Infant-directed speech in English and Spanish: Assessments of monolingual and bilingual caregiver VOT

Melanie S. Fish, Adrián García-Sierra, Nairán Ramírez-Esparza, Patricia K. Kuhl

University of Washington, Institute for Learning & Brain Sciences, 1715 Columbia Road N, Portage Bay Building, Box 357988, Seattle, WA 98195-7988, USA

University of Connecticut, Speech, Language & Hearing Sciences, 850 Bolton Road, Unit 1085, Storrs, CT 06269, USA

University of Connecticut, Department of Psychology, 406 Babidge Road, Unit 1020, Storrs, CT 06269, USA

Abstract

It has been shown that monolingual caregivers exaggerate acoustic speech cues in infant-directed speech (IDS), but less is known about the characteristics of IDS in late-second-language (L2) bilingual caregivers. Furthermore, there is inconsistency in the literature regarding voice onset time (VOT) of stop consonants in IDS. The present study explores VOT of English and Spanish stops in English monolingual and Spanish-dominant bilingual caregivers, in infant- versus adult-directed speech registers. Both monolinguals and bilinguals exaggerate VOT in IDS; however, different patterns are noted across consonant type and language context. Also, bilinguals produced English stops with Spanish-like and English-like properties, depending upon their L2 proficiency. The characteristics of late-L2 Spanish–English bilingual IDS may create a complex phonetic environment for infants, which may in turn affect the perception and later production of stop consonants in dual language-learning infants.

1. Introduction

It is widely recognized that the language environments of bilingual infants are more complex than those of their monolingual peers. One complexity may arise from experience with the phonological systems of two languages. It is possible for the phonemic categories of two languages to be at odds, as is the case for Spanish and English. In these two languages, phonemic boundaries overlap such that the same acoustic signal corresponds to different phonemes in each of the two languages; conversely, different acoustic signals correspond to the same phoneme across languages. This may present a challenge for bilingual infants’ speech perception and subsequent phonological category formation in each of their languages.

A further challenge to dual language learning may be variability in the acoustic renderings of phonetic units in language input. Consider infants born in the US to Mexican immigrants who are late second-language (late-L2) learners of English. Since the parents are Spanish-dominant, they will likely speak with a Spanish accent when addressing the infant in English, and produce phonetic units distinct from those produced by native speakers. Another example is a bilingual infant who receives native Spanish input from a bilingual parent who is a late-L2 learner of English, and receives native English input from a monolingual English-speaking parent with limited Spanish proficiency. In this case, the infant receives input from native speakers of both languages, but is also likely to hear nonnative-like input from both parents. In these scenarios, the acoustic characteristics of language input to the bilingual infant differ from those of his/her monolingual peers.

The complexities for infants exposed to late-L2 Spanish–English bilingual language environments are twofold, including: (1) variability of phonetic input arising from the independent phonological systems of each language—the phonemic overlap of some Spanish and English speech sounds results in the same acoustic signal representing different phonemes in each language, and different acoustic signals representing the same phoneme depending on the language context, and (2) variability of phonetic input arising from bilingual caregivers’ nonnative realizations of the speech sounds.
of each language. These factors may contribute to the extended period of phonemic category formation that has been noted in bilingual infants as compared to monolingual infants (Curtin, Byers-Heinlein, & Werker, 2011; see also Bosch & Sebastián-Gallés, 2003a, 2003b; García-Sierra et al., 2011; Sundara, Polka, & Genesee, 2006).

The present study investigates the acoustic characteristics of monolingual English versus bilingual Spanish–English (Spanish-dominant, late-L2 English) caregiver speech to their infants. Specifically, it centers on the ways in which bilingual caregivers manage the acoustic complexities imposed by their two languages in infant-directed speech (IDS) versus in adult-directed speech (ADS). Possible implications of bilingual language input for infants’ stop consonant perception and later production is also discussed, based on the acoustic characteristics of the late-L2 bilingual caregiver speech.

1.1. Stop consonants

To investigate the characteristics of bilingual caregiver speech, the present study focuses on voice onset time (VOT) of word-initial stop consonants. VOT can be defined as the time in milliseconds between the consonantal release and onset of the following vowel, or between the onset of voicing and consonantal release (Abramson & Lisker, 1970, 1972; Lisker & Abramson, 1970). VOT is considered one of the strongest phonetic cues used to discriminate between stop consonants that share a place of articulation because it robustly specifies voicing quality (see Llanos, Dmitrieva, Shultz, & Francis, 2013). In many languages, including English and Spanish, phonemic categories of stop consonants may be differentiated based on VOT ranges with two distributions, where one distribution represents voiced stops and the other represents voiceless stops (Fabiano-Smith & Bunta, 2012).

By convention, the time between consonant release and onset of phonation is denoted with positive VOT when phonation follows the release, and with negative VOT when phonation precedes the release. In English, the main acoustic cue that differentiates voiced from voiceless stops is positive VOT. Voiced consonants /b,d,g/ have short positive VOT of approximately 25–30 ms, while voiceless consonants /p,t,k/ have long positive VOT of more than 30 ms, accompanied by a puff of air (aspiration). In summary, English /b,d,g/ are short-lag, unaspirated stops; English /p,t,k/ are long-lag, aspirated stops. In Spanish, voiced consonants /b,d,g/ are produced with at least 50 ms of voicing preceding the consonantal release (i.e., negative VOT) and voiceless consonants /p,t,k/ are produced with short positive VOT of approximately 0–25 ms. Thus, Spanish /b,d,g/ are prevoiced stops; and Spanish /p,t,k/ are short-lag, unaspirated stops, with acoustic characteristics similar to English /b,d,g/. Consequently, (1) the same acoustic properties (short-lag, unaspirated) represent different phonemes in English (/b,d,g/) and in Spanish (/p,t,k/), and (2) different acoustic properties (e.g., short-lag, unaspirated in English; prevoiced in Spanish) represent the same phonemes (/b,d,g/) across languages.

The present investigation compares VOT of /p,t,b,d/ in bilingual and monolingual caregiver speech produced in IDS and ADS speech registers, in English and Spanish. The bilingual sample is comprised of Spanish-dominant late-L2 English learners. We expected bilinguals to produce VOT distinct from their monolingual counterparts, reflecting phonemic category assimilation (producing English long-lag /p,t/ with shorter VOT than monolinguals) or cross-language phonetic interference (producing English /b,d/ sometimes as short-lag, consistent with English, and sometimes as prevoiced, consistent with Spanish).

1.2. Late-L2 bilinguals

Infant studies suggest that a sensitive period “opens” at about 6 months of age, with evidence of perceptual narrowing near the end of the first year, indicating initial “closing” of the sensitive period (Kuhl, 2010). Phonetic learning continues after the closing of the sensitive period, but becomes much more difficult after puberty. At 12 months of age, monolingual infants show a sharp decline in their ability to discriminate foreign-language phonetic contrasts that they discriminated with ease at 6 months of age (Kuhl et al., 2008). Late-L2 bilinguals are those who learn a second language after a sensitive period, or a period of neural plasticity during which language learning is facilitated (Hagoort, 2006; Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005; Uylings, 2006). Most of the bilingual caregivers who participated in the present study were born outside the US and learned English outside the putative sensitive period for phonemic category formation; hence, we refer to them as late-L2 bilinguals. Although a few of the bilingual caregivers in this study were born in the US, they are presumed to have learned English when they entered kindergarten. For this reason, these participants are also referred to as late-L2 bilinguals, though it is important to consider that they are likely to have experienced more early English exposure than those who were born outside the US. It should be noted, however, that English proficiency of bilinguals born in the US did not differ greatly from those born outside the US. Please see Section 2.2 and Fig. 1 for more information regarding the bilingual participants’ language experience.

Research on the age of L2 acquisition has shown that the degree to which bilinguals speak an L2 with a nonnative accent depends on the age at which the L2 is acquired (e.g., Flege, Munro, & MacKay, 1995). Overall, the literature suggests that bilinguals may respond to overlapping phonemic categories in multiple ways, including (1) assimilation of phonemic categories, where perceptually similar phonetic properties of the L2 merge with those of the L1 to create a single underlying phonology (Flege, Schirru, & MacKay, 2003); and (2) cross-language interference between the L1 and L2, where speakers may rely on L1 phonetics to produce L2 phonemes and vice versa (Bosch & Ramon-Casas, 2011; Flege & Eefting, 1987a; Fowler, Sramko, Ostry, Rowland, & Halle, 2008; Sancier & Fowler, 1997). It has also been reported that proficient bilinguals (i.e., those who produce their L2 free of perceptible accent) may show an accent when producing native L1 sounds, suggesting bidirectional language interference (Bosch & Ramon-Casas, 2011; Flege & Eefting, 1987a; Fowler et al., 2008; Sancier & Fowler, 1997). Bidirectional interference between L1 and L2 has been reported for stop consonants in proficient Dutch–English, French–English, and Portuguese–English bilinguals (Flege & Eefting, 1987b; Fowler et al., 2008; Sancier & Fowler, 1997), as well as for
vowels in proficient Spanish–Catalan bilinguals (Bosch & Ramon-Casas, 2011). However, language context might play a significant role in cross-language interference (Antoniou, Best, Tyler, & Kroos, 2011; Antoniou, Tyler, & Best, 2012). For example, Antoniou, Best, Tyler, and Kroos (2010) found that Greek–English bilinguals’ productions of stops showed less bidirectional interference in a single language context than when switching spontaneously between languages (see also García-Sierra, Diehl, & Champlin, 2009; García-Sierra, Ramírez-Esparza, Silva-Pereyra, Siard, & Champlin, 2012 and Gonzales & Lotto, 2013 for perception of consonants as a function of language contexts in bilinguals).

Assimilation and cross-language interference are separate phenomena that denote the inability of bilinguals to attain native like productions in their L2, and may help explain accented speech. They are not mutually exclusive, however, as both phenomena can be observed within an individual. Of interest to the present study is the degree to which the late-L2 bilinguals show evidence of assimilation and cross-language interference. To this end, we compare the distributional properties of VOT contained in the productions of bilingual caregivers’ speech in English and Spanish versus monolingual caregivers’ speech in English. The present study aims to elucidate the characteristics of late-L2 bilingual caregiver speech that may contribute to the patterns of stop consonant acquisition in infants; hence, we focus on bilingual caregiver speech in IDS versus in ADS.

1.3. VOT and infant-directed speech

VOT measurements can be used to examine differences in consonant quality associated with IDS and ADS. VOT is a robust temporal cue that facilitates stop consonant differentiation across languages as well as within a language during phonological development (see Baran, Zlatin Laufer, & Daniloff, 1977). It has been shown that infants as young as 1–4 months old can discriminate syllables that differ only in VOT (i.e., voiced and voiceless stops) (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). Vowels spoken in an IDS register are believed to aid infant speech perception (e.g., Englund & Behne, 2005; Kuhl et al., 1997; Liu, Kuhl, & Tsao, 2003); it follows that a similar adaptation involving enhanced temporal cues in IDS might also apply to consonants. Previous research on the VOT of IDS conducted in small samples from monolingual populations has produced conflicting results. For example, Baran et al. (1977) evaluated recordings of three monolingual English mothers speaking to their 1-year-old infants in a laboratory setting, and observed no significant differences in the VOT of IDS and ADS. Sundberg and Lacerda (1999) reported shorter VOT for IDS than ADS in a recorded laboratory play situation with Swedish mothers and their 3-month-olds infants. Englund (2005) investigated Norwegian IDS in mothers speaking spontaneously to their infants in a natural interactional setting across the first six months of life, and found longer VOT for most stop consonants in IDS than in ADS.

When interpreting the findings of these studies, one must consider the potential effects of speaking rate on VOT, since IDS is produced at a slower rate than ADS (e.g., Garcia-Sierra, 1977). Intuitively, slowing the rate of speech should lengthen both the consonant and vowel in a CV-syllable (Wayland & Miller, 1994; see Englund, 2005); however, speaking rate has been shown to affect VOT in prevoiced and long-lag, but not short-lag, stop consonants (Allen & Miller, 1999; Kessinger & Blumstein, 1997; Miller, Green, & Reeves, 1986; Pind, 1995; see Beckman, Helgason, McMurray, & Ringen, 2011). In slowed speech, phonological contrast is exaggerated by selectively increasing the phonetic cue for the specified feature of prevoicing (e.g., in Spanish) or aspiration (e.g., in English) (Beckman et al., 2011). This factor may account for some of the inconsistencies across studies of VOT length in IDS. Englund (2005) controlled for speech rate effects; a repeated measures analysis of speech style, place of articulation, and voicing on syllable duration revealed no significant main effects or interactions. Thus, speaking rate was ruled out as the explanation for the increased VOT found in IDS as compared to ADS.

VOT of IDS and ADS in Spanish–English bilingual caregivers is of particular interest since increased VOT may have consequences for infants’ phonemic category formation and word learning. For example, due to the overlapping phonemic categories of Spanish and English stops, it is possible for increased VOT in some IDS productions of bilingual caregivers to change category membership across the two languages, which may in turn influence infants’ pace in phonemic category learning. The finding that there are significant differences between monolingual and bilingual VOT in IDS could contribute to our understanding of the role of early dual language input and related variation in language input on speech perception and later language abilities. The data from the present investigation will inform theories of how bilingual caregivers manage the complexity of speaking both languages in two distinct speech registers.

2. Methods

2.1. Participants

Participants were English monolingual primary caregivers (N = 25) and Spanish–English bilingual primary caregivers (N = 25). Their infants were 11 and 14 months old at the time...
of data collection. There were 14 bilingual 11-month-olds (6 female) and 11 bilingual 14-month-olds (5 female). There were 13 monolingual 11-month-olds (7 female) and 12 monolingual 14-month-olds (6 female). Caregivers included mothers (N = 47) and fathers (N = 3; 2 monolingual and 1 bilingual). Preliminary analyses showed that excluding male participants did not alter the results, so these participants were retained in the final sample. Six bilingual caregivers were born in Mexico, five in the US, three in Colombia, two in Puerto Rico, two in Venezuela, one in Chile, one in El Salvador, and one in Peru. The remaining four participants did not provide their country of origin. All 25 monolingual caregivers were Americans born in the US. The mean age was 28.33 (SD = 8.2) for bilingual and 33.80 (SD = 5.8) for monolingual caregivers. The sample was 61% white, 33% Hispanic, and 6% other.1

All participants were recruited as part of an ongoing large-scale study at the Institute for Learning & Brain Sciences at the University of Washington. Infants were full-term (37–43 weeks), of normal birth weight (6–10 lbs.), and with no major birth or postnatal complications, recurrent ear infections, or any known hearing impairments.

2.2. Caregiver language characteristics

Monolingual and bilingual caregivers were asked to complete a questionnaire that assessed level of exposure to and use of English, and confidence in using English and Spanish at different age ranges. Fig. 1 provides information about bilingual caregivers’ language exposure and use. Monolingual caregivers reported exclusive exposure to English from birth to present. A Spanish–English language background questionnaire (Garcia-Sierra et al., 2009, 2012) revealed that bilingual caregivers transitioned from Spanish language dominance in childhood to a more balanced use of English and Spanish in adulthood.

Confidence in speaking and understanding English and Spanish was assessed in bilingual caregivers using a two-step process. The first step was a self-report rating current overall confidence in speaking and understanding English and Spanish. Bilingual caregivers were asked to rate themselves on a 1–5 Likert scale (1 = “I cannot speak the language, I have a few words or phrases, and I cannot produce sentences”; 5 = “I have a native-like proficiency with few grammatical errors and I have good vocabulary”). The overall mean for bilingual caregivers’ confidence in speaking was 4.4 (SD = 0.72) for English and 4.7 (SD = 0.63) for Spanish. The overall mean for bilingual caregivers’ confidence in understanding was 4.5 (SD = 0.72) for English and 4.7 (SD = 0.63) for Spanish.

The second step was a self-report rating confidence in understanding and speaking English and Spanish over time. Bilingual caregivers were asked to rate themselves on a 1–5 Likert scale (1 = Not Confident, 2 = 25% Confident, 3 = 50% Confident, 4 = 75% Confident, 5 = 100% Confident) in three-year increments beginning at age nine. Fig. 2 shows that self-reports of bilingual caregiver confidence in understanding and speaking English as a function of age indicate an increase in confidence over time.

2.3. Assessing caregiver IDS and ADS: Stimuli

All participants were given a set of 12 English sentences that included three examples of each word-initial stop /p,t,b, d/. The sentences were given to parents in the form of a picture book with visual cues to accompany each target sentence (e.g., a picture of a pond for the sentence containing the target word pond). The bilingual caregivers were also given a similar list of 12 Spanish sentences that included four examples of /b/, /p/ and two examples of /d, t/.

The English sentences included high-frequency words with six initial bilabial and alveolar stops in voiced/voiceless minimal pairs (i.e., words that differ only in one phonological element), which contrasted in voicing of the initial stops. The Spanish sentences included high-frequency words with eight initial bilabial stops and four initial alveolar stops in voiced/voiceless minimal pairs, which contrasted in voicing of the initial stops. Velar stops were excluded from both lists because no minimal pairs of high frequency words beginning with /g,k/ in Spanish were identified. Similarly, more bilabial than alveolar stops were included in the Spanish list because fewer high-frequency minimal pairs exist in Spanish for /d, t/. The sentences were ordered such that minimal pairs did not occur sequentially, except in the case of puso/buzo. Target words never appeared in sentence initial position, and were each produced twice: first within the sentence (in-context) and again following a one second pause after the sentence (in-isolation). See Appendix A for the complete English and Spanish sentence lists, as well as a gloss for the Spanish sentences.

2.4. Language Environment Analysis

A Language Environment Analysis (LENA) device (LENA Foundation, Boulder, Colorado) was used to record caregiver speech to their infants at home. The LENA digital language processor (DLP) is a small device worn by the child in a specifically designed vest that collects first-person digital audio recordings. The device is unobtrusive, allowing IDS to be recorded in the subjects’ natural home environment at their convenience. Using the LENA method eliminates many environmental variables associated with a laboratory setting that may alter natural caregiver speech patterns.

Parents received two DLPs, each capable of 16 h of audio recording, and were instructed to record continuously during two weekdays and two weekend days, for 8 h each day. This resulted in a total of approximately 32 h of recorded audio data from each family. The present investigation reports the acoustic properties of the target words from the experimental sentences recorded by the DLP and described in Section 2.3.

2.5. Assessing caregiver IDS and ADS: Experimental procedure

The participants were asked to read the sentences in two separate experimental conditions—one designed to elicit ADS, and the other designed to elicit IDS. IDS was defined as the speech style used by caregivers when reading the sentences to their children at home, and ADS was defined as the
speech style used by caregivers when reading the sentences as if to another adult in the laboratory. In both settings, participants were instructed to read the sentence, pause for one second, then read the target word in-isolation, e.g., *John and I are going to the big pond...* (pause) *pond*. The target words were produced in-isolation to corroborate the measurements of the words produced in-context.

ADS recordings were collected in the laboratory. Caregivers were seated in a sound-attenuated recording booth and instructed to read the experimental sentences at a normal speed and volume, as though they were reading to another adult. Three productions of ADS were recorded for each participant. IDS recordings were collected in the participants’ homes, using the LENA system described in Section 2.4. Caregivers were instructed to read the sentences to their infants once a day for four days, in the register they would normally use when reading to their infants. Although having parents read the experimental sentences to their children at home is an improvement over recording IDS in the laboratory, this approach does not assess IDS in natural conversations. In previous studies, IDS has been evaluated behaviorally in natural interactions using LENA (Ramírez-Esparza, García-Sierra, & Kuhl, 2014). This, however, presents another set of challenges for data collection and interpretation. This report includes only the acoustic properties of the target words from the experimental sentences recorded by the DLP, described in Section 2.3.

In the ADS condition, participants repeated each sentence three times. The planned number of English ADS productions was 225 per stop consonant (/b,p,d,t/) per context condition, in-context and in-isolation (e.g., number of participants [25]; by number of words starting with /p/ in-context [3]; by number of times repeated [3]; $25 \times 3 \times 3 = 225$). In IDS, participants repeated sentence four times; therefore, the planned number of productions per consonant per context was 300 (i.e., $25 \times 3 \times 4 = 300$). The Spanish sentences contained an unequal distribution of stop consonants across place of articulation. The Spanish sentences included four productions of each bilabial consonant /b,p/ and two productions of each dental consonant /d,t/. The planned number of Spanish productions was 300 for each bilabial consonant in ADS (i.e., $25 \times 4 \times 3 = 300$), and 400 in IDS (i.e., $25 \times 4 \times 4 = 400$). The planned number of productions was 150 for each dental consonant in ADS (i.e., $25 \times 2 \times 3 = 150$), and 200 in IDS (i.e., $25 \times 2 \times 4 = 200$).

The total number of productions available for analysis was limited by factors such as environmental noise in the recording or failure of the participant to provide the target number of productions. The number of in-context and in-isolation target stop consonants available for analysis in English are listed in Table 1A (monolinguals’ and bilinguals’ productions with positive-VOT) and Table 1B (monolinguals’ and bilinguals’ productions with negative VOT). The number of in-context and in-isolation target stop consonants available for analysis in Spanish are listed in Tables 2A (bilinguals’ productions with positive-VOT) and 2B (bilinguals’ productions with negative VOT).

2.6. Assessing caregiver productions in IDS and ADS

2.6.1. Data preparation

The audio files were edited with Audacity software to extract three productions of each target word in ADS and up to four productions of each target word in IDS. Laboratory (ADS) and in-home (IDS) recordings were analyzed using the spectrogram in conjunction with the waveform display in PRAAT (Boersma & Weenink, 2002) to determine the VOT of each extracted in-context and in-isolation stop consonant in each production. VOT was measured from the abrupt increase in amplitude of the waveform (i.e., the release of stop closure) to the onset of voicing, which is characterized by low frequency periodic energy. Positive VOT was measured as the interval between...
the release of the stop and the onset of voicing of the following vowel. Negative VOT was measured as the interval of voicing occurring between 150–50 ms prior to the release of the stop. Negative VOT was most commonly observed in the Spanish productions, though a small number of English productions also indicated the existence of prevoicing (see Table 1B). There was a tendency across monolinguals to produce voiced stops with negative VOT, albeit in only 8% of total instances.

2.6.2. Speech rate

IDS is characterized by a slower speech rate than ADS (Allen & Miller, 1999; Kessinger & Blumstein, 1997; Miller et al., 1986; Pind, 1995). Since vowel duration increases as speech rate decreases, the present study uses vowel duration as a proxy for speech rate. It was hypothesized that vowel duration would be longer in IDS than in ADS; indeed, there was a significant increase in vowel duration in IDS as compared to ADS in bilingual and monolingual caregivers. Overall, our dataset showed no evidence of a substantial relationship between VOT and vowel duration. The correlations between VOT and vowel duration were weak for English productions in monolinguals ($r = -0.015$) and in bilinguals ($r = -0.106$), as well as for Spanish productions ($r = -0.181$). Please see online supplementary materials for complete analyses of vowel duration, VOT as a function of the following vowel, and correlations between vowel length and VOT.

2.7. Statistical analysis

2.7.1. VOT

Initial analyses employing infant age (i.e., 11 months and 14 months) as a between subject factor did not yield significant main effects or interactions for age. Therefore, all analyses reported below are collapsed across infant age. Examination of the raw VOT distributions revealed bimodal distributions of VOT for English and Spanish /b,d/ in both speech contexts: one with prevoicing (VOT less than 0 ms) and the other with short-lag VOT (0–30 ms). This bimodal distribution was most robust in bilingual caregivers’ productions of English voiced stops. However, monolingual caregivers also showed this

---

Table 1

Number of monolinguals’ and bilinguals’ English stop consonant productions included in the statistical analyses.

<table>
<thead>
<tr>
<th></th>
<th>Monolingual</th>
<th>Bilingual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In Context</td>
<td>In Isolation</td>
</tr>
<tr>
<td></td>
<td>ADS IDS</td>
<td>ADS IDS</td>
</tr>
<tr>
<td>/p/</td>
<td>171</td>
<td>239</td>
</tr>
<tr>
<td>/t/</td>
<td>171</td>
<td>240</td>
</tr>
<tr>
<td>/b/</td>
<td>159</td>
<td>233</td>
</tr>
<tr>
<td>/d/</td>
<td>150</td>
<td>238</td>
</tr>
</tbody>
</table>

Note: Although not typical, monolinguals produced a small number of English voiced stop consonants with prevoicing (negative VOT). Panel A depicts the number of English voiced stop consonants produced with positive VOT. Panel B shows the number of English voiced stop consonants produced with negative VOT.

Table 2

Number of bilinguals’ Spanish stop consonant productions included in the statistical analyses.

<table>
<thead>
<tr>
<th></th>
<th>Monolingual</th>
<th>Bilingual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In Context</td>
<td>In Isolation</td>
</tr>
<tr>
<td></td>
<td>ADS IDS</td>
<td>ADS IDS</td>
</tr>
<tr>
<td>/b/</td>
<td>21</td>
<td>7</td>
</tr>
<tr>
<td>/d/</td>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>

Note: Panel A depicts the number of Spanish voiced stop consonants produced with positive VOT. Panel B shows the number of Spanish voiced stop consonants produced with negative VOT.
bimodal distribution when producing the target words in isolation (See Fig. 4). In contrast, the expected unimodal distributions were observed for bilinguals’ productions of English and Spanish /p,t/, and for monolinguals’ productions of English /p,t/. These distribution differences had consequences for data analysis, detailed below. Results of the analyses of English and Spanish productions are reported separately, as are productions with positive versus negative VOT. For English, VOT from each qualifying production was entered into UNINOVA analyses using Group (Monolingual & Bilingual), Speech Style (ADS & IDS), Context (In-Context & In-Isolation), and Consonant (/p,t,b,d/) as fixed factors. For Spanish, VOT from each qualifying production was entered into UNINOVA analyses using Speech Style (ADS & IDS), Context (In-Context & In-Isolation), and Consonant (/p,t,b,d/) as fixed factors.

3. Results

Results are reported separately for English stop consonants (/p,t,b,d/) with positive VOT, English voiced stops (/b,d/) with negative VOT, Spanish stop consonants (/p,t,b,d/) with positive VOT, and Spanish voiced stops (/b,d/) with negative VOT. Main effects and significant interactions involving Speech Style are reported. Additional analyses related to vowel duration and VOTs as a function of the following vowel are available as online supplementary materials.

3.1. English /p,t,b,d/ with positive VOT

Refer to Table 3 for means and standard deviations of monolinguals’ and bilinguals’ productions of all English stop consonants with positive VOT. Fig. 3 shows VOT distributions of English productions of /p/ and /t/, respectively. Fig. 4 shows VOT distributions of English productions of /b/ and /d/, respectively.

Overall VOT of monolingual caregivers’ productions was significantly longer (M = 59.73 ms, SE = 0.528) than bilingual caregivers’ productions (M = 50.00 ms, SE = 0.592); F(1, 5677) = 153.197, p < 0.0001, $\eta^2_p = 0.026$. VOT of IDS (M = 61.10 ms, SE = 0.547) was significantly longer than ADS (M = 48.581 ms, SE = 0.574); F(1, 5677) = 248.136, p < 0.0001, $\eta^2_p = 0.042$. There was no significant main effect of context but, as expected, VOT differed significantly across consonants; /t/ (M = 96.66 ms, SE = 0.724), /p/ (M = 76.145 ms, SE = 0.724), /d/ (M = 26.52 ms, SE = 0.847), and /b/ (M = 20.00 ms, SE = 0.867); F(3, 5677) = 2253.00, p < 0.0001, $\eta^2_p = 0.543$.

The results showed a significant three-way interaction between Group, Speech Style, and Consonant F(3, 5677) = 16.32, p < 0.0001, $\eta^2_p = 0.003$. Both monolinguals and bilinguals produced English consonants with significantly longer VOT in IDS than in ADS. However, overall VOT and the size of the difference between VOT in IDS and ADS varied across group and consonant. Monolinguals produced IDS and ADS with significantly longer VOT than bilinguals for /p/ and /t/: F(1, 1611) = 117.292, p < 0.001, $\eta^2_p = 0.068$ and F(1, 1612) = 187.826, p < 0.0001, $\eta^2_p = 0.104$, respectively. The difference between IDS and ADS was significantly larger in monolinguals than bilinguals for the voiceless consonants /p/ and /t/: F(1, 1611) = 8.157, p < 0.01, $\eta^2_p = 0.005$ and F(1, 1612) = 28.515, p < 0.0001, $\eta^2_p = 0.017$, respectively. The difference between IDS and ADS was also significantly larger for voiceless stops /p,t/ than voiced stops /b,d/ in monolinguals: F(1, 1556) = 7.422, p < 0.01, $\eta^2_p = 0.005$ and F(1, 1560) = 36.557, p < 0.0001, $\eta^2_p = 0.023$, respectively. In contrast, VOT in IDS and ADS did not differ across monolingual and bilingual caregivers for /b/ and /d/: F(1, 1216) = 2.919, p > 0.05 and F(1, 1254) = 2.250, p > 0.10, respectively.

The interaction between Group, Speech Style, and Context was also significant F(1, 4.03) = 76.8, p = 0.045, $\eta^2_p = 0.001$. While both groups produced significantly longer VOT in IDS than in ADS in both contexts, VOT varied across group and context. Monolingual caregivers produced significantly longer VOT than bilingual caregivers when reading IDS and ADS target words in-context (IDS: F(1, 5677) = 174.63, p < 0.0001, $\eta^2_p = 0.031$; ADS: F(1, 5677) = 27.702, p < 0.0001,

Table 3

<table>
<thead>
<tr>
<th>Monolingual</th>
<th>In Context</th>
<th>In Isolation</th>
<th>In Context</th>
<th>In Isolation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADS</td>
<td>Mean (Std.)</td>
<td>IDS</td>
<td>Mean (Std.)</td>
<td>IDS</td>
</tr>
<tr>
<td>/p/</td>
<td>74.34 (19.76)</td>
<td>96.04 (32.54)</td>
<td>76.21 (22.87)</td>
<td>94.52 (39.13)</td>
</tr>
<tr>
<td>/t/</td>
<td>101.00 (19.8)</td>
<td>130.1 (43.32)</td>
<td>91.41 (19.75)</td>
<td>113.9 (43.87)</td>
</tr>
<tr>
<td>/b/</td>
<td>11.71 (6.05)</td>
<td>27.36 (15.54)</td>
<td>13.00 (8.31)</td>
<td>23.70 (14.72)</td>
</tr>
<tr>
<td>/d/</td>
<td>21.81 (9.7)</td>
<td>32.00 (11.56)</td>
<td>20.86 (14.50)</td>
<td>27.78 (11.31)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bilingual</th>
<th>In Context</th>
<th>In Isolation</th>
<th>In Context</th>
<th>In Isolation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADS</td>
<td>Mean (Std.)</td>
<td>IDS</td>
<td>Mean (Std.)</td>
<td>IDS</td>
</tr>
<tr>
<td>/p/</td>
<td>58.80 (26.00)</td>
<td>68.10 (39.00)</td>
<td>64.81 (27.00)</td>
<td>76.30 (48.00)</td>
</tr>
<tr>
<td>/t/</td>
<td>79.41 (27.16)</td>
<td>87.04 (40.86)</td>
<td>82.61 (25.16)</td>
<td>87.68 (46.86)</td>
</tr>
<tr>
<td>/b/</td>
<td>15.1 (9.51)</td>
<td>20.05 (18.01)</td>
<td>17.88 (10.10)</td>
<td>31.08 (55.27)</td>
</tr>
<tr>
<td>/d/</td>
<td>22.64 (10.32)</td>
<td>29.53 (22.17)</td>
<td>25.87 (13.91)</td>
<td>31.83 (38.61)</td>
</tr>
</tbody>
</table>

Note: Table depicts VOT means and standard deviations (in parentheses) for all English stop consonants produced in-context and in-isolation as a function of ADS and IDS. The right side of each of the panels shows the F-values and statistical significances when comparing ADS to IDS. The top panel shows VOT for monolingual caregivers and the bottom panel shows VOT for bilingual caregivers. $0.01 < p < 0.05$, $0.001 < p < 0.01$, $p < 0.001$. **0.001 < p < 0.001, ***p < 0.001.
Monolingual caregivers also produced significantly longer VOT than bilingual caregivers when reading IDS target words in isolation. For English stop consonants produced with positive VOT, ADS productions had significantly longer VOT than IDS productions in both groups (i.e., monolingual and bilingual caregivers), for all four consonants, and in both contexts (i.e., in-context and in-isolation). However, the amount of VOT increase in IDS is not constant across consonant, group, or context. Differences between ADS and IDS in monolingual productions of voiceless stops /p,t/ were significantly larger than monolingual productions of voiced stops /b,d/, bilingual productions of voiceless stops /p,t/, and bilingual productions of voiced stops /b,d/. Furthermore, the size of the significant increase in VOT across ADS and IDS productions in bilinguals did not differ across consonants. Monolinguals produced longer VOT in both IDS and ADS than bilinguals for words in isolation. The same pattern was noted for words in isolation when produced in IDS but not in ADS: no difference was found between monolinguals' and bilinguals' productions of words in isolation in ADS.

3.2. English /b,d/ with negative VOT

The results for stop consonants with negative VOT should be interpreted with caution, since the number of productions with negative VOT was significantly larger in bilinguals than in monolinguals. Specifically, monolinguals produced a total of 144 voiced stops with negative VOT (in-context = 36; in-isolation = 108), while bilinguals produced 566 voiced stops with negative VOT (in-context = 255; in-isolation = 341). Therefore, the group difference is likely biased by the discrepancy in number of speech samples in monolingual and bilingual groups. Please refer to Table 4 for means and standard errors.
### English Voiced Stop Consonants with Negative VOT

<table>
<thead>
<tr>
<th></th>
<th>ADS</th>
<th>IDS</th>
<th>ADS</th>
<th>IDS</th>
<th>ADS vs IDS</th>
<th>ADS vs IDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>In Context</td>
<td>Mean (Std.)</td>
<td>Mean (Std.)</td>
<td>Mean (Std.)</td>
<td>Mean (Std.)</td>
<td>F sig.</td>
<td>F sig.</td>
</tr>
<tr>
<td>/b/ Monolingual</td>
<td>67.21 (19.07)</td>
<td>115.0 (24.32)</td>
<td>68.00 (31.84)</td>
<td>138.5 (43.36)</td>
<td>28.78 ***</td>
<td>25.00 ***</td>
</tr>
<tr>
<td>/d/ Monolingual</td>
<td>122.06 (40.00)</td>
<td>158.3 (67.80)</td>
<td>107.7 (23.64)</td>
<td>173.4 (64.26)</td>
<td>0.94 ns</td>
<td>27.70 ***</td>
</tr>
<tr>
<td>In Isolation</td>
<td>Mean (Std.)</td>
<td>Mean (Std.)</td>
<td>Mean (Std.)</td>
<td>Mean (Std.)</td>
<td>F sig.</td>
<td>F sig.</td>
</tr>
<tr>
<td>/b/ Bilingual</td>
<td>90.34 (39.46)</td>
<td>86.46 (46.01)</td>
<td>110.7 (39.03)</td>
<td>122.5 (55.85)</td>
<td>0.20 ns</td>
<td>2.40 ns</td>
</tr>
<tr>
<td>/d/ Bilingual</td>
<td>80.70 (34.86)</td>
<td>74.16 (34.07)</td>
<td>105.2 (32.70)</td>
<td>130.6 (60.81)</td>
<td>0.50 ns</td>
<td>11.03 **</td>
</tr>
</tbody>
</table>

Note: Table depicts VOT means and standard deviations (in parentheses) for English voiced stop consonants produced in-context and in-isolation as a function of ADS and IDS. The right side of each of the panels shows the F-values and statistical significances when comparing ADS to IDS. The top panel shows VOT for monolingual caregivers and the bottom panel shows VOT for bilingual caregivers. *0.01 < p < 0.05; **0.001 < p < 0.01; ***p < 0.001.

---

**Table 4**

Monolinguals’ and bilinguals’ VOT means for English voiced stop consonants produced with negative VOT.
deviations of monolinguals’ and bilinguals’ productions of all English stop consonants with negative VOT. See Fig. 4 for frequency distributions of English productions of /b/ and /d/, respectively.

Overall VOT of monolingual caregivers’ productions of /b,d/ were significantly longer (M = −121.10 ms, SE = 5.65) than bilingual caregivers’ productions (M = −100.10 ms, SE = 2.26); F(1, 694) = 11.87, p < 0.0001, \( \eta^2_p = 0.017 \). VOT of IDS (M = −125.00 ms, SE = 5.00) was significantly longer than ADS (M = −96.23 ms, SE = 3.54); F(1, 694) = 22.18, p < 0.0001, \( \eta^2_p = 0.031 \). There was a significant main effect of context, where productions in-isolation (M = −122.00 ms, SE = 2.5) had longer VOT than productions in-context (M = −99.30 ms, SE = 5.55); F(1, 694) = 13.74, p < 0.0001, \( \eta^2_p = 0.019 \). There was also a main effect of consonant, with a significantly longer VOT for /d/ (M = −199.10 ms, SE = 5.14) than /b/ (M = −102.083 ms, SE = 3.26); F(3, 694) = 7.77, p < 0.005, \( \eta^2_p = 0.011 \).

The results showed a two-way interaction between Group and Speech Style F(1, 694) = 13.1, p < 0.0001, \( \eta^2_p = 0.018 \). Monolinguals’ VOT was significantly more negative in IDS (M = −146.42 ms, SE = 9.6) than ADS (M = −95.71, SE = 6.00), F(1, 694) = 20.1, p < 0.0001, \( \eta^2_p = 0.028 \). In contrast, bilingual caregivers’ productions in IDS and ADS were not significantly different F(1, 694) = 2.20, p = 0.14, \( \eta^2_p = 0.003 \). VOT of monolinguals was significantly more negative than bilinguals in IDS F(1, 694) = 18.85, p < 0.0001, \( \eta^2_p = 0.026 \). No significant differences between monolinguals and bilinguals were found for ADS F(1, 694) = 0.21, p = 0.885, \( \eta^2_p = 0.026 \).

3.2.1. Summary: English /b,d/ with negative VOT
For English voiced stops produced with negative VOT, monolinguals produced significantly longer prevoicing than bilinguals. Overall, /d/ was produced with more prevoicing than /b/. In terms of speech style, productions of IDS yielded longer negative VOT than productions of ADS. This pattern held true for monolinguals only; VOT for bilinguals’ productions of IDS versus ADS did not differ significantly. Furthermore, monolinguals produced more negative VOT in IDS than bilinguals; VOT in ADS did not differ across caregiver groups.

3.3. Spanish /p,t,b,d/ with positive VOT
The next step was to analyze bilingual caregivers’ production of stop consonants in Spanish. As previously mentioned, the productions of Spanish voiced stops showed both positive and negative VOT; for this reason, the statistical analyses are reported separately for positive and negative VOT. Productions of all Spanish stops with positive VOT are reported here. See Section 3.4 for productions of /b,d/ with negative VOT. Fig. 5A shows the VOT distributions for Spanish /p,t/ and Fig. 5B shows the VOT distributions for Spanish /b,d/. Note that only productions with positive VOT from Fig. 5A and B are included in this analysis. See Table 5 for means and standard deviations of bilinguals’ productions of all Spanish stops with positive VOT.

VOT of IDS and ADS productions were not significantly different F(1, 1761) = 0.228, p = 0.633. However, there was a significant main effect of context F(1, 1761) = 4.00, p = 0.046, \( \eta^2_p = 0.002 \): productions of the target words in-isolation (M = 20.8 ms, SE = 0.803) had longer VOT than productions in-context (M = 18.05 ms, SE = 1.13). The main effect of consonant was also significant F(3, 1761) = 7.102, p < 0.0001, \( \eta^2_p = 0.012 \), with /b/ showing the longest VOT (M = 21.70 ms, SE = 0.735), followed by /t/ (M = 20.67 ms, SE = 0.735), then /d/ (M = 18.51 ms, SE = 1.71), and /p/ having the shortest VOT (M = 16.85 ms, SE = 0.515). However, only /p/ and /t/ differed significantly in VOT, with /t/ having approximately 3 ms longer VOT than /p/ (p < 0.0001). There were no significant two- or three-way interactions.

3.3.1. Summary: Spanish /p,t,b,d/ with positive VOT
VOT of IDS and ADS productions were not significantly different. Bilinguals’ productions of Spanish /p,t,b,d/ with positive VOT were dependent upon the context of the target word (i.e., in-context versus in-isolation). All consonants fell within the short-lag range but only /p/ and /t/ differed significantly in VOT, with /t/ being significantly longer than /p/.

3.4. Spanish /b,d/ with negative VOT
Refer to Table 6 for means and standard deviations of bilinguals’ productions of Spanish voiced consonants with negative VOT. The VOT of IDS (M = −92.53 ms, SE = 1.67) was significantly longer than ADS (M = −79.30 ms, SE = 1.63); F(1, 1324) = 32.1, p < 0.0001, \( \eta^2_p = 0.024 \). Productions in-isolation (M = −113.00 ms, SE = 1.67) had longer VOT than productions in-context (M = −58.88 ms, SE = 1.63), F(1, 1324) = 534.20, p < 0.0001, \( \eta^2_p = 0.287 \). Consonant effects were not significant. The interaction between speech style and context was significant F(1, 1324) = 8.17, p < 0.01, \( \eta^2_p = 0.006 \): IDS productions had 20 ms more negative VOT than ADS productions in-isolation F(1, 1324) = 35.61, p < 0.0001, \( \eta^2_p = 0.026 \) and 6.5 ms more negative VOT than ADS productions in-context F(1, 1324) = 4.01, p = 0.045, \( \eta^2_p = 0.003 \). Bilingual caregivers’ IDS productions in-isolation were approximately 60.5 ms more negative than productions in-context F(1, 1324) = 330.0, p < 0.0001, \( \eta^2_p = 0.20 \). Likewise, bilingual caregivers’ ADS productions in-isolation were approximately 47.5 ms more negative than productions in-context F(1, 1324) = 210.0, p < 0.001, \( \eta^2_p = 0.137 \).

3.4.1. Summary: Spanish /b,d/ with negative VOT
Amount of negative VOT was speech style-dependent in bilingual caregivers: VOT of IDS showed significantly more prevoicing than VOT of ADS. Negative VOT was also context-dependent: productions in-isolation had significantly more prevoicing than productions in-context for both speech styles. Both in-isolation and in-context productions of IDS had significantly greater negative VOT than the corresponding productions of ADS.

4. Discussion
The goal of the present study was to compare IDS and ADS in monolingual English and late-L2 bilingual Spanish–English caregivers. VOT of word-initial stop consonants was the acoustic speech cue of interest. To our knowledge, there have been no reports evaluating VOT in late-L2 bilingual IDS; hence, we investigated how Spanish-dominant caregivers navigate the
phonetic complexity imposed by their two phonological systems when speaking to their infants in this speech register. Analyses were conducted as a function of individual conso-
nants. While some significant differences were found between voiced (/b,d/) and voiceless (/p,t/) stops, overall patterns of VOT within voiced and voiceless categories were consistent within group, language, speech style, and context. For this reason, consonants will be grouped by voicing for the purposes of this discussion. Similarly, though productions in-isolation tended to have longer VOT than productions in-context, both contexts elicited the same general patterns of VOT across consonant and speech style. Context effects will not be discussed here, except with regard to the discrepancy in late-L2 bilinguals’ ability to produce native English-like VOT across contexts.

The significant increase in VOT from ADS to IDS reported here could be a by-product of the slower speech rate of IDS. However, speaking rate typically affects VOT in prevoiced and long-lag stops, not short-lag stops. In slowed speech, the phonological contrast between prevoiced and short-lag (Spanish) or short-lag and long-lag (English) is enhanced by selective increase in the phonetic cue for the specified feature of prevoicing or aspiration (Beckman et al., 2011). This pattern was not observed in our results; furthermore, VOT and vowel duration (a proxy for speech rate) are weakly correlated. Please see online supplementary materials for details. Thus, we argue that emphasizing the distinction between voiced and voiceless stop consonants by exaggerating VOT is a feature of IDS. Future studies will need to corroborate this finding, as well as measure additional speech cues such as aspiration intensity, in order to form a more comprehensive picture of stop consonant production in IDS.

4.1. English /p,t,b,d/ with positive VOT

Both monolingual and bilingual caregivers showed a significant increase in VOT from ADS to IDS for all English stops produced with positive VOT. While it was hypothesized that /p,t/ would be exaggerated in IDS, the same effect for /b,d/ was unexpected. Though there are inconsistencies in the literature regarding length of VOT in IDS versus ADS (e.g., Baran et al., 1977; Englund, 2005; Sundberg & Lacerda, 1999), it has been suggested that emphasizing the distinction between voiced and voiceless stops may enhance infants’ perception of the voicing contrast and facilitate phonemic category formation (Baran et al., 1977; Englund, 2005). This prediction follows logically from the findings of several studies of vowel production in IDS: across languages, the distinction among vowels is exaggerated (i.e., their formant frequencies are farther apart), resulting in an acoustically expanded vowel space that may aid infant perception (Englund & Behne, 2005; Kuhl et al., 1997; Liu et al., 2003; see Cristia & Seidl, 2014; McMurray, Kovack-Lesh, Goodwin, & McEchron, 2013 for an opposing view). Lengthening /b,d/ in addition to /p,t/ in English IDS appears inconsistent with this hypothesis, since similar lengthening of voiced and voiceless consonants in IDS would not serve to differentiate voicing categories. Our data show, however, that the amount of VOT increase in IDS is not constant across voicing category and group. Differences between ADS and IDS in monolingual productions of voiceless stops /p,t/ were significantly larger than monolingual productions of voiced stops /b,d/, and significantly larger than bilingual productions of English voiceless /p,t/ and English voiced /b,d/. Furthermore, the size of the significant increase in VOT across

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Bilinguals’ VOT means for Spanish stop consonants produced with positive VOT.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spanish Stop Consonants with Positive VOT</strong></td>
<td><strong>Bilingual</strong></td>
</tr>
<tr>
<td><strong>In Context</strong></td>
<td><strong>In Isolation</strong></td>
</tr>
<tr>
<td><strong>Mean</strong> (Std.)</td>
<td><strong>Mean</strong> (Std.)</td>
</tr>
<tr>
<td>/p/</td>
<td>14.87 (9.45)</td>
</tr>
<tr>
<td>/t/</td>
<td>18.46 (8.34)</td>
</tr>
<tr>
<td>/b/</td>
<td>24.66 (29.78)</td>
</tr>
<tr>
<td>/d/</td>
<td>15.01 (4.83)</td>
</tr>
</tbody>
</table>

Note: Table depicts VOT means and standard deviations (in parentheses) for Spanish voiceless stop consonants produced in-context and in-isolation as a function of ADS and IDS. *0.01 < p < 0.05; **0.001 < p < 0.01; ***p < 0.001.

<table>
<thead>
<tr>
<th>Table 6</th>
<th>Bilinguals’ VOT means for Spanish voiced stop consonants produced with negative VOT.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spanish Stop Consonants with Negative VOT</strong></td>
<td><strong>Bilingual</strong></td>
</tr>
<tr>
<td><strong>In Context</strong></td>
<td><strong>In Isolation</strong></td>
</tr>
<tr>
<td><strong>Mean</strong> (Std.)</td>
<td><strong>Mean</strong> (Std.)</td>
</tr>
<tr>
<td>/p/</td>
<td>−55.67 (23.50)</td>
</tr>
<tr>
<td>/t/</td>
<td>−55.55 (18.42)</td>
</tr>
</tbody>
</table>

Note: Table depicts VOT means and standard deviations (in parentheses) for Spanish voiced stop consonants produced in-context and in-isolation as a function of ADS and IDS. *0.01 < p < 0.05; **0.001 < p < 0.01; ***p < 0.001.
ADS and IDS productions in bilinguals did not differ across consonants.

Both groups exaggerated English /b,d/ in IDS and, while significant, the magnitude of the VOT shift from ADS to IDS was small. As a result, the length of these consonants in IDS did not extend beyond the VOT boundary separating short-lag from long-lag stops; hyperarticulating /b,d/ does not impede the distinction between voiced and voiceless categories. Instead, the difference in the amount of VOT increase from ADS to IDS in monolinguals (i.e., /p,t/ having a large increase and /b,d/ having a small increase) emphasizes the distinction between voicing categories in IDS. This effect was not observed in bilinguals, perhaps due to their difficulty attaining authentic English long-lag /p,t/.

Proficiency in English and difficulty navigating the phonetic complexity imposed by their L2 may contribute to group differences in VOT of English stops in bilingual caregivers. There is general agreement that accented speech is the result of superimposing L2 phonology onto an existing L1 system (phonemic category assimilation), or relying on L1 phonetics to produce L2 sounds (cross-language phonetic interference) (Flege, 1981, 1987, 1991; Flege & Eefting, 1987a, 1987b; Flege & Port, 1981; Flege et al., 1995; MacKay, Flege, Piske, & Schirru, 2001). In this study, bilingual caregivers produced English /p,t/ with shorter VOT than monolinguals in both speech styles. It is interesting to note that, even with the exaggeration of VOT in IDS, bilinguals’ overall VOT for English /p,t/ in ADS was shorter than monolinguals’ overall VOT for English /p,t/ in ADS. Moreover, the size of VOT lengthening between IDS and ADS was significantly larger in monolinguals than bilinguals for the voiceless stop consonants /p,t/. These findings may be consistent with the phonetic category assimilation hypothesis, as late-L2 bilinguals are unable to produce authentic long-lag English stops even in a speech register characterized by enhanced phonetic properties (IDS) (Flege, 1991; Flege & Eefting, 1987a, 1987b). Alternatively, it could be the result of late-L2 bilinguals’ difficulty attaining the nonnative cue of aspiration.

An interesting context effect was found for bilinguals’ productions of English stops, which may also be related to their status as late-L2 speakers of English. While bilinguals produced English /p,t/ with shorter VOT than monolinguals in both speech registers, they produced English /p,t/ with VOT comparable to that of monolinguals in one condition: ADS productions in-isolation. This finding indicates that, while these late-L2 bilinguals may be capable of producing English long-lag stops in a native-like manner, they did not do so in connected (in-context) speech, nor in IDS.

4.2. English /b,d/ with negative VOT

A small number of monolingual productions of English /b,d/ showed negative VOT; consequently, these results should be interpreted with caution. While production of prevoicing by English monolingual caregivers is somewhat unexpected, it may be attributed to natural variation in pronunciation, as prevoicing is not phonemic in English (MacKain, 1982). Interestingly, monolinguals showed significantly more negative VOT in IDS than ADS, a pattern not observed in bilingual caregivers. In addition, the amount of prevoicing in English /b,d/ was significantly greater in monolinguals than bilinguals in IDS, but not in ADS. The meaning of these results is unclear. However, it may be a consequence of experimental design, as most occurrences of negative VOT in monolinguals occurred in-isolation. It is possible that caregivers over-enunciated the target words in this context. As expected, a larger number of bilingual productions of English /b,d/ showed negative VOT. This is likely due to cross-language interference, as bilinguals alternated between producing English /b,d/ as short-lag (English-like) and prevoiced (Spanish-like). The effect of L2-proficiency on bilingual caregivers’ English VOT is explored in the online supplementary materials.

4.3. Spanish /p,t,b,d/

For the purposes of this discussion, we report only productions of Spanish /p,t,b,d/ that showed VOT typical of Spanish stops (i.e., prevoiced /b,d/; short-lag /p,t/). A very small number of Spanish productions of /b,d/ showed cross-language interference (i.e., Spanish voiced stops produced as short-lags), which can be explained by the fact that bilinguals were mostly Spanish-dominant.

Bilinguals showed an increase in VOT from ADS to IDS for Spanish prevoiced /b,d/ but not for short-lag /p,t/. This result is different from the effect noted for monolingual productions of English consonants. Monolinguals showed a small increase in VOT of short-lag /b,d/ from ADS to IDS and a large increase in VOT of long-lag /p,t/ across speech style; bilinguals (in Spanish) showed no increase in VOT of short-lag /p,t/ from ADS to IDS, but they did show an increase in prevoicing of /b,d/ across speech style. In this way, monolinguals and bilinguals show similar patterns of VOT exaggeration in IDS when speaking their native language, but in opposite directions: monolinguals showed an increase of VOT duration in the positive direction while bilinguals showed it in the negative direction. This finding strengthens the claim that caregivers emphasize the distinction between voicing categories of stop consonants when speaking to their infants. We speculate that the lack of VOT lengthening in Spanish /p,t/ is due to the phonetic boundaries imposed by each language; conversely, lengthening of long-lag /p,t/ and prevoiced /b,d/ occur due to the lack of phonemic boundaries in the high positive and negative VOT ranges, respectively. We were unable to evaluate monolingual Spanish or balanced Spanish–English bilingual caregivers; however, it will be important to include these groups in future investigations of VOT in Spanish IDS to fully understand and interpret the findings of this study.

4.4. Implications for VOT perception and production in bilingual infants

It has been shown that bilingual infants may take longer than their monolingual peers to develop strong representations of the sounds of their two native languages; furthermore, this protracted period of category formation may result in more flexible phonological representations that persist over time (Bosch & Ramon-Casas, 2011). The manner in which late-L2 Spanish–English bilinguals manage the complexities of the phonemic overlap in English and Spanish stop consonants, and their increased phonetic variation in stop consonant production in IDS, may contribute to the extended period of language
commitment in dual language-learning infants reported by some studies (e.g., Ferjan Ramirez, Ramirez, Clarke, Taulu, & Kuhl, 2016; García-Sierra, Ramírez-Esparza, & Kuhl, 2016; García-Sierra et al., 2011).

Though phonetic variability is characteristic of late-L2 bilingual speech, it has been shown that bilingual infants are more capable of interpreting phonetically variable input than monolinguals (Mattock, Polka, Rivachew, & Krehm, 2010). Indeed, several recent studies have strengthened the claim that monolingual and bilingual infants develop adaptive speech processing skills specific to their language environments. For example, Fennell and Byers-Heinlein (2014) investigated the ability of 17-month-old English monolingual and English-French bilingual infants to learn minimal pairs of novel words that differ in voicing of the initial stop consonant, when produced by monolingual versus bilingual speakers. Infants learned the minimal pair distinction only when the speaker matched their native language environment, suggesting that bilingual learning strategies are flexible enough to associate the same word with multiple acoustic variations. The fact that infants are sensitive to subtle pronunciation differences is not unique to bilingual infants. For example, Durrant, Delle Luche, Cattani, and Floccia (2014) showed that monolingual infants’ representations of phonetic information can also be influenced by their language input. Specifically, Durrant and colleagues showed that monolingual infants from multidialectal backgrounds treated different pronunciations of the same words as equal, while monolingual infants from monodialectal environments treated them as different. These findings suggest that variable phonetic input in the form of accented or multidialectal speech can impact the specificity of lexical representation in infancy.

As evidenced by these and other studies, exposure to variable phonetic environments (as in bilingual or multidialectal households) may be associated with more flexible representations of phonetic information (see Byers-Heinlein & Fennell, 2014). Neural measures of infant speech perception suggest that it takes longer for bilingual infants to establish neural representations for the phonetic units of their two languages than for monolingual infants to establish representations for their single language (Bosch & Sebastián-Galé, 2003a, 2003b; García-Sierra et al., 2011; Sundara et al., 2006). For example, García-Sierra et al. (2011) assessed speech discrimination abilities in 6- to 9-month old and 10- to 12-month-old bilingual infants using an English and Spanish speech contrast. These infants showed patterns of brain activation similar to those of monolinguals, but at a later age. Specifically, bilingual infants showed responses at 10–12 months that resembled those reported in monolingual infants at 7 months (Rivera-Gaxiola, Klarman, García-Sierra, & Kuhl, 2005a; Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2005b). These findings are consistent with the hypothesis that bilingual infants may experience a longer learning period due to: (1) increased variability in the acoustic representations of phonetic units, and (2) reduced input in each of the native languages during the sensitive period (Kuhl, 1987; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Kuhl et al., 2006; Polka & Werker, 1994; Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Tees, 1984). The results of this study document the phonetic variability associated with language input provided by native Spanish speaking caregivers who are late-L2 learners of English.

Phonetic variability in language input may also have consequences for speech production. It can be hypothesized that production of stop consonants in bilingual children might also show a different pace in development as compared to monolinguals. Bilingual children are confronted with the unique challenge of experiencing variable phonetic input while developing phonological categories and phonetic representations of stop consonants in each of their languages. A study by Kehoe, Lieó, and Rakow (2004) characterized productions of German and Spanish stop consonants in four bilingual 2- and 3-year-olds to assess cross-language interference in bilingual phonetic/phonological development. German stop consonants are similar to English stop consonants, with short-lag /b,d,g/ and long-lag /p,t,k/. Some children showed a delay in the acquisition of the German voicing system (i.e., short- versus long-lag), which may result from the higher processing load imposed by the tripartite VOT distinction in German-Spanish bilinguals, as compared to the bipartite distinction of monolinguals. In other children, cross-language interference of voicing features (i.e., producing German stops with Spanish-like VOT and vice versa) was observed. This effect may be due to the influence of the dominant language or the susceptibility of bilingual children’s developing phonological systems to the impact of the second language, regardless of language dominance (Kehoe et al., 2004).

The findings of Kehoe et al. (2004) were corroborated by an investigation by Fabiano-Smith and Bunta (2012), which assessed VOT of voiceless bilabial and velar stops in English monolingual, Spanish monolingual, and Spanish–English bilingual 3-year-olds. The authors found that bilingual children tended to produce Spanish and English voiceless stops with similar VOT, rather than differentiating them as short-lag and long-lag, respectively. Furthermore, the bilingual children showed assimilation effects by producing English stops with shorter VOT than monolinguals. As for sequential bilingual children, McCarthy, Mahon, Rosen, and Evans (2014) postulated that the perception and production of stops is initially driven by their experience with L1, but that they form new phonemic categories for L2 with increased language experience. As a result, sequential bilingual children may take longer to commit to the sounds of their native language (Kuhl et al., 2008).

These findings suggest that language learning is guided by an implicit form of computational, or statistical, learning that enables children to extract patterns of distributional properties in their input (Saffran, Aslin, & Newport, 1996). For example, Rost and McMurray (2009) tested 11-month-old infants’ ability to recognize minimal pairs when the words were produced by multiple speakers as opposed to a single speaker. The results showed that infants were better at recognizing the new words when they were produced by multiple speakers. The authors concluded that when the input contains the appropriate statistical structure for a given learning mechanism, infants benefit from learning it. In the case of dual language-learning infants, the statistical structure of their phonetic input can be more complicated since the phonemic categories of the languages being learned can be at odds. For example, it has been postulated that infants exposed to English and Spanish learn phonetic distinctions at a different pace than monolinguals due to higher processing loads imposed in learning a triple VOT.
distinction (i.e., prevoicing, short-lags, and long-lags). In contrast, monolingual speakers of English or Spanish extract distributional patterns from dual VOT distinctions, imposing a lighter computational load.

More research is needed to explore cognitive load associated with learning more than one voicing distinction. One potential area of investigation is the development of stop consonant perception in languages that have more than one phonemic voicing distinction. For example, it would be informative to compare monolingual Hindi-learning infants and Spanish–English bilingual infants because Hindi has a four-way VOT contrast; voiceless unaspirated, voiced unaspirated, voiceless aspirated, and voiced aspirated (Benguere & Bhatia, 1980). The complex pattern of Hindi stop consonants might result in similar patterns of phonetic learning in monolingual Hindi-learning infants and bilingual Spanish–English infants. Another area of research that may elucidate bilinguals’ cognitive strategies in establishing phonetic representations is further exploration of bilinguals exposed to languages with similar voicing distinctions. German and English parse VOT in similar ways; therefore, it can be predicted that infants exposed to German and English develop phonetic categories at the same pace as monolingual infants exposed only to English or to German. This prediction follows logically from the findings reported by García-Sierra et al. (2016) indicating that monolingual and bilingual infants with similar amounts of language input show similar brain activation patterns of speech discrimination.

Overall, the results of the studies described above suggest that bilingual language input may potentially affect children’s phonologial and phonetic development in both languages. These effects may be due to assimilation effects, cross-language interference and/or language dominance, the nonnative accents of their caregivers, or some combination of these factors. This complex input may lead to a protracted period of phonemic category formation in infancy, as well as more flexible phonological representations that may persist into adulthood (Bosch & Ramon-Casas, 2011; Fabiano-Smith & Bunta, 2012; Garcia-Sierra et al., 2009, 2012; Gonzales & Lotto, 2013). We suggest that the phonetic complexities associated with stop consonants in Spanish and English, combined with the variability in late-L2 caregiver input, may help account for the potentially extended period of phonetic and phonological development in both languages. Further studies to investigate potential links between variability in the realization of phonemic categories and infants’ phonetic perception and later language development in both monolingual and bilingual children.

4.5. Limitations and Future directions

This study has limitations that should be addressed in future studies. First, ADS could be assessed differently to better mirror the assessment of IDS. In this study, participants read the target words in filler sentences ‘as though’ they were speaking to an adult. Future studies should ask participants to read the sentences to familiar adults, since average utterance length has been shown to be considerably shorter when addressing familiar adults than when addressing unfamiliar adults in ADS (Johnson, Lahey, Ernestus, & Cutler, 2013). The acoustic characteristics of speech sounds also change depending on speech rate (Allen & Miller, 1999; Kessinger & Blumstein, 1997; Miller et al., 1986; Pind, 1995) and listener familiarity (Johnson et al., 2013), which were not controlled here. However, the present investigation showed a weak correlation between VOT and vowel duration (a proxy for speech rate; see online supplementary materials). Thus, while speaking to unfamiliar persons can result in longer utterances and longer vowels, VOT was shown to be independent of these factors.

Another limitation of the present study is that productions of target stop consonants were obtained by asking participants to read a set of experimental sentences containing minimal pairs, rather than through spontaneous speech. Future studies would benefit from extracting ADS and IDS from spontaneous speech using comparable methodologies (see Ramírez-Esparza, García-Sierra, & Kuhl, in press; Ramírez-Esparza et al., 2014). Although assessing VOT from spontaneous speech can be challenging, it would provide a better picture of the phonetic input received in everyday natural social interactions. Finally, interpretation would be greatly enhanced by assessment of Spanish monolingual caregivers.

5. Conclusions

The present investigation explored the phonetic environments of monolingual and bilingual infants in order to better understand the characteristics and potential consequences of a complex language environment. Overall, bilingual caregivers showed more variability in the acoustic characteristics of stop consonants, producing VOT in IDS consistent with both assimilation and cross-language interference. This may be attributed to the overlapping phonological categories of Spanish and English, as well as the caregivers’ status as late-L2 bilinguals who have not attained native-like English proficiency. The unique characteristics of bilingual caregivers’ stop consonant productions in both languages may have implications for dual language-learning infants, including a protracted period of phonemic category formation and more flexible phonological representations that persist into adulthood.

Acknowledgements

This research was supported by a National Science Foundation Science of Learning Program grant to the LIFE Center [SMA-0835854], Kuhl, PI. We thank D. Padden for her support on data analysis and manuscript preparation.

Appendix A

English (A.1) and Spanish (A.2) sentence lists. The English gloss for the Spanish list is shown in italics; it is provided here for the reader, but was not included as part of the original stimuli.

Appendix A.1

1. John and I are going to the big pond…pond.
2. The little tot ran up the hill…tot.
3. My friend Paul is taller than me…Paul.
4. The big park near my house has swings…park.
5. The small dot at the center of the bull’s-eye is white…dot.
6. I saw a big tear in your eye…tear.
7. I will drive the teal car... teal.
8. The loud bark of my dog woke me up... bark.
9. The soccer ball is my favorite!... ball.
10. The big deer was in the woods... deer.
11. Mothers and babies have a unique bond... bond.
12. I will deal the cards next time... deal.

Appendix A.2
1. La pala es mi juguete preferido... pala.
2. El dardo pegó en el blanco... dardo.
3. La babá del bebé mojó su babero... baba.
4. El día y la noche son como el sol y la luna... día.
5. Mira la papa que está en el jardín... papa.
6. El panda es un oso que vive en China... panda.
7. La tía Margarita me cargó en sus brazos... tía.
8. El tren bala es el más rápido del mundo... bala.
9. La baba del bebé mojó su babero... baba.
10. The big deer was in the woods... deer.
11. Él puso mucha energía para llegar a la meta... puso.
12. He was delayed coming home because of traffic... delayed.

Appendix B. Supplementary data
Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.wcn.2017.04.003.

References


