Stop Voicing Perception in the Societal and Heritage Language of Spanish-English Bilingual Preschoolers: The Role of Age, Input Quantity and Quality

Simona Montanari[1], Jeremy Steffman[2] & Robert Mayr[3]

[1]Department of Child and Family Studies, California State University, Los Angeles

[2]Linguistics and English Language, University of Edinburgh

[3]Centre for Speech and Language Therapy and Hearing Science, Cardiff Metropolitan University

Under revision with the *Journal of Phonetics*

Corresponding author:

Simona Montanari, Ph.D.

Child & Family Studies

California State University, Los Angeles

5151 State University Drive

Los Angeles, CA, 90032, USA

smontan2@calstatela.edu

+1-818-270-3583

Funding: This work was supported by the National Institutes of Health/National Institute of Deafness and Other Communication Disorders [R15DC019493, 2021-2023]

Abstract

This is the first study to examine stop voicing perception in the societal (English) and heritage language (Spanish) of bilingual preschoolers. The study a) compares bilinguals' English perception patterns to those of monolinguals; b) it examines how child-internal (age) and external variables (input quantity and quality) predict English and Spanish perceptual performance; and c) it compares bilinguals' perception patterns across languages. Perception was assessed through a forced-choice minimal-pair identification task in which children heard synthesized audio stimuli that varied systematically along a /p-b/ and /t-d/ Voice Onset Time (VOT) continuum and were asked to match them with one of two pictures for each contrast. The results of Bayesian mixed-effects logistic regression analyses indicate that the bilinguals' category boundary for English stops was impacted by their experience with Spanish, with more short-lag VOT tokens being perceived as voiceless consistent with Spanish VOT. Age solely predicted English perceptual skills, whereas input quantity was the only moderator of Spanish perceptual performance. Finally, the bilingual children showed separate stop voicing contrasts in each language, although perceptual performance was already more mature in English by preschool age. Implications for theories of bilingual speech learning and the role of sociolinguistic variables are discussed.

Keywords: speech perception; voice onset time; bilingualism; heritage language; preschoolers; Spanish; English

1. Introduction

Despite an abundance of research on speech perception in simultaneous bilingual infants (see e.g. Fennell et al., 2016) and adult second/foreign language learners (e.g. Ingvalson et al., 2014), few studies have examined how young bilinguals perceive speech sounds in the preschool years, when important preliteracy skills whose emergence is dependent on speech perception abilities are being developed (Lyytinen et al., 2015; Nittrouer & Burton, 2005; van der Leij, 2013). Understanding speech development in young bilinguals is an issue of growing concern in the United States as the number of children who speak a language other than English at home, particularly Spanish, has more than doubled in the past three decades (National Academies of Sciences, Engineering, and Medicine [NASEM], 2017). Furthermore, over 70% of Spanish-speaking bilingual children come from families that are at or under 185% of the federal poverty line (NASEM, 2017), an issue that may place them in contexts that are less favorable to language learning. Since the emergence of preliteracy skills is dependent on speech perception abilities that are developed from early infancy and that are affected by the quality of the home language environment (Nittrouer & Burton, 2005), understanding speech development in Spanish-English bilingual children has important educational consequences.

Investigating bilingual perception at preschool age also has theoretical implications. Studies on simultaneous bilinguals have documented protracted development as they develop categories in two languages that are characterized by competing phonetic and distributional properties (Bosch & Sebastián-Gallés, 2003). Research examining age of acquisition effects in the speech perception abilities of adult L2 learners (e.g., Amengual, 2016; Baker et al., 2008; Bosch & Ramon-Casas, 2011; Højen & Flege, 2006; Tsukada et al., 2005) has also demonstrated that speakers who have not had regular and extensive exposure to L2 before age 2-3 may never attain native-like perceptual

performance. Thus, in order to understand how phonetic representations emerge in bilingual speakers, it is important to study bilingual perception at an age that has not been studied before. Furthermore, while a fair amount of work exists on the *production* abilities of Spanish-English bilingual preschoolers (see Montanari et al., 2018), we currently know very little about how these children *perceive* speech sounds in both languages. Given the links that have been documented between perception and production (Kuhl et al., 2008), studying the perceptual skills of these young bilinguals is crucially important to fully understand the development of multiple phonological systems.

This study focuses on stop voicing perception in Spanish-English bilingual preschoolers and has three principal objectives. It aims to (1) compare bilingual children's English perception patterns to those of English monolingual peers; (2) examine the role of specific child-internal (age) and external variables (input quantity and quality) as predictors of English and Spanish perceptual performance; and (3) compare bilinguals' perception patterns across languages. Investigating bilingual speech perception at preschool age can shed light on our understanding of how young bilingual children build contrastive categories corresponding to each of their languages despite reduced input in each language. This, in turn, allows us to extend and refine current models of bilingual speech perception and inform current curricular and instructional approaches for this population.

1.1 The Development of Phonemic Categories in Two Languages

Children are born as universal listeners, as they can discriminate sounds that are and are not part of their native language's phonetic inventory from their first months of life (Kuhl, 2004). However, with increased exposure to their native language environment, universal sound discrimination declines and language-specific speech perception abilities are developed. This perceptual

attunement is thus marked by diminished sensitivity to non-native contrasts and increased sensitivity to native phonetic contrasts, with improved discrimination of the native language's vowels by 6-8 months and of consonants by 10-12 months (Polka & Bohn, 2011; Polka & Werker, 1994; Stager & Werker, 1997; Werker & Tees, 1984). Some studies have shown that bilingual children follow a similar developmental trajectory as monolingual children, transitioning from a phase of universal sound discrimination to a phase of increased attunement to the sounds of the languages they are exposed to, with improved discrimination of the vowels and consonants of both of their languages by the end of their first year (Albareda-Castellot et al., 2011; Burns et al., 2007; Sundara et al., 2008). Other studies, however, have documented a temporary delay in bilingual infants' attunement to native categories when the distributional properties of a category are different between the two languages. Bosch and Sebastián-Gallés (2003), in particular, examined Catalan-Spanish bilingual infants' discrimination of the Catalan mid-vowel /e/-/ɛ/ contrast. This contrast is phonemic in Catalan, whereas Spanish only uses an /e/ that falls between the Catalan /e/ and /ɛ/ in vocalic space. The authors found that the children discriminated the contrast at 4 months; they failed to discriminate it at 8 months, and then discriminated it again at 12 months. This U-shaped developmental trajectory was interpreted as evidence that bilingual infants require additional accumulated exposure to the two languages in order to track the competing distributional properties of these vowel categories and learn them. Neuroimaging evidence also shows differences in the neural responses to speech sounds in bilingual and monolingual infants (Ferjan Ramirez et al., 2016; Garcia-Sierra et al., 2011), possibly pointing to an extended phase of universal sound discrimination with a later transition to language-specific discrimination among bilinguals. Thus, perceptual performance may differ somewhat between monolingual and bilingual infants due to the difficulty of acquiring competing phonetic properties.

Studies examining speech sound development in young sequential bilinguals, who begin to learn a second language (L2) in early childhood, can also inform us on bilingual speech sound perception. McCarthy et al. (2014) assessed English L2 perception longitudinally in a sample of sequential Sylheti L1/English L2 bilingual children who began to be formally exposed to English in preschool. The authors tested the perception of the English voicing contrast in word-initial stops, which requires sensitivity to the fine phonetic distinctions in voice onset time (VOT), tracking the developmental trajectory of this contrast from the first year of preschool at around 4;4 to a year later and examining the extent to which English stops showed influence from native (i.e., Sylheti) VOT patterns. The results showed that, initially, after only 7 months of consistent English exposure, English perception patterns were different from monolinguals' and appeared affected by children's existing Sylheti phonemic categories. However, by Time 2, after an additional year of English experience, the bilingual children's productions suggested more refined phonemic categories that were no longer significantly different from those of monolingual peers. The authors speculated that phonemic categorization may be initially affected by the first language (L1) or the language that children know most in young bilinguals. However, phonemic categories in L2 can be acquired and refined with language experience.

In contrast, other perception studies in children and adults show that L2 categorical perception is subject to L1 influence and is difficult to modify despite early intensive exposure to L2. For instance, Ramón-Casas et al. (2023), who tested the perception of the Catalan /e/-/ɛ/ contrast by Catalan-dominant and Spanish-dominant 4.5-year-old bilingual children, found that the former reliably outperformed the latter in identifying correct and mispronounced words containing the /e/-/ɛ/ contrast, which, again, is phonemic in Catalan but not in Spanish. In addition, the Spanish-dominant children displayed a considerable level of variability in performance

compared to the Catalan-dominant children. Interestingly, language dominance was determined by the language of the main caregiver (usually the mother). This means that the Spanish-dominant children had mothers who spoke Spanish. Yet, they were exposed to Catalan through other family members, friends, as well as educators in daycare before preschool, and, since they were in their second year of preschool, they had also been regularly and consistently exposed to Catalan from the teachers in this program. Therefore, their exposure to Catalan was not recent nor irregular – in fact, participants who did not have this regular, even if unbalanced, exposure to both Catalan and Spanish before entering preschool at age 3 were excluded from the study. These findings suggest that even early and intensive exposure to an L2 may not be enough to prevent phonemic representations from L1 possibly influencing L2 speech perception leading to deviations from monolingual patterns.

Darcy and Krüger (2012) obtained similar results with 10-year-old sequential Turkish L1/German L2 bilingual children in Germany who started to learn German, the societal language, between 2 and 4 years of age. The children were tested on their discrimination of four different German vowel contrasts. Two of these (/aː/-/iː/ and /eː/-/ɛ/) mapped into two separate vowel categories in Turkish; the other two contrasts (/iː/-/ɪ/ and /iː/-/eː/) mapped to a single Turkish vowel category (/i/). The results showed that the bilingual children's perception patterns for the first two contrasts were equivalent to those of German monolingual peers. In contrast, the bilinguals were significantly less accurate in their perception of /iː/-/ɪ/ and /iː/-/eː/ than the monolinguals, contrasts that were shown to be perceptually similar for Turkish monolingual speakers. These difficulties were attributed to an influence of the L1 on L2 phonological structure and were interpreted as evidence that even young bilinguals with early intensive exposure to L2 perceptually assimilate L2 phonemes to L1 phonemes that they judge to be most similar.

Netelenbos and Li's (2013) study of VOT perception in English-speaking children enrolled in a French immersion program in Canada extends Ramón-Casas et al.'s (2023) and Darcy and Krüger's (2012) findings to bilingual children who learn an L2 through immersion education. Children in these programs receive instruction in the L2 in all subjects beginning from grade 1, and are thus "immersed" in an L2 environment in a school setting. The authors tested grade 1, 3 and 5 children on their perception of both the English and French /p/-/b/ contrast. This contrast differs between French and English, since French uses short-lag VOT for the voiceless stop and lead VOT (i.e. prevoicing) for the voiced stop; on the other hand, in English /p/ is realized with long-lag VOT, while /b/ typically has short-lag VOT, especially, word-initially (e.g. Ahn 2018, et alia.). This contrast is thus challenging because it involves a different voiced-voiceless boundary in each language (i.e., acoustically, French /p/ is most comparable to English /b/). The results showed that children had different categorical boundaries for the voiced-voiceless contrast in French and English and their categorical boundary in French was native-like across the three grades. At the same time, the children were less consistent and accurate in identifying prevoiced French /b/, since acoustically, this phoneme differs from English /b/. Crucially, children performed similarly across grades, providing evidence that increased accumulated exposure did not improve perceptual performance in L2.

Overall, studies on the development of phonemic categories in two languages by simultaneous and sequential bilinguals show that children may have difficulty in developing categories that are characterized by competing phonetic and distributional properties. This may result in protracted perceptual development for simultaneous bilinguals and in perceptual patterns that deviate from monolingual patterns for sequential bilinguals despite several years of L2 experience and regular, daily use. These results are mirrored in production studies, which show that

young simultaneous and sequential bilinguals typically display cross-linguistic interaction as they develop speech sounds in two languages (Fabiano-Smith & Bunta, 2012; Kehoe et al., 2004; Mayr & Montanari, 2015; Mayr & Siddika, 2018; Montanari et al., 2018).

1.2 Input Quantity and Quality and Bilingual Perception

It appears that both *input quantity* and *quality* may affect bilingual perceptual abilities. In terms of input quantity, McCarthy et al. (2014) found that 19 months of regular and consistent exposure to English in preschool was sufficient for their participants to develop native-like English perception patterns that differed from the L1-influenced patterns the children displayed a year earlier. Ramón-Casas et al. (2023) also found that the children with higher accumulated exposure to Catalan (i.e. the Catalan-dominant bilinguals who had Catalan-speaking mothers) reliably outperformed the children who heard more Spanish (the Spanish-dominant bilinguals) in their perception of a Catalan-specific contrast. At the same time, studies have shown that even years of regular and consistent exposure to an L2 do not guarantee native-like L2 perception, especially when it comes to sounds that are characterized by phonetic and distributional properties that differ from the L1. For instance, the Spanish-dominant bilinguals in Ramón-Casas et al. (2023) had heard some Catalan from early in life and had been consistently and regularly exposed to it through preschool for a year before the study. Similarly, the sequential Turkish L1/German L2 bilingual 10-year-olds in Darcy and Krüger (2012) had started to learn German between 2 and 4 years of age and had been exposed to German for 7 years on average through schooling and mainstream society. Likewise, the children in Netelenbos and Li's (2013) study had learned French prior to age 6 and were fully schooled through French, thus their L2 exposure had been regular and consistent for years (especially for the children in grade 3 and 5). Clearly, research is inconclusive as to the amount of input or exposure that is needed in L2 to develop native-like perception. In fact, studies examining

age of acquisition effects in adult L2 learners' speech perception (e.g., Amengual, 2016; Baker et al., 2008; Bosch & Ramon-Casas, 2011; Højen & Flege, 2006; Tsukada et al., 2005) suggest that speakers who have not had regular and extensive exposure to L2 before age 2-3 may never attain native-like perceptual performance.

Studies of both simultaneous and sequential bilinguals have also shown that the quality of the speech that bilingual children hear in their environment can impact their speech perception skills. Kalashnikova and Carreiras (2022), who tested the relation between input quality and the perception of native and non-native phonemes in monolingual and bilingual 5- and 9-month-old infants, found not only that both monolinguals and bilinguals showed increased discrimination of the non-native contrast at 5 months, well before completing their perceptual attunement, but also that the extent to which individual mothers exaggerated vowels in their infant-directed speech (i.e., increasing input quality) significantly related to 9-month-old infants' speech perception performance. In other words, monolingual and bilingual infants who heard input of higher quality (i.e., characterized by exaggerated vowels) were also more ahead in their perceptual attunement than infants who heard input less conducive to language learning.

When it comes to sequential bilinguals, Ramón-Casas et al. (2023) also attribute their non-native and variable L2 performance to L2 input quality. The authors argue that the children in McCarthy et al. (2014), who displayed native-like L2 perception 19 months after preschool entry, were regularly exposed to non-accented L2 English and did not use L1 Sylheti in the school environment (since this was limited to the home environment), a factor that increased the quality of their L2 exposure. On the other hand, the participants in Darcy and Krüger (2012), Netelenbos and Li (2013), and Ramón-Casas et al. (2023) came from sociolinguistic contexts in which L2 input was influenced by L1 leading to lower quality of L2 exposure. Specifically, the children in Darcy and

Krüger (2012) attended a Turkish-German dual language school where they interacted daily in both languages. They also used Turkish extensively and were pressured to maintain it as a marker of ethnic identity, which could have promoted L1 activation and L1-to-L2 interaction effects. The participants of Netelenbos and Li's (2013) study were schooled exclusively in L2 French. However, this was not the societal language, and hence children heard it solely from their teachers in school, often "a single instructor for the entire schooling year, who may or may not be a native speaker of French" (Netelenbos & Li, 2013, p. 2383). In this context, the children may not only have heard English-accented French but they also often used English to interact with peers in the French classroom. The children in Ramón-Casas et al. (2023) came from a completely different sociolinguistic setting (e.g., officially bilingual Barcelona), "an extensive language contact context where [however] Spanish-accented Catalan is present in many areas and neighborhoods in which Spanish is also extensively used (Lleó, Benet & Cortés, 2007; Lleó et al., 2008; Mora & Nadeu, 2012)" (Ramón-Casas et al., 2023, p. 172). In this setting, the authors argue, Spanish-dominant bilinguals may be very likely to hear Spanish-accented Catalan in their home and social environment and experience Spanish-to-Catalan interaction effects, especially given the phonological and lexical proximity between the languages.

Ramón-Casas et al. (2023) attempted to assess the quality of the Catalan input that their Spanish-dominant bilingual children received and examine whether it was related to both their perceptual and production performance. Unable to estimate the children's level of exposure to accented speech in a reliable manner, the authors used the native language of children's grandparents as a measure of exposure to native or accented speech, since the Catalan produced by older generations has been shown to have more native-like phonological features (Mora & Nadeu, 2012). Thus, children with native Catalan-speaking grandparents were assumed to be exposed to

more native Catalan, whereas children with Spanish-speaking grandparents were expected to hear more Spanish-accented Catalan. A comparison of the former's perceptual performance with that of the latter did not reveal significant differences. However, Spanish-dominant bilingual participants with native Catalan-speaking grandparents were significantly better in /ɛ/-word production than children with Spanish-speaking grandparents. Furthermore, access to native vs. non-native input appeared to explain the high range of variability in production detected in the Spanish-dominant group. The authors speculate that accented input may result in the building and maintenance of inaccurate representations as well as in high levels of variability in production, despite extended exposure to the language. As Ramón-Casas et al. (2023, p. 172) put it: "When the input to the learner presents inconsistencies at the phonological level, the representation of the corresponding categories needs to be modified in order to accommodate this non-native variability (see, for example, Durrant, Delle Luche, Cattani & Floccia, 2015), leading to differences in categorization and increasing the likelihood of more variable productions." Overall, the study points to the importance of input quality as a variable contributing to bilingual perception abilities.

1.3 The Present Study

This study contributes to the literature on bilingual speech perception by examining the perceptual performance of Spanish-English bilingual preschoolers with English and Spanish stop voicing contrasts. To our knowledge, this is the first study that examines Spanish-English bilingual perception at *preschool age* by children who have been exposed to both Spanish and English from their first years of life and are transitioning to English-only schooling. At this age, children begin to develop important preliteracy skills – such as phonemic awareness, i.e., the ability to identify and manipulate individual sounds in spoken words – that are dependent upon their ability to discriminate speech sounds (Nittrouer & Burton, 2005). Our first goal is to compare the bilinguals'

perception of English /p-b/ and /t-d/ to that of English-speaking monolingual peers to assess the extent to which exposure to an additional phonological system may exert influence on English sound discrimination. Since children may take years to perceive segments categorically as adults (Feng & Peng, 2023), it is important to compare bilingual performance to that of monolingual peers who are also in the process of refining their categorical perception. We have limited the bilingual-monolingual comparisons to English (and not Spanish), because the participants are growing up in the United States and, in order to be successful in school and in life, they are expected to develop English perceptual skills that are on par with those of monolingual English-speaking peers. On the other hand, since monolingual Spanish-speaking preschoolers are almost nonexistent in our context, we felt it was inappropriate to compare the participants' Spanish perceptual performance to that of Spanish monolingual peers living in Mexico, as Spanish input in a heritage language setting will undoubtedly differ from that in a societal language context with consequences on speech sound development (Bayram et al., 2021).

We focus on stop consonants since both English and Spanish contrast voiced and voiceless categories, but the phonetic realizations of the two categories differ between the two languages as assessed through voice onset time: English uses long-lag VOT for voiceless stops and short-lag VOT for voiced ones, whereas in Spanish voiceless stops are realized with short-lag VOT and voiced stops with lead VOT (i.e. prevoicing). We are not aware of studies of VOT perception in Spanish-English bilingual children but production studies in this population reveal that VOT patterns are affected not only by developmental factors but also by crosslinguistic interactions. In particular, the literature has revealed that the voicing contrast in English is developed before the one in Spanish due to the aerodynamic challenges that prevoicing poses (Deuchar & Clark, 1996; Macken & Barton, 1979, 1980). At the same time, bilingual school-age children have been found to

significantly differentiate only English and Spanish voiceless stops, while producing voiced stops in both languages with similar short-lag (i.e., English-like) values (Konefal & Fokes, 1981; Mayr & Montanari, 2015; Muru & Lee, 2017, Procter et al., 2015). Production studies of bilingual children learning languages with voicing systems similar to those of English and Spanish (Kehoe et al., 2004; Khattab, 2003) confirm these results. We therefore expect that the voicing contrast may pose challenges even in perception in Spanish-English bilingual development.

We focus on stop perception at two places of articulation, bilabials and coronals – thus excluding velars, in order to strike an optimal balance between examining place of articulation effects and making the task feasible for preschoolers. While we originally pilot tested a task that assessed perception across all places of articulation, the children failed to participate in a meaningful way due to the length of the test. Moreover, prosodically-matched English and Spanish stimuli were also more readily available for bilabials and coronals than for velars (see section 2.2. above). Hence, similarly to other perception studies in young populations (McCarthy et al., 2014; Ramón-Casas et al., 2023), we limited our study to two rather than all three stop categories.

Next, we examine the extent to which age, input quantity and quality predict English and Spanish perception patterns. We take *caregiver-reported language exposure* as a measure of input quantity. In addition, since Spanish input was reliably provided by first-generation native Spanish-speaking parents, grandparents, and babysitters, while English was spoken by native English-speaking parents, teachers, and siblings, we take the *number of native input providers in each language* as a measure of input quality. Bilingual and monolingual research has indeed found that hearing a language from multiple native speakers is more supportive of language development than the same number of hours of language exposure from fewer and non-native speakers, since multiple native speakers expose children to a wider range of sound repertoires, lexical items and syntactic

constructions (Huttenlocher et al., 2010; Place & Hoff, 2011, 2016). Hence, we assumed that more native input providers meant more opportunities to hear different sounds, words and sentences, increasing the quality of the input.

Lastly, we compare bilinguals' perception patterns across languages to assess the extent to which children have developed contrasting categories in each of their languages. The study thus addresses the following research questions:

1. To what extent do Spanish-English bilingual children and age-matched English monolingual children differ in their perception of the English /p-b/ and /t-d/ contrasts? Do Spanish /p-b/ and /t-d/ contrasts affect the bilinguals' perception of the same contrasts in English?
2. To what extent do age, input quantity (in terms of caregiver-reported language exposure) and input quality (in terms of number of native input providers in each language) predict bilingual children's English and Spanish perception patterns?
3. To what extent does bilinguals' perception of Spanish stimuli compare to their perception of English stimuli with acoustically identical VOT values? Have children developed separate voicing contrasts in Spanish and English?

We ground our study in Flege's Speech Learning Model (SLM, Flege, 1995, 2002; see also the revised model, the SLM-r, Flege & Bohn, 2021), one of the most influential models of L2 speech perception. While the SLM-r model was proposed for L2 adult learners, the model is also applicable to explain perceptual performance in children who are learning two languages early in life. The SLM proposes that L2 speech perception is guided by the degree of similarity between L1 and L2 phones and the age of the learner. As to the similarity between L1 and L2 phones, the model proposes that the ease with which L2 phones are perceived accurately is dependent on the extent to which they map onto L1 categories. When L2 phones are similar to those of the L1, L1 categories

act as attractors and L1 and L2 sounds form merged representations which may have compromise values that differ from those of monolinguals in both the L1 and L2. When L2 phones are dissimilar to sounds in the L1, however, they will be perceived more accurately because learners will form new speech sound categories, although these may still differ from those of monolinguals, in particular if bilinguals strive to render two sounds maximally distinct cross-linguistically, resulting in cross-linguistic dissimilation or polarization effects. Based on the SLM-r model predictions and on the findings of the studies reviewed above, we thus hypothesize, for RQ 1, that the bilingual children will have more difficulty with the perception of English voiced stops than monolinguals since voiced stops in English acoustically overlap with voiceless stops in Spanish.

At the same time, the SLM-r model predicts that L1 phonetic categories are less robust/entrenched at a young age, resulting in more limited interaction between L1 and L2 categories and thus more accurate L2 perception in children than adults (Baker et al., 2008; Tsukada et al., 2005). Therefore, we hypothesize, for RQ 2, that age will predict increased stop voicing perception in English, as older bilingual children will have accumulated more English exposure compared to younger children (as the children in McCarthy et al., 2014, at Time 2). Likewise, based on the same model's predictions, we hypothesize, for RQ3, that the bilingual children will have developed separate voicing contrasts in Spanish and English despite some interaction effects between phonological systems. Finally, we are unable to put forward a conclusive hypothesis as to the extent to which input quantity and quality will predict bilinguals' English and Spanish perception patterns (RQ2) given that studies have produced mixed results as to the relevance of these factors.

2. Methods

2.1 Participants

A total of 60 children aged between 3;6 and 5;6 participated in the study, 28 of whom were Spanish-English bilinguals and 32 were English monolinguals. The children were recruited and tested by trained Child Development majors as part of their coursework for a language development course at a public, 4-year urban university in Southern California. The children were included in the study only if they had no documented history of hearing, speech, language, cognitive, or neurological deficits based on parental reports. At the time of the study, the bilinguals (8 males and 20 females) were 54.7 months old on average (*SD* = 6.4) and were matched in age to the monolinguals (15 males and 17 females), whose mean age was 51.1 months (*SD* = 6.1) ($t(59) = 1.56$, $p > .05$). The number of male and female participants was not significantly different between the bilinguals and monolinguals ($\chi^2(1) = 2.116$, $p = .146$). The bilingual children were primarily of Mexican origin and were raised in homes where they had regularly and consistently been exposed to both languages from early in life (i.e., before age 3), as they all had family members who spoke both Spanish and English. Thus, they could be considered simultaneous bilinguals (Paradis et al., 2021), although they differed in how much they heard each language. The English monolingual children heard mainly English from their input providers, but due to the bilingual nature of the community in which they lived, they also had limited exposure to Spanish. Nevertheless, they fit the description of "functional monolinguals" with no active use or knowledge of Spanish (Best & Tyler, 2007). This information was gathered through a questionnaire in which caregivers reported their child's birth date, gender, amount of exposure to English and Spanish, and number of native input providers for each language. Amount of Spanish and English exposure was measured on a Likert scale from 1 to 5, with 1 representing "child hears mostly Spanish," 2 "child hears more

Spanish than English," 3 "child hears as much Spanish as English," 4 "child hears more English than Spanish," and 5 "child hears mostly English." Based on this information, the bilingual children obtained a mean score of 3.32 (*SD* = 0.72), whereas the monolinguals scored 4.56 (*SD* = 0.5), a difference that was statistically significant ($t(59) = 7.79$, $p < .001$) and confirmed that the bilinguals were exposed to Spanish and English in equal measure, whereas the English monolinguals heard primarily English.

The number of native input providers in each language was measured by asking parents to report which native speakers spoke Spanish and English, respectively, to their child, with the possibility of including multiple input sources among "mother, father, siblings, grandparents, babysitter, teacher, media, and other" (thus, between 1 and 8 sources). Based on this information, the bilingual children were exposed to an average of 3.93 native input sources in Spanish (*SD* = 1.65, range 1-7) and 4.14 native input sources in English (*SD* = 1.21, range 1-6), a difference that was not statistically significant ($t(27) = 0.55$, $p > .05$). The English monolinguals, on the other hand, heard English through an average of 4.59 (*SD* = 1.32, range 2-7) native input providers. Bilinguals and monolinguals did not differ on the amount of native English input providers ($t(59) = 1.37$, $p > .05$), which suggests that all children were exposed to the societal language through a comparable number of native input sources. However, the difference between the number of native Spanish input providers for the bilinguals and of English input providers for the English monolinguals was statistically significant ($t(59) = 1.73$, $p = .044$), indicating that bilingual children heard Spanish, their heritage language, from fewer interlocutors. Table 1 summarizes the data for the bilingual and monolingual participants.

Table 1. Monolinguals and bilinguals' number, gender, age, language exposure patterns, and number of native English and Spanish input providers.

	Age in months (*M, SD*)	**Language Exp[1] (*M, SD*)**	**Number of English providers (*M, SD*)[2]**	**Number of Spanish providers (*M, SD*)**
Monolinguals *N* = 32 (17F, 15M)	51.1 (6.1)	4.56 (0.5)	4.59 (1.32)	NA
Bilinguals *N* = 28 (20F, 8M)	54.7 (6.4)	3.32 (0.72)	4.14 (1.21)	3.93 (1.65)

[1] Measured on a 1 to 5 scale with 1 representing "child hears mostly Spanish," 2 "child hears more Spanish than English," 3 "child hears as much Spanish as English," 4 "child hears more English than Spanish," and 5 "child hears mostly English."

[2] Measured by including multiple native sources among "mother, father, siblings, grandparents, babysitter, teacher, media, and other."

The bilingual children were further divided into two groups based on whether they heard more Spanish (language exposure scores of 1 to 3) or more English (language exposure scores of 4 and 5). Figure 1 shows the distributions of exposure scores and ages for bilingual children. Note that age, unlike exposure, was treated as a continuous numerical variable as there is a spread of ages across the age range. For the exposure variable, there was an overwhelming majority of 3 and 4 scores, leading us to bifurcate the variable. This avoids reliance on very little data (one participant) for the effect at the lower end of the exposure score scale. Although we think this approach is justified in our data, we also confirmed that a critical interaction effect involving exposure (in Section 3.2.2.) also obtains when exposure is treated as a continuous variable.

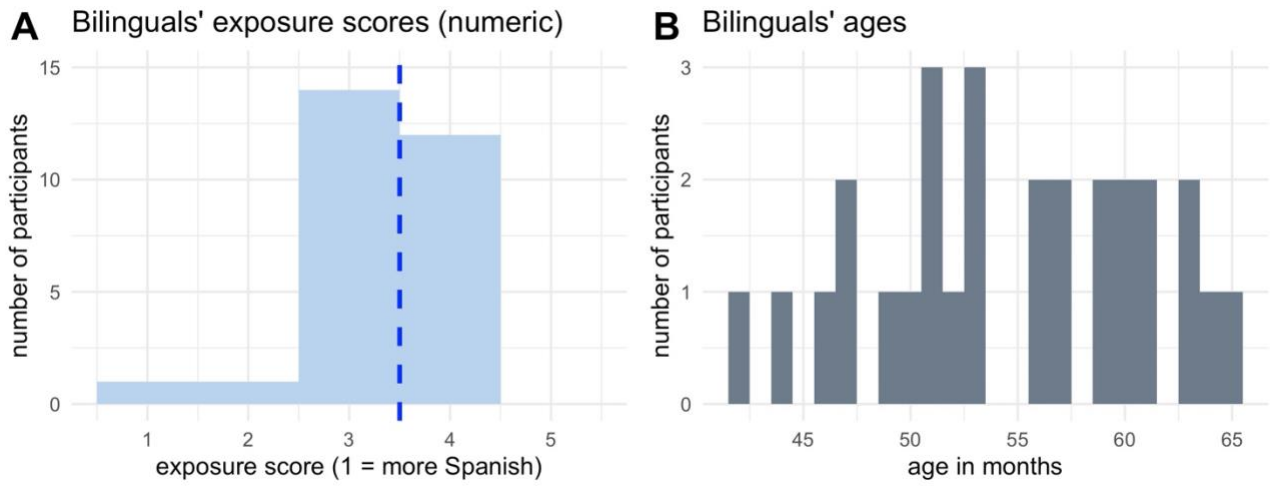

Figure 1. Histograms of bilinguals' exposure scores (panel A) and ages (panel B). The dashed vertical line in panel A represents the location at which the exposure variable was bifurcated. The width in each bin in panel B is one month.

The high Spanish exposure group (HIGH-SPAN) group included 16 children (5 males and 11 females) and the high English exposure group (HIGH-ENG) consisted of 12 children (3 males and 9 females). The HIGH-SPAN children were 55.4 months old on average (*SD* = 6.68), while the HIGH-ENG children were 55.1 months old (*SD* = 6.17), a difference that was not statistically significant ($t(27) = 0.26$, $p > .05$). The HIGH-SPAN children obtained a mean language exposure score of 2.8 (*SD* = 0.54), whereas the HIGH-ENG children scored 4.00 (*SD* = 0), a difference that was statistically significant ($t(27) = 7.53$, $p < .001$) and confirmed the higher Spanish exposure for the HIGH-SPAN group as compared to the HIGH-ENG children. In terms of the number of native Spanish and English input providers, however, the two groups did not differ. The HIGH-SPAN children heard Spanish from an average of 4.13 (*SD* = 1.36, range 2-6) native input sources, and the HIGH-ENG group from 3.67 (SD = 2.02, range 1-7), a difference that was not statistically significant ($t(27) = 0.72$, $p > .05$). Likewise, the HIGH-SPAN group heard English from an average of 4 native input sources (*SD* = 1.37, range 1-6), and the HIGH-ENG children scored 4.33 (SD = 0.98, range 3-6), which, again, was not statistically significant ($t(27) = 0.72$, $p > .05$). Thus, while

the HIGH-SPAN and HIGH-ENG children differed in the amount of exposure they received in Spanish and English, they heard both languages from a comparable number of native interlocutors. Table 2 reports information for the HIGH-SPAN and HIGH-ENG bilingual groups.

Table 2. Number, gender, age, language exposure patterns, and number of native English and Spanish input providers for bilingual children with higher Spanish exposure (HIGH-SPAN) and higher English exposure (HIGH-ENG).

	Age in months (*M, SD*)	**Language Exp[1] (*M, SD*)**	**Number of English providers (*M, SD*)[2]**	**Number of Spanish providers (*M, SD*)**
HIGH-SPAN N = 16 (5M, 11F)	54.4 (6.68)	2.8 (0.54)	4.0 (1.37)	4.13 (1.36)
HIGH-ENG N = 12 (3M, 9F)	55.1 (6.17)	4.0 (0)	4.33 (0.98)	3.67 (2.02)

[1] Measured on a 1 to 5 scale with 1 representing "child hears mostly Spanish," 2 "child hears more Spanish than English," 3 "child hears as much Spanish as English," 4 "child hears more English than Spanish," and 5 "child hears mostly English."

[2] Measured by including multiple native sources among "mother, father, siblings, grandparents, babysitter, teacher, media, and other."

2.2 Materials

In order to assess the children' perception of the Spanish and English /p-b/ and /t-d/ contrasts, we created a child-friendly forced-choice minimal-pair picture identification task in each language in which children heard an auditory stimulus that varied systematically along the VOT continuum and were asked to match it with one of two pictures representing a minimal pair. For English, we used the minimal pairs *penny*/*Benny* and *toe*/*doe.* For Spanish, we used the minimal pairs *peso* (i.e., Mexican currency)/*beso* ("kiss") and *tos* ("cough")/*dos* ("two"). We selected these words because they were matched prosodically across Spanish and English, with *penny/Benny* and *peso/beso* being bi-syllabic items stressed on the first syllable and *toe/doe* and *tos/dos* being stressed monosyllabic words. The stops in both languages were also matched in terms of vowel context, with /p/ and /b/

being followed by a mid-front vowel and /t/ and /d/ by a mid-back vowel. The words were also selected for their imageability and because they could easily be taught to children if they were not already familiar with them. While it is true that some of these words are more frequent than others, we familiarized the children to them before the experiment and the children could participate in the experiment only if they showed they could identify each test item. See section 2.3 for more details. Figure 2 shows the pictures that were used both for the word familiarization and the experiments.

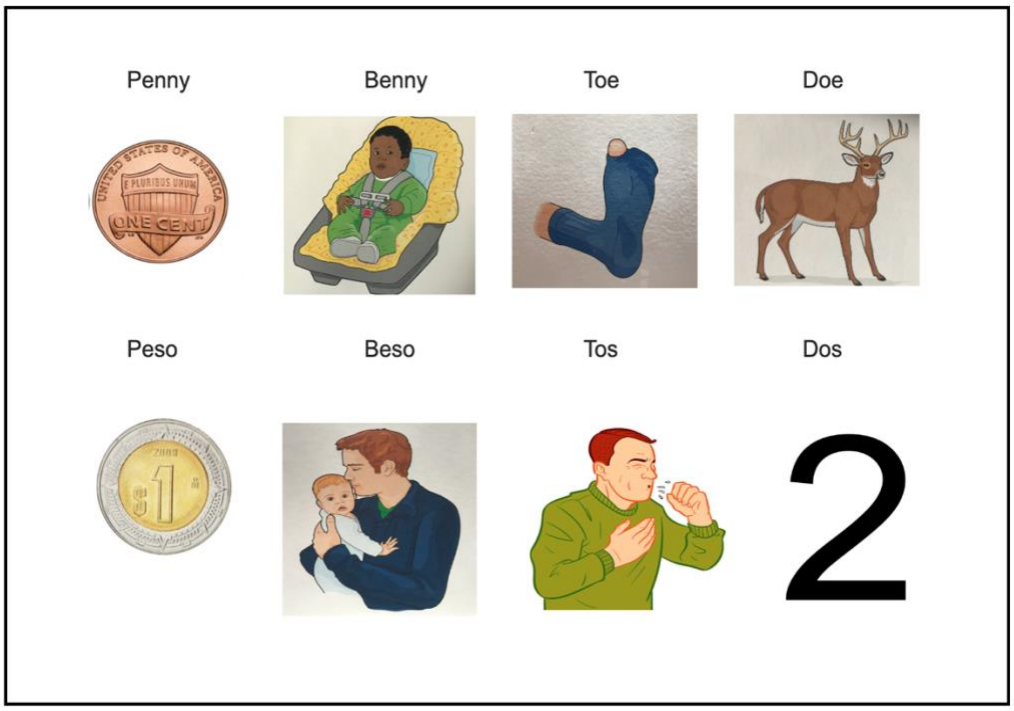

Figure 2. Pictures representing the stimuli.

Next, we created four VOT continua: An English /b/~/p/ continuum ranging between the words *Benny* and *penny*, an English /d/~/t/ continuum ranging between the words *doe* and *toe*, a Spanish /b/~/p/ continuum ranging between the words *beso* and *peso*, and a Spanish /d/~/t/ continuum ranging between the words *dos* and *tos.* Two different female speakers were recorded, one for each language; both were native speakers of American English and Mexican Spanish, respectively. The Mexican Spanish speaker had been in the US for less than 6 months and knew

little English, limiting the possibility of interaction with English. The native American English speaker was not a speaker of Spanish. Both speakers were recorded in a quiet location using a Yeti Blue USB Microphone for English, and a Tascam DR-07X for Spanish. The recordings were then digitized at 44.1 kHz and 32 bit depth.

VOT was manipulated using a Praat script (Winn, 2020) and the "cutback and replacement" approach described in Winn (2020), which increases the duration of voice onset time and simultaneously decreases the duration of the following vowel, given the well-documented inverse relation between the duration of these two cues (e.g., Allen & Miller 1999; Summerfield 1981). As is required in the "cutback and replacement" method, the starting tokens for manipulation in each case were words with initial /b/ and /d/ (in each language), for which the vowel was cut and replaced with aspiration noise to create /p/ and /t/. Note that vowel duration thus varied in these stimuli too, although we assumed that VOT would be the primary cue to the contrast (e.g., Williams, 1977; Toscano & McMurray, 2012), and therefore we refer to the continua in terms of their VOT values.

We set the VOT endpoint values for the continua based on approximate ranges for VOT between voiced and voiceless stops in each language. We opted to use the same VOT range (90 ms), but to locate it differentially along a VOT continuum for each language, corresponding to what we predicted would be good exemplars of voiced and voiceless stops based on previous speech production research which shows VOT norms for both languages (e.g., Dmitrieva et al., 2015). For Spanish, the range was -40 (prevoiced) to 50 ms VOT. For English, the range was 0 to 90 ms VOT. For each language, we used the same range for each place of articulation. Coronal stops tend to have longer VOT than bilabial stops, at least for stops that are aspirated with long lag VOT, as are English voiceless stops (e.g., Nearey & Rochet 1974; Port & Rotunno, 1994). Similarly,

monosyllabic words (i.e. our coronal-initial stimuli) tend to have longer VOT than bi-syllabic words (i.e. our bilabial-initial stimuli) (Yu et al., 2015). Hence, we predicted that, for the same VOT range, perception of voicing would be place-dependent, with longer VOT being required for a voiceless coronal percept than a voiceless bilabial percept. We return to this point in modeling and in discussing the results below. Our approach differs from an alternative in which the same VOT values could be used for both languages, though we opted to create language-specific ranges because both Spanish voiceless stops with long-lag VOT and English voiced stops with substantial pre-voicing may sound unnatural to listeners. Each continuum was created to have 10 equidistant VOT steps within each language's range. We wanted to ensure that the experiment was not too long so as to avoid generating fatigue effects in the participants. Accordingly, in order to reduce the total number of stimuli presented we next subset these continua by excluding the next-to-endpoint steps from presentation, i.e., steps 2 and 9 from the range of steps 1-10 (see e.g., Kingston et al., 2016, for a similar approach). This trimming of the steps used allowed us to keep clear endpoints (steps 1 and 10), and sample more densely from ambiguous regions of the continuum (steps 3-8). There were thus a total of 8 steps on each of the four continua presented to listeners.

2.3 Procedure

The procedure was a child-friendly forced-choice minimal-pair picture identification (2AFC) task in each language. The experiment was run through Qualtrics (Qualtrics, Provo UT). In a given trial, children heard an auditory stimulus and were asked to match it to one of two pictures that were presented on the screen. Specifically, the task was presented through a parrot (a red parrot in the English task and a yellow one in the Spanish task) who, children were told, was learning new words. Children were asked to listen to the word and point to the correct picture to help the parrot learn this word. The placement (left vs. right) of the correct picture was counterbalanced across

trials. The task was administered by a research assistant who played the auditory stimulus when the child was attentive and clicked on the picture selected by the child. There were a total of 24 trials (8 for /b/-/p/, 8 for /d/-/t/ and 8 for a vowel contrast not reported here). The trials were randomly presented, and, in order to maintain the child's attention, they were interspersed with puzzles that gradually revealed a treat for the parrot. The experiment took between 4 and 5 minutes to complete. Children were familiarized with the stimuli in their homes one week before the experiment. Right before the administration of the task, the children were presented with pictures of each minimal pair and were asked to identify each test word. Only children who identified 100% of the stimuli could participate in the experiment. Half of the bilingual children completed the experiment in Spanish first, whereas the other half completed it in English first. The English monolingual children completed the task only in English.

2.4 Exclusion Criteria

We excluded children who did not evidence any sensitivity to changing VOT along the continuum. This was accomplished by running individual logistic regression analyses for each participants' perception of each contrast in each language. The regression analysis predicted the log odds of a voiceless response, as a function of (only) scaled continuum step. Given how the variables are coded, a positive estimate in the individual models represents an increase in voiceless responses as VOT increases: the expected effect. We adopted a very lax criterion for exclusion in the sense that we excluded only participants who showed zero estimates (flat categorization across the continuum), or who showed a negative estimate for VOT, which indicates more voiced responses at longer VOT values. We reasoned that either of these patterns would represent either a lack of attention in the task or complete lack of perception of the contrast of interest. This procedure was carried out on a by-contrast and by-language basis, such that, for example, a participant could have

their data for perception of /b/~/p/ excluded, but their perception of /d/~/t/ included. By this method we excluded 15.3% of the data, that is 27 contrast + language pairings out of 176 in total (this total being the sum of 28 bilinguals with two contrasts each, in two languages equaling 112, plus 32 monolinguals with two contrasts equaling 64). This can be further broken down to 16% for bilinguals' perception of English stimuli (9/56), 14% for monolinguals' perception of English stimuli (9/64) and 16% bilinguals' perception of Spanish stimuli (9/56).

2.5 Statistical Analysis

We use Bayesian mixed effects models to analyze categorization responses, using *brms* (Bürkner, 2017) as implemented in R and R Studio (R Core Team 2021; Posit Team 2022). Models were fit to listeners' categorization responses with a logistic link function and voiceless /p/ and /t/ mapped to 1 and voiced /b/ and /d/ mapped to 0. We present a series of different analyses below which consider different variables. The models all had in common that they used random intercepts for participants as well as "maximal" by-participant random slopes, that is, random slopes for all fixed effects and interactions for which participants are exposed to multiple levels of a variable.

Models were fit to draw 4,000 samples using a no-U-turn sampler in each of four Markov chains, with a burn-in period of 1,000 samples, retaining 75% of samples for inference. The adapt delta parameter was set to 0.99. Priors for both the intercept and fixed effects in all models were set to be weakly informative and normally distributed with a mean of 0 and standard deviation of 1.5 in log-odds space. In reporting results from the models, we give the median of the estimated posterior for a given effect and 95% credible intervals (CrI). These intervals are the upper and lower bounds of the distribution which contains 95% of posterior. When 95% CrI exclude the value of zero, this indicates that the model has estimated an effect with a consistent directionality and reliably non-zero value. We also report a metric which indexes the percentage of the posterior which has a given

sign, referred to as the "probability of direction" (henceforth pd), computed with the *bayestestR* package (Makowski et al., 2019a). A posterior distribution centered precisely on a value of zero (hence, no evidence for an effect) would have a pd value of 0, while a strongly skewed distribution will have a pd value approaching 100. We consider pd values greater than 95 to represent "credible" evidence for the existence of a particular effect (see e.g., Makowski et al., 2019b), which is useful to consider in addition to CrI as it provides a more graded index for the evidence of this effect. The figures presented in this paper are predictions from each model, which show the model fit to variables of interest along with estimated CrI. In reporting the results we include the estimates for credible effects in the text, and the Appendix contains the model summaries in full. Data visualization in subsequent figures is extracted as conditional effects (plotted using that functionality in *brms*).

3. Results

3.1 Bilingual Children's Perception of English Stops as Compared to Monolinguals

Our first analysis focused on the influence of language experience in the perception of the English stimuli. To this end, we compared bilingual to monolingual language groups, both of whom completed the English experiment. In modeling, we predicted listeners' response as a function of *continuum step* (scaled and centered), *contrast* (coded with /b/~/p/ mapped to -0.5 and /d/ ~ /t/ mapped to 0.5), and *language experience* (coded with bilingual mapped to -0.5 and monolingual mapped to 0.5). We included the interaction of all fixed effects, and included *continuum step* and *contrast* as by-participant random slopes (not including *language experience* because it is a between-subjects manipulation).

The model – reported in Table A1 in the Appendix – finds a credible effect of *continuum step*, as would be expected, showing that as VOT increases along the continuum, the perception of

the voiceless stops /p/ and /t/ increases (β = 2.79, 95%CrI = [2.24,3.44], pd = 100). There was also a credible interaction between *continuum step* and *contrast* (pd = 100), which was examined further using the estimate slopes function from the *modelbased* package (Makowski et al., 2020), estimating the effect of continuum step for each contrast. This assessment showed a larger effect of *continuum step* for coronal stops (β = 3.46) as compared to bilabial stops (β = 2.06), visible as steeper categorization functions for the former contrast in Figure 3.

Language experience, the main predictor of interest, also showed a credible main effect, which did not interact with either other fixed effect (β = -0.61, 95%CrI = [-1.21,-0.02], pd = 98). The effect is reflected in Figure 3, which shows, for both contrasts, that monolinguals show overall decreased voiceless responses as compared to bilinguals. This effect is consistent with predictions based on language experience: bilingual experience with Spanish stops predicts that, overall, positive VOT values (all of the English continuum) should be mapped to the voiceless category, whereas English speaking monolinguals should tend to map positive short lag VOT to a voiced category, thus giving overall fewer voiceless responses. The first analysis thus shows that bilinguals' perception of these contrasts differs from monolinguals,' likely reflecting their experience with Spanish stop voicing contrasts.

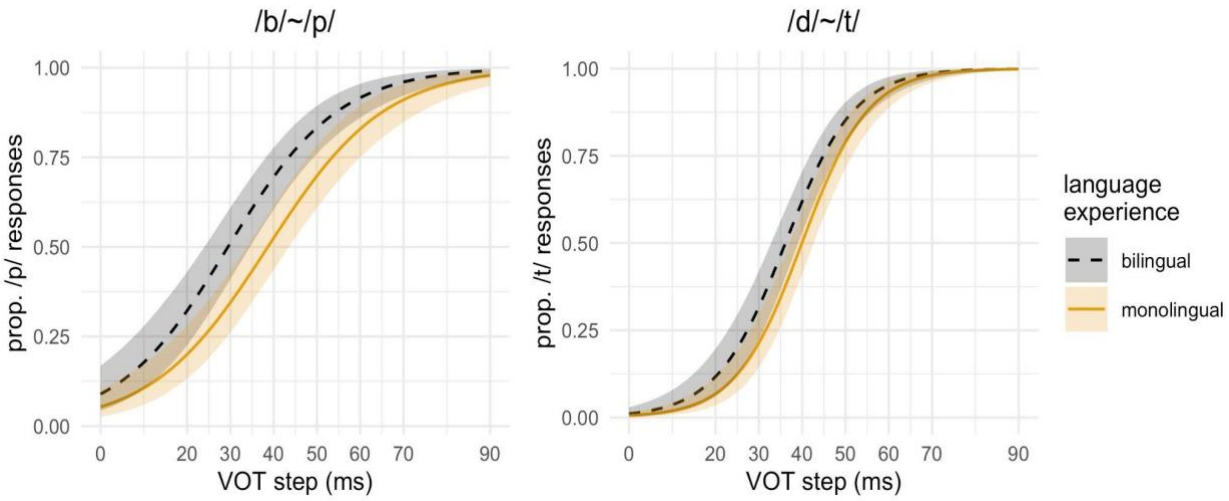

Figure 3. Monolingual and bilingual children's categorization for the English /b/~/p/ and /d/~/t/ continua along the VOT continuum (x axis), plotted as estimated by the model. Line type and coloration shows language group.

3.2 Bilingual Children's Perception of English and Spanish Stops

3.2.1 Moderators of Bilingual Children's English Stop Perception

We considered the following moderators as variables which might explain patterns in bilinguals' perception: a) *age* (scaled and centered); b) *caretaker-reported exposure,* treated as a binary variable, with the HIGH-ENG bilingual group mapped to -0.5, and the HIGH-SPAN one mapped to 0.5; and c) *number of native input providers* for the language of interest (scaled and centered). For this analysis we consider just the bilingual speakers' perception of the English stimuli. In Section 3.2.2 we examine their perception of the Spanish stimuli. All possible interactions among these predictors were included. We also included *contrast* as before, coded in the same manner as the previous analysis. Random slopes were all possible by-participant slopes, that is, *continuum step*, *contrast*, and the interaction between the two.

The model for bilinguals' perception of English stimuli – reported in Table A2 in the Appendix – found the expected effect of *continuum step* ($\beta = 3.31$, 95%CrI = [2.28,4.41], pd =

100). In addition to this main effect, only one interaction was credible (β = -0.93, 95%CrI = [-1.98,0.14], pd > 96), that being the interaction between *age* and *contrast*. This effect is visible in Figure 4 panel A, which plots categorization as a function of age (at three levels of the scaled variable for each contrast). The model estimates are plotted at three ages (4, 4;6, 5), which are selected for the purpose of visualizing the effect, although, recall that the age variable was treated as a continuous one (not binned). See Figure 1 for the distribution of ages for participants. As shown in Figure 4, for /b/ ~ /p/ only, there is a difference across ages, with older children producing more aspirated responses to the /b/ ~ /p/ continuum and showing steeper, i.e., more mature, categorization for this contrast. However, there is no difference across ages for the /t/~/d/ contrast. This is confirmed using the estimate slopes function from the modelbased package (Makowski et al., 2020), whereby (scaled) age showed a credible influence for /b/~/p/ (β = 0.94, 95%CrI = [0.15,1.79], pd = 99), but not /d/~/t/ (β = 0.03, 95%CrI = [-0.76,0.78], pd = 53). The lack of a main effect of age is notable in that it shows that age-related influences in perception are not uniform across contrasts, a point which we return to in the discussion section. Tests of this finding across different ages and contrasts will be important in further exploring this effect. It is possible that since the difference between /t/ and /d/ is acoustically larger than the one between /p/ and /b/ (e.g., Nearey & Rochet 1974; Port & Rotunno, 1994), the identification of the voicing contrast is less challenging with segments that differ more physically. We will return to this hypothesis in the discussion.

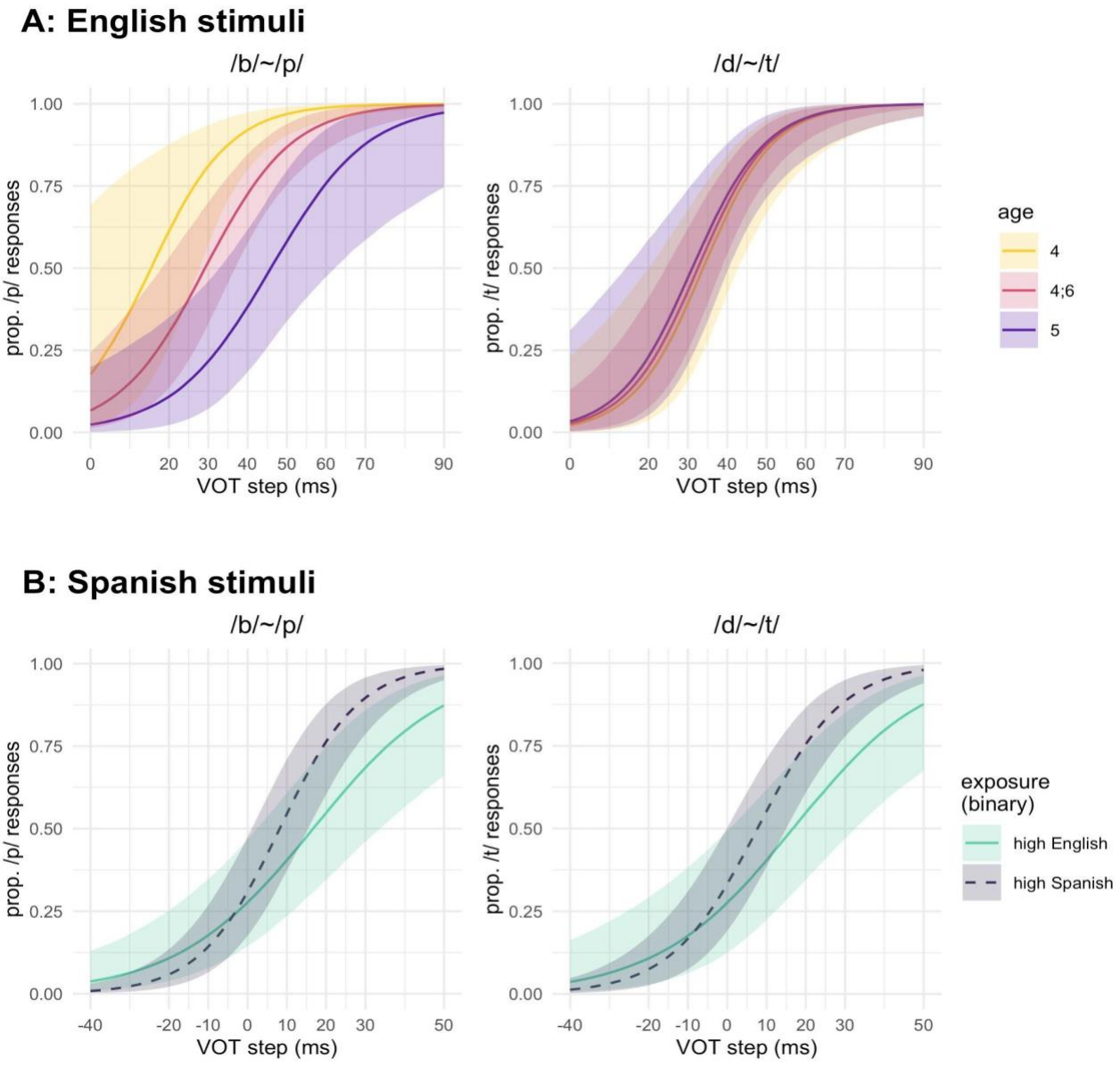

Figure 4. Bilingual children's categorization for the /b/~/p/ and /d/~/t/ continua along the VOT continuum (x axis) for English (A) and Spanish stimuli (B), plotted as estimated by the model. Line type and coloration shows age (A) and language exposure group (B).

3.2.2 Moderators of Bilingual Children's Spanish Stop Perception

In this section we performed a complementary analysis to that in Section 3.2.1. in which we consider the effects of each of the moderators on the perception of the Spanish stimuli. The model – reported in Table A3 in the Appendix – was fit in the same way as with the English stimuli with the

same predictors. As with previous models, this one finds an expected effect of *continuum step* showing that voiceless responses increase as VOT increases (β = 2.07, 95%CrI = [1.48,2.78], pd = 100). There was an additional credible interaction between *language exposure* and *continuum step* (β = 0.99, 95%CrI = [-0.08,2.99], pd = 96). This same interaction was credible in an alternative model which treated exposure as a continuous variable, as noted above (pd = 99). There was not a further three-way interaction with contrast (pd = 61) suggesting it is uniform across places of articulation. This interaction was examined visually first by plotting the model fit for each contrast as a function of language exposure. These effects are shown in Figure 4 panel B. As can be seen in the figure, listeners with high Spanish exposure evidence steeper and more categorical fits along the continuum: Bilinguals who heard more Spanish show more categorical perception of Spanish stops. This was confirmed using the estimate slopes function from the *modelbased* package (Makowski et al. 2020). Slope estimates for the effect of continuum step were taken for both exposure groups showing larger (steeper) values for the HIGH-SPAN bilinguals (β = 2.72), as compared to the HIGH-ENG bilinguals (β = 3.62). This shows that, within the bilingual group, exposure to Spanish predicts better categorical perception of the continuum.

3.3 Comparison between Bilingual Children's English and Spanish Stop Perception

In our final analysis, we carried out a direct comparison of bilinguals' perception of Spanish and English stimuli. To render this comparison as direct as possible, we selected acoustically identical VOT values from both the English and Spanish continua. Recall that the English continuum ranged from 0 to 90 ms VOT, and the Spanish continuum ranged from -40 to 50 ms VOT. The overlapping ranges across continua are thus 0-50 ms. However, because of our pairing of the continua to exclude the next-to-endpoint steps, the actual values which are shared across continua are 0, 20, 30, and 50 ms. We subset the data to contain just these values. The model – reported in Table A4 in the

Appendix – thus predicted listeners' responses as a function of *continuum step* (re-scaled with the new range), *stimulus language* (English mapped to -0.5, Spanish mapped to 0.5), and *contrast* (/b/~/p/ mapped to -0.5, /d/~/t/ mapped to 0.5). Because this analysis involves sub-setting and combining two sets of data which were previously analyzed separately, this might raise concerns about selective cherry-picking of the data and the potential to find evidence for an effect which is not supported in the data set as a whole (e.g., analogous to a Type I error in the frequentist framework). We opted to take this sub-setting approach because we felt it was the best way to compare bilinguals' perception of the two languages. If the data were not subset in this way, differences in language would also mean differences in VOT range, which would make it impossible to isolate these respective effects. The notion of a Type I error does not directly translate into the Bayesian framework. However, we can be wary of an effect that is not credible when the whole continuum is analyzed and becomes credible when a subset of the continuum is. As will be discussed below, this is only relevant for one particular effect, as all other comparisons of interest involve a new variable which was not subject to a previous statistical analysis: stimulus language for bilingual children. Because of the necessity of using different VOT ranges for each language continuum, this effect can only be properly assessed (in our view) by sub-setting and recombining the data as we have done.

The results are shown in Figure 5. There was a main effect of *continuum step* as expected ($\beta = 1.78$, 95%CrI = [1.33,2.36], pd = 100). There was no main effect of contrast (pd = 84). *Stimulus language* showed a credible main effect whereby overall, listeners gave more voiceless responses to the Spanish continuum as compared to the (same VOT) English continuum ($\beta = 1.48$, 95%CrI = [0.58,2.43], pd = 100). This is visible in Figure 5A for both contrasts and is in line with the fact that positive VOT values (as pointed out before) are mapped to the voiceless category in Spanish but not

in English (in which low positive values are mapped to the voiced category). This finding can be interpreted as evidence that the bilinguals have developed separate categorical boundaries for the Spanish and English stop voicing contrasts, with the Spanish boundaries being positioned at lower VOT values compared to the English ones, in particular at approximately 20 ms for /p-b/ (as opposed to around 28 ms for English) and at around 13 ms for /t-d/ (as opposed to approximately 35 ms for English - these values can be seen at the 0.50 point on the *y* axis, where the proportion of responses shifts from /b/ to /p/ and from /d/ to /t/). There was evidence for an interaction of *stimulus language* with *continuum step* (pd = 95); this was inspected using the estimate slopes function, which shows overall larger effects of *continuum step* for the English stimuli ($\beta = 2.09$) as compared to Spanish ($\beta = 1.46$), visible in Figure 5A as the somewhat steeper categorization of the English continuum. Overall, this result can be interpreted as evidence that children display better categorization in English as compared to Spanish. There was additional evidence for an interaction between *stimulus language* and *contrast* (pd = 96). Figure 5B visualizes the interaction by showing model predictions, collapsed across the continuum, for each combination of these two variables. Pairwise comparisons were extracted using the *emmeans* package (Lenth, 2021), which showed evidence for a difference as a function of place of articulation for the English stimuli ($\beta = 0.88$, 95%CrI = [-0.04.,1.83], pd = 97) but not for the Spanish stimuli ($\beta = -0.21$, 95%CrI = [-1.14,0.66], pd = 69). The presence of a detected difference between the two places of articulation for the English stimuli in the combined analysis should be considered in relation to the effect of place of articulation in the full analysis of bilinguals' perception of the English continuum. In that model, the main effect for place of articulation was not credible, though it shared the directionality of the present effect (pd = 73). The detected effect here thus suggests place-based differences are evident in the 0-50ms VOT range, which is a fairly ambiguous range for the English voicing distinction.

However, importantly, this difference does not generalize to the continuum as a whole. In this light we suggest that this effect should be interpreted somewhat cautiously as it is coming from a subset of the English stimuli data. Nevertheless, the presence of a credible interaction shows an asymmetry across languages. This potentially reflects place-based differences in the long-lag VOT of English voiceless stops noted in Section 2.2: /t/ has characteristically longer VOT than /p/ (e.g., Nearey & Rochet 1974; Port & Rotunno, 1994). If overall longer VOT at the coronal place of articulation is needed for a voiceless (aspirated) /t/ percept, we would expect to see fewer aspirated responses for coronal, as opposed to bilabial, stops (within the same VOT range), as we see here. The credible interaction shows that this effect does not appear to translate into perception of VOT in Spanish, for which place-based variation in VOT in short-lag stops is not systematic, including in bilingual Spanish speech (Balukus & Coops, 2014). Again, this suggests that the bilingual children have developed separate categorical boundaries – and separate voicing contrasts – for Spanish and English stops.

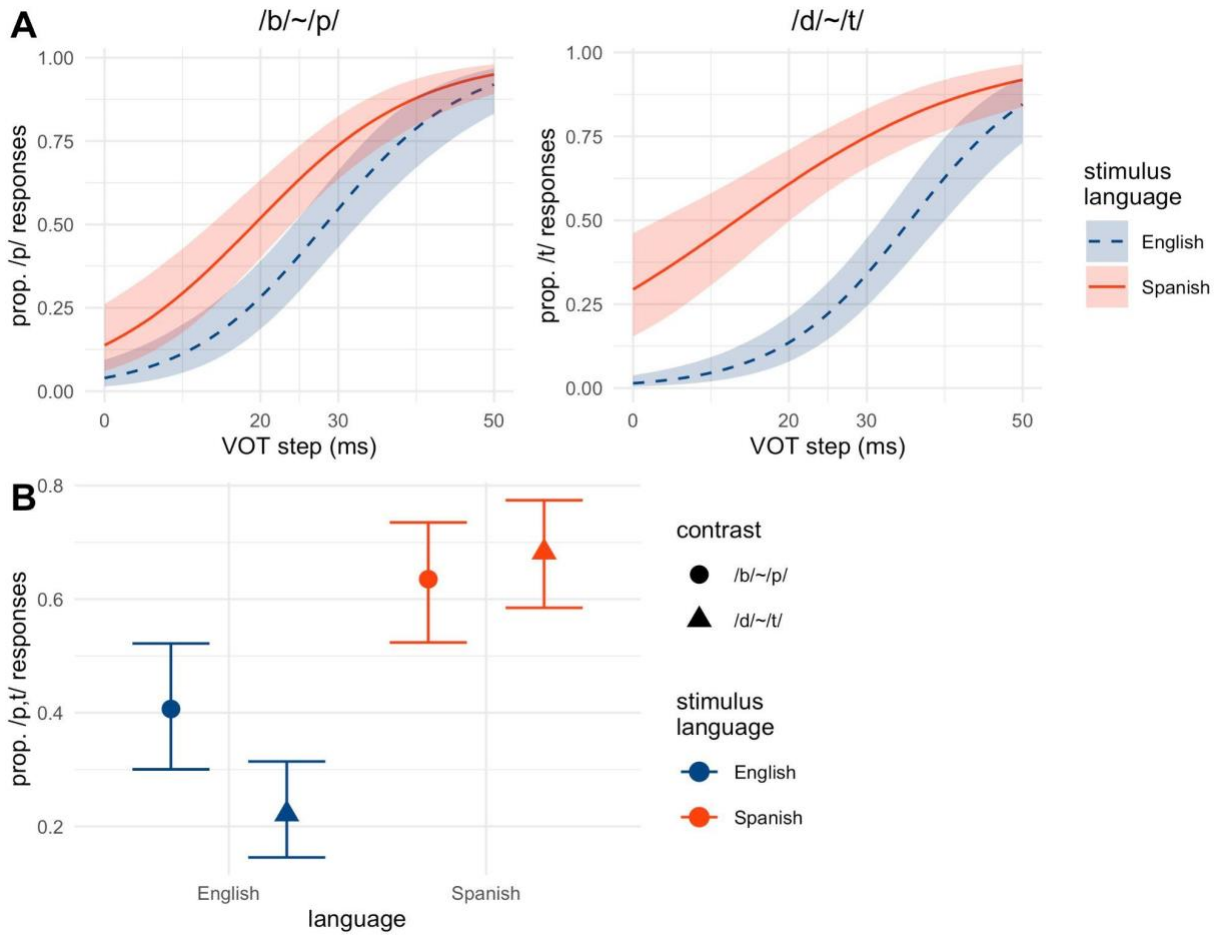

Figure 5. Panel A: Bilingual children's categorization for the English and Spanish /b/~/p/ and /d/~/t/ continua at the values shared by both languages' continua (0, 20, 30 and 50 ms), plotted as estimated by the model. Line type and coloration shows stimulus language. **Panel B:** Model estimates for overall /p,t/ responses (collapsed across the continua), with shape indicating the contrast and coloration indicating the stimulus language.

4. Discussion

This study aimed at contributing to the literature on bilingual speech perception by examining the perceptual performance of Spanish-English bilingual preschoolers with English and Spanish stop voicing contrasts. Our first goal was to compare the bilinguals' perception of English /p-b/ and /t-d/ to that of English-speaking monolingual peers to assess the extent to which exposure to the Spanish phonological system may have exerted influence on English categorical perception. Next, we examined the extent to which both child-internal (i.e. age) and external factors (i.e. input quantity

and quality) predicted English and Spanish perception patterns. We took caregiver-reported language exposure as a measure of input quantity and the number of native input providers in each language as a measure of input quality. Finally, we compared bilinguals' perception of Spanish stimuli to their perception of English stimuli with acoustically identical VOT values to assess whether children had developed separate voicing contrasts in Spanish and English. To our knowledge, this is the first study that examines Spanish-English bilingual perception at preschool age, when children who have been exposed to both languages from their first years of life transition to English-only schooling and are taught preliteracy skills that are dependent upon their speech perception abilities (Nittrouer & Burton, 2005).

The first analysis revealed that bilinguals' perception of English /p-b/ and /t-d/ differed from the monolinguals', reflecting their experience with Spanish stop voicing contrasts. Specifically, bilinguals heard more voiceless stops compared to the monolinguals, in line with the fact that, in Spanish, positive VOT values (all of the English continuum) are mapped to the voiceless category, whereas English speaking monolinguals map positive short lag VOT to a voiced category, thus giving overall fewer voiceless responses. Thus, the bilinguals heard some of the voiced stops as voiceless, displaying some difficulty with the perception of English voiced stops. These results are consistent with the predictions of Flege's SLM-r model (Flege & Bohn, 2021), which posits difficulty with the perception of L2 categories that are similar to those of the L1 since the latter act as attractors possibly producing merged representations for L1 and L2 sounds. For our participants, Spanish voiceless stops overlapped, acoustically, with English voiced stops, therefore interfering with the latter's accurate identification. Overall, these results show that early and intensive exposure to English – recall that all of our participants were reliably and consistently exposed to English

before age 3 – was not sufficient to prevent cross-linguistic interaction from the Spanish phonological system on the perception of stop voicing contrasts in English.

The analysis of the moderators of bilingual children's English stop perception indicated that age – but not current input quantity nor input quality – was a predictor of English perceptual performance, at least for the /p-b/ contrast, with older children producing more aspirated responses and showing steeper, i.e. more mature, categorization for this contrast. These results suggest that age – which can be thought of as representing cumulative language experience coupled with a more stable sound system – may be one of the strongest predictors of performance in the societal language. After all, as children get older, they will be educated in English and they will interact increasingly more with members of the wider English-speaking community, thereby expanding their opportunities to both hear and speak the societal language. Our results are in line with those of McCarthy et al. (2014), who also found that after 19 months of regular and consistent exposure to English, the societal language, in preschool, their Sylheti L1/English L2 5-year-old participants had developed native-like English perception patterns that were no longer influenced by L1 as a year earlier. Note that the children in McCarthy et al. (2014) were 52.7 months on average at the beginning of the study – the same age as our participants. Therefore, it is possible that with increasing age and exposure to English (and perhaps more experience managing two phonological systems), the children in our study may also develop native-like perceptual patterns in the societal language as cross-linguistic interaction from Spanish subsides.

Interestingly, age effects were only found for English /p-b/ and not for /t-d/. We can only speculate as to why children of different ages performed equally well with the identification of the coronal voicing contrast. A possibility is that, since the difference between /t/ and /d/ is acoustically larger than that between /p/ and /b/, the identification of the voicing contrast is less challenging with

segments that differ more physically. Recall that /t/ has characteristically longer VOT than /p/ (e.g., Nearey & Rochet 1974; Port & Rotunno, 1994). Thus, if children need, say, approximately 80 ms for a /t/ percept vs. 50 ms for a /p/ percept (Chodroff et al., 2022), assuming /d/ and /b/ have similar short-lag VOT values of approximately 10-15 ms, the /t-d/ contrast will be acoustically more distinct than the /p-b/ one; hence, children may be able to identify the coronal contrast earlier than the bilabial one. Williams (1979a, 1979b) found indeed that the categorical boundary for /b/-/p/ changes systematically with age (19 ms for 8-10-year-olds, 21 ms for 14-16-year-olds, and 25 ms for adults). On the other hand, Zlatin and Koenigsknecht (1975, 1976) found no difference in category boundary for /d/-/t/ for English-speaking 2-year-olds, 6-year-olds and adults. Recent acquisition theories also propose the concept of coronal underspecification, according to which the coronal place of articulation is the language universal default place of articulation for phonemes, with the mastery of coronal sounds (i.e., produced with the tongue tip or blade) occurring earlier than that of sounds produced at the labial (i.e., produced with the lips) or dorsal (i.e., produced with the dorsum of the tongue) place of articulation (Cummings et al., 2020). Thus, it is possible that our results reflect a different developmental trajectory for the acquisition of the bilabial and coronal voicing contrasts in line with this proposal.

While age was the only predictor of perceptual performance in the societal language, language exposure (i.e. current input quantity) was the only predictor of categorical perception in Spanish, the heritage language. This means that it was the extent to which children heard Spanish at the time of the study that resulted in steeper, i.e. more mature categorical perception, with children with higher Spanish exposure being better Spanish perceivers than those receiving less Spanish input. These results are important as they highlight the different role that child-internal and external variables may play in the acquisition of the societal vs. the heritage language. While in the context

of the societal language age will imply increased accumulated exposure to – and hence better performance in that language, for the heritage language, it is the amount of *current* exposure that predicts better performance. Indeed, input in a heritage language typically decreases as children get older (Oller et al., 2011); hence, it is not older children but children with higher Spanish input who are the best Spanish perceivers. Extant studies on bilingual perception have mostly focused on perceptual skills in the societal language (English in McCarthy et al., 2014, German in Darcy and Krüger, 2010, and Catalan in Ramón-Casas et al., 2023). Thus, this study makes an original contribution to the literature by showing that child-internal and external factors may affect categorical perception differently in the societal vs. the heritage language.

Surprisingly, in contrast to findings in the language development literature, input quality was not a predictor of perceptual abilities in either language in this study. It is possible that our measure of input quality – the number of native input providers in each language – was not sufficiently fine-grained, and hence it did not capture the quality of the input. Recall that it was measured based on a maximum of 8 possible sources (i.e., "mother, father, siblings, grandparents, babysitter, teacher, media, and other"), but the categories "siblings" and "grandparents" included multiple speakers, and we have no information on how much time each source spent with each child. Moreover, it is unclear how parents interpreted the concept of "native speaker" and whether native input providers were experiencing L1 attrition. Lastly, recall that the participants did not differ in the number of Spanish and English sources; hence the results could just be the outcome of limited variation in this measure.

At the same time, we do not exclude the possibility that input variability (provided by multiple interlocutors) is more important for lexical and grammatical development than for speech development. Indeed, the studies that have documented a link between input variability and

children's language outcomes have focused on vocabulary (Place & Hoff, 2011) and syntax (Huttenlocher et al., 2010; Place & Hoff, 2016). However, some studies of phonological development have shown that hearing input from fewer speakers may benefit children, at least in certain cases. Mayr and Montanari (2015) found indeed that their Italian-Spanish-English trilingual participants benefited from being exposed to Spanish from a single source for their Spanish productions, since less variable and ambiguous input limited the amount of cross-linguistic interaction and facilitated the adoption of speaker-specific patterns. Studies of L2 speech learning confirm that children develop L2 phonological skills best when hearing less variable L2 input (Alshangiti et al., 2019; Evans et al., 2016; Giannakopoulou et al., 2017). Indeed, despite robust evidence of the benefits of high-variability phonetic training for adults learning an L2 (e.g. Bradlow et al., 1999; Logan et al., 1991), studies on the effectiveness of high- vs low-variability phonetic training have shown that children who are trained in L2 through a single speaker make more gains in both L2 perception and production than children who are trained through multiple speakers (Alshangiti et al., 2019; Evans et al., 2016; Giannakopoulou et al., 2017). This is because children, who find it harder than adults to adapt to multiple talkers (Magnuson & Nusbaum, 2007), may more readily remember how a particular speaker produces a certain sound and use this information to shape their own perception and production (Alshangiti et al., 2019). Clearly, more data are needed to confirm this hypothesis.

Our last analysis compared bilinguals' perception of Spanish stimuli to their perception of English stimuli with acoustically identical VOT values (0, 20, 30, and 50 ms) to assess whether children had developed separate voicing contrasts in Spanish and English. Three main findings emerged from this analysis. First, the same children heard more voiceless percepts in the Spanish than in the same English VOT continuum, again in line with the fact that, in Spanish, positive VOT

values are mapped to the voiceless category, whereas in English positive short lag VOT values are mapped to the voiced category. This finding can be interpreted as evidence that the children had developed separate categorical boundaries for the Spanish and English stop voicing contrasts, with the Spanish boundary being positioned at lower VOT values compared to the English one. These results are in line with Netelenbos and Li (2013), who also found that their English-speaking participants educated via French immersion displayed a French VOT boundary located around the 5 ms range for /p-b/ and an English VOT boundary around the 25 ms range for the same contrast. Second, the fact that children provided fewer aspirated responses for coronal, as opposed to bilabial stops in English but not in Spanish provides further evidence of differentiation. Indeed, in English, the children required longer VOT at the coronal place of articulation to hear a voiceless (aspirated) /t/, in line with VOT values for these stops. However, this did not occur in Spanish in which place-based variation in VOT in short-lag stops is not robust. Overall, these findings suggest that, despite some difficulty with identifying some English voiced stops (parallel to Netelenbos and Li's, 2013, participants' imperfect identification of French /b/), the bilingual children in our study had developed different voicing contrasts in Spanish and English, and hence separate phonological systems for their two languages.

A last finding from this analysis was that children displayed steeper categorization in English despite the same VOT Spanish continuum, which suggests more mature categorization in English. This finding can have two possible explanations. First, the literature suggests that the voicing contrast in English is developed before the one in Spanish since the prevoicing vs. short-lag distinction is more costly to acquire than the short-lag vs. long-lag distinction (Deuchar & Clark, 1996; Macken & Barton, 1979, 1980). Therefore, it is possible that the observed asymmetry in English and Spanish perception is simply due to developmental factors. Another possibility is that

the results reflect children's higher exposure to, and increasing dominance in English, the societal language. Indeed, they were on average 4;7 at the time of the study, and, by this age, they had been exposed to English regularly and consistently through preschool and society at large, whereas their use of Spanish remained limited to the few members of their family and community. Clearly, only comparisons with the perceptual patterns of Spanish monolingual peers will reveal whether the observed asymmetry in English and Spanish bilingual perception is due to child-internal (i.e. developmental) factors or child-external variables, such as input quantity and majority/minority language status.

5. Conclusion, Implications and Directions for Future Research

In conclusion, this study is the first to document the perception abilities of bilingual preschoolers in both the societal and heritage language. Our study shows that language experience affects perceptual performance, with bilinguals' identification of English stop voicing contrasts being affected by their experience with Spanish stops despite regular exposure to the societal language from early in life. Unlike extant literature which has primarily focused on bilingual children's perceptual skills in the societal language (McCarthy et al., 2014, Darcy & Krüger, 2010, and Ramón-Casas et al., 2023), our study also provides evidence that child-internal (i.e. age) and external factors (i.e. input quantity) play different roles on perceptual performance in the societal and heritage language. While age solely predicts perceptual skills in English, the societal language, input quantity is the only moderator of how well children perceive the sounds of Spanish, their heritage language. Overall, the results suggest that children who have been exposed to two languages from early in life have separate stop voicing contrasts in each of their languages by preschool age, although perceptual performance is more mature in the societal language by this age.

Our findings have both theoretical and practical implications. First our study provides further support for Flege's SLM-r model (Flege & Bohn, 2021), showing that bilingual preschoolers experience more difficulty with the perception of English categories (i.e. voiced stops) that acoustically overlap with Spanish categories (i.e. voiceless stops). While the SLM-r model was proposed for L2 adult learners, the results of this study suggest that the model is also applicable to explain perceptual performance in children who are learning two languages early in life. But while the SLM-r model predicts that L1 categories are less entrenched at a young age, resulting in more limited L1-to-L2 interaction, our study finds that interaction effects can be seen even in the case of regular L2 exposure from early in life (i.e. before age 3). Indeed, in line with the model's hypothesis that L1 and L2 phonetic categories interact with one another dynamically throughout the lifespan, our results show that interaction may just be the natural outcome of the coexistence of two or more phonological systems.

Our study also has educational implications. Recall that at preschool age children develop preliteracy skills (such as phonemic awareness) that are dependent upon their ability to perceive speech sounds. Specifically, two recent longitudinal studies in pre-readers showed, by recording brain event-related potentials (ERPs), that auditory processing and speech perception at preschool age predict both phonological and pre-reading skills, as well as later reading and writing skills at school age (Lyytinen et al., 2015; van der Leij, 2013). The literature also shows that living in low-SES environments – as many Spanish-English bilingual children in the US (NASEM, 2017) – may delay speech discrimination; poor speech discrimination reduces the distinctness of phonological representations, making them more difficult to remember, recall and articulate. This, in turn, affects the development of phonemic awareness, indirectly contributing to differences in reading acquisition (Nittrouer & Burton, 2005). Thus, studying bilingual children's perception skills before

formal schooling begins can inform current curricular and instructional approaches, possibly improving their reading and educational outcomes. Specifically, the findings of our study show that language experience (i.e. hearing Spanish) affects perceptual performance in English; that certain English sounds are more difficult to be perceived than others for Spanish-English bilinguals (i.e. voiced stops); and that different child-internal and external factors differently affect perceptual performance in English and Spanish. Thus, educators and policy makers should expect bilingual-monolingual differences in how children hear English sounds as they enter into preschool, and adjust curricular programs and instructional practices accordingly, for example by focusing on voiced stops with respect to voiceless ones. At the same time, it should be expected that with age, children will improve their perceptual performance in English, whereas only increasing exposure to Spanish will improve their Spanish perceptual skills.

As with all studies, our investigation has some limitations. First, we did not include a Spanish monolingual control group, and we were unable to determine whether the steeper (i.e. more mature) categorization in English was due to developmental, input-related or sociolinguistic factors. We also only included certain child-internal and external factors as moderators of perceptual performance, and some of our measures (i.e. input quality) were perhaps not sufficiently fine-grained to capture what they were meant to capture. Future research should include more sophisticated measures of input quality as well as of other variables that may be related to speech perception, including language exposure and use (i.e. language input and output), language proficiency (i.e. lexical and grammatical measures), as well as broader sociolinguistic variables such as maternal education and acculturation. Finally, future studies should track bilingual children's perceptual performance in both the societal and heritage language over time to examine

the extent to which increasing interaction with the mainstream culture refines English perception skills while possibly affecting the same skills in Spanish.

References

Ahn, S. (2018). The role of tongue position in laryngeal contrasts: An ultrasound study of English and Brazilian Portuguese. *Journal of Phonetics, 71,* 451–467.

Albareda-Castellot, B., Pons, F., & Sebastián-Gallés, N. (2011). The acquisition of phonetic categories in bilingual infants: New data from an anticipatory eye movement paradigm. *Developmental Science, 14*(2), 395-401.

Allen, J.S., & Miller, J.L. (1999). Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words. *The Journal of the Acoustical Society of America*, *106*(4), 2031-2039.

Alshangiti, W., Evans, B.G., & Wibrow, M. (2019). Learning to speak in a second language: Does multiple talker production training benefit production of English vowels in Arabic children? In S. Calhoun, P. Escudero, M. Tabain, & P Warren. (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences,* Melbourne, Australia (pp. 2193-2197).

Amengual, M. (2016). The perception and production of language-specific mid-vowel contrasts: Shifting the focus to the bilingual individual in early language input conditions. *International Journal of Bilingualism, 20(*2), 133–152.

Baker, W., Trofimovich, P., Flege, J.E., Mack, M., & Halter, R. (2008). Child-adult differences in second-language phonological learning: The role of cross-language similarity. *Language and Speech, 51*(4), 317-342.

Balukas, C., & Koops, C. (2015). Spanish-English bilingual voice onset time in spontaneous code-switching. *International Journal of Bilingualism*, *19*(4), 423-443.

Bayram, F., Kubota, M., Luque, A., Pascual y Cabo, D., & Rothman, J. (2021). You can't fix what is not broken: Contextualizing the imbalance of perceptions about heritage language bilingualism. *Frontiers in Education,* vol. 6. Published online 29 April 2021. doi: 10.3389/feduc.2021.628311

Bradlow, A.R., Akahane-Yamada, R.A., Pisoni, D.B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics 61*, 977–985.

Best, C.T., & Tyler, M.D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34). John Benjamins.

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech, 46*(2-3), 217-243.

Bosch, L., & Ramon-Casas, M. (2011). Variability in vowel production by bilingual speakers: Can input properties hinder the early stabilization of contrastive categories? *Journal of Phonetics, 39*(4), 514–526.

Bürkner, P.C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1-28.

Burns, T.C., Yoshida, K.A., Hill, K., & Werker, J.F. (2007). The development of phonetic representation in bilingual and monolingual infants. *Applied Psycholinguistics, 28*(3), 455-474.

Cheung, H., Chung, K.K.H., Wong, S.W.L., McBride-Chang, C., Penney, T.B., & Ho, C.S.-H. (2010). Speech perception, metalinguistic awareness, reading, and vocabulary in Chinese–English bilingual children. *Journal of Educational Psychology, 102*(2), 367–380.

Chodroff, E., Bradshaw, L., & Livesay, V. (2022). Subsegmental representation in child speech production: Structured variability of stop consonant voice onset time in American English and Cantonese. Published online 05 August 2022. *Journal of Child Language,* pp. 1-29. https://doi.org/10.1017/S0305000922000368

Cummings, A.E., Ogiela, D.A., & Wu, Y.C. (2020). Evidence for [coronal] underspecification in typical and atypical phonological development. *Frontiers in Human Neuroscience, 14,* 580697.

Darcy, I., & Krüger, F. (2012). Vowel production and perception in Turkish children acquiring L2 German. *Journal of Phonetics, 40,* 568–581.

Deuchar, M., & Clark, A. (1996). Early bilingual acquisition of the voicing contrast in English and Spanish. *Journal of Phonetics, 24*(3), 351-365.

Dmitrieva, O., Llanos, F., Shultz, A.A., Francis, A.L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics, 49,* 77-95.

Durrant, S., Delle Luche, C., Cattani, A., & Floccia, C. (2015). Monodialectal and multidialectal infants' representation of familiar words. *Journal of Child Language, 42*(2), 447–465.

Evans, B.G., & Alshangiti, W. (2018). The perception and production of British English vowels and consonants by Arabic learners of English. *Journal of Phonetics, 68,* 15-31.

Fabiano-Smith, L., & Bunta, F. (2012). Voice onset time of voiceless bilabial and velar stops in 3-year-old bilingual children and their age-matched monolingual peers. *Clinical Linguistics and Phonetics, 26*(2), 148-163.

Feng, Y., & Peng, G. (2023). Development of categorical speech perception in Mandarin- speaking children and adolescents. *Child Development, 94*(1), 28-43.

Fennell, C.T., Tsui, A.S-M., & Huddon, T.M. (2016). Speech perception in simultaneous bilingual infants. In E. Nicoladis, & S. Montanari, S. (eds), *Bilingualism across the lifespan: Factors moderating language proficiency* (pp. 43-62). American Psychological Association.

Ferjan Ramirez, N., Ramirez, R.R., Clark, M., Taulu, S., & Kuhl, P.K. (2016). Speech discrimination in 11-month-old bilingual and monolingual infants: A magnetoencephalography study. *Developmental Science, 20*(1), e12427.

Flege, J. (1995). Second-language speech learning: Theory findings, and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 229–273). York Press.

Flege, J. (2002). Interactions between the native and second language phonetic systems. In P. Burmeister, T. Piske, & A. Rohde (eds.), *An integrated view of language development: Papers in honor of Henning Wode* (pp. 217-244). Wissenschaftlicher Verlag.

Flege, J.E., & Bohn, O.-S. (2021). *The Revised Speech Learning Model (SLM-r).* Preprints. doi:10.13140/RG.2.2.27529.06249

Garcia-Sierra, A., Rivera-Gaxiola, M., Percaccio, C.R., Conboy, B.T., Romo, H., Klarman, L., Ortiz, S., & Kuhl, P.K. (2011). Bilingual language learning: An ERP study relating early brain responses to speech, language input, and later word production. *Journal of Phonetics, 39*(4), 546-557.

Giannakopoulou, A., Brown, H., Clayards, M., & Wonnacott, E. (2017). High or low? Comparing high and low-variability phonetic training in adult and child second language learners. *PeerJ, 5*, e3209.

Højen, A., & Flege, J. (2006). Early learners' discrimination of second-language vowels. *Journal of the Acoustic Society of America, 119*(5), 3072-3084.

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L.V. (2010). Sources of variability in children's language growth. *Cognitive Psychology, 61*(4), 343-365.

Ingvalson, E.M., Ettlinger, M., & Wong, P.C.M. (2014). Bilingual speech perception and learning: A review of recent trends. *International Journal of Bilingualism, 18*(1), 35–47.

Kalashnikova, M., & Carreiras, M. (2022). Input quality and speech perception development in bilingual infants' first year of life. *Child Development, 93*(1), e32-e46.

Kehoe, M.M., Lleó, C., & Rakow, M. (2004). Voice onset time in bilingual German-Spanish children. *Bilingualism: Language and Cognition, 7*(1), 71-88.

Khattab, G. (2003). Age, input, and language mode factors in the acquisition of VOT by English-Arabic bilingual children. *Proceedings of the International Congress of Phonetic Sciences, 15,* 3213-3216.

Kingston, J., Levy, J., Rysling, A., & Staub, A. (2016). Eye movement evidence for an immediate Ganong effect. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(12), 1969.

Konefal, J.A., & Fokes, J. (1981). Voice onset time: The development of Spanish-English distinction in normal and language disordered children. *Journal of Phonetics, 9,* 437–444.

Kuhl, P.K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5,* 831–8 43.

Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory

expanded (NLM-e). *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 363*(1493), 979–1000.

Lenth, R.V. (2021). *emmeans: Estimated marginal means, aka least-squares means*. R package version 1.7.1-1. https://CRAN.R-project.org/package=emmeans

Lleó, C., Benet, A., & Cortés, S. (2007). Some current phonological features in the Catalan of Barcelona. *Catalan Review, 21*(1), 279–300.

Lleó, C., Cortés, S., & Benet, A. (2008). Contact-induced phonological changes in the Catalan spoken in Barcelona. In P. Siemund, & N. Kintana (eds.), *Language contact and contact languages* (pp.185–212). John Benjamins.

Logan, J.S., Lively, S.E., & Pisoni, D.B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America, 89*(2), 874-86.

Lyytinen, H., Erskine, J., Hämäläinen, J., Torppa, M., & Ronimus, M. (2015). Dyslexia—early identification and prevention: Highlights from the jyväskylä longitudinal study of dyslexia. *Current Developmental Disorders Reports, 2*(4), 330–338.

Macken, M., & Barton, D. (1979). The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language, 7,* 41–74.

Macken, M., & Barton, D. (1980). The acquisition of the voicing contrast in Spanish: A phonological study of word-initial stop consonants. *Journal of Child Language, 7,* 433–458.

Magnuson, J.S., & Nusbaum, H.C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance, 33*(2), 391.

Makowski, D., Ben-Shachar, M.S., & Lüdecke, D. (2019a). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, *4*(40), 1541.

Makowski, D., Ben-Shachar, M.S., Chen, S.A., & Lüdecke, D. (2019b). Indices of effect existence and significance in the Bayesian framework. *Frontiers in Psychology*, *10*, 2767.

Makowski, D., Ben-Shachar, M.S., Patil, I., & Lüdecke, D. (2020). *Estimation of model-based predictions, contrasts and means*. CRAN.

Mayr, R., & Montanari, S. (2015). Cross-linguistic interaction in trilingual phonological development: The role of the input in the acquisition of the voicing contrast. *Journal of Child Language, 42*(5), 1006-1035.

Mayr, R., & Siddika, A. (2018). Inter-generational transmission in a minority language: Stop consonant production by Bangladeshi heritage children and adults. *International Journal of Bilingualism,* 22(3), 255-284.

McCarthy, K., Mahon, M., Rosen, S., & Evans, B.G. (2014). Speech perception and production by sequential bilingual children: A longitudinal study of voice onset time acquisition. *Child Development, 85*(5), 1965-1980.

Montanari, S., Mayr, R., & Subrahmanyam, K. (2018). Bilingual speech sound development during the preschool years: The role of language proficiency and cross-linguistic relatedness. *Journal of Speech, Language, and Hearing Research, 61*, 2467-2486.

Mora, J.C., & Nadeu, M. (2012). L2 effects on the perception and production of a native vowel contrast in early bilinguals. *International Journal of Bilingualism, 16*(4), 484–500.

Muru, A., & Lee, S.A. (2017). Development of phonetic categories of stop consonants in Spanish-English bilingual children. *Clinical Archives of Communication Disorders, 2*(1), 60-68.

National Academies of Sciences, Engineering, and Medicine. (2017). *Promoting the educational success of children and youth learning English: Promising futures.* The National Academies Press.

Nearey, T.M., & Rochet, B.L. (1994). Effects of place of articulation and vowel context on VOT production and perception for French and English stops. *Journal of the International Phonetic Association, 24*(1), 1-18.

Netelenbos, N., & Li, F. (2013). The production and perception of voice onset time in English-speaking children enrolled in a French-immersion program. *Proceedings of the 14th Annual Conference of the International Speech Communication Association* (INTERSPEECH), Lyon, France (pp. 2380–2383).

Nittrouer, S., & Burton, L. (2005). The role of early language experience in the development of speech perception and phonological processing abilities: Evidence from 5-year-olds with histories of otitis media with effusion and low socioeconomic status. *Journal of Communication Disorders, 38,* 29–63.

Oller, D.K., Jarmulowicz, L., Pearson, B.Z., & Cobo-Lewis, A.B. (2011). Rapid spoken language shift in early second-language learning: The role of peers and effects on the first language. In A. Y. Durgunoğlu & C. Goldenberg (eds.), *Language and literacy development in bilingual settings* (pp. 94–120). The Guilford Press.

Paradis, J., Genesee, F., & Crago, M. (2021). *Dual language development and disorders* (3rd ed.). Brookes.

Place, S., & Hoff, E. (2011). Properties of dual language exposure that influence 2-year-olds' proficiency. *Child Development, 82*(6), 1834–1849.

Place, S., & Hoff, E. (2016). Effects and noneffects of input in bilingual environments on dual language skills in 2 ½-year-olds. *Bilingualism: Language and Cognition, 19,* 1023–1041.

Polka, L., & Bohn, O.S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics, 39,* 467–4 78. wocn.2010.08.007

Polka, L., & Werker, J.F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 421–435.

Port, R.F., & Rotunno, R. (1979). Relation between voice-onset time and vowel duration. *The Journal of the Acoustical Society of America*, *66*(3), 654-662.

Posit team (2022). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. URL http://www.posit.co/.

Procter, A., Bunta, F., & Aghara, R. (2015). Stop VOT productions by young bilingual Spanish-English children and their monolingual peers. In M. Yavas (ed.), *Unusual productions in phonology: Universals and language-specific considerations* (pp. 226-241). Psychology Press.

Qualtrics Experiment Management. Provo, UT. Version October 2022. www.qualtrics.com

Ramón-Casas, M., Cortés, S., Benet, A., Lleó, C., & Bosch, L. (2023). Connecting perception and production in early Catalan–Spanish bilingual children: Language dominance and quality of input effects. *Journal of Child Language, 50*(1), 155–176.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. UR https://www.R-project.org/.

Stager, C.L., & Werker, J.F. (1997). Infants listen for more phonetic detail in speech perception than in word-l earning tasks. *Nature, 388,* 381– 382.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 7*(5), 1074.

Sundara, M., Polka, L., & Molnar, M. (2008). Development of coronal stop perception: Bilingual infants keep pace with their monolingual peers. *Cognition, 108*(1), 232-242.

Toscano, J.C., & McMurray, B. (2012). Cue-integration and context effects in speech: Evidence against speaking-rate normalization. *Attention, Perception, & Psychophysics, 74*(6), 1284-1301.

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics, 33,* 263–290.

U.S. Department of Health and Human Services, Administration for Children and Families, & Office of Head Start. (2014). *Head Start program fact sheet, Fiscal year 2014.* http://eclkc.ohs.acf.hhs.gov/hslc/data/factsheets/docs/hs-program-fact-sheet-2014.pdf

van der Leij, A. (2013). Dyslexia and early intervention: What did we learn from the Dutch Dyslexia Programme? *Dyslexia, 19*(4), 241–255.

Werker, J.F., & Tees, R.C. (1984). Cross-language speech-perception— Evidence for perceptual reorganization during the 1st year of life. *Infant Behavior and Development, 7,* 49– 63.

Williams, L. (1977). The voicing contrast in Spanish. *Journal of Phonetics, 5*(2), 169-184.

Williams, L. (1979a). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception and Psychophysics, 21*, 289-297.

Williams, L. (1979b). The modification of speech perception and production in second-language learning. *Perception and Psychophysics, 26,* 95-104.

Winn, M.B. (2020). Manipulation of voice onset time in speech stimuli: A tutorial and flexible Praat script. *The Journal of the Acoustical Society of America*, *147*(2), 852-866.

Yu, V.Y., De Nil, L.F., Peng, E.W. (2015). Effects of age, sex and syllable structure on voice onset time: Evidence from children's voiceless aspirated stops. *Language and Speech, 58*(Pt2), 152-167.

Zlatin, M., & Koenigsknecbt, R. (1975). Development of the voicing contrast: Perception of stop consonants. *Journal of Speech, Language and Hearing Research, 18*(3), 541-553.

Zlatin, M., & Koenigsknecbt, R. (1976). Development of the voicing contrast: A comparison of voice onset time in stop perception and production. *Journal of Speech, Language and Hearing Research,* 19(1), 93-111.

Appendix

Table A1. Model summary for the analysis of language experience

Predictors	*Log-Odds*	*CrI (95%)*
Intercept	0.87	0.56 – 1.22
step.scaled	2.77	2.24 – 3.44
language experience	-0.61	-1.21 – -0.02
contrast	0.05	-0.59 – 0.68
step.scaled:lang exp	0.03	-0.88 – 0.91
step.scaled:contrast	1.38	0.60 – 2.44
lang exp:contrast	0.30	-0.82 – 1.40
step.scaled:lang exp:contrast	0.28	-1.01 – 1.55

Table A2. Model summary for the analysis of the moderators of bilinguals' perception of English stimuli

Predictors	*Log-Odds*	*CI (95%)*
Intercept	1.41	0.83 – 2.05
step.scaled	3.28	2.28 – 4.41
contrast	-0.29	-1.26 – 0.69
lang exp	-0.25	-1.33 – 0.86
age	0.44	-0.16 – 1.06
inputs	-0.37	-1.15 – 0.50
step.scaled:contrast	0.91	-0.39 – 2.36
step.scaled:lang exp	1.06	-0.75 – 2.67

contrast: lang exp	-0.12	-1.76 – 1.53
step.scaled:age	0.51	-0.59 – 1.61
contrast:age	-0.92	-1.98 – 0.14
lang exp:age	-0.39	-1.56 – 0.80
step.scaled:inputs	-0.68	-2.01 – 0.62
contrast:inputs	0.66	-0.60 – 1.89
lang exp:inputs	0.52	-1.03 – 2.03
age:inputs	0.41	-0.38 – 1.22
step.scaled:contrast: lang exp	0.78	-1.29 – 2.83
step.scaled:contrast:age	-0.44	-1.93 – 0.99
step.scaled:lang exp: age	0.85	-0.96 – 2.64
contrast:lang exp:age	1.46	-0.27 – 3.28
step.scaled:contrast: inputs	-0.10	-1.70 – 1.47
step.scaled: lang exp :inputs	-0.96	-3.00 – 1.12
contrast: lang exp :inputs	0.48	-1.59 – 2.54
step.scaled:age:inputs	-0.12	-1.36 – 1.08
contrast:age:inputs	-0.04	-1.26 – 1.23
lang exp:age:inputs	-0.62	-2.08 – 0.92
step.scaled:contrast: lang exp:age	-0.51	-2.66 – 1.65
step.scaled:contrast: lang exp:inputs	-0.29	-2.59 – 2.11
step.scaled:contrast: age:inputs	-0.08	-1.61 – 1.46

step.scaled:lang exp: age:inputs	-1.18	-3.14 – 0.93
contrast:lang exp: age:inputs	0.18	-1.87 – 2.23
step.scaled:contrast: lang exp:age:inputs	-0.40	-2.72 – 1.96

Table A3. Model summary for the analysis of the moderators of bilinguals' perception of Spanish stimuli

Predictors	*Log-Odds*	*CI (95%)*
Intercept	-0.48	-1.23 – 0.28
step.scaled	2.07	1.48 – 2.78
contrast	0.01	-1.22 – 1.24
lang exp	0.43	-0.90 – 1.68
age	-0.08	-0.85 – 0.68
inputs	-0.02	-0.70 – 0.63
step.scaled:contrast	-0.11	-1.16 – 0.93
step.scaled:lang exp	0.99	-0.08 – 2.09
contrast:lang exp	-0.02	-1.95 – 1.87
step.scaled:age	-0.02	-0.70 – 0.66
contrast:age	-0.39	-1.68 – 0.89
lang exp:age	-0.70	-2.09 – 0.68
step.scaled:inputs	-0.26	-0.85 – 0.28
contrast:inputs	-0.32	-1.39 – 0.80
lang exp:inputs	0.42	-0.77 – 1.65
age:inputs	0.14	-0.46 – 0.69

step.scaled:contrast: lang exp	-0.22	-1.91 – 1.45
step.scaled:contrast:age	0.56	-0.56 – 1.68
step.scaled:lang exp: age	1.14	-0.04 – 2.43
contrast:lang exp:age	-0.55	-2.56 – 1.50
step.scaled:contrast: inputs	-0.19	-1.15 – 0.74
step.scaled:lang exp :inputs	-0.24	-1.30 – 0.82
contrast:lang exp :inputs	-1.88	-3.69 – 0.05
step.scaled:age:inputs	-0.04	-0.54 – 0.46
contrast:age:inputs	-0.36	-1.37 – 0.60
lang exp:age:inputs	0.13	-0.91 – 1.20
step.scaled:contrast: lang exp:age	0.86	-1.07 – 2.72
step.scaled:contrast: lang exp:inputs	-0.09	-1.75 – 1.48
step.scaled:contrast: age:inputs	-0.08	-0.94 – 0.82
step.scaled:lang exp: age:inputs	-0.17	-1.16 – 0.74
contrast:lang exp: age:inputs	0.33	-1.41 – 1.98
step.scaled:contrast: lang exp:age:inputs	-0.83	-2.43 – 0.67

Table A4. Model summary for the analysis of bilinguals' perception of Spanish and English stimuli

Predictors	*Log-Odds*	*CrI (95%)*
Intercept	-0.07	-0.40 – 0.24
contrast	-0.33	-1.01 – 0.33
step.scaled	1.78	1.33 – 2.36
stimulus language	1.48	0.58 – 2.43
contrast:step.scaled	-0.21	-1.11 – 0.69
contrast:stim language	1.10	-0.14 – 2.37
step.scaled:stim language	-0.62	-1.44 – 0.16
contrast:step.scaled:stim language	-0.64	-2.04 – 0.73