English vowel perception in Spanish-English 1 bilingual preschoolers

Multiple-talker input is only beneficial for children with high language exposure levels

Citation for published version:
Montanari, S, Steffman, J & Mayr, R 2024, 'English vowel perception in Spanish-English 1 bilingual preschoolers: Multiple-talker input is only beneficial for children with high language exposure levels', *Journal of Speech, Language, and Hearing Research*, vol. 67, no. 10, pp. 3643-3659. https://doi.org/10.1044/2024_JSLHR-24-00044

Digital Object Identifier (DOI):
10.1044/2024_JSLHR-24-00044

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Journal of Speech, Language, and Hearing Research

English Vowel Perception in Spanish-English Bilingual Preschoolers:
Multiple-Talker Input is only Beneficial for Children with High Language Exposure Levels

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, The University of Edinburgh

[3]Centre for Speech, Hearing and Communication Research, Cardiff Metropolitan University

Conflict of interest statement:

We have no conflict of interest to disclose

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

Abstract

Purpose: This study examines English vowel perception in Spanish-English bilingual preschoolers, comparing bilinguals' perception patterns to those of monolinguals, and examining how child-internal (age) and external variables (input quantity and input diversity) predict perceptual performance.

Method: Sixty children between 3;6 and 5;6 of age participated in the study, 28 of whom were Spanish-English bilinguals and 32 English monolinguals. Perception was assessed through a forced-choice minimal-pair identification task in which children heard synthesized audio stimuli (i.e., "sheep" and "ship") that varied systematically along the /i-ɪ/ continuum and were asked to match them with one of two pictures. The data were analyzed with Bayesian mixed-effects logistic regression analyses, modeling responses as a function of continuum step, language background (monolingual or bilingual), age, English exposure (i.e., input quantity), and number of English input providers (i.e., input diversity).

Results: The results indicate that, despite displaying non-native English stop voicing perception in a previous study, the same bilingual children showed English /i-ɪ/ perception patterns that did not differ from those of monolinguals. While age did not predict vowel perception, input quantity and diversity jointly interacted to moderate how well children perceived the /i-ɪ/ contrast. Specifically, diverse input promoted perceptual performance in children who received high levels of English exposure – and who presumably had more advanced English language skills, whereas it limited perceptual performance in children with more limited English exposure and skills.

Conclusions: This study shows that bilingual children can show monolingual-like perception patterns for some sounds while displaying non-target perception for others. This is the first study

to demonstrate that language exposure mediates the role of input diversity on speech sound development, suggesting that varied input can be more or less beneficial for speech sound development based on the learner's language learning stage.

Keywords: speech perception; vowel perception; Spanish-English bilingual children; input quantity; input diversity

1. Introduction

While a wealth of studies has examined the speech perception abilities of simultaneous bilingual infants (see Sundara, 2022, for a review), few investigations have been devoted to how bilingual children perceive speech sounds at preschool age, when children develop important pre-literacy skills that lay the foundation for their future academic success. Understanding preschoolers' speech perception abilities is important for several reasons. First, children's ability to perceive speech sounds is related to both their pre-reading skills, such as their ability to identify and manipulate individual sounds in spoken words (i.e., phonemic awareness), and to their later reading and writing skills (Lyytinen et al., 2015; van der Leij, 2013). Thus, studying bilingual children's English perception skills before formal schooling begins has important educational consequences. Second, the number of dual language learners – i.e., children who speak a language other than English at home – has grown significantly in the U.S. in the past decades (for example, in states like California they represent more than 60% of children aged 0-8 years), suggesting that these children constitute a large segment of the U.S. child population (Migration Policy Institute, 2019). Furthermore, the vast majority of dual language learners live in low-income families (Migration Policy Institute, 2019), a situation that may have consequences for the quality of parent-child interactions and, in turn, young children's cognitive and language outcomes (Perkins et al., 2013). Hence, investigating young bilinguals' English perception abilities at preschool age can inform current curricular and instructional approaches, possibly mitigating the negative effects of poverty and reducing gaps in achievement between bilingual and monolingual children before they begin formal schooling.

Investigating speech perception at preschool age also has theoretical implications. The literature on bilingual perception in infancy shows that simultaneous bilinguals who learn prosodically similar languages (like Spanish and Catalan) may experience protracted development as they develop categories in two languages that are characterized by competing phonetic and distributional properties (Sundara & Scutellaro, 2011). The limited research with preschool-age bilinguals also suggests that they may display perceptual patterns that differ from monolingual patterns despite several years of experience and regular, daily use of each of their languages (Montanari et al., 2023; Ramón-Casas et al., 2023). Studies that examine 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) further confirm that speakers without regular and extensive exposure to L2 before age 2 or 3 may never attain native-like perceptual performance. Thus, examining speech perception at an age when children are expected to refine their phonetic categories can shed light on the development of phonetic representations in bilingual speakers between infancy and adulthood.

This study focuses on English vowel perception in Spanish-English bilingual preschoolers. Specifically, it aims to (1) compare bilingual children's perception patterns to those of English monolingual peers; and (2) examine the role of specific child-internal (age) and external variables (input characteristics such as input quantity and input diversity) as predictors of English perceptual performance. Investigating English perception in young preschoolers can shed light on the consequences of reduced input and the impact of specific child and input characteristics on English perceptual performance, contributing to the refinement of current models of bilingual speech perception and of curricular and instructional approaches for this population.

1.1 Bilingual Speech Sound Perception: The Relevance of Age, Input Quantity and Input Diversity

Considerable research has tracked the time course of first-language speech perception development, revealing that children transition from a phase of universal sound discrimination –

during which they are sensitive to native and non-native speech sounds – to a phase of increased attunement to and improved discrimination of the sounds of their native language (Polka & Bohn, 2011; Polka & Werker, 1994; Werker & Tees, 1984). Studies of simultaneous bilingual infants have shown that they follow a similar developmental trajectory as monolingual children, at least when they are learning prosodically-different languages that can be discriminated against at birth (Sundara et al., 2008). However, when learning prosodically-similar languages such as Spanish and Catalan or English and German, bilingual children show a temporary delay in their attunement to native categories, suggesting that they may require additional accumulated exposure to the two languages in order to track the competing distributional properties of those categories (Bosch & Sebastián-Gallés, 2003; Sundara & Scutellaro, 2011). 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), indicating that bilingual infants may rely on their auditory perceptual sensitivities (characteristic of early universal sound discrimination) rather than language-specific experience for an extended period of time during infancy to distinguish speech sound categories.

Studies examining bilingual perception at preschool age also show that children may display perceptual patterns that differ from monolingual patterns despite early and consistent exposure to both languages. For instance, Montanari et al. (2023), who examined stop voicing perception in both the societal (English) and heritage (Spanish) language of bilingual preschoolers (aged 4;7 on average), found that perceptual performance in English was affected by Spanish experience. Specifically, the bilingual children's category boundaries for English /b-p/ and /d-t/ were positioned at lower Voice Onset Time (VOT) values as compared to those of English monolingual peers, consistent with the lower VOT values that characterize the Spanish

/b-p/ and /d-t/ categorical boundaries. Similarly, McCarthy et al. (2014), who assessed English /b-p/ and /g-k/ perception in Sylheti/English bilingual children, documented perception patterns that were different from those of English monolinguals and were affected by Sylheti categories at age 4;5, before the children had spent a significant amount of time in English-language preschool. Ramón-Casas et al. (2023) found differences in vowel perception (the Catalan /e-ɛ/ contrast) even in groups of bilingual children, such that Catalan-dominant 4.5-year-old bilinguals outperformed Spanish-dominant bilingual peers in identifying correct and mispronounced words containing the /e-ɛ/ contrast despite the latter having regular and consistent exposure to Catalan from early in life (i.e., before age 3). Yu et al. (2019) documented bilingual-monolingual differences even in the neural indices of English vowel discrimination in monolingual and bilingual children between 3 and 47 months. Specifically, the authors found that Spanish-dominant children displayed neural responses to the American English vowel contrast /ɛ-ɪ/ that were different from those of their monolingual English peers. However, English-dominant bilinguals' neural indices did not differ from those of monolinguals, suggesting that language exposure modulates how speech is processed in the brain. Overall, the results of these studies demonstrate that, in the presence of two phonological systems, phonemic representations in one language may influence those in the other language leading to deviations from monolingual patterns, especially in the presence of limited input.

Studies have shown that different child-internal and external variables may be related to perceptual performance. As to child-internal factors, age has been shown to predict perceptual skills in the societal language, so that older children are better speech sound perceivers than younger children. For instance, in Montanari et al. (2023), age was the only predictor of perceptual performance in the societal language, English, whereas it did not moderate perception abilities in Spanish, the heritage language. The authors argued that, as children get older, they are educated in English and they interact increasingly more with members of the wider English-speaking community, thereby expanding their opportunities to both hear and speak the societal language and improve their perceptual abilities. On the contrary, input in a heritage language typically decreases with age (Oller et al., 2011), so that older children may not necessarily be the best perceivers of that language. McCarthy et al. (2014) also found that children's perceptual performance in the societal language, English, improved with age. Indeed, after 19 months of regular and consistent exposure to English in preschool, the perception patterns of their Sylheti/English bilingual participants differed from the Sylheti-influenced patterns the children displayed a year earlier and were no longer significantly different from those of monolinguals. The authors speculated that phonemic categorization may be initially affected by language dominance in young bilinguals. However, phonemic categories in the weaker language can be acquired and refined with language experience.

As to child-external variables, it appears that several input characteristics may be related to perceptual performance. One of these is input quantity, or the amount of exposure that children receive to a language. In Montanari et al. (2023), input quantity solely predicted perceptual performance in the heritage language, Spanish, whereas it did not moderate English perception. That is, children with higher Spanish exposure at the time of the study – but not older children – were better Spanish perceivers than those receiving less Spanish input, suggesting that, for the heritage language, input quantity is more important than age. Ramón-Casas et al. (2023) also found that the children with higher 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. Likewise, Yu et al. (2019) demonstrated that language exposure moderates vowel discrimination between 3 and 47 months of age, with only bilingual children with higher English exposure (as compared to those with more Spanish exposure) processing the English contrast similarly to monolinguals. Thus, not surprisingly, it appears that the more input is received in a language the better the perceptual abilities in that language.

Another input characteristic that may possibly affect perceptual performance is the quality of the speech that bilingual children hear in their environment. Measuring input quality in terms of the characteristics of infant-directed speech, Kalashnikova and Carreiras (2022) found that the extent to which individual mothers exaggerated vowels in their infant-directed speech (i.e., therefore increasing the quality of their input) significantly related to the perception of native and non-native phonemes in monolingual and bilingual 5- and 9-month-old infants. Specifically, infants who heard input characterized by exaggerated vowels were more ahead in their perceptual attunement than infants who heard less-diverse input that was less conducive to language learning. Ramón-Casas et al. (2023) measured input quality in terms of exposure to native (vs. accented) speech, assuming that accented input may result in children's building and maintenance of inaccurate representations. The authors found that bilingual preschoolers with more exposure to native Catalan were significantly better in /ɛ/-word production as compared to children who were more exposed to Spanish native input. However, the effect of native input exposure was not found for the perception of the Catalan /e-ɛ/ contrast. Montanari et al. (2023) measured input quality in terms of input diversity (that is, the number of input providers that children had in each language), since bilingual and monolingual research has found that hearing a language from different speakers is more supportive of language development than the same number of hours of language exposure from fewer speakers (Place & Hoff, 2011, 2016;

Huttenlocher et al., 2010). The authors tested the relation between input diversity and stop voicing perception in English and Spanish in their bilingual participants, but found that this measure was not a predictor of perceptual abilities in either language. Thus, the extant evidence is inconclusive as to whether input quality – as measured by different constructs – contributes to perceptual performance.

It is important to point out that studies that have shown that children benefit from input diversity have focused on vocabulary (Place & Hoff, 2011; Place & Hoff, 2016) and grammar (Huttenlocher et al., 2010; Place & Hoff, 2016). Specifically, Place and Hoff (2011) found that both the number of different English speakers as well the number of English conversational partners predicted Spanish-English bilingual toddlers' English vocabulary once the relative amount of English exposure was kept constant. Huttenlocher et al. (2010) further documented a syntactic advantage for 14-to-46-month-old English monolingual children who heard language from multiple interlocutors compared to those with fewer input sources. Likewise, Place and Hoff (2016) found that the number of input providers predicted Spanish-English bilingual toddlers' lexical and grammatical skills, although this was the case only for the heritage language (i.e., Spanish). Thus, it is possible, as speculated in Montanari et al. (2023), that hearing input from different speakers may be more important for lexical and grammatical development than for speech development. This is because less variable and ambiguous input may in some cases limit the amount of cross-linguistic interaction in speech and facilitate the adoption of speaker-specific patterns. Mayr and Montanari (2015), who studied two simultaneous trilingual sisters who heard Italian and English from multiple speakers but Spanish from a single interlocutor, provides evidence for this hypothesis. Specifically, the authors found extensive cross-linguistic interaction in Italian and English, whereas the girls' Spanish productions were largely unaffected by the other two languages and mirrored those of the sole Spanish input provider. Mayr and Montanari (2015) speculated that a single, unambiguous input source may lead to more firmly entrenched storage of speaker-specific phonetic information (Allen & Miller, 2004; Smith & Hawkins, 2012), which, in turn, may limit interactions from other phonological systems.

Studies of L2 speech learning in children also point to a possible advantage of hearing less variable L2 input. Indeed, research on the effectiveness of high- vs. low-variability phonetic L2 training have shown that, while adults benefit from high-variability training (e.g. Bradlow et al.,1999; Logan et al., 1991), children who are trained in L2 through a single speaker achieve higher L2 perception and production outcomes than children who are trained through different speakers (Alshangiti et al., 2019; Evans & Alshangiti, 2018; Giannakopoulou et al., 2017). This is because children find it harder than adults to adapt to multiple talkers (Magnuson & Nusbaum, 2007), and, when hearing an L2 from fewer speakers, they may more readily remember how these speakers produce a certain sound and use this information to shape their own perception and production (Alshangiti et al., 2019). Literature on non-native speech learning in adults with weak vs. strong phonological skills further shows that multiple-talker phonetic training enhances learning only for individuals with strong perceptual abilities (Perracchione et al., 2011). Learners with weaker perceptual abilities are actually impaired by multiple-talker training relative to a low-variability condition, where a novel phonological contrast is introduced by a single speaker. The authors speculated that in an environment with high acoustic-phonetic variability, there is no trial-by-trial consistency or predictability in the stimuli's phonetic features, and speakers with lower perceptual aptitude must employ additional processing resources to recognize speech sounds (Ingvalson et al., 2013; Perracchione et al., 2011). In sum, given these inconclusive findings regarding the advantages and disadvantages of input diversity for speech development, the question remains as to whether bilingual perceptual performance in early childhood is facilitated or hindered by input that is more diverse.

1.2 The Present Study

This study contributes to the literature on bilingual speech perception by examining English vowel perception in the same Spanish-English bilingual children who participated in Montanari et al. (2023). Our first goal is to compare the bilinguals' perception of the English /i-ɪ/ contrast to that of English monolingual peers to assess the extent to which exposure to an additional phonological system may exert influence on English sound discrimination. Recall that children's ability to perceive speech sounds is related to both their pre-reading skills, such as their ability to identify and manipulate individual sounds in spoken words (i.e., phonemic awareness), and to their later reading and writing skills (Lyytinen et al., 2015; van der Leij, 2013). Since Spanish-English dual language learners already lag behind English-speaking monolingual peers in important English language and literacy skills by the time they reach kindergarten (Hoff, 2013), examining their ability to hear speech sounds as compared to monolinguals before formal schooling begins can have important educational implications. Moreover, assessing speech perception at preschool age – when children fine-tune their phonetic categories – can provide insights on how phonetic representations develop in bilingual speakers between infancy and adulthood.

We focus on the English /i-ɪ/ vowel contrast because extensive work has been done on its perception by Spanish speakers (Flege 1991; Flege, Bohn & Jang, 1997; Fox, Flege & Munro, 1995; Morrison, 2008, 2009). These studies demonstrate that, at early stages of English acquisition, L1-Spanish listeners may perceive both /i/ and /ɪ/ as exemplars of the same Spanish category /i/ and therefore fail to discriminate them. More recent research confirms that even early Spanish-English bilinguals, whose exposure to English started in childhood, may assimilate

English /ɪ/ to Spanish /i/ (Baigorri et al., 2019). At the same time, the first-language speech perception development literature shows that vowel perception emerges earlier than consonant perception (Kuhl, 2004), with perceptual attunement for vowel categories occurring around 4–6 months of age but around 10–11 months for consonants (Polka & Bohn, 2011; Polka & Werker, 1994; Stager & Werker, 1997; Werker & Tees, 1984). Furthermore, vowel perception is continuous while consonant perception is categorical (Tyler et al., 2014), with boundaries between phonological categories being less sharp, and within-category discrimination better for vowels than consonants (Fry et al., 1962). These acoustic differences may result in differences in the timing in which native-like perception of vowels and consonants is achieved. Since the same children who participated in this study displayed English consonant perception patterns that differed from those of English monolinguals and that were affected by their experience with Spanish (Montanari et al., 2023), it is thus of interest to examine whether the same children show the same difficulty with an English vowel contrast.

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), which proposes that when L2 phones are similar to those of the L1, L1 categories act as attractors and L1 and L2 sounds may form merged representations that have compromise values that differ from those of monolinguals. We also ground our study in Best's L2 Perceptual Assimilation Model (PAM-L2, Best & Tyler, 2007). According to the model, L2 contrasts that are assimilated to a single native category are particularly hard to discriminate perceptually. We thus hypothesize, based on the SLM-r and PAM-L2 predictions as well as on the results of Montanari et al. (2023), that the discrimination of the English /i-ɪ/ contrast will be difficult as both sounds may be perceived as exemplars of the same Spanish category /i/.

Next, we examine the extent to which 1) age, 2) input quantity and 3) input diversity predict English perceptual patterns. While the SLM-r and PAM-L2 predict difficulties for the children's perception of the English vowels in this study due to their overlap with a single Spanish category, the SLM-r also proposes 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, given the participants' young age, we hypothesize that children will be able to improve their perceptual performance, with older children performing better than younger ones, since the former will have had more opportunities to hear and discriminate English phonemes (as in McCarthy et al., 2014, and Montanari et al., 2023). As to input quantity, we take *caregiver-reported English exposure* as a measure of this variable, hypothesizing that more English input will also be related to better perceptual performance (as in Ramón-Casas et al., 2023, and Yu et al., 2019). Finally, we take the *number of English input providers* as a measure of input diversity. English input was primarily provided by parents, teachers, and siblings who spoke English natively, i.e. they had acquired it from early in life. Thus, we speculated that these speakers possibly exposed children to higher-quality input characterized by a native accent that would have allowed children to build accurate and less variable representations. Nonetheless, given the mixed findings related to the benefits of input diversity for speech perception, we are unable to put forward a conclusive hypothesis as to whether hearing English from different speakers does indeed facilitate or hinder English perceptual performance.

2. Methods

2.1 Participants

Sixty children between 3;6 and 5;6 of age participated in the study. 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. Twenty-eight of the participants were Mexican-origin, Spanish-English bilinguals who had been raised in homes where they had regular and consistent exposure to both languages from early in life (i.e., before age 3), as they all had family members who spoke both Spanish and English (no other home languages were reported). Thus, they could be considered simultaneous bilinguals (Paradis et al., 2021), and they spoke both languages to different degrees. Thirty-two children were English monolinguals with no active use or knowledge of Spanish, although they also had limited exposure to Spanish due to the bilingual nature of their community. The bilingual children (8 males, 20 females; mean age: 54.7 months, $SD = 6.4$) were matched to the monolingual children (15 males,17 females; mean age: 51.1 months, $SD = 6.1$) in gender ($X^2(1) = 2.116$, $p = .146$), age ($t(59) = 1.56$, p >. 05), and SES as they came from low- to middle-income families from the same community. No child had a documented history of hearing, speech, language, cognitive, or neurological deficits according to parental reports.

Table 1 summarizes the information for the participants, including their amount of English exposure and the number of English input providers, as reported by the parents in a questionnaire. The amount of English exposure was measured in terms of relative exposure to English and Spanish 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," with higher scores thus representing higher exposure to English. While measures of absolute language exposure (i.e., 12 hours of exposure to Spanish and one hour to English) may better capture the number of hours a child hears a language, our goal was to differentiate children with more vs. less English input and we felt that our measure of relative language exposure was a time-efficient and parent-friendly method to reliably estimate a child's overall bilingual experience (Calandruccio et al., 2021). The number of English input providers was measured by asking parents to report which speakers spoke English to their child, with the possibility of including multiple English sources among "mother, father, siblings, grandparents, babysitter, teacher, media, and other" (thus, between 1 and 8 sources). While we could not assess the extent to which each English source spoke natively, the vast majority of the parents, teachers, and siblings who provided English input were native speakers; hence, we assumed that more English providers meant more opportunities to hear different English sounds, words and sentences, increasing the diversity, and thus the quality of the English input. Based on this information, the bilingual children's English exposure score was 3.32 (*SD* = 0.72), whereas the monolinguals' was 4.56 (*SD* = 0.5), a difference that was statistically significant ($t(59) = 7.79$, $p < .001$) and confirmed that the monolinguals heard more English than the bilinguals. As to input diversity, the bilingual children were exposed to an average of 4.14 English sources (*SD* = 1.21, range 1-6), whereas the monolinguals heard English through an average of 4.59 input providers (*SD* = 1.32, range 2-7). This difference was not statistically significant ($t(59) = 1.37$, $p > .05$), suggesting that both bilinguals and monolinguals were exposed to the societal language through a comparable number of speakers.

INSERT TABLE 1 HERE

The bilingual children were further divided into two groups based on their English exposure (HIGH-ENG, language exposure score of 4; LOW-ENG, language exposure scores of 1 to 3). Figure 1 shows the distributions of exposure scores and ages for bilingual children. Age,

Commented [SM1]: Can we use the same figure as in the JPhon paper?

Commented [JS2R1]: I think so!

unlike exposure, was treated as a continuous numerical variable as there was a spread of ages across the age range. However, for the exposure variable, there was an overwhelming majority of 3 and 4 scores, leading us to bifurcate the variable. This avoided reliance on very little data for the effect at the lower end of the exposure score scale. While it is true that the LOW-ENG children obtained a mean English exposure score (2.8, $SD = 0.54$) that did not seem to differ widely from the one obtained by the HIGH-ENG children (4.00, $SD = 0$), this difference was statistically significant ($t(27) = 7.53$, $p < .001$) and confirmed that the LOW-ENG children heard as much Spanish as English, whereas the HIGH-ENG group was exposed to more English than Spanish. The HIGH-ENG children (3 males, 9 females; mean age: 55.1 months, $SD = 6.17$) were matched to the LOW-ENG group (5 males, 11 females; mean age: 55.4 months, $SD = 6.68$) in age ($t(27) = 0.26$, $p > .05$). Similarly, the two groups did not differ in terms of the number of English input providers, as the HIGH-ENG children scored 4.33 (SD = 0.98, range 3-6), whereas the LOW-ENG children 4 ($SD = 1.37$, range 1-6), a difference that was not statistically significant ($t(27) = 0.72$, $p > .05$). Thus, while the two groups were exposed to English to different extents, they heard this language from a comparable number of speakers. Table 2 reports information for the two bilingual groups.

INSERT FIGURE 1 HERE

INSERT TABLE 2 HERE

2.2 Materials and Procedure

In order to assess the children' perception of the English /i-ɪ/ vowel contrast, we created a child-friendly forced-choice minimal-pair picture identification (2AFC) task in which children heard an auditory stimulus that varied systematically along the /i-ɪ/ continuum and were asked to match it with one of two pictures representing the minimal pair *sheep/ship.* The experiment was run through Qualtrics (Qualtrics, Provo UT).

A phonetic continuum ranging from *sheep* to *ship* was created by manipulating vowel formants and vowel duration using natural tokens of the target words produced by a female monolingual speaker of American English as initial values. The speaker was recorded in a quiet location using a Yeti Blue USB Microphone, and the recordings were then digitized at 44.1 kHz and 32 bit depth. Manipulation of the vowel continuum was carried out using a Praat script, employing the Burg Method of LPC decomposition and resynthesis (Winn, 2019). The formant values for each endpoint of the continuum were based on the naturally produced formants for /i/ and /ɪ/. The resynthesis process estimated source and filter models for one endpoint (the word with /i/). The filter model's F1, F2, and F3 were then interpolated linearly (in Bark space, Traunmüller, 1990) to the values of the other continuum endpoint, resulting in 10 intermediate filter steps. Phase-locked higher frequencies from the starting base file (/i/) that were lost in the process of LPC resynthesis were restored to all steps, improving the naturalness of the continuum. The result was a 10-step continuum ranging from /i/ to /ɪ/. We next manipulated vowel duration to co-vary with F1, F2 and F3 based on the speaker's natural productions, whereby /i/ was longer than /ɪ/ (in agreement with large-scale studies measuring vowel duration, Hillenbrand et al., 1995). The duration manipulation thus interpolated between endpoint duration values in 10 linearly equidistant steps in the way that duration and vowel formants would normally co-vary (e.g., the /i/ endpoint had the longest duration, the /ɪ/ endpoint had the shortest duration). In order to reduce the number of trials in an attempt to minimize fatigue effects, we selected 8 steps by excluding step 2 and step 9 (keeping the 6 most central steps in the continuum and the endpoints).

In a given trial, children heard an auditory stimulus and were asked to match it to the picture of a sheep or a ship presented on the screen. Since we also assessed the perception of /b-p/ and /t-d/ as reported in Montanari et al. (2023), there were a total of 24 trials (8 for /i-ɪ/, 8 for /b/-/p/, and 8 for /d/-/t/). The trials were randomly presented, and the placement (left vs. right) of the correct picture was counterbalanced across trials. While having only one trial per stimulus means that we cannot assess how stable a given child's perception of a particular token is (via multiple repetitions of each token), the benefit is to avoid fatigue and keep attention throughout the presentation of the stimuli. McCarthy et al. (2014) used an adaptive procedure with a similar intent of limiting fatigue effects, in their case to 40 trials. Thus, while we acknowledge that this smaller number of trials leaves us unable to assess intra-participant token reliability, the attentional benefits of keeping the experiment short outweigh this downside. Across participants, we are still able to assess variations as a function of language experience and other factors, and estimate uncertainty around our the effects, using Bayesian modeling (described below).

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. The experiment took between 4 and 5 minutes to complete. Children were familiarized with the stimuli one week before the experiment and could participate in the experiment only if they were able to identify the words before the task.

2.3 Analyses

We present two analyses, following Montanari et al. (2023). First we compared perception patterns across bilingual and monolingual participants, testing if the former differed from the latter in their categorization of the continuum (section 3.1.). Then, we considered moderators for the bilingual children's English perception, analyzing how age, English exposure (i.e. input quantity), and number of English input providers (i.e., input diversity) predicted categorization of

Commented [SM3]: One reviewer says: "I have some concerns with the methods and the results. It seems that the children heard each token on the /i-I/ continuum once. I realize that there are constraints when working with preschoolers, but only having one response per token may not capture how reliable children's responses are or how robust their categories are."
Jeremy, can you try to address this perhaps saying this is the typical methodology? We can also state that McCarthy et al. (2014) used a similar methodology and remind the reader about our exclusion criteria.

the continuum for the bilingual group only (section 3.2.). We excluded participants who did not show sensitivity to the vowel acoustic continuum. This was done by running individual-level regression analyses predicting each participants' categorization as a function of continuum step. We excluded participants who showed a zero estimate (no effect of step, i.e., flat categorization across all steps) or a negative estimate (the opposite of the predicted effect given the coding of variables, i.e., more /ɪ/ responses with /i/-like formants and duration). We reasoned that these patterns represented either a lack of attention in the task or a complete lack of perception of the contrast. This led us to exclude 9% of the data, specifically, approximately 8% of bilinguals' responses (2 participants) and 10% of monolinguals' responses (3 participants).

The participants' data were modeled using Bayesian logistic regression, as implemented in brms (Bürkner, 2017). In the first analysis, participants' response was predicted as a function of *continuum step* (centered and scaled) and *group*, which was contrast-coded with bilingual mapped to -0.5, and monolingual mapped to 0.5. The interaction of these two fixed effects was included as well. Random effects were by-*participant* random intercepts with a random slope for *continuum step* (no slope was included for *group* because it is a between-subjects variable). In the moderator based analysis, fixed effects in the model were *continuum step*, *age* (centered and scaled), *number of English input providers* (i.e. input diversity, also centered and scaled), and *English exposure* (i.e. input quantity), which was coded as binary due to the fact that the variable was almost entirely 3 and 4 values on the scale (see also Montanari et al., 2023). This binary variable was contrast-coded with HIGH ENG mapped to -0.5, and LOW ENG mapped to 0.5. The random effect structure of the model in the moderator analysis was the same as the first model. Priors for both models were weakly informative normal priors, for both the intercept and fixed effects, specified with a mean of 0 and a standard deviation of 1.5 (in log-odds).

Full model summaries are given in the appendix. In reporting effects in the text of the paper we provide the median posterior estimate for an effect and the 95% credible intervals for the estimate. These intervals are the upper and lower bounds containing 95% of the posterior estimate. When 95% CrI exclude the value of zero, this indicates that the model has estimated an effect with a consistent directionality and consistently non-zero value. We also report "probability of direction" (henceforth pd), computed with the *bayestestR* package (Makowski et al., 2019a). This metric indexes the percentage of the posterior which has a given sign: a posterior distribution centered on a value of zero (hence, no evidence for an effect) would have a pd value of 50, 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., 2019a, 2019b).

3. Results

3.1 Bilingual Children's Perception of the English Vowel Contrast as Compared to Monolinguals

The first analysis found an expected effect of *continuum step*, showing that increases in continuum step (numerically from low to high) resulted in an increase in the log-odds of a /i/ response ($\beta = 2.04$, CrI = [1.61,2.54], pd = 100). We found no credible evidence for a difference between monolingual and bilingual children, either in the main effect of *group* ($\beta = -0.18$, CrI = [-1.05,0.68], pd = 66), or interaction with *continuum step* ($\beta = -0.07$, CrI = [-0.82,0.68], pd = 57). The data thus suggests no clear difference between the groups' categorization of the vowel continuum, and hence a different pattern compared to the one we documented in Montanari et al. (2023) for stop consonants. These results are shown in Figure 2.

INSERT FIGURE 2 HERE

3.2 Moderators of Bilingual Children's English Vowel Perception

The moderator-based analysis found credible evidence for the existence of a three-way interaction between *continuum step*, *English exposure*, and *number of English input providers* (β = -1.81, CrI = [-3.88,0.26], pd = 96). No other fixed effects or interactions in the model were credible (pds < 92), save for the expected effect of *continuum step* (β = 2.50, CrI = [1.71,3.49], pd = 100). Of note, there was no overall difference between the HIGH-ENG and LOW-ENG groups (pd = 89).

To examine the three-way interaction further, the conditional effects from the model are plotted in Figure 3. Panel A shows categorization for the HIGH-ENG and LOW-ENG groups (indicated by line color and type) at three levels of the variable "number of English input providers" (panels); fewer input sources are shown on the left panel (i.e., 2 sources of English input), more input sources on the right (4 and 6, respectively in the mid and right panel). As can be seen, the HIGH-ENG and LOW-ENG groups are similar in their categorization of the continuum at low- to mid-levels of input diversity (characterized 2 and 4 English input sources, respectively, left and mid panels). However, for the LOW-ENG exposure group only (rightmost panel), increases in input diversity (i.e. 6 English input sources) result in shallower (i.e., less mature) and less-anchored categorization of the continuum. For the HIGH ENG group the pattern goes in the opposite direction, with higher input diversity resulting in slightly steeper (i.e., more mature) categorization. Panel B shows the model estimates for the slope of the continuum step variable, where larger values correspond to steeper categorization. The value of the estimates becomes smaller, such that a smaller (shallower) effect of step is observed as input diversity increases for the LOW-ENG group. Overall, these results suggest that input diversity promotes perceptual performance at relatively high levels of English exposure - and more advanced stages of language learning, when children can possibly use input from multiple speakers to *refine* their perceptual categories. However, at low levels of exposure and earlier stages of learning, hearing input from multiple speakers may make it harder for children to distinguish between new perceptual categories, thus limiting perceptual performance.

INSERT FIGURE 3 HERE

4. Discussion

This study aimed at contributing to the literature on bilingual speech perception by examining English vowel perception in the same Spanish-English bilingual children who participated in Montanari et al. (2023). Since these children displayed English consonant perception patterns that differed from those of monolinguals and that were affected by their experience with Spanish, our first goal was to compare the bilinguals' perception of the English /i-ɪ/ contrast to that of English monolingual peers to assess the extent to which exposure to an additional phonological system influenced English vowel identification. Extensive work has shown that L1-Spanish listeners struggle with the perception of English /ɪ/ as they may assimilate it to Spanish /i/ (Baigorri et al., 2019). At the same time, vowel perception emerges earlier than consonant perception in infancy (Kuhl, 2004) and is continuous rather than categorical (Tyler et al., 2014), suggesting that bilingual children may develop native-like perception of vowels earlier than of consonants. Next, we examined the extent to which both child-internal (i.e. age) and external factors (i.e. input quantity and diversity) predicted English perceptual patterns. We took caregiver-reported English exposure as a measure of input quantity and the number of English input providers as a measure of input diversity, which we used as a proxy for input quality. To our knowledge, this is the first study that examines the role of child and input characteristics on the English perceptual performance of Spanish-English bilingual preschoolers.

The first analysis revealed that bilinguals' perception of the English /i-ɪ/ contrast did not differ from the monolinguals', suggesting that experience with the Spanish vowel system did not affect English vowel perception. This was unexpected as the same children displayed non-native

English consonant perceptual patterns that were affected by their experience with Spanish. Specifically, when presented with the English /b-p/ and /d-t/ continua, the bilingual children heard overall more voiceless stops compared to the monolinguals, in line with the fact that, in Spanish, positive VOT values (all of the English continua) are mapped to the voiceless category, whereas English speaking monolinguals map positive short lag VOT to a voiced category. Our results also did not align with the predictions of Flege's SLM-r model (Flege & Bohn, 2021) and Best's L2 Perceptual Assimilation Model (PAM-L2, Best & Tyler, 2007), which posit difficulty with the perception of L2 categories that are similar to those of the L1 or assimilated to a single native category, postulating that both /i/ and /ɪ/ will be perceived as exemplars of the same Spanish category /i/ (Baigorri et al., 2019). One possibility is that the English /i-ɪ/ contrast is less susceptible to cross-linguistic interaction than the stop voicing contrasts. After all, /ɪ/ is significantly lower than English and Spanish /i/ in vocalic space, is associated with higher F1 values in acoustic terms, and, by four and a half, children might have had sufficient exposure to English – and to the bimodal /i-ɪ/ distribution – to contrast /ɪ/ from both English and Spanish /i/ and form a separate category for this English-specific phoneme. On the contrary, stop voicing perception leads to high levels of cross-linguistic interaction because the same VOT values – short-lag VOT – must be perceived as voiced in English but as voiceless in Spanish (Montanari et al., 2023). Production studies in Spanish-English bilingual children confirm difficulties with VOT patterns due to one phonological system affecting the other (Konefal & Fokes, 1981; Mayr & Montanari, 2015; Montanari et al., 2018; Muru & Lee, 2017, Procter et al., 2015). Hence, our results may stem from the articulatory, acoustic, and phonological differences between the vowel and consonant contrasts that have been the focus of study.

Another possibility is that vowel perception emerges earlier than consonant perception, as shown in the first-language speech perception development literature (Kuhl, 2004), and that bilinguals' English exposure had been sufficient for vowel – but not consonant – perceptual attunement. For instance, studies on such attunement – when children transition from universal to language-specific speech-sound discrimination, characterized by increased sensitivity to native phonetic contrasts but decreased sensitivity to non-native contrasts (Kuhl, 2004) – show that this process occurs around 4–6 months for vowel categories and around 10–11 months for consonants (Polka & Bohn, 2011; Polka & Werker, 1994; Stager & Werker, 1997; Werker & Tees, 1984). These studies suggest that vowel discrimination may emerge earlier than consonant discrimination, since, compared to consonants, vowels are of higher acoustic intensity; they are more extended temporally (Ladefoged, 2005); they are more clearly heard in the womb, and attunement for vowels begins with speech exposure before birth (Granier-Deferre et al., 2011). A host of infant, child, and adult psycholinguistic evidence also suggests that vowels and consonants are processed by distinct neural mechanisms (Caramazza et al., 2000), and that their role in language is functionally different (Bonatti et al., 2005). Thus, it is possible that our bilingual participants – who had heard both Spanish and English from early in life – had already fully attuned to English vowels but not to consonants, establishing native-like perceptual performance with the former but not with the latter. The fact that age did not moderate vowel perception confirms this hypothesis.

The analysis of the moderators of bilingual children's English vowel perception also produced surprising results, with both input quantity and input diversity jointly contributing to English perceptual performance in an unexpected way. Specifically, the HIGH- and LOW-ENG children did not differ in their perceptual performance when exposed to low- to mid-levels of input diversity. However, hearing input from multiple speakers enhanced the perceptual skills of children who heard more English and who, presumably, had more advanced language skills. On the contrary, input diversity had the opposite effect in the context of limited English exposure, with LOW-ENG children displaying less mature perceptual performance when hearing English from multiple speakers. The results are puzzling because they show that input diversity can be both beneficial and detrimental for perceptual performance based on the learner's language exposure patterns and, presumably, learning stage.

We speculate that input diversity promotes perceptual performance at relatively high levels of English exposure – and more advanced stages of language learning, when children can possibly use input from multiple speakers to *refine* their perceptual categories. However, at low levels of exposure and earlier stages of learning, hearing input from multiple speakers may make it harder for children to distinguish between, and form, new perceptual categories, thus limiting perceptual performance. Recall that adults with weak perceptual abilities – who may possibly be comparable to learners at early stages of learning – have also been shown to be impaired by multiple-talker as opposed to single-speaker phonetic training (Perracchione et al., 2011). A recent study by Lev-Ari (2018) that used computational simulations further supports this and our findings. These simulations showed that while adults exposed to increased input variability were better at phonemic categorization, learners at earlier stages of development did not benefit from input diversity. In particular, adults who had larger social networks, i.e. were exposed to higher levels of input variability, were better at vowel perception in noise than speakers with smaller social networks, irrespective of cognitive abilities. Thus, having a larger social network increased the variability in the input, and this greater input variability led to better phoneme categorization. However, a comparison of "child" and "adult" simulations showed that this positive effect of input variability did not apply at the earliest stage of learning, as it rendered the categories harder to distinguish. Specifically, when the vowel categories were not yet known and still needed to be learned, learners exposed to more input variability incorrectly formed large categories that comprised several target vowels. Thus, input variability limited performance in these early stages of learning, because it made it harder to distinguish between new vowels. In contrast, input that was more homogeneous, i.e., less diverse, was more useful for identifying and separating new vowels. Overall, the results of our study are important as they reconcile previous conflicting findings as to the benefits/drawbacks of input variability – and of high- vs low-variability phonetic training – for speech sound development. Indeed, our study is the first to show, using real perception data from bilingual children rather than simulation experiments with adult learners, that varied input can be more or less beneficial for speech sound development based on the learner's language learning stage.

It remains unclear why input diversity contributed to English vowel perception but not to consonant perception (Montanari et al., 2023). Unlike consonants, vowels carry intrinsic pitch (Russo et al., 2019) and they are exaggerated in infant-directed speech (Kalashnikova & Carreiras, 2022). Presumably, children are exposed to speech containing exaggerated vowels – but not exaggerated consonants – throughout their first years of life. It is thus possible that infant-directed speech makes vowels more noticeable to children with respect to consonants, promoting earlier attunement to vowel categories, as discussed above. Then hearing infant-directed speech – and exaggerated vowels – from multiple speakers may further promote vowel perception, especially as children refine their emerging perceptual categories. Indeed, since the absolute formant values for any vowel vary a great deal from talker to talker (Lehiste & Meltzer, 1973) and listeners have to be sensitive to all frequency values within specific ranges, as any of them may represent a phonemic category for a particular speaker, greater variability in the input ensures that more of the vocalic space – and a wider range of vowels – are sampled, ultimately leading to the formation of more robust representations (Lev-Ari, 2018).

5. Conclusion, Implications and Directions for Future Research

In conclusion, this study focused on Spanish-English bilingual preschoolers' English vowel perception and the role of age, input quantity and diversity on perceptual performance, revealing for the first time, the significance of input diversity – mediated by language exposure – for speech sound development. The first analysis showed that, despite displaying stop consonant perceptual patterns that differed from those of English monolingual peers (Montanari et al., 2023), the same children produced monolingual-like perception of the English /i-ɪ/ contrast, suggesting that experience with the Spanish vowel system did not affect their English vowel perceptual performance. The second analysis provided evidence that while a child-internal factor such as age was not related to English vowel perception, child-external variables such as input quantity and diversity jointly interacted to moderate how well children perceived the English /i-ɪ/ contrast. Specifically, input diversity promoted perceptual performance in children who received high levels of English exposure and who possibly had more advanced English language skills, suggesting that these children possibly used input from multiple speakers to refine their perceptual categories. On the contrary, children at low levels of English exposure and earlier stages of learning did not benefit from hearing input from multiple speakers, a result that may indicate that input diversity may make it harder for these children to distinguish between and form new perceptual categories.

Our findings have both theoretical and practical implications. First, the results appear at odds with the predictions of Flege's SLM-r model (Flege & Bohn, 2021) and Best's L2 Perceptual Assimilation Model (PAM-L2, Best & Tyler, 2007), which posit difficulty with the perception of L2 categories (English /ɪ/) that are similar to those of the L1 (Spanish /i/). However, these models were mainly proposed for L2 adult learners whose L1 categories are more robust and established compared to those of young children. Indeed, the SLM-r model proposes that L1 categories are less entrenched at a young age, resulting in more limited interaction between L1 and L2 categories and thus more accurate L2 perception and production in children than in adults (Baker et al., 2008; Tsukada et al., 2005). In this light, then, and considering that the same children did display non-native English stop consonant perception patterns (Montanari et al., 2023), our results are not fully at odds with the predictions of both models. That is, while the coexistence of two phonological systems may often result in cross-linguistic interaction, this may be attenuated when children are young enough to separately attune to the sound inventories of two languages, note bimodal distributions in each input, and form separate speech sound categories for each of their languages.

Our study also has educational implications. Recall that speech perception at preschool age predicts 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). Furthermore, living in low-SES environments – as many Spanish-English bilingual children in the US (National Academies of Sciences, Engineering, and Medicine, 2017) – may delay speech discrimination, affecting the development of phonemic awareness, and indirectly contributing to poor reading outcomes (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 success. Collectively, the findings of this study and of Montanari et al (2023) show that Spanish-English bilingual preschoolers may display English perceptual patterns that differ from those of monolinguals for some sounds but not others; that some English stops are more difficult to be perceived than some English vowels; and that input diversity can facilitate vowel perception at higher stages of acquisition, while being detrimental at lower developmental phases. Thus, educators and policy makers should adjust curricular programs and educational practices accordingly, for example by focusing instruction on the sounds that have been found to be difficult for this population (i.e., voiced stops, Montanari et al., 2023), by limiting multiple-speaker exposure for children at low stages of English acquisition, while increasing the number of English interlocutors for children with more advanced language skills.

As with all studies, our investigation has some limitations. First, we only focused on English perception, as Spanish does not have a vowel contrast that parallels English /i-ɪ/. Future studies should examine similarities and differences between vowel perception in both the societal and heritage language to examine the full range of bilingual children's perception abilities and the extent to which increasing exposure to and involvement in the mainstream culture influences the development and maintenance of children's Spanish perceptual skills. We also only studied one vowel contrast and we are unable to determine whether the children will show native-like perceptual patterns with other English vowels as well. Furthermore, we included rough measures of input quantity and diversity as moderators of perceptual performance, when many more variables may be related to this skill. Future research should include more sophisticated measures of input quantity (i.e., absolute language exposure) and quality (i.e., level of exposure to native vs. accented speech) and examine the role of other variables on speech perception, including language input and output, vocabulary and grammatical measures, as well as broader sociolinguistic variables such as maternal education and acculturation. While we have no evidence that gender affects speech perception, the current study also involved more girls than boys and it is important that future research includes a more balanced number of male and female participants. Finally, future studies should track bilingual children's perceptual performance from infancy to school-age to examine the extent to which increasing exposure to and involvement in the mainstream culture shapes and refines English perception skills.

Acknowledgements

This work was supported by the National Institutes of Health/National Institute of Deafness and Other Communication Disorders [R15DC019493, 2021-2024]

Data Availability Statement

We will make the data available to users only under a data-sharing agreement that provides for: (1) a commitment to using the data only for research purposes; (2) a commitment to securing the data using appropriate computer technology; (3) a commitment to destroying or returning the data after analyses are completed or after a maximum of one year after the data was shared; and d) a commitment to cite the original funding source in the acknowledgement for publications.

References

Allen, J., & Miller, J. (2004). Listener sensitivity to individual talker differences in voice-onset time. *Journal of the Acoustical Society of America, 116*, 3171–3183.

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.

Baigorri, M., Campanelli, L., & Levy, E. (2019). Perception of American–English vowels by early and late Spanish–English bilinguals. *Language and Speech, 62*(4), 681–700.

Baker, W., Trofimovich, P., Flege, J., 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.

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

Bonatti, L., Peña, M., Nespor, M., & Mehler, J., (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science, 16*(6), 451-459.

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.

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.

Bradlow, A., Akahane-Yamada, R., Pisoni, D., & 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.

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

Caramazza, A., Chialant, D., Capasso, R., & Miceli, G. (2000). Separable processing of consonants and vowels. *Nature, 403,* 428-430.

Calandruccio, L., Beninate, I., Oleson, J., Miller, M.K., Leibold, L.J., Buss, E., & Rodriguez, B.L. (2021). A simplified approach to quantifying a child's bilingual language experience. *American Journal of Audiology, 30*(3), 769-776.

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

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

Flege, J. (1991). The interlingual identification of Spanish and English vowels: Orthographic evidence. *Quarterly Journal of Experimental Psychology, 43*(3), 701–731.

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., & Bohn, O.-S. (2021). The revised Speech Learning Model (SLM-r). In R. Wayland (ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press.

Flege J., Bohn O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25*(4), 437–470.

Fox, R., Flege J., & Munro, M. (1995). The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *Journal of the Acoustical Society of America, 97*(4), 2540–2551.

Fry, D., Abramson, A., Eimas, P., & Liberman, A. (1962). The identification and discrimination of synthetic vowels. *Language and Speech, 5,* 171–189.

Garcia-Sierra, A., Rivera-Gaxiola, M., Percaccio, C., Conboy, B., Romo, H., Klarman, L., Ortiz, S., & Kuhl, P. (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.

Granier-Deferre, C., Bassereau, S., Ribeiro, A., Jacquet, A.-Y., & Decasper, A. (2011). A melodic contour repeatedly experienced by human near-term fetuses elicits a profound cardiac reaction one month after birth. *PLoS One, 6*(2), e17304.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustic Society of America, 97,* 3099-3111.

Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: Implications for closing achievement gaps. *Developmental Psychology, 49*(1), 4-14.

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.

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

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

Ladefoged, P. (2005). *Vowels and consonants: An introduction to the sounds of languages* (2nd ed.). Blackwell.

Lehiste, I., & Meltzer, D. (1973). Vowel and speaker identification in natural synthetic speech. *Language and Speech, 16,* 356-64.

Lev-Ari, S. (2018). The influence of social network size on speech perception. *Quarterly Journal of Experimental Psychology, 71*(10), 2249 –2260.

Logan, J.S., Lively, S.E., & Pisoni, D.B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *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.

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.

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.

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.

Migration Policy Institute. (2019). Young dual language learners in the United States and by state. www.migrationpolicy.org/programs/data-hub/charts/us-state-profiles-young-dlls

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.

Montanari, S., Steffman, J., & Mayr, R. (2023). Stop voicing perception in the societal and heritage language of Spanish-English bilingual preschoolers: The role of age, input quantity and input diversity. *Journal of Phonetics.* Published online 8 October 2023. https://doi.org/10.1016/j.wocn.2023.101276

Morrison, G.S. (2008). L1-Spanish speakers' acquisition of the English /i /—/ɪ/ contrast: Duration-based perception is not the initial developmental stage. *Language and Speech, 51,* 285-315.

Morrison, G.S. (2009). L1-Spanish speakers' acquisition of the English /i/—/ɪ/ contrast II: Perception of vowel inherent spectral change. *Language and Speech, 52,* 437–462.

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.

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.

Perkins, S.C., Finegood, E.D., & Swain, J.E. (2013). Poverty and language development: Roles of parenting and stress. *Innovations in Clinical Neuroscience, 10*(4), 10-19.

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,* 467–478.

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.

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.

Russo, F., Vuvan, D., & Thompson, W.F. (2019). Vowel content influences relative pitch perception in vocal melodies. *Music Perception: An Interdisciplinary Journal, 37*(1), 57-65.

Sundara, M. (2022). Development of speech perception and production in bilingual preverbal infants. In B. Goldstein (ed.), *Bilingual language development and disorders in Spanish-English speakers* (3rd ed.) Brookes Publishing.

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.

Sundara, M., & Scutellaro, A. (2011). Rhythmic distance between languages affects the development of speech perception in bilingual infants. *Journal of Phonetics, 39(4),* 505-513.

Smith, R., & Hawkins, S. (2012). Production and perception of speaker-specific phonetic detail at word boundaries. *Journal of Phonetics, 40,* 213–233.

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America, 88,* 97-100.

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.

Tyler, M., Best, C., Faber, A., & Levitt, A.G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica, 71*(1), 4–21.

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.

Winn, M.B. (2019). GUI-based wizard for creating realistic vowel formant continua (v. 38). http://www.mattwinn.com/praat/Make_Formant_Cont inuum_v38.txt

Yu, Y., Tessel, C., Han, X., Campanelli, L., Vidal, N., Gerometta, J., Garrido-Nag, K., Datta, H., & Shafer, V. (2019). Neural indices of vowel discrimination in monolingual and bilingual infants and children. *Ear and Hearing, 40*(6), 1376–1390.

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

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

[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 English input sources among "mother, father, siblings, grandparents, babysitter, teacher, media, and other."

Table 2. Number, gender, age, English exposure, and number of English input providers for bilingual children with higher (HIGH-ENG) and lower (LOW-ENG) English exposure.

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

[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 input sources among "mother, father, siblings, grandparents, babysitter, teacher, media, and other."

Figure 1. Model fit for categorization along the vowel continuum (x axis), plotting the proportion of /i/ responses on the y axis. Participant group is indicated by line coloration and line type.

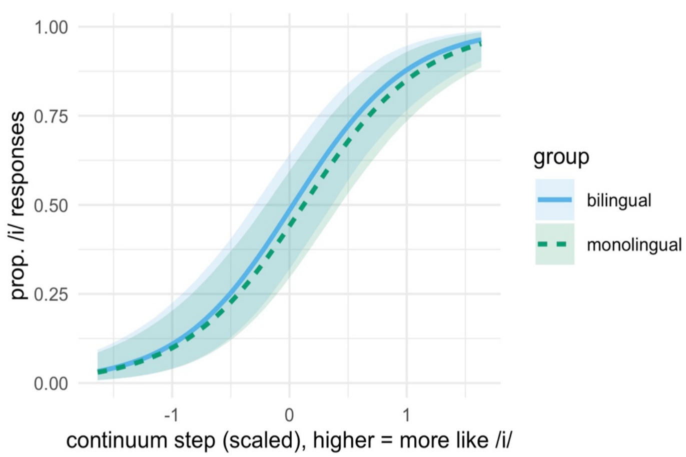
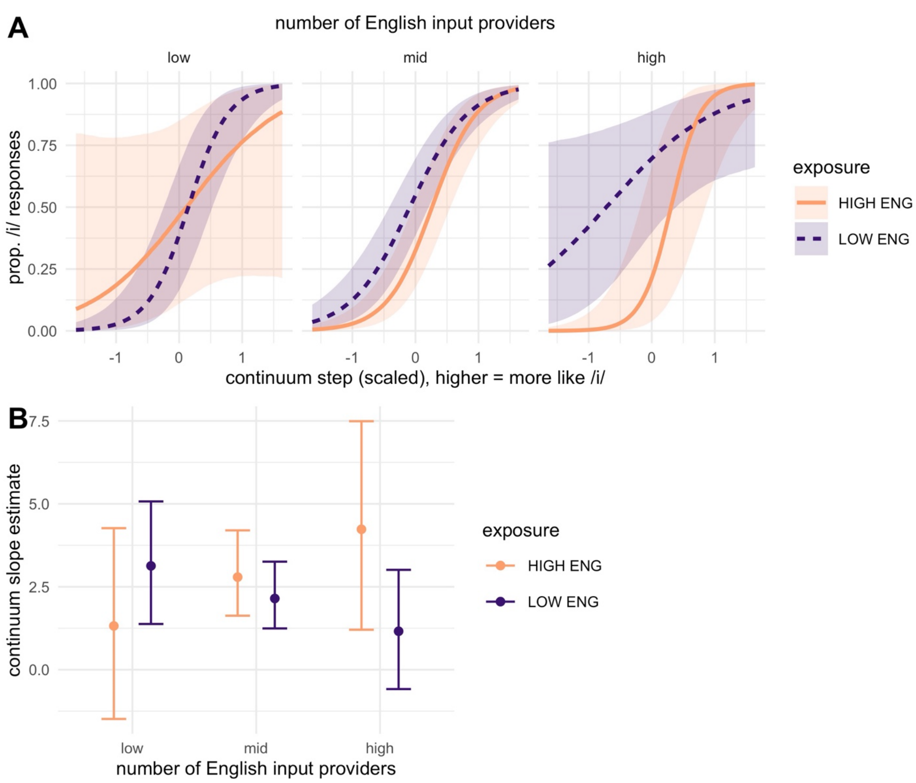


Figure 2. Panel A model fit for categorization along the vowel continuum (x axis), plotting the proportion of /i/ responses on the y axis. The three panels show low, mid, and high levels of input diversity, sampled at three points along the continuous scaled variable in the model corresponding to 2 (low), 4 (mid) and 6 (high) sources of English input. Exposure group is indicated by line coloration and line type. Panel B shows the estimate for continuum step as a function of binned input diversity (x axis) and exposure group (coloration).



Appendix

Table 1: Model summary for the analysis comparing monolinguals and bilinguals

Predictors	*Log-Odds*	*CrI (95%)*
Intercept	-0.15	-0.65, 0.32
step.scaled	2.02	1.60, 2.52
language group	-0.18	-1.06, 0.72
step.scaled:lang exp	0.07	-0.82, 0.68

Table 2: Model summary for the moderator analysis

Predictors	*Log-Odds*	*CrI (95%)*
Intercept	-0.22	-1.11,0.66
continuum step (scaled)	2.50	1.68,3.46
exposure (binary)	0.94	-0.62,2.50
age (scaled)	0.61	-0.30,1.49
number of inputs (scaled)	0.05	-1.06,1.25
step:exp	-0.62	-2.09,0.86
step:age	0.01	-0.85,0.90
exp:age	-0.04	-1.64,1.60
step:inputs	0.15	-1.03,1.32
exp:inputs	0.82	-1.21,2.77
age:inputs	0.43	-0.71,1.53
step:exp:age	0.81	-0.76,2.37
step:exp:inputs	-1.82	-3.83,0.19
step:age:inputs	-0.07	-1.19,1.05
exp:age:inputs	-0.81	-2.75,1.22
step:exp:age:inputs	-0.87	-2.84,1.04