Differential perception of vowels and stop consonants in Spanish-English bilingual preschoolers

Citation for published version:
Montanari, S, Steffman, J, Mayr, R, Flores, P & Feeney, C 2023, Differential perception of vowels and stop consonants in Spanish-English bilingual preschoolers. in R Skarnitzl & J Volín (eds), *Proceedings of the 20th International Congress of Phonetic Sciences.*, 86, Guarant International, Prague, pp. 42-46, 20th International Conference of Phonetic Sciences (ICPhS), Prague, Czech Republic, 7/08/23. <https://guarant.cz/icphs2023/86.pdf>

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Proceedings of the 20th International Congress of Phonetic Sciences

DIFFERENTIAL PERCEPTION OF VOWEL AND STOP CONTRASTS IN SPANISH-ENGLISH BILINGUAL PRESCHOOLERS

Simona Montanari[a], Jeremy Steffman[b], Robert Mayr[c], Patricia Flores[a] & Carissa Feeney[a]

[a]California State University, Los Angeles, [b]University of Edinburgh, [c]Cardiff Metropolitan University

ABSTRACT

Despite extensive work on perceptual development in bilingual infants, we know little about speech perception in bilingual preschoolers. This study investigates English vowel and stop voicing perception in 20 Spanish-English bilingual preschoolers and 20 English monolingual peers. Perception was assessed through a forced-choice minimal-pair identification task in which children heard synthesized audio stimuli that varied systematically along an /i-ɪ/ vowel continuum and a /b-p/ 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 indicated no credible difference in vowel perception between monolinguals and bilinguals. In contrast, the bilinguals' category boundary for English /b-p/ was impacted by their experience with Spanish, with more short-lag VOT tokens being perceived as voiceless in line with Spanish VOT. We interpret the observed asymmetry as stemming from differences in L1-L2 mappings across the contrasts.

Keywords: speech perception; Spanish-English bilinguals; preschoolers; voice onset time; vowels

1. INTRODUCTION

Despite an abundance of research on speech perception in simultaneous bilingual infants (see e.g. [1]) and adult second/foreign language learners (e.g. [2]), 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 [3]. McCarthy et al. [4], who assessed the perception of the English voicing contrast in bilabial and velar stops in Sylheti-English bilingual children, found that English perception patterns at preschool entry (i.e., at age 4;4) were different from monolinguals' and appeared affected by children's existing Sylheti phonemic categories. However, a year later, the bilingual children's productions were no longer significantly different from those of their monolingual peers, suggesting more refined phonemic categories. The authors speculated that phonemic categorization may be initially affected by language dominance in young bilinguals. However, phonemic categories in the other language can be acquired and refined with language experience. The limited research on speech perception in sequential bilingual children confirms that even in the case of early exposure to a second language (L2) (e.g., in infancy or early childhood), native-like perception may take years to develop. Darcy and Krüger [5], who tested the perception of four different German vowel contrasts in 10-year-old Turkish/German bilingual children who began to be exposed to German between age 2 and 4, found that the children had difficulties with the discrimination of those contrasts that were perceptually similar for Turkish monolingual speakers. These findings were interpreted as evidence of perceptual assimilation of German phonemes to similar Turkish phonemes even in the case of early bilingualism. Similarly, Tsukada et al. [6], who compared the perception of 4 English vowel contrasts in Korean-English bilingual school-age children who had spent up to 6 years in the United States, found that the children discriminated English vowels more accurately than Korean-English bilingual adults but less accurately than monolingual English-speaking children. Overall, these studies suggest that even early and intensive exposure to two languages may not be enough to prevent phonemic representations from one language influencing speech perception in the other [7]. However, it is unclear whether some L2 sound contrasts are more difficult to perceive than others for bilingual preschoolers

This study expands the limited literature on bilingual speech perception at preschool age by examining L2 speech perception on two types of contrast, a vowel contrast and a stop voicing contrast, in Spanish-English bilingual children. We ask: 1) To what extent does vowel or VOT continuum step predict categorization responses? 2) To what extent do Spanish-English bilingual children and age-matched English monolingual children differ in their categorization responses on the English /i-ɪ/ vowel continuum? 3) To what extent do they differ in their categorization responses on the English /b-p/ VOT continuum?

2. METHOD

2.1. Participants

The participants are 20 Spanish-English bilingual (4 males and 16 females, mean age 4;9) and 20 English monolingual (12 males and 8 females, mean age 4;6) children growing up in Los Angeles, California, matched in age (*t*(38)=1.78, *p*=.084). The bilingual children are raised in homes where they have regularly and consistently heard both languages from early in life, although they are exposed to English more overall and through a larger variety of speakers/sources as compared to Spanish, which they mainly hear from family members. On that basis, they were classed as English-dominant. The monolinguals hear mainly English from their input providers, but due to the bilingual nature of the community in which they live, they also have limited exposure to Spanish. Nevertheless, based on the information provided they fit the description of "functional monolinguals" with no active use or knowledge of Spanish [8]. 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.

2.2. Stimuli

Two phonetic continua were created, manipulating vowel formants and vowel duration, and VOT alongside other cues to voicing. These continua ranged between "sheep" and "ship" for the /i-ɪ/ contrast, which is known to pose difficulties for Spanish monolinguals [9,10], and "penny" and "Benny" for the /p-b/ contrast. To obtain initial values, natural tokens of the target words were recorded by a female native speaker of American English.

Manipulation of the vowel and voicing contrast continua was carried out using a Praat scripts. For the vowels, we employed the Burg Method of LPC decomposition and resynthesis ([11], www.mattwinn.com/praat/Make_Formant_Continuum_v38.txt). 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, [12]) 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, [13]). 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). For the stop voicing contrast, VOT, F0 and the duration of the following vowel were manipulated over a total of 10 steps [14]. VOT ranged from 0 (short-lag VOT) to 90 ms (long-lag VOT). The duration of the following vowel was shortened incrementally as VOT was lengthened. F0 also varied along the continuum, showing a localized increase following longer VOT (+30 Hz for the longest continuum step relative to the base file), which interpolated to a slight dip (-5 Hz relative to the base file). The duration of the F0 perturbation was set to be 75 ms, which is a typical range for voicing-induced F0 changes in vowels [15].

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 for both continua (keeping the 6 most central steps in the continuum and the endpoints).

2.3. Procedures

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 vowel or VOT continuum and were asked to match it with one of two pictures representing a minimal pair ("sheep" and "ship" for /i-ɪ/ and "penny" and "Benny" for /p-b/). In the experiment, they were exposed to the 8 tokens from each contrast in random order alongside other materials not reported here, for a total of 24 trials. Children were familiarized with the stimuli one week before the experiment and had to show they knew the words in order to participate.

2.4. Analyses

We analyzed the data with Bayesian mixed-effects logistic regression analyses of categorization data (carried out using *brms* [16]). We analyzed the stop and vowel data separately. In each model we predicted binomial responses (/p/ and /i/ mapped to 1 in the stop and vowel model, respectively). This was predicted as a function of continuum step (scaled), gender (female mapped to -0.5, male to 0.5), a background variable that some literature has found to be predictive of bilingual development (see [17] for a review), and language background (bilingual mapped to -0.5, monolingual to 0.5). A quadratic term for continuum step was also included to allow the model to capture potentially larger effects in the middle of the continuum (as compared to either end, as with the linear continuum step term). All interactions between fixed effects were included as well. The random effects consisted of a random intercept for speaker and by-speaker slope for continuum step. We used weakly informative normal priors for both intercept and fixed effects: normal (0,1.5) in log-odds space. In reporting effects, we give the 95% credible interval (CrI), which indexes the range in which 95% of the estimated posterior falls. When this interval *excludes* the value of zero it indicates a robust effect: that is, a non-zero effect estimate with a consistent directionality. We also report the probability of direction “pd” [18], which is more intuitively interpreted as strength of evidence for an effect. This metric gives the percentage of a distribution with a given positive/negative sign. It ranges from 50 (distribution centered on 0, no effect) to 100 (strong evidence for an effect). Pd values in excess of 97.5% correspond to 95% CrI excluding 0, which we take as a reliable effect [19]. Pd values approaching this threshold constitute weaker evidence for an effect. Full model summaries are given in Table 1.

We excluded participants who did not show sensitivity to the acoustic continua we used. 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 /b/ responses for longer VOT, or more /ɪ/ responses with /i/-like formants and duration). Either of these indicates potential inattention to the task, or lack of perception of the contrast at all. This was done on a by-contrast basis, leaving 17 bilinguals and 18 monolinguals for the /b/~/p/ contrast and 19 bilinguals and 18 monolinguals for the /ɪ/~/i/ contrast.

3. RESULTS

3.1. Stop results

The model of stop voicing perception found an expected effect of (linear) *VOT continuum step* (β = 2.57, 95%CrI = [1.82,3.43], pd = 100), showing that listeners increase /p/ responses with increasing VOT. There was not a main effect of *gender* (pd = 63), but crucially, there was a main effect of *language background* (β = -1.12, 95%CrI = [-2.27,-0.00], pd = 98), showing that monolinguals show decreased /p/ responses overall (Fig. 1A). We can consider this in terms of the category boundary for each group, operationalized as the point on the continuum at which listeners give 50% /p/ responses. The category boundary is shifted numerically lower for bilingual speakers (about 30 ms) as compared to monolinguals (about 40 ms). Put differently, bilinguals require overall shorter VOT to perceive /p/, and perceive short-lag VOT as /p/ more often. Neither the interaction between *gender* and *continuum step*, nor *language background* and *continuum step* was credible pd = 63 and pd = 88, respectively).

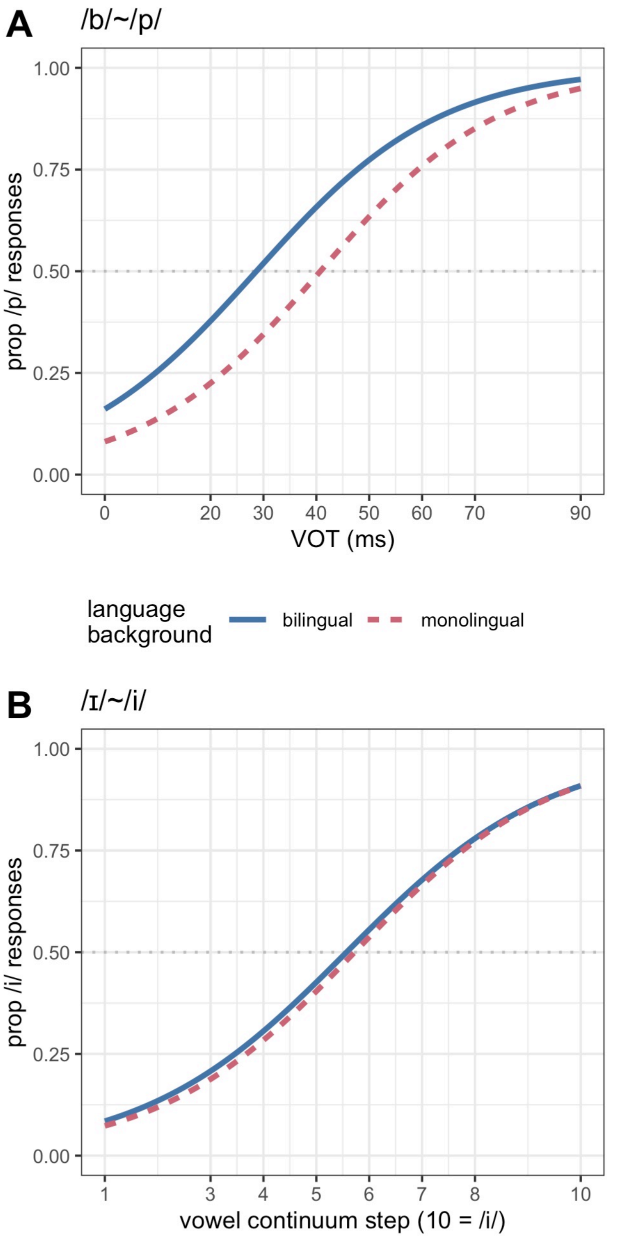

Figure 1: Categorization for the /b/~/p/ continuum (A) and /ɪ/~/i/ continuum (B), along the continuum (x axis). Lines are logistic fits to the data. Line type and coloration shows language group. The dotted horizontal line placed at 0.50 on the y axis represents the category boundary.

3.2. Vowel results

The model of vowel perception found an expected effect of the *formant continuum* (β = 2.24, 95%CrI = [1.78, 3.21], pd = 100), showing that listeners' increase /i/ responses with changes in formants and vowel duration. There was not a main effect of *gender* (pd = 84). Unlike with the stop voicing model, there was no effect of *language background* (pd = 55), as shown in Figure 1B. The interaction between *gender* and *continuum step*, and *language background* and *continuum step* were not credible (pd = 94, and pd = 68 respectively). The weaker evidence for the former interaction was inspected via the *estimate slopes* function in [20], showing a larger effect of step for male speaker versus female speaker (male β = 2.86, 95%CrI = [1.93, 4.06]; female β = 1.93, 95%CrI = [1.26, 2.82]).

stop model	β	error	95%CrI
Intercept	0.79	0.32	[0.19,1.44]
continuum step	2.57	0.42	[1.82,3.47]
language backgr.	-1.12	0.57	[-2.27,-0.00]
gender	0.15	0.57	[-0.98,1.28]
step:lang backgr.	0.67	0.57	[-0.46,1.79]
step:gen	0.24	0.67	[-1.05,1.60]
gen:lang backgr.	1.54	0.92	[-0.29,3.35]
gen:step:lang backgr.	0.04	1.06	[-2.07,2.09]
step.quad	-0.02	0.27	[-0.54,0.52]
step.quad:lang backgr.	0.25	0.49	[-0.70,2.78]
step.quad:gen	1.70	0.54	[0.70,2.78]
vowel model	β	error	95%CrI
Intercept	-0.24	0.40	[-1.04,0.53]
continuum step	2.24	0.36	[1.78,3.21]
language backgr.	-0.12	0.76	[-1.62,1.37]
gender	0.75	0.76	[-0.75,2.26]
step:lang backgr.	-0.28	0.61	[-1.49,-0.91]
step:gen	0.94	0.61	[-0.23,2.17]
gen:lang backgr.	0.31	1.23	[-2.09,2.74]
gen:step:lang backgr.	0.02	1.09	[-2.18,2.10]
step.quad	0.23	0.26	[-0.28,0.76]
step.quad:lang backgr.	-0.47	0.47	[-1.42,0.44]
step.quad:gen	-0.06	0.51	[-1.05,0.95]

Table 1: Model summaries for the stop model (top) and vowel model (bottom). Estimates and errors are given with 95% CrI. "step.quad" indicates the quadratic term for continuum step.

4. DISCUSSION AND CONCLUSION

This study sought to extend our understanding of speech perception abilities in bilingual preschoolers. To this end, we investigated the perception of the /i-ɪ/ vowel contrast and the /b-p/ stop voicing contrast in Spanish-English bilinguals aged 4-5 years, and compared it to that of their English monolingual peers. The results revealed an asymmetry across the two contrasts, with monolinguals and bilinguals differing in their perception of the stop voicing contrast, but not the vowel contrast. The bilinguals categorized shorter VOT tokens as voiceless than the monolinguals and hence the two sets of preschoolers differ in their category boundary between voiced and voiceless bilabial stops. The preschoolers' gender, in turn, did not affect categorization responses on either contrast. How can we explain these findings?

To begin with, the direction of the category boundary difference on the VOT continuum suggests an influence of the bilinguals' first language (L1) as Spanish contrasts prevoiced and short-lag VOTs [21]. These results are in line McCarthy et al.'s findings [4]. However, in their study, the children were L1-dominant when interactions surfaced in L2 English VOT patterns. In contrast, in the present study, the bilingual preschoolers were dominant in L2 English, and hence, crucially, what this study adds is that cross-linguistic interactions in speech perception are not limited to bilinguals' weaker language.

At the same time, the results also suggest that L1-L2 interactions do not occur across the board since the bilinguals' perception of the English-specific /i-ɪ/ vowel contrast matched that of their monolingual peers, with identical location of the category boundary. What may have caused interactions on the VOT continuum, but not the vowel one, is that the former involves a difference in phonetic implementation for the *same* phonological contrast. On the other hand, the vowel contrast involves acquisition of the novel L2 category /ɪ/, which requires perceptual attunement to its temporal and spectral properties. As such, it is a cross-linguistically more dissimilar sound than the two stops, which according to the *Speech Learning Model* (SLM, [22]) and its recent revision, the SLM-r ([23]) makes it comparatively easier to acquire – in particular since early bilinguals are more likely to establish a novel L2 category than late bilinguals according to the model.

Taken together, this paper extends the limited literature on speech perception in bilingual preschoolers by documenting that subtle differences from age-matched monolinguals may occur on cross-linguistically similar contrasts and may even manifest in the dominant language. At the same time, despite these differences, the bilinguals managed to differentiate both contrasts successfully, and hence the data reported here do not raise any concerns in terms of the development of the children's preliteracy skills. Future work is needed that investigates which perceptual dimensions are particularly prone to L1-L2 interactions in this population, and to what extent increased use of the L2 may shift L1 perceptual boundaries.

5. REFERENCES

[1] Fennell, C.T., Tsui, A.S-M., & Huddon, T.M. 2016. Speech perception in simultaneous bilingual infants. In: Nicoladis, E., & Montanari, S. (eds), *Bilingualism across the Lifespan: Factors Moderating Language Proficiency*. American Psychological Association, 43-62.

[2] Ingvalson, E.M., Ettlinger, M., & Wong, P.C.M. 2014. Bilingual speech perception and learning: A review of recent trends. *Int. J. Biling.* 18, 35–47.

[3] 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. *J. Commun. Disord.* 38, 29–63.

[4] 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. Dev.* 85, 1965-1980.

[5] Darcy, I. & Krüger, F. 2012. Vowel perception and production in Turkish children acquiring L2 German. *J. Phon.* 40*,* 568-581.

[6] Tsukada, K., Birdsong, D., Bialystok, E. et al. 2005. A developmental study of English vowel production and perception by native Korean adults and children. *J. Phon.* 33, 263–290.

[7] Simonet, M. 2010. Dark and clear laterals in Catalan and Spanish: Interaction of phonetic categories in early bilinguals. *J. Phon. 38*, 663–678.

[8] Best, C.T. & Tyler, M.D. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In: Bohn, O.-S. & Munro, M.J. (eds), *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*. John Benjamins, 13–34.

[9] Morrison, G.S. 2008. L1-Spanish speakers' acquisition of the English /i/-/ɪ/ contrast: Duration-based perception is not the initial developmental stage. *Lang Speech* 51, 285-315.

[10] Morrison, G.S. 2009. L1-Spanish speakers' acquisition of the English /i/—/ɪ/ contrast II: Perception of vowel inherent spectral change. *Lang Speech* 52, 437-62.

[11] Winn, M.B. (2019). GUI-based wizard for creating realistic vowel formant continua (v. 38). http://www.mattwinn.com/praat/Make_Formant_Continuum_v38.txt

[12] Traunmüller, H. 1990. Analytical expressions for the tonotopic sensory scale. *J. Acoust. Soc. Am.* 88, 97-100.

[13] Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.* 97, 3099-3111.

[14] Winn, M. B. (2020). Manipulation of voice onset time in speech stimuli: A tutorial and flexible Praat script. *J. Acoust. Soc. Am.*147, 852-866.

[15] Hombert, J. M. (1975). *Towards a theory of tonogenesis: An empirical, physiologically and perceptually based account of the development of tonal contrasts in languages*. University of California, Berkeley.

[16] Bürkner, P. C. (2021). Bayesian Item Response Modeling in R with brms and Stan. *Journal of Statistical Software*, 100(5), 1-54.

[17] Iglesias, A., & Rojas, R. 2022. Bilingual language development of English learners: Estimating the growth of two languages. In: Goldstein, B. (ed.), *Bilingual Language Development and Disorders in Spanish-English Speakers.* Brookes, 9-44.

[18] Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *Journal of Open Source Software*, 4(40), 1541.

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

[20] Makowski, D., Ben-Shachar, M. S., Patil, I., & Lüdecke, D. (2020). *Estimation of Model-Based Predictions, Contrasts and Means.* CRAN.

[21] Cho, T., & Ladefoged, P. 1999. Variation and universals in VOT: Evidence from 18 languages. *J. Phon.* 27, 207–229.

[22] Flege, J.E. 1995. Second language speech learning: Theory, findings, and problems. In: Strange W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. York Press, 233–272.

[23] Flege, J.E., & Bohn, O.S. 2021. Flege, J.E. & Bohn, O.S. 2021. The revised Speech Learning Model (SLM-r). In: Wayland, R. (ed.), *Second Language Speech Learning: Theoretical and Empirical Progress*. Cambridge University Press, 3–83.