2002), and cultural affiliations (Scobie, 2006). For example, Scobie (2006) noted regional variation of Shetland Island English VOT, which was noticeable between heritage islanders and non-heritage islanders, indicating group membership.

Bilingual speakers are known to produce social markers to their L2 speech if they have acquired pragmatic competence in the L2 as well as the L2 structure to be used (Bialystok, 1993). Pragmatic competence is defined by Bialystok (1993) as: (1) the speaker’s ability to use language for different purposes; (2) the listener’s ability to get past the language and understand the speaker’s real intentions (e.g. indirect speech acts, irony and sarcasm); (3) the command of the rules by which utterances come together to create discourse. The participants in the present study were judged by the researcher to have pragmatic competence, as determined by their ability to make linguistic register and style choices per listener (i.e. their parents, versus me) and by occasion (speaking to me with different register during meals versus English lessons), as well as their ability to make jokes in English. If the children have acquired L2 VOT, they should be able to show cultural affiliation through phonetic variation of their VOTs.

To sum up, adult acquisition of an L2 will likely result in incomplete attainment of native-like stop consonant VOTs (Flege, 1991, 1995), but early bilinguals have often been observed to attain native-likeness. Later bilingual children may or may not have mixed success of L2 VOT acquisition. The ability to determine whether such children get ultimate attainment or display interlanguage influence is muddled by prior studies’ categorization of such children with earlier bilingual children. Thus, it is unclear whether 11 y.o. L1 Korean L2 English speakers with 2 years of immersion would have developed native-like L2 VOTs. Further, VOT may be used to index cultural affiliations. Since the ranges of VOTs in the participants’ languages partially overlapped, they might use this feature to index the affiliated cultures. Accordingly, this study considered whether this feature was manipulated by cultural index.

2. Study Questions

The data in Yang (1993) and the overlap of word-initial English and Korean voiceless stops provides the basis for several possible acquisition patterns by Korean learners of English. One possible pattern would be for learners to automatically transfer the Korean phoneme most closely matching the English phonemic production. For all three voiceless stops, the Korean phonemes closest to the English segment VOT are the long-lag. A different possible acquisition pattern would be for learners to develop an English segment based upon all overlapping Korean phonemes. An English segment based on combining the short-lag and long-lag Korean [p] or [t] would likely be shorter than a native-English speaker’s [p] or [t]. However, the combination of short-lag and long-lag Korean [k] could be both shorter and longer than a native-English speaker’s [k]. Based on Korean speakers’ ability to distinctly perceive fortis stops from short-lag and long-lag consonants, it is not expected that a learner would transfer Korean fortis [p], [t] or [k] to their English [p], [t] or [k].
After initial acquisition stages, child bilinguals show development of L2 VOTs beyond mere L1 transfer. Bilingual children of various AoA and exposure have been seen to produce intermediate L2 VOTs rather than native-like L2 VOTs. Interlanguage influence and L1 dominance are implicated as possible causes for non-native-like VOTs, beyond mere delays in development. Regardless, the paucity of studies distinguishing later bilingual children from earlier bilingual children means that strong predictions about the development of voiceless stop VOTs by 11 y.o. bilingual children are not possible. Since the children included in this study are in an L2 immersion environment, primarily using English in school and socially, they may show L2 English VOTs that more closely approximate monolingual English children. However, with an AoA of 9 y.o. and 2 years of exposure, the children may also show a pattern of intermediate attainment of VOTs between native-like and the transfer patterns predicted above.

If a child has both pragmatic competence and native-like acquisition of the appropriate feature, they may index their cultural affiliations through the VOTs. The bilingual children participating in this study have been determined to have L2 pragmatic competence. As such, they may be able to index American or Korean culture in words that are affiliated with either culture through the aspiration of the word-initial stops of culturally-affiliated words. This study considered whether the children produced different VOTs depending upon the cultural index of the word. Their VOTs for American cultural indexes would then be expected to more closely match native-English speakers’. For Korean indexes, their VOTs would be expected to less closely match native-English speakers’, at least by average. One possibility is the VOTs of Korean indexes would be outside the range of native-speaker-like L2 VOTs and more consistent with their L1 VOTs. For /p/ and /t/, this would mean that averages would be shorter. However, for /k/ this could mean that Korean indexing /k/ VOTs are either shorter or longer than American indexing VOTs, since Korean short-lag VOTs can be shorter than English, but Korean long-lag VOTs can be longer than English. It is possible a child with intermediary VOTs can demonstrate cultural indexes through their VOTs. In that case, it would be far more difficult to predict what pattern would differentiate their American index from their Korean indexes. However, a child without native-like L2 VOTs might not be able to contextually alter their L2 VOTs, in which case no significant differences between tokens indexing American and Korean cultures should be present. Based upon these considerations, this study seeks to answer the following questions:

(1) Will 11 y.o. Korean-English bilingual children produce native-like English VOTs in word-initial voiceless stops?
(2) Will cultural affiliation of words motivate changes in the VOTs of word-initial stops in bilingual children with native-like phonological proficiency?

Of further interest are possible differences that may result from the identity of the participants as twins. Little research has been conducted on speech differences between normally developing twins. Myriad studies have investigated twins’ speech as early indicators of abnormalities (Rutter, Thorpe, Greenwood, Northstone & Golding, 2003; Levy, McLaughlin, Wood, Hay & Waldman, 1996; Caspi et al., 2004). Studies of very young (2 y.o.) normally developing twins indicates that individuation in speech can occur, which may be due to environmental factors (McEwen et al., 2007). While studies of psychological development have more thoroughly considered how being a twin affects emotional development and wellbeing (Thorpe, 2003), no studies on the impact of being a twin on normal language acquisition or development were revealed through a thorough literary review.

3. Methodology
3.1. Participants

This case study considered two sequential bilingual Korean-English children. The children selected for this study were 11 y.o. twin sisters, born in South Korea. Broadly, the language background of the girls was equivalent. They are native-speakers of Korean with Basic Interpersonal Communicative Skills (BICS) in English (meaning that they can effectively communicate in a variety of settings, but not necessarily academic settings). The participants, *HEK and *HJK (pseudonyms), were being tutored by the researcher in Cognitive Academic Language Proficiency (CALP) to improve their English grammar and writing. As assessed by native-speaker judgements, they have near-native-like pronunciation, if not
necessarily of their English voiceless stop consonant VOTs. The girls lived in the U.S. for about 2 years, attending elementary school in South Carolina and returned to South Korea shortly after the completion of the study\(^1\). They also held many friends and hobbies in common and participated in the same extracurricular activities.

Although the girls were very similar in a number of ways, it was clear from observational notes that the twins had distinct personalities. They were extensively observed by the researcher over a period of 10 months. While they were friends with each other and maintained overlapping friend groups, they had different interests, hobbies, and non-mutual friends. In particular, one twin demonstrated more extroversion and a greater interest in American topics. As a result of their desire to show individualization, the participant children should produce different VOTs from one another to define themselves as independent individuals or to index their Korean-ness to varying degrees.

3.2. Stimuli

Stimuli were made up of lexical items that were deemed to have a strong cultural connotation to either Korea or America, as determined by the researcher through prior experience with the participants, as well as through research about Korean-American culture via Youtube videos. An example of a word that had a Korean connotation and began with /t/ was ‘tea,’ as Koreans strongly associate tea and other food items with their Korean culture. An example of a word that had an American connotation and began with /t/ is teacher, as the participants associated their current U.S. schooling with American culture. Stimuli were selected by both general and participant specific cultural affiliations, for a total of 19 lexical items. For a complete list of lexical stimuli, see Table 1.

### Table 1

**Lexical tokens by cultural index and word-initial stops**

<table>
<thead>
<tr>
<th></th>
<th>American</th>
<th>Korean</th>
</tr>
</thead>
<tbody>
<tr>
<td>/p/</td>
<td>pancake</td>
<td>pancake (referring specifically to pajeon)</td>
</tr>
<tr>
<td></td>
<td>parks</td>
<td>pickled</td>
</tr>
<tr>
<td></td>
<td>posters</td>
<td>k-Pop</td>
</tr>
<tr>
<td></td>
<td>people</td>
<td>Korean people</td>
</tr>
<tr>
<td>/t/</td>
<td>teacher</td>
<td>tutor</td>
</tr>
<tr>
<td></td>
<td>tan</td>
<td>tired</td>
</tr>
<tr>
<td></td>
<td></td>
<td>tea</td>
</tr>
<tr>
<td>/k/</td>
<td>cookie</td>
<td>k-pop</td>
</tr>
<tr>
<td></td>
<td>candy</td>
<td>Korean people</td>
</tr>
<tr>
<td></td>
<td>South Carolina</td>
<td>Korean drama/k-drama</td>
</tr>
<tr>
<td></td>
<td>American Culture</td>
<td></td>
</tr>
<tr>
<td></td>
<td>kitkat</td>
<td></td>
</tr>
</tbody>
</table>

All stimuli were stressed nouns beginning with word-initial voiceless stops followed by a vowel. Stimuli were grouped by onset consonant, /p/, /t/ and /k/, and by cultural affiliation of the word. Ten words associated with American culture were selected, with an 11th word, “kitkat,” identified and used _in situ_. Ten words associated with Korean culture, including items repeated in /p/ and /k/ groups, were used; some items were used for more than one stop consonant. Chosen lexical items were used in planned elicitation questions (Thomas, 2002, 2010). For example, one planned question was: “What is your favorite Korean drama?”

\(^1\) Possible effects of Southern dialect were not investigated. Recent research (Lide, 2014) has shown that southern female youth acquire less southern accent.
3.3. Materials and procedure

Participants were recorded on a Dell Inspiron 15 laptop in Windows OS and were recorded using Praat software, digitized at 44100 Hertz. Recordings were completed by the researcher in the participants’ home. Parents were present but not involved in the recording process. The stimuli were presented to participants in an elicitation task. Follow-up questions were asked if the target stimulus was not part of the response utterance. One-word responses were also not accepted, since these may not necessarily trigger the cultural index sought. The girls were interviewed one at a time, with the other child not present in the room during the time of the interview.

3.4. Analysis

Results were coded by word (lexical stimuli), by consonant (/p/, /t/, and /k/) and by speaker (HJK and HEK). Lexical items were coded as having a Korean cultural index or an American cultural index, as described above. VOT measurements were taken for each stop consonant of interest in the target words. VOT was measured from the last burst to the beginning of glottal pulses using Praat software (c.f. Cho & Ladefoged, 1999). Time was recorded in milliseconds as positive or negative values. However, since the class of voiceless stops was under investigation in this study, there were no instances of negative VOT. These results will be compared to native-English speaker VOTs presented in Yang (1993) to determine whether the child has attained native-like L2 VOTs. One analysis for determining whether participants’ VOTs were native-like is to compare their averages to native-speaker ranges, following Stöltgen et al. (2015). However, this analysis may not be as conclusive for Korean learners of English, whose full phonemic inventory includes English averages. So, participants’ ranges were also compared to native-speaker ranges. To compare VOT by consonant across speaker and cultural index of words (American or Korean), results of the coding and measurements were analyzed using a Generalized Linear Model.

4. Results and Discussion

4.1. Native-like-ness of L2 VOT acquisition

The recordings for the interview produced over 30 minutes of recording and 101 tokens for analysis from the target items, with 29 total /p/ tokens, 31 total /t/ tokens, and 41 total /k/ tokens. All included tokens occurred intrasententially, in stressed nouns (see Table 1). HEK produced 32 total tokens, and HJK produced the other 69 tokens. Utterances that were whispered, were mumbled, or used creaky voice were excluded, since these types of speech affect VOT analysis. Overall, the average /p/ VOT for both participants was 57ms. The average /t/ VOT for both participants was 94ms. The average /k/ VOT for both participants was 86ms. While the average and range of VOTs for these children varied from those of native-English speaking children in Yang (1993), t-tests did not result in any significant differences between any of the participants’ consonants and native-speakers’ consonants.

For each child, word-initial voiceless stop VOTs were compared to English and Korean VOTs reported by Yang (1993). HEK produced only 32 tokens included in analysis. Her average word-initial [p] VOT was 83ms (range: 44ms-141ms), her average [t] VOT was 119ms (range: 53ms-248ms), and her average [k] VOT was 96ms (range: 43ms-179ms). This included two outlier tokens of [t] (241ms and 248ms), which were both greater than 2 standard deviations longer than the longest /t/ VOTs reported in Yang (1993). HEK also produced nine tokens that were excluded from measurement. Her sister HJK produced considerably more measurable tokens than her sister, with 69 total tokens, and no outliers or excluded tokens. HEK’s average word-initial [p] VOT was 47ms (range: 13ms-131ms), her average [t] VOT was 78ms (range: 25ms-143ms), and her average [k] VOT was 81ms (range: 32ms-151ms). The small number of tokens, the outliers, and the excluded tokens might be explained by the fact HEK was interviewed after HJK, after their bedtime, and appeared to be visibly sleepy.
Participant children’s word-initial VOT production

Figure 2. VOTs of voiceless stops in word-initial positions of participants HEK and HJK are compared to the English and Korean speakers VOTs reported by Yang (1993). All realizations of individual segments are depicted in the same colors as Figure 1.

For a visual comparison of reported VOT averages and ranges across both languages to each child’s English VOT average and range for the consonants in consideration, consider Figure 2. As is shown, both HEK’s and HJK’s average voiceless stop VOTs were within the range of monolingual English children’s in Yang (1993), although HJK’s average [p] VOT was only marginally so. HEK’s [p] minimum, maximum, and average VOTs closely mirrored native-speakers’ in Yang (1993), differing by at most 5ms for the three measurements. Her minimum [t] timing approximated monolingual English children as well, and, excluding the two outliers, her longer [t] VOTs were within the maximum of those by the native-speakers. Her average [t] VOT was longer than native-speaker children and also the furthest from the native-speaker average compared to her other stops. Similarly to her other minimum timings, HEK’s minimum [k] timing was very close to native-speakers. However, HEK’s maximum [k] VOT was 48ms longer than native-speakers’. Her average was also longer than native-speakers’, but only by 8ms. Overall, when excluding HEK’s outlier tokens, her minimum, maximum, and average stop VOTs largely mirror native-speaker children’s stop VOTs. Particularly for her [p] and [t], these patterns suggest that HEK may have acquired native-like L2 voiceless stop VOTs. The notable exception is her maximum [k] VOT, which is her only measurement more than 10ms different from native-speakers. This measurement seems to pattern most similarly to the long-lag maximum noted by Yang (1993). In fact, HEK’s maximum and average English [t] VOT were closer to the Korean long-lag timings, differing from native-speakers’ by 3ms and 1ms, respectively.

HEK’s sister HJK showed a different pattern to her word-initial voiceless stop VOTs. With one exception, all of HJK’s minimum, maximum, and average VOTs were lower than native-speaker children’s as reported by Yang (1993). HJK’s minimum [p] VOT was 34ms lower while her maximum was only 11ms lower. Her average [p] VOT was 30ms lower than native-speakers’ average. HJK’s [t] patterned similarly, with differences of 29ms, 50ms, and 17ms, respectively, between her minimum, maximum, and average VOTs and native-speakers’. Only HJK’s [k] had a different pattern, with a minimum VOT 13ms lower than native-speakers’ and maximum 20ms higher than native-speakers’. By comparison, her average [k] VOT was only 7ms lower. While not significantly different from native-speaker voiceless stop VOTs, HJK’s VOTs do not appear to pattern well with native-speakers’ VOTs. HJK’s VOTs also appear to have significantly longer ranges than Korean long-lag VOTs. However, HJK’s range of English VOTs resembles a combination of Korean short-lag and long-lag VOTs. For example, the combined range of Korean short-lag and long-lag /p/ VOTs reported by Yang (1993) is
14ms to 124ms. When compared to HJK’s [p] VOT range (13ms-131ms), there are only differences of 1ms and 7ms, respectively. Likewise, the combined range of Korean short-lag and long-lag /t/ VOTs (15ms-113ms) is more similar to HJK’s [t] VOT range (25ms-143ms) than the English [t] VOTs (54ms-193), with smaller differences of 10ms and 30ms. The range of Korean short-lag and long-lag /k/ VOTs when combined (31ms-182ms) further supports this comparison: the differences from HJK’s range are 1ms and 31ms, respectively.

Overall, HEK seemed to have acquired native-like L2 VOTs, although strong conclusions about her development are not possible given the limited number of tokens that she produced. Both her average timings and VOT ranges (excluding outliers of 2 /t/ VOT tokens) show that she well approximates native-speakers. While the difference between her [k] VOTs and native-English speakers’ VOTs reported in Yang (1993) was not significantly different, these similarities seem to indicate that HEK’s /k/ VOT may still be in transition from L1 transfer.

On the other hand, HJK did not seem to produce VOTs that resembled native-speakers’ as closely. On average, HJK tended to have earlier VOTs than native-speakers or her sister. Instead, her VOTs appeared to more closely approximate one of the predicted possible L1 Korean transfer patterns (the affixing of the Korean short-lag and long-lag VOTs) than native-like English. Notably though, HJK’s VOT ranges are all intermediary between the English and Korean ranges. While HJK’s current VOT values can be interpreted as incomplete attainment following Flege (2005), it is also probable that she is still in the process of acquisition. Given that she is estimated to have roughly the same L2 language background that provided her twin with exposure adequate to acquire native-like L2 VOTs, it is probable that HJK’s L2 VOT is in transition. With additional exposure, she may acquire the native-like L2 VOTs that her twin appeared to have.

Most surprising of these results is the ostensibly different development of HEK and HJK. While HEK seemed to have developed native-like L2 VOTs, HJK seemed to have L2 VOTs in transition. Comparing the acquisition patterns of the twins, it is clear that there was variability in their VOT usage at this point during their acquisition. In fact, there is a significant difference between the average VOT for the /p/ consonant of HEK and HJK, with standard deviations of 34ms and 45ms, respectively. HEK and HJK showed no significant differences in /t/ or /k/. While these differential patterns were unexpected, the small number of tokens produced by HEK make further comparisons of the twins problematic. Besides the outlier values of her [t] timings, the small number of tokens both in total and for cultural indexes made inferential statistics unworlable. In total, HEK only produced 4 [p] tokens with an American index and 4 [k] tokens with a Korean index. Therefore, further analyses focus on HJK’s VOT values.

4.2. Sociophonetic VOT variation by cultural context

A generalized linear model (with a model significance of p<.001) was used to predict consonant VOTs across cultural contexts for HJK. Consonant type was a strong predictor of VOT (p<.001), which is expected given the different average VOTs for these consonants in English. Tokens from words that indexed American culture are hereafter referred to as “American indexes,” while tokens from words that indexed Korean culture will be referred to as “Korean indexes.” For American indexes, HJK’s [p] VOT had an average of 59ms (range: 22ms-131ms). Her [t] American indexes had an average VOT of 67ms (range: 25ms-143ms), and her American indexes with initial [k] had an average VOT of 73ms (range: 52ms-106ms). By contrast, her [p] Korean indexes had an average word-initial VOT of 40ms (range: 13ms-87ms). Her average [t] VOT for Korean indexes was 104ms (range: 87ms-122ms), and her average [k] VOT for Korean indexes tokens was 84ms (range: 32ms-151ms).

Figure 3 shows the ranges of HJK’s consonant aspiration timings for Korean and American culturally-affiliated words. As is evidenced above, HJK’s Korean indexes showed a greater variability, particularly for the /t/ VOT. Her /t/ VOT also showed greater differences across cultural index than did the /p/ onset and /k/ onset. Comparison of HJK’s /p/ timing by cultural index to timing of native-speakers in English and Korean showed a similar variability of minimum, maximum, and average, which was higher in American indexes and English than in Korean indexes and Korean. Contradictorily, HJK’s /t/ timing displayed great variability, with no distinct pattern emerging, though it should be noted that HJK’s American /t/ indexes included two outliers (111ms and 143ms). If the outliers are excluded, HJK’s American index maximum is 94ms. A comparison of these ranges reveals that /t/ cultural indexes
minimally overlap. On the other hand, HEK’s /k/ timings by cultural index patterned again with differences between her L1 and L2. The range of her American indexes was smaller than the range of her Korean indexes, just as the range of English /k/ VOTs recorded in Yang (1993) is smaller than the range of combined Korean short-lag and long-lag /k/ VOTs. The averages also matched this pattern, if considering the Korean long-lag average. HJK’s cultural indexes do appear to show patterns of differentiation. Overall, these differences indicate that HJK might have been displaying socially motivated differences in voiceless stop consonant VOTs that indexed American and Korean cultures. Nevertheless, conclusive judgement of HJK’s cultural indexes is not possible, as t-tests did not reveal a statistically significant difference between cultural affiliation between each consonant.

![HJK's VOTs by cultural index](image)

**Figure 3.** HJK’s word-initial voiceless stop VOTs were divided by words that index American culture and words index Korean culture. For comparison, the ranges and averages of English VOT, and the ranges and averages of Korean short-lag and long-lag VOTs (combined) were included (as adapted from Yang, 1993).

### 5. Conclusions

On the whole, the participants seemed to produce L2 VOTs that had developed beyond mere L1 transfer. Both HEK’s average /p/, /t/, and /k/ average timings and the ranges of her timings resemble native-like L2 VOTs. HJK cannot be said to have developed L2 VOTs that appeared to be as native-like as HEK’s. Although her average timings would indicate native-like L2 VOTs, the minimum timings in her VOT ranges indicate that she still applies L1 timings. A larger number of tokens could have better supported this conclusion. Unfortunately, the participants returned to South Korea shortly after data collection. Further, experimental testing of L2 VOT could have been made separately from tests of cultural indexical choices. Such testing may have involved the participants reading phrase-initial and phrase-medial words with different phonetic conditions. This could provide additional insights about the effects of phonemic environment not investigated in this study.

The second research question—will cultural affiliation of words motivate changes in the VOTs of word-initial stops in bilingual children with native-like phonological proficiency—is inconclusively
supported. While HJK did not demonstrate full attainment yet, she seems to demonstrate sociophonetic variation between words affiliated with Korean culture and words associated with American culture. However, testing of sociophonetic variation by children and bilinguals should separately be investigated to understand how each group can manipulate their phonology to index their cultures. This area of research seems to be ripe for further investigation.

Finally, the twins’ different acquisition showed an interesting result from equivalent exposure. Future studies of twins, particularly language learners, should be conducted to explain these results. Comparing the acquisition patterns of the twins, there seemed to be variability in VOT production at this point in their acquisition process. The diverse results of the twins epitomize the variety of results from research on “later” bilingual children. While further research is needed, these results suggest that later child bilingual participants could acquire fully native-like L2 VOTs. Surprisingly, one twin seemed to demonstrate sociolinguistic variation in her L2 VOTs, despite having intermediate attainment. This interpretation would support an even stronger reading of Bialystok’s (1993) pragmatic competence, in that mere competence could allow sociolinguistic variation of features not fully acquired. These pilot case study results tentatively suggest that language learners can show sociolinguistic variation of L2 VOTs, even during L2 development.

References


