University of Utah
UNDERGRADUATE RESEARCH JOURNAL

VOICE ONSET TIME IN YOUNG SPANISH-ENGLISH BILINGUALS

Student's Full Name
Ellie A. Kaiser (Tanya Flores)
Department of World Languages and Cultures

Literature Review

One of the most fascinating lines of research in the field of linguistics today is the investigation of bilingual speech production and perception. While in many cases the prototypical linguistic subject has been a monolingual native speaker, there is now an increased interest in what bilinguals can tell us about the human capacity for language – even beyond the subfield of Second Language Acquisition. One of the major questions about bilingualism is whether bilinguals effectively function as two monolinguals, or whether they have a coexistent mental system for both languages – or somewhere in between these two ends of the spectrum.

To begin to answer this question, we can look at the pronunciation of bilinguals in both languages. In more technical terms, we can study the phonetics (the physical properties of how sounds are articulated and produced) of their language to understand their phonology (the mental representations of sounds that are the beginning of making meaning from the arbitrary sound-meaning relationship). In particular, we tend to look at situations where a particular sound (or more technically, a phoneme) exists in both of the languages but is produced slightly differently in each. Flege and Hillenbrand (1984) hypothesized that in this case, second language (L2) learners would be unable to produce the sound authentically (meaning like monolinguals) in their L2 because the similarities in the sounds would interfere with their ability to distinguish the differences in the sounds (Flege and Eefting, 1987). Flege and Eefting (1987) explained some possibilities of how bilinguals may mentally map the sound systems of the two languages: they may have a "merged" system, meaning that they have a separate mental category for the sound in each language, or they may have a "coexistent" system in which each sound falls into the same mental category. Within the coexistent system, they may have mental rules which change the pronunciation in each language, but ultimately the sounds would have the same underlying representation. They hypothesize that a merged system (two separate phonetic categories) is beneficial to helping bilinguals produce the sounds in an authentic manner.

For English and Spanish bilinguals, some of the most commonly studied sounds are the unvoiced stop consonants /p t k/ and their voiced /b d g/ counterparts, respectively. Sounds are called voiced when the vocal folds vibrate during their production, and vice versa. All of these sounds are called stops because they involve a complete obstruction of airflow followed by a release burst. Phonologically (according to our mental maps of what sound differences are meaningful), both English and Spanish have a voiced and unvoiced stop consonant category, but phonetically (in terms of the actual production of said sounds), the difference is articulated quite differently. Historically, linguists have looked at different acoustical cues to find what property of the sound it is that actually differentiates these categories. Glottal buzz, presence of aspiration, and amount of articulatory force are some of the acoustic properties that were studied, but these were inconsistent and hard to study. In 1964, Lisker and Abramson were the first to explore Voice Onset Time (VOT) as a measurement of the voicing distinction (Banov, 2014). VOT is the time between the release burst and the beginning of the pronunciation of the following sound (Zampini & Green, 2001). Since the 1970s, VOT has been the most common way of measuring the voicing distinction in bilinguals (Banov, 2014).

In English, /p/, /t/, and /k/ are produced as long-lag stops: they have VOTs of more than 30-35 ms and are said to be "aspirated". Their voiced counter parts /b/, /d/, and /g/ are produced as short-lag stops with VOTs less than 35 ms. In Spanish, however, the unvoiced /p/, /t/, and /k/ are short-lag stops with VOTs less than 35 ms, while the voiced /b/, /d/, and /g/ are prevoiced with a negative VOT – vibration of the vocal folds begins before the release burst. Therefore, in terms of VOT, there is a phonetic overlap between English /b d g/ and Spanish /p t k/ (Zampini & Green, 2001).

Additionally, the idea of a "critical period" is important for understanding bilingual language. It is the idea that humans are only capable of learning a language to a monolingual standard by a certain age (often around the onset of puberty). Not only is there a lot of argument

around when and why the critical period ends, but there is a lot of evidence for rejecting this theory (Flege & Eefting, 1987), leading to many scholars arguing instead for a "sensitive" period. In their 1987 study on the perception and production of stops by English-Spanish bilinguals, Flege and Eefting compared early childhood bilinguals to late childhood bilinguals and found that age of learning English alone could not account for all of the variation in mean VOT, but they did suggest that learning an L2 at an earlier age may assist in the creation of a merged phonetic system. Thornburgha and Ryallsb (1998) found that bilinguals who learned English after the age of 12 differentiated voiced from unvoiced stops to a greater degree, but that there wasn't a significant correlation between age of learning and the overall mean VOT difference.

Therefore, although the age at which one acquires their second language does have an effect on one's mean VOT and differentiation between voiced and unvoiced stop, it is not the only variable of importance here. Some studies have found significant differences between genders, but the exact effect is inconclusive: Banov (2014) found females to have more monolingual-like VOTs than males in English /t/ and /k/, while Thornburgha and Ryallsb (1998) found that males contrasted voicing to a greater degree than females did. The quality of the L2 input may also have an effect (Flege & Eefting, 1987). Grosjean 1992 (as cited in Zampini & Green 2001) suggested that a failure to control for such factors may be what has lead to such varying results about bilingual production. It has also been shown that even among monolingual English speakers, there are systematic individual differences in mean VOTs (Allen & DeSteno, 2003).

Another factor that has been identified as crucial in bilingual language production is that of language mode – perhaps bilinguals have different mental settings that can be activated by different situations and knowledge about the person or people they are speaking to. Different studies controlled for this factor in different ways, sometimes by starting half of the participants in one language mode and half in the other, sometimes by collecting data for each language on a different way, and sometimes by leading participants to believe that they had been recruited for two completely different studies (Banov, 2014; Flege & Eefting, 1987; Green & Zampini, 2001).

To complicate matters further, Zampini and Green have conducted several studies highlighting the importance of an entirely different acoustic cue for differentiating voiced and unvoiced stops in Spanish: Voiceless Closure Interval (VCI), which is a measure of "the amount of time that elapses between the obstruction of the airflow and its release" (2001, p. 23). While in English, word-initial stops show no difference in VCI (Crystal & House, 1988), VCI alone is enough to distinguish unvoiced and voiced word-initial stops in Spanish (Martínez Celdrán, 1993) (as cited in Zampini & Green, 2001). Because VCI is essentially a measurement of the silence produced by the blockage of air to produce a stop consonant, it is indistinguishable in waveform from other unmeaningful silences. Thus, we can only measure it when it occurs in continuous speech after at least one other word. Zampini and Green's work found that bilinguals and L2 learners may have different degrees of success in manipulating VOT vs. VCI to create the voicing distinction in Spanish. This shows that both the voicing contrast and the question of bilingual phonetic categories is complicated and multi-faceted.

A final important piece of the puzzle is that of language acquisition. While the previous literature has focused on production of the voicing distinction in adult bilinguals, there is also some research on how this production is developed in both languages and in bilinguals at young ages. Most studies agree that the English short vs. long-lag distinction is acquired early – around the age of 2 years (Deuchar & Clark, 1996), while the Spanish short lag vs. prevoiced distinction is acquired later – Macken and Barton (1979, 1980 as cited in Deuchar & Clark, 1996) found that child speakers of Mexican Spanish did not have an adult-like contrast even by 4 years old, but that a contrast in the short-lag range was still possible at this age. Other studies found that by four years old, children *have* developed an adult-like voicing distinction including prevoiced

stops. This difference has been hypothesized to be due to phonological differences in Spanish dialects, and the relatively late acquisition of prevoiced stops is suspected to be due to its low perceptual salience and difficulty to produce (Eilers, Oller, & Benito Garcia 1984).

At the end of his thesis, Ivan Banov (2014) brings up some important questions about the nature of this line of study. First of all, the custom in the field is to compare bilinguals and L2 learners to "native" monolinguals, but how valuable is that standard? It is becoming increasingly recognized that the global standard is *not* monolingualism, and perhaps we should adjust our standards accordingly. Second of all, we know from studies such as Allen and DeSteno (2003) that individual monolinguals have systematically longer or shorter mean VOTs, but we judge bilinguals by the interpersonal averages. This begs the question of what the acceptable ranges of native VOT production are and where bilingual VOTs fit into this. Finally, Banov suggests that it might be useful to have a panel of monolinguals evaluating whether bilinguals' productions sound native-like to see whether any systematic differences we see are significant enough to be perceived by monolingual speakers. While the present study does not seek specifically to address these issues, they are still important to keep in mind as we evaluate the purpose and scope of this work. The crucial distinction to make here is that while we compare bilingual production to monolingual production, we do so to answer interesting questions about how bilingual minds work rather than to judge bilingual differences as inadequacies.

More research is needed to resolve some of the inconsistencies that have been found. While Deuchar and Clark (1996) and Eilers, Oller, and Benito-Garcia (1984) have analyzed the acquisition of these sounds by young bilinguals, they found results that differed from each other and from other studies. Additionally, newer research by Green and Zampini (2001) has brought up many concerns that they believed may have accounted for the great differences in results that have been found: fluency in L1 and L2, age of L2 acquisition, quality of L2 input, and language mode. They also looked at the additional acoustic factor of Voiceless Closure Interval, but this hasn't been analyzed in young children's acquisition. Because there is so much variability in these factors and therefore the results of studies, collecting data from a variety of sociocultural backgrounds and contexts can help add to the picture. The present study takes many of these factors into account and collected language background from the parents to better be able to understand what effect their linguistic experience may have. Additionally, this study will analyze natural speech rather than word lists or word repetition that some past studies have used. Finally, this study takes into account VCI in addition to VOT, as well as the cognate status of a word because recent theories of bilingual systems have suggested that a cognate words may activate both languages' sound systems to a greater extent; Brown and Amengual (2015) found that for Spanish-English bilinguals, words starting with a dental stop (/t/ and /d/) in Spanish that were cognates with an English word were pronounced with more English-like VOTs.

The Present Study

The present work began as part of a larger study analyzing various aspects of the language development of young (3-4 years old) hard-of-hearing Hispanic children who are exposed primarily to Spanish in their homes but are enrolled in English-speaking schools (T. Flores, personal communication July 2017) . Unfortunately, because of difficulties with participant turnover and data collection, the original goal of comparing the pronunciations of hard-of-hearing and normal hearing children was unable to be realized. Instead, data from the "control" group of normal hearing bilinguals will be analyzed here. The previous research on acquisition of the voicing contrast in Spanish-English bilinguals suggests that the ages of 3-4 are critical in the development of the voicing contrast in Spanish, but many contradictory results have been found. Thus, more research is needed on these age groups to begin to understand the complexities of bilingual voicing contrasts and this analysis is valuable even without the comparison to hard-of-hearing groups.

Because VCI can only be measured in non-utterance-initial position in continuous speech, the first data collection didn't have enough tokens so it will not be analyzed in depth here. Also, there were too few cognate words to make a meaningful comparison so this data will not be discussed extensively here either. Thus, this analysis will measure the VOT of 10-15 tokens of /p t k/ and /b d g/ for six bilingual children. Prevoicing will also be measured both as part of the VOT and separately.

Participants

The participants in this study are six children – four female and two male – attending preschool in the Salt Lake Valley who speak Spanish in their homes and are exposed to English in the classroom and community. Their teacher is a Spanish-English bilingual. They were all between 3 and 4 years of age and had varying degrees of language dominance. None of them had any hearing or language deficits. Participants 13F, 14M, 17F, and 18F all produced data in both Spanish and English. 12F and 19M, however, only produced data in English and Spanish, respectively.

Methodology

Natural speech was collected from the children using "frog stories" (short picture books with illustrations but no words) about various adventures involving a boy and his pet frog. For each language a different story was used, and children were directed in that language to explain what was happening on each page of the book. This data was collected alongside a more structured language elicitation mechanism that was not used for the analysis in the present study. In most cases, data from different languages was collected on different days to avoid the possible conflating effect of language mode.

The program Praat was used to create waveforms and spectrograms and to annotate these onto textgrids. First, an orthographic transcription was made of all the files and the children that produced the most tokens of word-initial /p t k/ and /b d g/ in each language were selected for this study. Four children had at least ten tokens in each language and were thus selected for this analysis. The participant with the fewest tokens was 17F, who had only ten tokens in English. Additionally, all of these tokens were used during the participant's Spanish frog story; their English frog story had no usable tokens. Thus, 17F was the only participant for whom all the data was collected in the same session and with the same frog story. The first 10-15 tokens for each participant were collected. If the child produced a stop as its approximant allophone or some other non-stop form, or if the token was produced in such a highly coarticulated manner that it was impossible to distinguish the consonant, that token was skipped.

Cognate status was determined by the judgments of the author, an L1 speaker of English and advanced L2 learner of Spanish. They were then corroborated by the research advisor, a Spanish-English bilingual. There was some question of the cognate status of the word "gato" ("cat"), which will be discussed later.

In addition to these four participants, two more were chosen who had at least 15 tokens in one of their languages but few or no tokens in the other language. 12F had many tokens in Spanish but none in English; 19M had many tokens in Spanish but very few in English.

VCI was measured only when the token occurred in running speech and there was not an audible pause before the word, as the length of the silence is only significant when it is part of the production of the consonant, not when it is merely a pause in speech. For this reason, tokens that occurred after the filler word "um" were also omitted because the use of a filler word would be expected to preclude a pause (and this is what was observed in many cases).

In some cases, prevoiced tokens were produced with a release burst that moved straight into the vowel. In this case, VOT was reported as a negative number with the same absolute value as the prevoicing. When the consonant was prevoiced but also had a short-lag stop, both

the prevoicing and the VOT were recorded as positive numbers. Thus the unique energy segments can be analyzed separately.

Results

As expected for natural language data from children so young, there was very little data; nevertheless, some tentative patterns emerge. At this point, no patterns in VCI or cognate status could be found. There were very few tokens to begin with, and because VCI is only meaningful in running speech and a very small portion of the words were cognates, neither of these measurements suggested any meaningful patterns. It is also possible that VCI has not yet been acquired by these children as a means for distinguishing their Spanish consonants. While Green and Zampini's work on VCI didn't include any research on acquisition by children, they do suggest that it is perceptively less salient than VOT, so it is possible that children don't begin to utilize this acoustic cue until later.

The following table shows the mean VOTs for each subject in each language for the various consonants. The consonants are paired with their voiceless and voiced counterparts according to place of articulation to make it clear what the consonant-specific voicing distinction looks like. While some of these means are based on as many as 10 tokens, others are an average of only one token, so they should not be taken as robust statistical measures, but merely as suggestions of a larger pattern. Because prevoiced tokens are by convention described with negative VOTs (because voicing begins *before* the release burst), they would not make meaningful averages when combined with positive values. Thus, they were excluded from this chart but indicated by an asterisk. They will be discussed separately later.

Table 1

Results for Mean VOT by Language and Subject

	Mean English VOTs (ms)						**Mean Spanish VOTs (ms)**					
Subject	**p**	**b**	**t**	**d**	**k**	**g**	**p**	**b**	**t**	**d**	**k**	**g**
12F	145.5	15.0	80.5	18.5	60.0	26.0						
13F	43.0	17.5*		14.7	66.5		15.0	*	27.0		26.1	
14M		10.0*	102.0	17.4	94.0		10.4	*		21.0	20.0	
17F	15.9		59.0			34.0	43.0		33.0		38.2	
18F	135.0	13.5		16.3*	14.0	14.0	10.8				17.0	70.0
19M							9.7		13.0		16.5	

*The asterisk indicates a value for which prevoiced tokens were produced but omitted from this chart. Values for the prevoiced tokens are discussed below.

Conclusions

Now preliminary results will be discussed for each subject. To give a preliminary sense of the data, each child's mean VOT by consonant will be compared to the means discussed by Lisker and Abramson's data (as cited in Deuchar & Clark 1996):

Table 2

Lisker and Abramson's Mean VOTs by Language and Consonant

English			
Voiced	/p/ 20-120 ms	/t/ 30-105 ms	/k/ 50-135 ms
Voiceless	/b/ 0-5 ms	/d/ 0-25 ms	/g/ 0-35 ms
Spanish			
Voiced	/p/ 0-15 ms	/t/ 0-15 ms	/k/ 15-55 ms
Voiceless	/b/ -235 - -60 ms	/d/ -170 - -75 ms	/g/ -165 - -45 ms

Subject 12F, who had no Spanish data, appears to have a fairly clear distinction for voiced and voiceless consonants. In each case, her unvoiced consonants are much larger than her voiced consonants. All of her means are within the monolingual English ranges suggested by Lisker and Abramson (1964) except for /p/ and /b/, which are both slightly above the upper range suggested (as cited in Deuchar & Clark 1996).

All of 13F's mean English VOTs were within the ranges suggested by Lisker and Abramson, except for /b/. Interestingly, she did produce some prevoiced /b/ tokens in English, which is not expected for monolingual English. Her voiceless Spanish VOTs were also within these ranges, and she produced the expected prevoicing for /b/. Unfortunately, she didn't produce any tokens of the other voiced Spanish stops to compare. The only place of articulation for which she had both a voiced and voiceless token was the bilabial stop in English. For this pair, the voiceless stop /p/ is much larger than the voiced /b/, showing evidence of a monolingual-like voicing distinction.

All of 14M's English means are within the Lisker and Abramson ranges, except for their /b/ tokens. Like 13F, he did produce some /b/ tokens with prevoicing, which is not to be expected for English. His voiceless Spanish tokens were also within these ranges, but only his /b/ tokens and not his /d/ tokens presented the expected prevoicing. Both his English and Spanish /p b/ pair demonstrate a close-to-monolingual-adult voicing distinction.

While 17F's English /t/ and /g/ means were within Lisker and Abramson's ranges, her /p/ mean was shorter than the bottom bracket of this range – in fact, her average for this consonant was much shorter than those of the rest of the subjects. Her Spanish /p/ and /t/ were both above the top bracket of the range, while her /k/ fell within the range. Unfortunately, she did not produce enough tokens to see how she is producing the voicing distinction at any place of articulation.

18F's English and /p/ and /b/ were both larger than the Lisker and Abramson ranges, her /d/ and /g/ were within the expected range, and the /k/ mean was much shorter than the bottom of the range. Both her /p/ and /k/ tokens in Spanish were within the expected ranges, but her /g/ was much higher than the range and was actually higher than her /k/ average; this means that the voicing distinction actually went in the opposite direction from what we expect. Interestingly, this average was based on only two tokens of /g/, both for the word "gato". One of these tokens had a VOT of 17ms, and the other had a much greater VOT of 123ms. This second VOT was so

large that it sounded much more like “cato”. It is debatable whether this word is truly a cognate, but this very English-like VOT for the Spanish word suggests that some amount of cross-linguistic influence was occurring in the mind of this speaker. 18F produced something close to a monolingual-adult like voicing distinction for the English /p b/ pair, but for the /k g/ pair, there was no difference in mean VOT.

19M, who only produced Spanish data, also only produced voiceless tokens, so it isn’t possible to see how he articulated the voicing distinction in either language. However, his means for /p t k/ were well within the Lisker and Abramson ranges.

An interesting surprise in the data was that most words in English beginning with a voiced dental fricative [ð] (one of the two sounds orthographically represented by “th”) were pronounced by the children as the voiced stop [d]. While this is an accepted variant in many dialects of English, it is most likely not the standard in this community. It is more likely that this is an interlanguage development as the [ð] phoneme has been known to be difficult to acquire and is learned relatively late by monolingual English children. However, it is also possible that the children’s teachers or parents or other caregivers may have this variant in their speech, influencing the pronunciation. If it is the case that this is a stage in their language development but that they are moving toward a [ð] pronunciation, then the pronunciation of this class of words won’t tell us much about the subjects’ underlying phonology. Nevertheless, it still appears phonetically as a [d] token, and because of its overwhelming prevalence, it was decided to include these tokens in the data. At further phases in the larger study, it may be possible to analyze whether these tokens are produced significantly differently from other [d] tokens, which could indicate whether these groups of words truly have the same underlying phonological representation.

Discussion

Despite the small amount of tokens, it is still possible to provide some descriptive patterns in this data. By this age, most of the children appeared to producing some kind of voicing distinction in both English and Spanish. Many produced the characteristically long VOTs of English voiceless stops, but only three of the children produced any prevoiced tokens, and only two children produced these in Spanish, where they were expected. However, the other subjects had very few voiced Spanish tokens to begin with, so it wouldn’t be fair to conclude that they systematically don’t produce prevoiced Spanish stops, although if more data collection shows this to be the case, it would be consistent with some past studies that have found prevoicing hasn’t been acquired by this age. Unfortunately, there were so few voiced Spanish tokens that it isn’t clear whether these children might have a voicing contrast within the short-lag range.

Additionally, there are some anomalies in the data that stick out. For example, 18F’s voiced velar stop /g/ had a significantly longer mean VOT than her voiceless velar stop /k/, which is exactly the opposite of what we would expect. The token that made the mean voiced VOT so high was for the word “gato” (“cat” in English). While this word wouldn’t usually be considered a cognate because it is too dissimilar from the English word, the child’s pronunciation of it sounded distinctly more like “cat-o” (based on my judgments as an advanced L2 learner of Spanish). Thus, this may be an example of some sort of morphological codeswitching that could explain why this token broke from patterns to such a great degree. This goes to show that this data is very messy and involves many confounding factors that will require more data and more analysis to disentangle.

Limitations and Future Directions

The most obvious limitation of this study is that it is merely a preliminary result from the first round of data and was not able to be continued from there. There are very few tokens, especially for the voiced Spanish consonants. Additionally, this paper doesn’t include any results from cognate status, VCI, or the language background information that was collected

from participants' families. Finally, because of time and resource constraints, we were not able to control for language mode as well as we would have liked to.

The larger study of which this work is a part will continue, but because of difficulties with participant retention, the researcher decided to no longer pursue this particular project. Instead, she is directing her attention to other aspects of bilingual stop pronunciation in other groups.

References

Allen, J.S., Miller, J.L., & DeSteno, D. (2003). Individual talker differences in voice-onset time. *Journal of the Acoustical Society of America*, *113*(1), 544-552.

Banov, I. K. (2014). *The Production of Voice Onset Time in Voiceless Stops by Spanish-English Natural Bilinguals.* Retrieved from BYU ScholarsArchive. (4340)

Brown, E. B., & Amengual, M. (2015). 2015. *Studies in Hispanic and Lusophone Linguistics, 8*(1), 59-83. http://doi.org/10.1515/shll-2015-0003.

Deuchar, M., & Clark, A. (1996). Early bilinguals acquisition of the voicing contrast in English and Spanish. *Journal of Phonetics, 24*, 351-365.

Eilers, R. E., Oller, D. K., & Benito-Garcia, C. R. (1984). The acquisition of voicing contrasts in Spanish and English learning infants and children: a longitudinal study. *J. Child Lang*, 313-336.

Flege, J. E., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics, 15*(1), 67-83. http://ezproxy.lib.utah.edu/docview/58190349?accountid=14677

Kaiser, E. A. (2018). *Voice Onset Time in Young Spanish-English Bilinguals.* Unpublished manuscript, University of Utah.

Thornburgha, D. F., & Ryallsb, J. H. (1998). Voice onset time in spanish-english bilinguals: early versus late learners of english. *Journal of Communication Disorders, 31*(3). Retrieved from https://doi.org/10.1016/S0021-9924(97)00053-1

Zampini, M. L. & Green, K. P. (2001). The voicing contrast in English and Spanish: The relationship between perception and production. In J. L. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (23-48). Malden, MA: Blackwell.