Coronal stop production in bilinguals

Alexander Aldrich, Joseph V. Casillas & Miquel Simonet

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Introduction

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Background and motivation

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Method

Participants

The data include 42 participants from 3 populations: monolingual English speakers, monolingual Spanish speakers, and bilingual Spanish-English speakers. All participants were females between the ages of 18 and 23. The monolingual English speakers were recorded in English and the monolingual Spanish speakers were recorded in Spanish. The Spanish-English bilinguals were recorded in both of their languages.

Monolingual English speakers. The study includes 8 monolingual English speakers. They were undergraduate students at the University of Arizona, born and raised in the US Southwest. The English speakers were functionally monolingual, though they reported having taken introductory Spanish courses. They were not able to maintain a basic conversation in Spanish. All of the participants in this group reported English as their native language and verified not having been exposed to any other languages while growing up.

Monolingual Spanish speakers. The monolingual Spanish group comprised 8 speakers that were recruited from the *Universitat de les Illes Balears* campus community and were born and raised on the island of Majorca, Spain. They reported that, although they had studied some English in Spain, they were not able to maintain a basic conversation in this language. The participants of this group also speak Catalan. We refer to them as monolingual Spanish speakers in the context of this study only for the sake of clarity. Importantly, there are no reported differences in the phonetic realization of voice timing between the Spanish and Catalan, nor are there place differences between the coronal stops.

Bilingual speakers. The Spanish-English bilinguals (n = 26) came from Southern Arizona and Northern Mexico. There are two samples from this population. The coronal data set includes 17 speakers and the bilabial data set includes 9 speakers. The Spanish-English bilinguals were undergraduate students at the University of Arizona in Tucson, Arizona. The bilinguals were brought up by Spanish-speaking families and were schooled mostly in English. They reported using English and Spanish daily, both in the classroom as well as with their friends and relatives.

The bilingual participants completed the Bilingual Language Profile (BLP, Gertken, Amengual, & Birdsong, 2014) in order to assess language dominance. The BLP calculates a weighted average of language dominance based on the individual history, use, proficiency, and attitudes of the bilinguals with regard to their languages. The measure ranges from −218 to 218 with values near the extremes implying dominance in one of the languages. Values close to 0 are taken as an indication of balanced bilingualism. In the present study, Spanish was arbitrarily assigned to positive values. Figure XXX plots language dominance (Panel A) and language use and proficiency data (Panel B) derived from the BLP. The bilingual group had a mean dominance score of X (SD = X), suggesting rather balanced bilingualism (Panel A). Participants that reported using Spanish more often also tended to report being more proficient in that language; the converse was also true for English (Panel B).

NEED BLP DATA FROM ALDRICH (or something similar).

Materials

Coronal data set

There were 48 target words (English: k = 24; Spanish: k = 24) that contained voiced and voiceless coronal stops in word initial position. For each language, 12 targets began with /d/ and 12 began with /t/, equally divided between stressed and unstressed syllables. All stops were followed by a low vowel (/a/ for Spanish and /æ, ɑ/ for English).

The participants completed a delayed repetition task in which they heard the target words presented in a carrier phrase ("*x is the word*" or the Spanish equivalent "*x es la palabra*"). The auditory stimuli were recordings of 6 male native speakers: 3 native English speakers and 3 native Spanish speakers. These recordings served as the auditory stimuli repeated out loud by the participants in the delayed repetition task. Words not containing coronal stops were considered distractors (k = 20). Praat Boersma & Weenink (2022) was used to present the sentences randomly in auditory form and the speakers were asked to listen to the entire sentence and then repeat it out loud at their own pace.

The monolingual English speakers and bilinguals were recorded in a sound attenuated booth. The monolingual Spanish speakers were recorded in a quiet classroom on the campus of the *Universitat de les Illes Balears* in Majorca, Spain. The monolingual English speakers were recorded in English and the monolingual Spanish speakers were recorded in Spanish. The Spanish-English bilinguals were recorded in both of their languages in a single session with all English and Spanish items presented in a single, randomized block. The full data set included 3600 tokens (24 target words per language × 3 repetitions × 25 participant). Eighty-one items (2.25%) were discarded due to mispronunciations or extraneous noise. A Shure SM10A dynamic head-mounted microphone with a Sound Devices MM-1 microphone pre-amplifier captured the acoustic signal and it was saved to a Marantz PMD660 digital speech recorder. The signal was digitized at 44.1 kHz and 16-bit quantization.

Bilabial data set

INCLUDE INFO FROM ALDRICH's STUDY

Metrics

The audio files were low-pass filtered at 11.025 kHz. Synchronized waveform and spectrographic displays were used to mark the onset of modal voicing and of the stop burst, as well as the offset of the first vowel. The onset of voicing was operationalized as the upwards zero-crossing of the first periodic pattern in the oscillogram and the offset of the vowel was marked at the downwards zero-crossing of the final periodic pattern. VOT was calculated as the difference (in ms) between the onset of modal voicing and the onset of the burst. Spectral moment measures were calculated from a 6 ms window beginning at the onset of the burst. Specifically, we extracted relative center of gravity, standard deviation, skewness, kurtosis, as well as F1 and F2 of the following vowel and relative intensity.

Procedure

I DONT HAVE THIS INFO. NEED FOR BOTH DATA SETS

Statistical analyses

All analyses were conducted in `R` (R Core Team, 2019, version 4.2.1). We use Bayesian multilevel models fitted in `Stan` using `brms` (Bürkner, 2017; Bürkner, 2018, version 2.17.0).[1] For all models, the criterion was standardized, or converted to z-scores, in order to facilitate comparability between metrics. Continuous predictors were also standardized and categorical predictors were sum-to-zero coded. Thus for all models the intercept represents the outcome variable at the grand mean in standardized units. We used regularizing, weakly informative priors in all models (specifics below) with 4,000 iterations (2,000 warm-up) running on 16 processing cores. We quantify our uncertainty regarding a given effect by reporting point estimates derived from the posterior predictive distribution, including the 95% Highest Density Credible Intervals (HDI). Additionally, we assume a negligible effect size of ±0.1 (Cohen, 1988, 2013; Kruschke, 2018) in order to establish a Region of Practical Equivalence (ROPE), for which we assess the proportion of the HDI that falls within this interval. Finally, we report the Maximum Probability of Effect (MPE), or the Probability of Direction, as the proportion of the posterior distribution that is of the median's sign (See Appendix E of the supplementary materials for additional information).

Results

The results are divided into 3 subsections dealing with (1) monolingual data, (2) bilingual data, and (3) POA data. Each subsection presents the results of 6 metrics: VOT, relative intensity,

---

[1] Bayesian Data Analysis (BDA) is an alternative to frequentist statistical analysis. See Appendix E for a brief overview, as well as Schoot & Depaoli (2014) and Vasishth, Nicenboim, Beckman, Li, & Kong (2018) for tutorials and in depth explanations related to BDA in the psychological and speech sciences.

center of gravity, kurtosis, standard deviation, and skewness. Complete model summaries are available in the supplementary materials.

Experiment 1: Monolinguals

We modeled VOT and the burst metrics of the coronal data as a function of language (English, Spanish), voicing (/d/, /t/), standardized F1 and F2, and item repetitions. The model used regularizing, weakly informative priors (Gelman, Simpson, & Betancourt, 2017; Vasishth et al., 2018). Specifically, all parameters were assumed to be distributed as normal with a standard deviation of 5, i.e., $Normal(\mu = 0, \sigma = 5)$. The standard deviation parameters for random effects and the model residual error (sigma) were truncated to exclude negative values. Figure 1 plots VOT and the burst metrics as a function of language (English, Spanish) and voicing (/d/, /t/). For all plots the y-axis represents the outcome variable in standardized units. The x-axis indicates the language and voiced/voiceless pairs are represented by color.

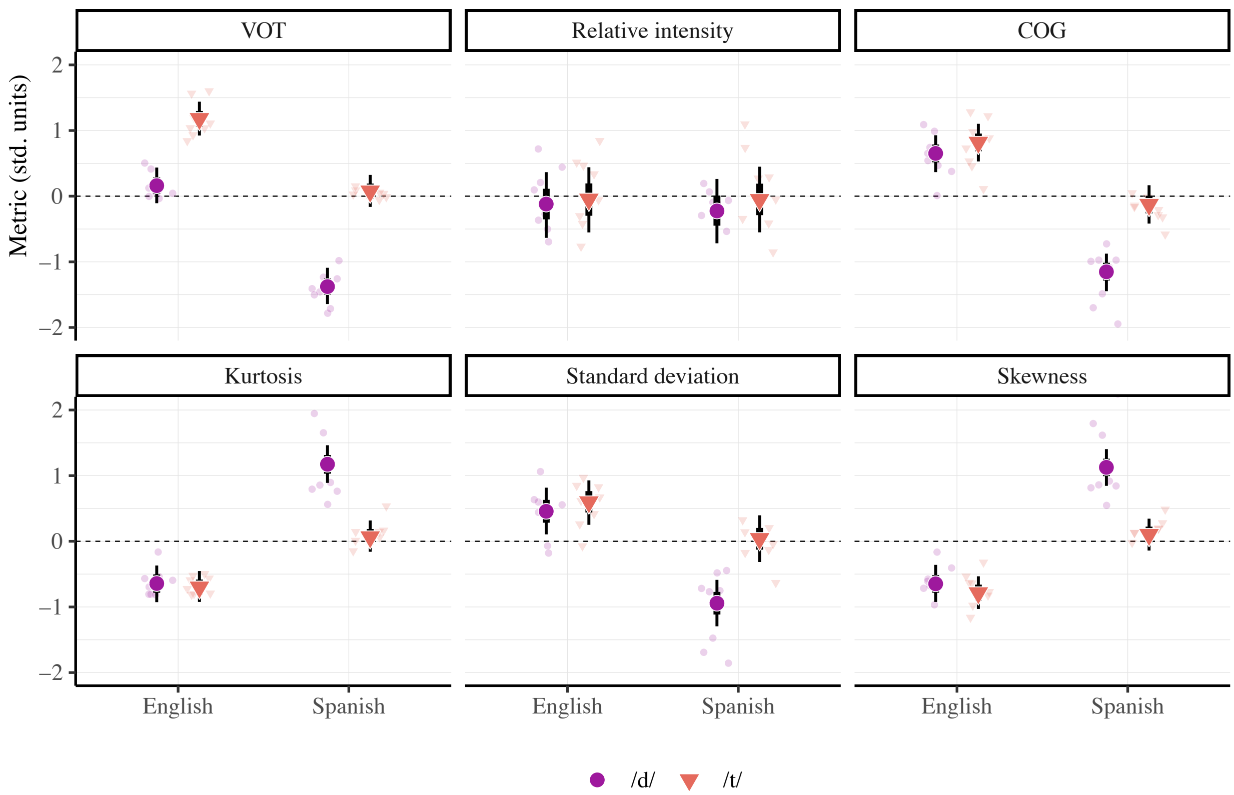

Figure 1: VOT and burst metrics of coronal stops (/d/, /t/) from monolingual speakers as a function of language (English, Spanish). Transparent points represent participant averages. Solid points indicate posterior means and 95% and 66% credible intervals.

Voice-onset time. English stops had higher VOT than Spanish stops (β = 0.660, HDI = [0.556, 0.763], ROPE = 0, MPE = 1) and the voiceless segments had higher VOT than the voiced segments (β = −0.617, HDI = [−0.697, −0.544], ROPE = 0, MPE = 1). We also find moderate evidence for an interaction between the two predictors (β = 0.108, HDI = [0.033, 0.180], ROPE = 0.406, MPE = 0.995). Specifically, the voicing difference between Spanish coronals was slightly larger than that of the English coronals. VOT did not vary as a function of the formant frequencies of the following vowel (F1: β = −0.010, HDI = [−0.041, 0.020], ROPE = 1, MPE = 0.752; F2: β = 0.010, HDI = [−0.023, 0.042], ROPE = 1, MPE = 0.73). We also compared the short-lag stops of each language, English /d/ and Spanish /t/, and found no evidence that the segments differed from each other (β = 0.085, HDI = [−0.181, 0.335], ROPE = 0.489, MPE =

0.745), as nearly half the HDI fell within the predetermined region of practical equivalence. The VOT data is plotted in the first panel of Figure 1. The complete model summary and the short-lag stop comparison are available in Table 2 and Table 3, respectively, in the supplementary materials.

Relative intensity. The relative intensity data is plotted in the top middle panel of Figure 1. English and Spanish stops differed little with regard to relative intensity. The model provided no compelling evidence for a group effect (β = 0.026, HDI = [−0.171, 0.246], ROPE = 0.666, MPE = 0.602), nor for voicing (β = −0.060, HDI = [−0.192, 0.055], ROPE = 0.742, MPE = 0.83) or F2 effects (β = −0.080, HDI = [−0.177, 0.015], ROPE = 0.676, MPE = 0.948). The height of the following vowel did appear to modulate relative intensity of the burst to some degree (β = −0.172, HDI = [−0.252, −0.085], ROPE = 0.026, MPE = 1), such that higher F1 values were associated with lower relative intensity. That being said, approximately 2.6% of the HDI fell within the region of practical equivalence. There was no evidence supporting a language × voicing interaction (β = 0.025, HDI = [−0.094, 0.157], ROPE = 0.906, MPE = 0.657). Finally, the short-lag stop comparison of Spanish /t/ and English /d/ provided no evidence suggesting the segments differed from each other (β = −0.069, HDI = [−0.546, 0.431], ROPE = 0.326, MPE = 0.608).

Center of gravity. The center of gravity (COG) data is plotted in the first row, third panel of Figure 1. English stops had a higher COG than Spanish stops (β = 0.687, HDI = [0.555, 0.810], ROPE = 0, MPE = 1), and, overall, voiceless segments had a higher COG than the voiced segments (β = −0.298, HDI = [−0.354, −0.244], ROPE = 0, MPE = 1), however, the two predictors interacted (β = 0.214, HDI = [0.157, 0.266], ROPE = 0, MPE = 1). The interaction was driven by a large COG difference between Spanish coronals (Spanish /d/ vs. Spanish /t/: β =

−1.024, HDI = [−1.184, −0.868], ROPE = 0, MPE = 1) that was not present in the English coronals (English /d/ vs. English /t/: β = −0.169, HDI = [−0.327, −0.012], ROPE = 0.171, MPE = 0.983)). No other predictors had an effect on COG. The post-hoc comparison of the short-lag stops from English and Spanish provided compelling evidence for a between-language difference (β = 0.778, HDI = [0.496, 1.048], ROPE = 0, MPE = 1). Figure 12 and Table 3 provide the complete analysis.

Kurtosis. The kurtosis data is plotted in the second row, first column of Figure 1. English stop bursts had lower kurtosis with regard to Spanish stop bursts (β = −0.646, HDI = [−0.755, −0.542], ROPE = 0, MPE = 1), and the voiced segments presumably had a higher kurtosis than the voiceless segments (β = 0.290, HDI = [0.221, 0.361], ROPE = 0, MPE = 1), though there was evidence of a language × voicing interaction (β = −0.263, HDI = [−0.332, −0.195], ROPE = 0, MPE = 1). Specifically, kurtosis was higher in the voiced stop bursts of Spanish (Spanish /d/ vs. Spanish /t/: β = 1.106, HDI = [0.905, 1.292], ROPE = 0, MPE = 1), but there was no evidence of a voicing difference in the English data (English /d/ vs. English /t/: β = 0.054, HDI = [−0.149, 0.247], ROPE = 0.659, MPE = 0.712). Neither F1 (β = −0.027, HDI = [−0.067, 0.015], ROPE = 1, MPE = 0.903) nor F2 (β = 0.001, HDI = [−0.038, 0.043], ROPE = 1, MPE = 0.509) had any influence on kurtosis. The post-hoc comparison of short-lag stops provided compelling evidence for a between-language difference (β = −0.713, HDI = [−0.969, −0.456], ROPE = 0, MPE = 1). See Figure 12 and Table 3 of the supplementary materials.

Standard deviation. With regard to standard deviation, we observe the same pattern found in the COG data (See third panel, first row and second panel, second row in Figure 1). That is, there was a difference between English and Spanish (β = 0.490, HDI = [0.343, 0.648], ROPE = 0, MPE = 1), as well as between voiced and voiceless segments (β = −0.282, HDI = [−0.355,

−0.204], ROPE = 0, MPE = 1), though, again, the two predictors interacted (β = 0.210, HDI = [0.141, 0.293], ROPE = 0, MPE = 1). Specifically, there was only a voicing difference for Spanish (Spanish /d/ vs. Spanish /t/: β = −0.985, HDI = [−1.193, −0.772], ROPE = 0, MPE = 1), where we see lower standard deviation values in voiced stops. No such difference is observed between the English coronals (English /d/ vs. English /t/: β = −0.143, HDI = [−0.365, 0.066], ROPE = 0.338, MPE = 0.91). Height and frontedness of the following vowel had no effect on standard deviation in the stop burst (F1: β = −0.001, HDI = [−0.051, 0.051], ROPE = 1, MPE = 0.509; F2: β = 0.007, HDI = [−0.044, 0.061], ROPE = 1, MPE = 0.595). Our comparison of Spanish /t/ and English /d/ showed that the short-lag stops differed in kurtosis to some degree (β = 0.417, HDI = [0.075, 0.753], ROPE = 0.011, MPE = 0.994), with only 1% of the HDI falling within our pre-determined ROPE (See Figure 12 and Table 3).

Skewness. The analysis of skewness of the stop burst showed a similar pattern as the one observed in the analysis of kurtosis. If we compare the first and third panels (second row) of Figure 1, we observe a language effect (β = −0.665, HDI = [−0.774, −0.561], ROPE = 0, MPE = 1), such that Spanish stops show higher skewness values, as well a voicing effect (β = 0.290, HDI = [0.222, 0.362], ROPE = 0, MPE = 1), which is driven by a language × voicing interaction (β = −0.221, HDI = [−0.292, −0.155], ROPE = 0, MPE = 1). Specifically, the Spanish voiced coronal had higher skewness of the burst than the voiceless counterpart (Spanish /d/ vs. Spanish /t/: β = 1.023, HDI = [0.827, 1.214], ROPE = 0, MPE = 1), and this difference is not present in the English data (English /d/ vs. English /t/: β = 0.139, HDI = [−0.049, 0.34], ROPE = 0.343, MPE = 0.92). Again, F1 and F2 did not affect skewness values in the burst (F1: β = −0.016, HDI = [−0.058, 0.022], ROPE = 1, MPE = 0.771; F2: β = −0.011, HDI = [−0.051, 0.030], ROPE = 1, MPE = 0.708). Upon comparing the short-lag stops from English and Spanish, we find compelling evidence for a between-language difference (β = −0.748, HDI = [−0.995, −0.484], ROPE = 0, MPE = 1).

Interim discussion.

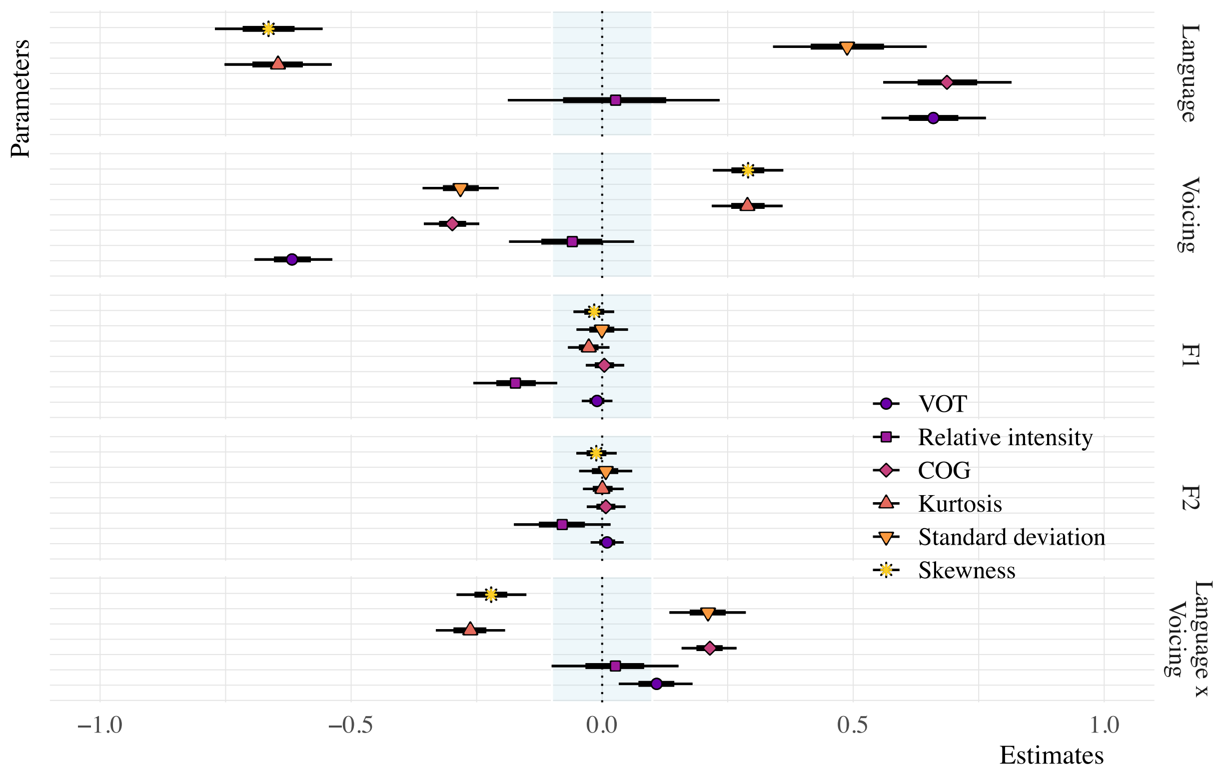

Figure 2: Posterior medians ±95% and 66% credible intervals for VOT and burst metrics of monolingual coronal stops. Individual point shapes and colors represent the six metrics analyzed. In this analysis, 'language' indicates between-language comparisons of different speakers.

The analyses of the monolingual coronal stops are summarized in the forest plot in Figure 2 and Table 2 of the supplementary materials. Unsurprisingly, VOT was realized in a language-specific manner. Spanish voiced and voiceless stops were produced with lead-VOT and short-lag VOT, respectively; English voiced and voiceless stops were produced with short-lag VOT and long-lag VOT, respectively. Relative intensity did not distinguish between voicing contrasts in either language. Spectral moments, on the other hand, were particularly useful for Spanish coronals. As shown in Figure 1, all metrics, with the exception of relative intensity, could be used to differentiate between Spanish voiced and voiceless segments (See also Figure 13 and Table 4 of the supplementary materials). Skewness and kurtosis patterned similarly, as did COG and standard deviation. The spectral moments did not vary as a function of voicing for the English coronals. We also found evidence for between-language differences with regard to the short-lag stops for all metrics except VOT and relative intensity (See Table 3).

Experiment 2: Bilinguals

For the analysis of the bilingual coronal data, we again modeled VOT and the burst metrics as a function of language (English, Spanish), voicing (/d/, /t/), standardized F1 and F2, and item repetitions. In other words, we carried out the same analysis using data from bilingual individuals. In this case, however, the fixed effect *language* is a within-participant factor. That is, the bilinguals provided data from each of their two languages. Figure 3 plots VOT and the burst metrics as a function of language (English, Spanish) and voicing (/d/, /t/). For all plots, the y-axis represents the outcome variable in standardized units, the x-axis indicates the language, and voicing is represented by color.

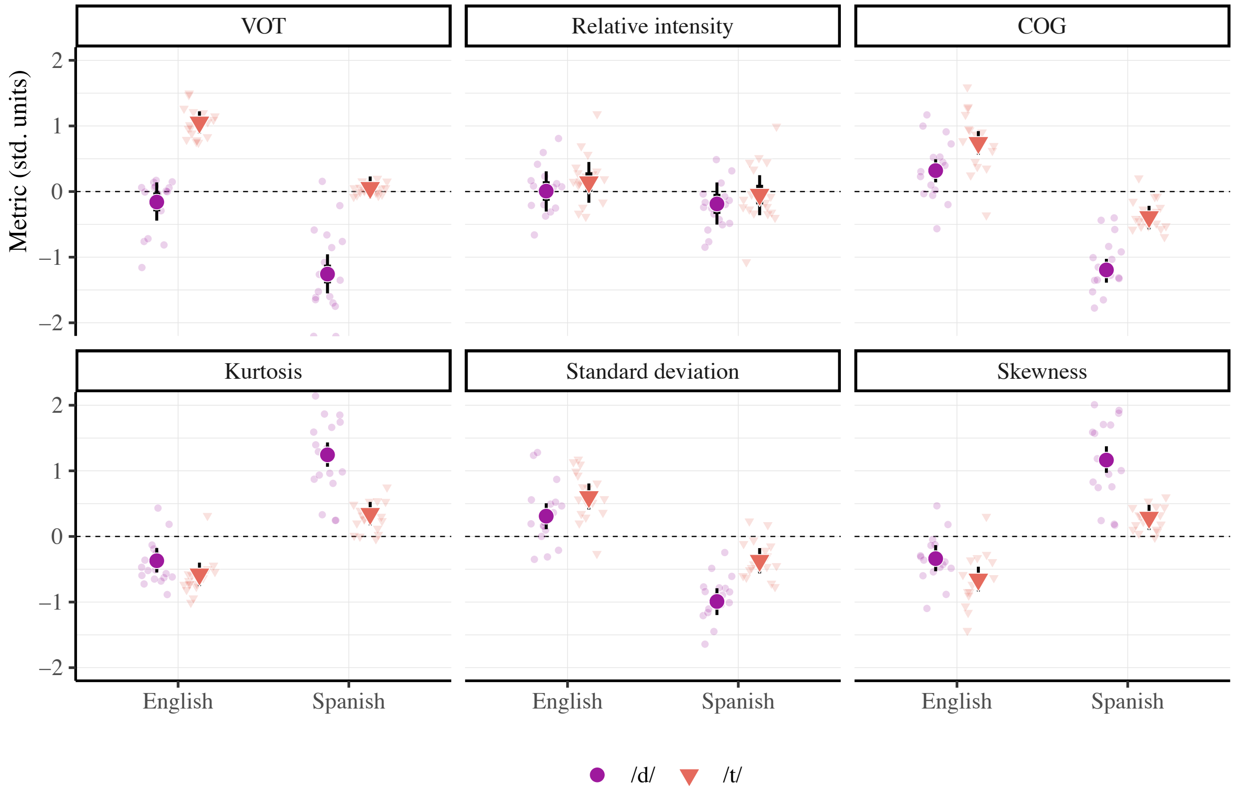

Figure 3: VOT and burst metrics of coronal stops (/d/, /t/) from bilingual speakers as a function of language (English, Spanish). Transparent points represent participant averages. Solid points indicate posterior means ±95% and 66% credible intervals.

Voice-onset time. VOT for English stops was higher than for Spanish stops (β = 0.523, HDI = [0.481, 0.566], ROPE = 0, MPE = 1), as was the case for voiceless segments in comparison with voiced segments (β = −0.633, HDI = [−0.738, −0.522], ROPE = 0, MPE = 1). The two predictors did not interact (β = 0.025, HDI = [−0.017, 0.067], ROPE = 1, MPE = 0.883). Upon comparing the short-lag stops of each language, English /d/ and Spanish /t/, we found minimal evidence that the segments differed from each other (β = −0.219, HDI = [−0.45, −0.001], ROPE = 0.133, MPE = 0.971). Approximately 12% of the HDI fell within the predetermined region of practical equivalence, though given the model, the data, and our prior assumptions, there is a 97% chance the effect is negative, i.e., that English /d/ has lower VOT than Spanish /t/. This is likely due to the bilinguals' tendency to pre-voice in English. The VOT data is plotted in the first panel of Figure 3. The complete model summary is available in Table 5. The short-lag stop comparisons are illustrated in Figure 14 and Figure 15 of the supplementary materials.

Relative intensity. We plot the relative intensity data in the top middle panel of Figure 3. There was no evidence for a group effect (β = 0.097, HDI = [0.030, 0.161], ROPE = 0.543, MPE = 0.997), nor for voicing (β = −0.070, HDI = [−0.147, 0.011], ROPE = 0.787, MPE = 0.959) or F2 (β = −0.073, HDI = [−0.134, −0.014], ROPE = 0.821, MPE = 0.991). We found a relationship between height of the following vowel and relative intensity of the burst (β = −0.224, HDI = [−0.303, −0.139], ROPE = 0, MPE = 1). Concretely, higher F1 values were associated with lower relative intensity. There was no evidence supporting a language × voicing interaction (β = 0.002, HDI = [−0.066, 0.063], ROPE = 1, MPE = 0.512). We found no evidence for a between-language difference of short-lag stops (β = 0.053, HDI = [−0.141, 0.257], ROPE = 0.652, MPE = 0.702).

Center of gravity. COG was higher in English stops than in Spanish stops (β = 0.661, HDI = [0.631, 0.692], ROPE = 0, MPE = 1). Voiceless segments had higher COG than the voiced segments (β = −0.308, HDI = [−0.340, −0.270], ROPE = 0, MPE = 1). We found weak evidence for a language × voicing interaction (β = 0.096, HDI = [0.067, 0.126], ROPE = 0.62, MPE = 1). The difference between voiced and voiceless coronals was slightly larger in Spanish than it was in English. No other predictors had an effect on COG. The COG data is plotted in the first row, third panel of Figure 3.

There was evidence supporting a moderate difference between short-lag /t/ of Spanish and /d/ of English (β = 0.707, HDI = [0.62, 0.799], ROPE = 0, MPE = 1).

Kurtosis. We plot the kurtosis data in the second row, first column of Figure 3. Kurtosis was higher in English stop bursts than Spanish stop bursts (β = −0.635, HDI = [−0.668, −0.603], ROPE = 0, MPE = 1), and also in voiced segments (β = 0.276, HDI = [0.237, 0.313], ROPE = 0, MPE = 1), though there was a language × voicing interaction (β = −0.173, HDI = [−0.204, −0.143], ROPE = 0, MPE = 1). In this case, we observed that kurtosis was higher in the voiced stop bursts of Spanish (Spanish /d/ vs. Spanish /t/: β = 0.897, HDI = [0.798, 1.003], ROPE = 0, MPE = 1), but the voicing difference was much smaller in the English data (English /d/ vs. English /t/: β = 0.207, HDI = [0.11, 0.3], ROPE = 0, MPE = 1). Neither F1 (β = −0.018, HDI = [−0.056, 0.022], ROPE = 1, MPE = 0.808) nor F2 (β = −0.009, HDI = [−0.048, 0.029], ROPE = 1, MPE = 0.666) had any influence on kurtosis. We found evidence supporting a difference between short-lag /t/ of Spanish and /d/ of English (β = −0.718, HDI = [−0.812, −0.621], ROPE = 0, MPE = 1).

Standard deviation. Regarding standard deviation, we again observe the same pattern found in the COG data (See third panel, first row and second panel, second row in Figure 3). In other words, there was a difference between English and Spanish (β = 0.568, HDI = [0.534, 0.604], ROPE = 0, MPE = 1), as well as between segments (β = −0.230, HDI = [−0.282, −0.178], ROPE = 0, MPE = 1). The estimate for the language × voicing interaction falls nearly entirely within our predetermined ROPE, though given the data, the model, and our prior assumptions, there is high probability that the effect it positive (β = 0.082, HDI = [0.046, 0.115], ROPE = 0.866, MPE = 1). Height and frontedness of the following vowel had no effect on standard deviation in the stop burst (F1: β = 0.000, HDI = [−0.043, 0.046], ROPE = 1, MPE = 0.506; F2: β = 0.004, HDI = [−0.038, 0.050], ROPE = 1, MPE = 0.571). Our post-hoc comparison of short-lag stops supports a between-language difference based on standard deviation (β = 0.677, HDI = [0.552, 0.802], ROPE = 0, MPE = 1). **Skewness.** With regard to skewness, we again see a similar pattern as the one observed in the analysis of kurtosis (See first and third panels in the second row of Figure 3). We find an effect of language effect (β = −0.611, HDI = [−0.642, −0.576], ROPE = 0, MPE = 1), such that Spanish stops show higher skewness values, and voicing (β = 0.298, HDI = [0.264, 0.337], ROPE = 0, MPE = 1), which is driven by a language × voicing interaction (β = −0.140, HDI = [−0.169, −0.105], ROPE = 0, MPE = 1). The Spanish voiced /d/ had higher skewness than voiceless /t/ (Spanish /d/ vs. Spanish /t/: β = 0.875, HDI = [0.774, 0.978], ROPE = 0, MPE = 1), and this difference is much smaller in the English data (English /d/ vs. English /t/: β = 0.317, HDI = [0.221, 0.41], ROPE = 0, MPE = 1). Again, F1 and F2 had no affect on skewness (F1: β = −0.029, HDI = [−0.068, 0.012], ROPE = 1, MPE = 0.921; F2: β = −0.008, HDI = [−0.047, 0.031], ROPE = 1, MPE = 0.649). Finally, our post-hoc analysis comparing short-lag stops provided compelling evidence for between-language differences in the burst (β = −0.627, HDI = [−0.72, −0.536], ROPE = 0, MPE = 1).

Interim discussion.

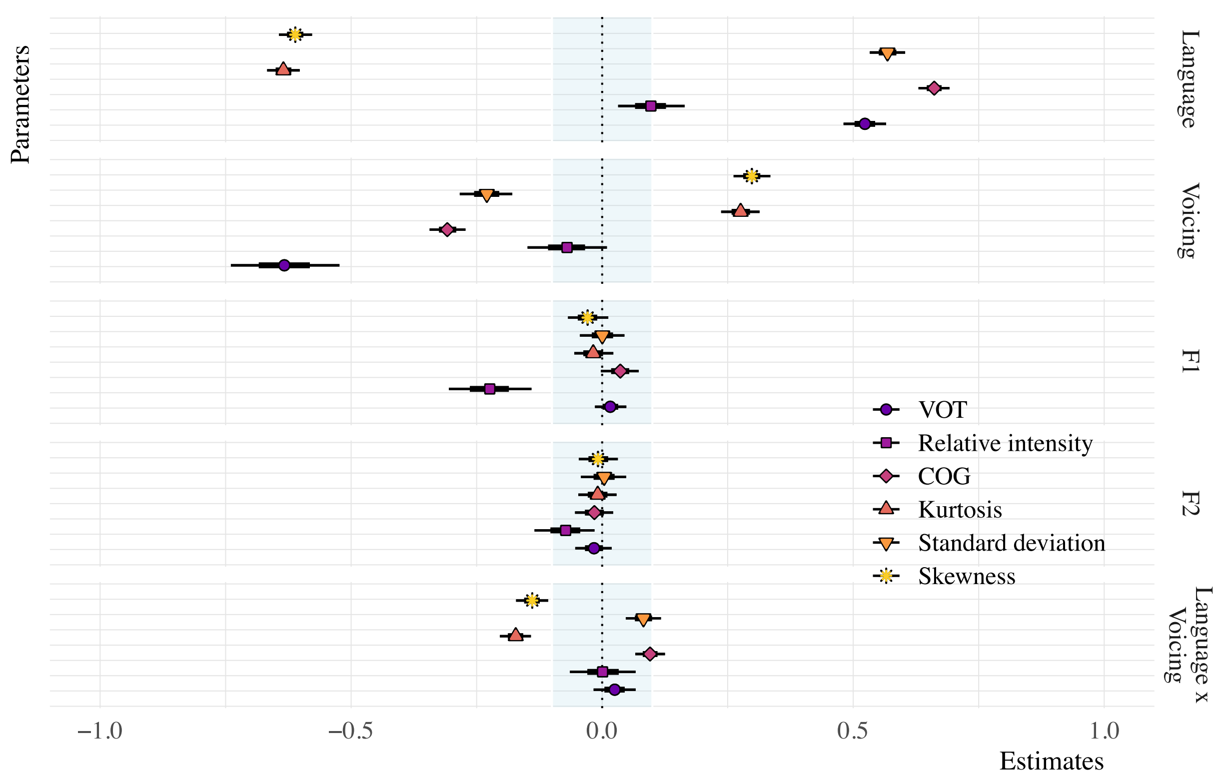

Figure 4: Posterior medians ±95% and 66% credible intervals for VOT and burst metrics of bilingual coronal stops. Individual point shapes and colors represent the six metrics analyzed. In this analysis, 'Language' refers to a within-participant comparison between English and Spanish.

The analyses of the bilingual coronal stops are summarized in the forest plot in Figure 4 and Table 5 of the supplementary materials. Again, we see VOT realized in a language-specific manner. The bilinguals produced Spanish voiced and voiceless stops with lead-VOT and short-lag VOT, respectively, and the English voiced and voiceless stops with short-lag VOT and long-lag VOT, respectively. The voicing contrasts of English and Spanish did not vary based on relative intensity, though all the spectral moments did for the Spanish data. Similar to monolingual data, skewness and kurtosis patterned similarly, as did COG and standard deviation. In many regards the results of the bilingual analyses were virtually identical to those of the monolingual analyses (compare Figure 1 with Figure 3); however, we do observe several important differences. First, we find more variability overall, i.e., wider posterior distributions, in the bilingual data. Second, the bilingual data showed more pre-voicing in English /d/, which was not present to the same extent in the monolingual data. Finally, small voicing distinctions in four of the spectral moments (center of gravity, standard deviation, skewness, and, to a lesser extent, kurtosis) were present in the bilinguals' production of English coronals. Figure 5 plots the posterior distributions of the /d/-/t/ difference estimates for each spectral moment and relative intensity in English and Spanish (See also Table 7 of the supplementary materials). We see the same general pattern for each language, though the difference estimates are much larger in the Spanish data. The bilingual participants appear to have carried over some of the burst characteristics of their Spanish coronals into their English. Finally, post-hoc comparisons of Spanish and English short-lag stops suggested that the bilinguals, like their monolingual counterparts, had distinct burst characteristics according to all metrics except VOT and relative intensity (See Table 6).

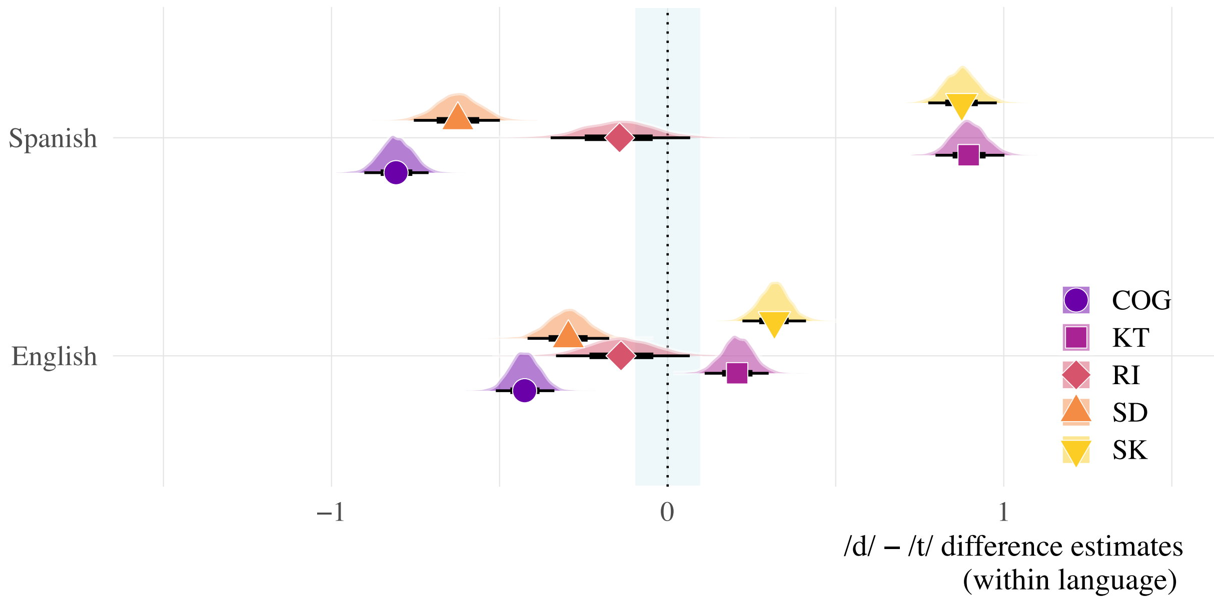

Figure 5: Posterior distributions of difference estimates between voiced and voiceless stops for Spanish and English. Each point represents the posterior mean difference ±95% HDI of the posterior of the voiceless segment subtracted from the posterior of the voiced segment. The light blue region represents a ROPE of ±0.1.

Experiment 3: Bilingual POA data

The final set of analyses compared a subset of the bilingual coronal stop data with bilabial data from a separate group of bilinguals (see Method section). We analyzed VOT and the same burst metrics as the previous analyses. In this case, however, the outcome variables were modeled as a function of language (English, Spanish), place of articulation (POA: bilabial, coronal), standardized F1 and F2, and item repetitions. We only included the voiceless stops (/p/, /t/) from each language. Importantly, the fixed effect *POA* is a between-participant factor. That is, one set of bilingual individuals provided the bilabial data in English and Spanish, and another set of bilingual individuals (those from the previous experiment) provided the coronal data in English and Spanish. Figure 6 plots VOT and the burst metrics as a function of language (English, Spanish) and POA (bilabial, coronal). For all plots, the y-axis represents the outcome variable in standardized units, the x-axis represents POA, and language is represented by color.

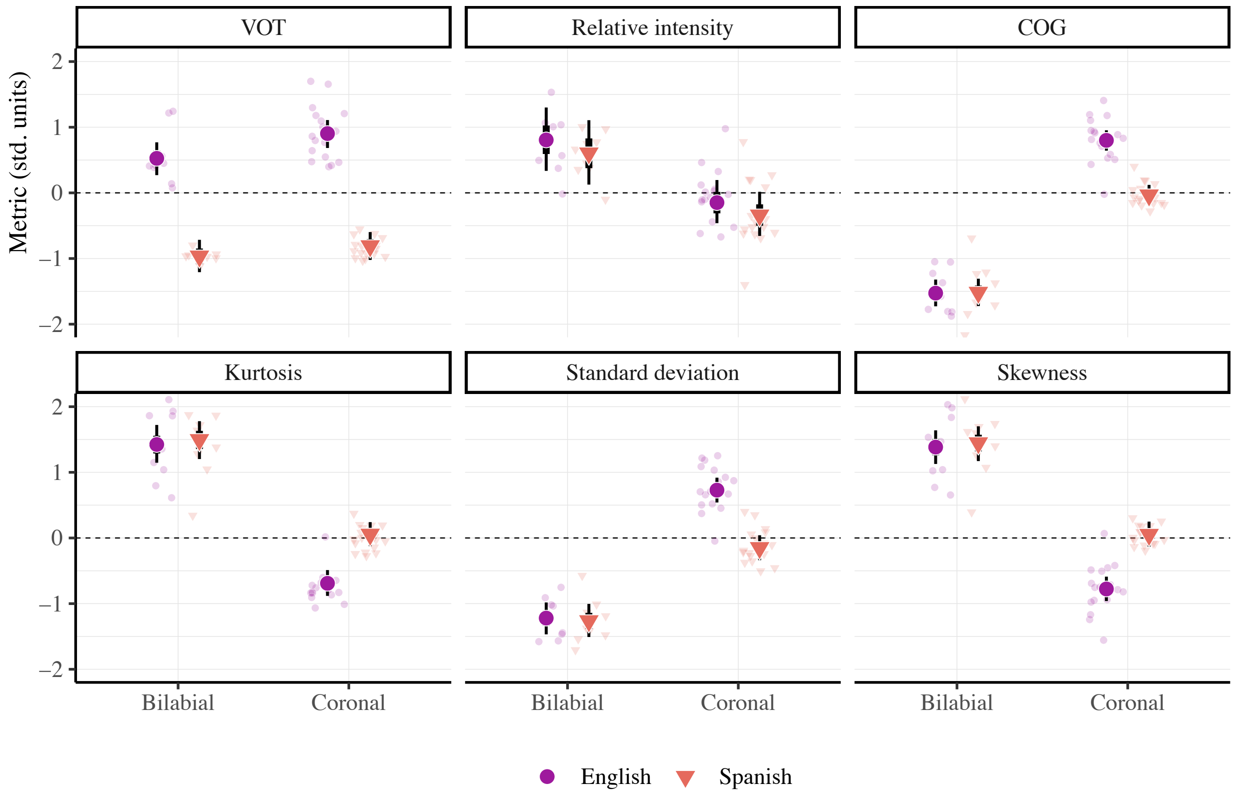

Figure 6: VOT and burst metrics of voiceless stops from bilingual speakers as a function of place of articulation (bilabial, coronal) and language (English, Spanish). Transparent points represent participant averages. Solid points indicate posterior means ±95% and 66% credible intervals.

Voice-onset time. The VOT data is plotted in the first panel of Figure 6. The complete model summary is available in Figure 7 below, as well as Table 8 of the supplementary materials. There was a strong effect of language (β = 0.802, HDI = [0.750, 0.852], ROPE = 0, MPE = 1). English voiceless stops had higher VOT than Spanish voiceless stops regardless of POA (β = 0.134, HDI = [0.029, 0.236], ROPE = 0.248, MPE = 0.992). We did not find compelling evidence for a language × POA interaction (β = 0.055, HDI = [0.004, 0.105], ROPE = 0.983, MPE = 0.984), nor were there any effects for the formant structure of the following vowel (F1: β = 0.013, HDI = [−0.025, 0.050], ROPE = 1, MPE = 0.751; F2: β = 0.000, HDI = [−0.046, 0.043], ROPE = 1, MPE = 0.505).

Relative intensity. Relative intensity of the burst was slightly higher in English than in Spanish (β = 0.099, HDI = [0.042, 0.155], ROPE = 0.518, MPE = 1), though nearly half of the posterior mass fell within our pre-established region of practical equivalence. Relative intensity was higher in bilabial stops than in coronals stops (β = −0.475, HDI = [−0.680, −0.269], ROPE = 0, MPE = 1). The two predictors did not interact (β = −0.003, HDI = [−0.056, 0.057], ROPE = 1, MPE = 0.542). We found evidence for a relationship between height of the following vowel and relative intensity of the burst (β = −0.222, HDI = [−0.304, −0.136], ROPE = 0, MPE = 1). Concretely, higher F1 values were associated with lower relative intensity. There was no evidence suggesting frontedness of the following vowel had any effect (β = −0.059, HDI = [−0.116, −0.001], ROPE = 0.945, MPE = 0.978). The relative intensity data is plotted in the second panel of Figure 6 and the complete model summary is available in Figure 7 or Table 8 of the supplementary materials.

Center of gravity. There was an effect of language (β = 0.206, HDI = [0.177, 0.236], ROPE = 0, MPE = 1) and POA (β = 0.953, HDI = [0.862, 1.041], ROPE = 0, MPE = 1) and the two predictors interacted (β = 0.211, HDI = [0.181, 0.239], ROPE = 0, MPE = 1). As illustrated in the third panel of Figure 6, COG was higher in English /t/ than in Spanish /t/ (Spanish /t/ vs. English /t/: β = 0.835, HDI = [0.781, 0.888], ROPE = 0, MPE = 1), but there was no such difference in the bilabials (Spanish /p/ vs. English /p/: β = −0.01, HDI = [−0.112, 0.093], ROPE = 0.975, MPE = 0.574). The following vowel did not affect COG of the bursts (F1: β = 0.019, HDI = [−0.008, 0.049], ROPE = 1, MPE = 0.905; F2: β = −0.010, HDI = [−0.041, 0.020], ROPE

= 1, MPE = 0.743). The model summary is presented in the forest plot in Figure 7 and in Table 8 of the supplementary materials.

Kurtosis. There was an effect of language (β = −0.204, HDI = [−0.237, −0.173], ROPE = 0, MPE = 1) and POA (β = −0.888, HDI = [−1.006, −0.774], ROPE = 0, MPE = 1) and the two predictors interacted (β = −0.169, HDI = [−0.200, −0.137], ROPE = 0, MPE = 1). As illustrated in the first panel of the second row in Figure 6, kurtosis was higher in Spanish /t/ than in English /t/ (Spanish /t/ vs. English /t/: β = −0.747, HDI = [−0.809, −0.692], ROPE = 0, MPE = 1), but there was no such difference in the bilabials (Spanish /p/ vs. English /p/: β = −0.071, HDI = [−0.186, 0.039], ROPE = 0.701, MPE = 0.891). The following vowel did not affect kurtosis in the bursts (F1: β = 0.015, HDI = [−0.016, 0.047], ROPE = 1, MPE = 0.831; F2: β = −0.015, HDI = [−0.047, 0.017], ROPE = 1, MPE = 0.822). The model summary is available in Figure 7 (See also Table 8 of the supplementary materials).

Standard deviation. Similar to the COG and kurtosis analyses, we again see an effect of language (β = 0.230, HDI = [0.186, 0.271], ROPE = 0, MPE = 1) and POA (β = 0.766, HDI = [0.671, 0.870], ROPE = 0, MPE = 1) and the two predictors interacted (β = 0.209, HDI = [0.169, 0.250], ROPE = 0, MPE = 1). As illustrated in the middle panel of the second row in Figure 6, standard deviation was higher in English /t/ than in Spanish /t/ (Spanish /t/ vs. English /t/: β = 0.876, HDI = [0.802, 0.951], ROPE = 0, MPE = 1), but there was no such difference in the bilabials (Spanish /p/ vs. English /p/: β = 0.042, HDI = [−0.107, 0.187], ROPE = 0.784, MPE = 0.722). The following vowel did not affect standard deviation of the bursts (F1: β = −0.012, HDI = [−0.050, 0.030], ROPE = 1, MPE = 0.725; F2: β = 0.008, HDI = [−0.035, 0.049], ROPE = 1, MPE = 0.662). The model summary is available in Figure 7, as well as Table 8 of the supplementary materials.

Skewness. Finally, for skewness we find an effect of language (β = −0.222, HDI = [−0.255, −0.188], ROPE = 0, MPE = 1) and POA (β = −0.888, HDI = [−0.994, −0.781], ROPE = 0, MPE = 1) and the two predictors interacted (β = −0.192, HDI = [−0.226, −0.159], ROPE = 0, MPE = 1). As illustrated in the final panel of the second row in Figure 6, skewness was higher in Spanish /t/ than in English /t/ (Spanish /t/ vs. English /t/: β = −0.829, HDI = [−0.889, −0.765], ROPE = 0, MPE = 1), but there was no such difference in the bilabials (Spanish /p/ vs. English /p/: β = −0.061, HDI = [−0.176, 0.06], ROPE = 0.747, MPE = 0.845). The following vowel did not affect skewness of the bursts (F1: β = −0.024, HDI = [−0.058, 0.009], ROPE = 1, MPE = 0.924; F2: β = 0.009, HDI = [−0.025, 0.042], ROPE = 1, MPE = 0.696). The model summary is available in Figure 7 below and Table 8 of the supplementary materials.

Interim discussion.

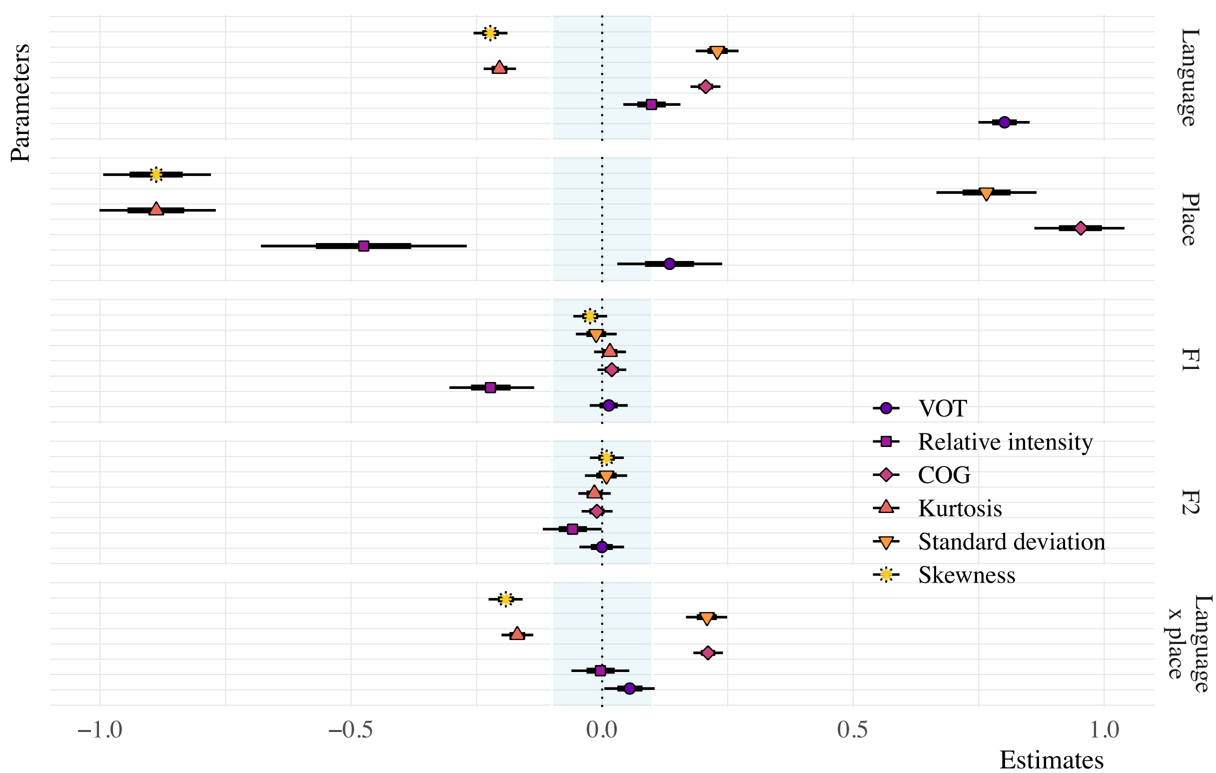

Figure 7: Posterior medians ±95% and 66% credible intervals for VOT and burst metrics of bilingual voiceless bilabial and coronal stops. Individual point shapes and colors represent the

six metrics analyzed. In this analysis, 'Place' refers to a between-participants comparison, as the bilabial data comes from a separate group of bilingual individuals.

The analyses of the bilingual bilabial and coronal voiceless stops are summarized in the forest plot in Figure 7 and Table 8 of the supplementary materials. The analyses revealed four important findings. First, VOT was realized in a language-specific manner that was was similar for bilabials and coronals. As with the previous analyses, Spanish voiceless stops were realized with short-lag VOT and English voiceless stops were realized with long-lag VOT. Second, relative intensity can capture place of articulation differences, which are realized similarly in English and Spanish. Third, we again find that COG and standard deviation pattern similarly, as do kurtosis and skewness, across languages for bilabial and coronal stops. And, finally, we see that English and Spanish coronals differ with regard to all four spectral moment measurements, where as bilabials do not. Figure 8 plots the posterior distributions of the Spanish-English difference estimates for VOT, relative intensity, and each spectral moment for bilabials and coronals (See also Table 9). One can observe clear between language differences for VOT in bilabials and coronals (i.e., short-lag vs. long-lag realizations), but only for coronals (orange triangles) in the spectral moments. Thus, all four spectal moment measurements capture the subtle articulatory differences (dental vs. alveolar) between English and Spanish coronals stops. No such fine-phonetic differences are attested in bilabials, and, unsurprisingly, we find no between-language differences at bilabial place.

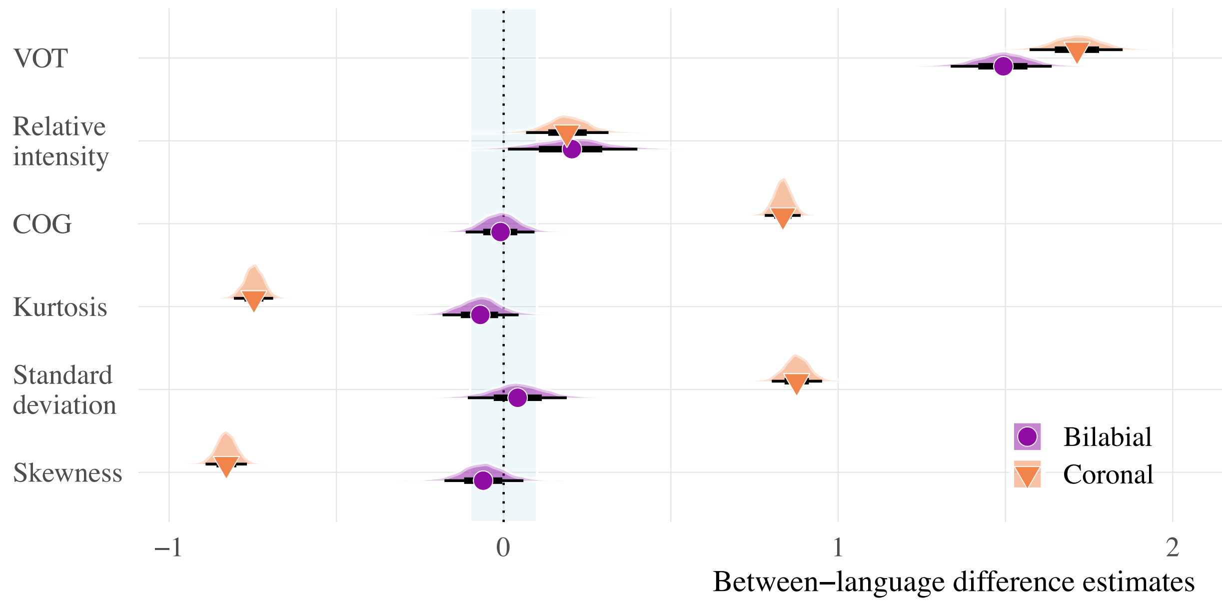

Figure 8: Posterior distributions of difference estimates between Spanish and English for bilabial and coronal stops. Each point represents the posterior mean difference ±95% HDI of the posterior of the Spanish values subtracted from the posterior of the English values. The light blue region represents a ROPE of ±0.1.

Discussion

Summary of findings

Interpretation and implications

Conclusion

References

Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer [computer program]*. Retrieved from http://www.praat.org/

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395–411. https://doi.org/10.32614/RJ-2018-017

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

Gelman, A., Simpson, D., & Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy*, *19*(10), 1–13. https://doi.org/10.3390/e19100555

Gertken, L. M., Amengual, M., & Birdsong, D. (2014). *Assessing Language Dominance with the Bilingual Language Profile* (P. Leclercq, A. Edmonds, & H. Hilton, Eds.). Bristol: Multilingual Matters.

Kruschke, J. K. (2018). Rejecting or accepting parameter values in bayesian estimation. *Advances in Methods and Practices in Psychological Science*, *1*(2), 270–280.

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Schoot, R. van de, & Depaoli, S. (2014). Bayesian analyses: Where to start and what to report. *European Health Psychologist*, *16*(2), 75–84.

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, *71*, 147–161.

Supplementary materials

Appendix A: Vowels

Prior to the stop analyses we scrutinized the formant structure, F1 and F2, of the vowel following the coronal stops. The purpose of this analysis was to determine if the low /a/ and /æ/ vowels of Spanish and English, respectively, differed from each other. This analysis was taken as a precautionary measure with the objective of determining whether or not coarticulation effects may be present in the stop metrics due to the possible different spectral envelopes of the next segment.

The F1 and F2 data were analyzed using separate Bayesian multilevel models. The criterion (F1 or F2) were standardized and modeled as a function of language (English, Spanish), phoneme (/d/, /t/), and item repetition. Language and phoneme were sum coded (-1, 1). The random effects structure included by-subject intercepts with random slopes for phoneme and item repetition, as well as by-item intercepts with a random slope for item repetition. The model included weakly informative priors with the mean centered at 0 and a standard deviation of 2.

The model suggested weak evidence for a language effect on F1 ($\beta = 0.008$, HDI = [−0.221, 0.203], ROPE = 0.693, MPE = 0.536) and F2 ($\beta = 0.222$, HDI = [−0.007, 0.462], ROPE = 0.121, MPE = 0.967). Together, the point estimates suggest the spectral centroid of the Spanish vowel was slightly higher and more posterior with respect to that of the English vowel. There was no evidence for a phoneme effect on F1 ($\beta = 0.124$, HDI = [0.007, 0.235], ROPE = 0.328, MPE = 0.981), nor on F2 ($\beta = 0.129$, HDI = [−0.042, 0.318], ROPE = 0.353, MPE = 0.917). Given the possibility of a small effect of language on the spectral envelope, standardized F1 and F2 were used in subsequent analyses of the coronal stops to control for any coarticulatory effects on the burst. Figure 9 plots the F1 × F2 data and Figure 10 plots the model summary. A complete summary of the F1 and F2 models is available in Table 1.

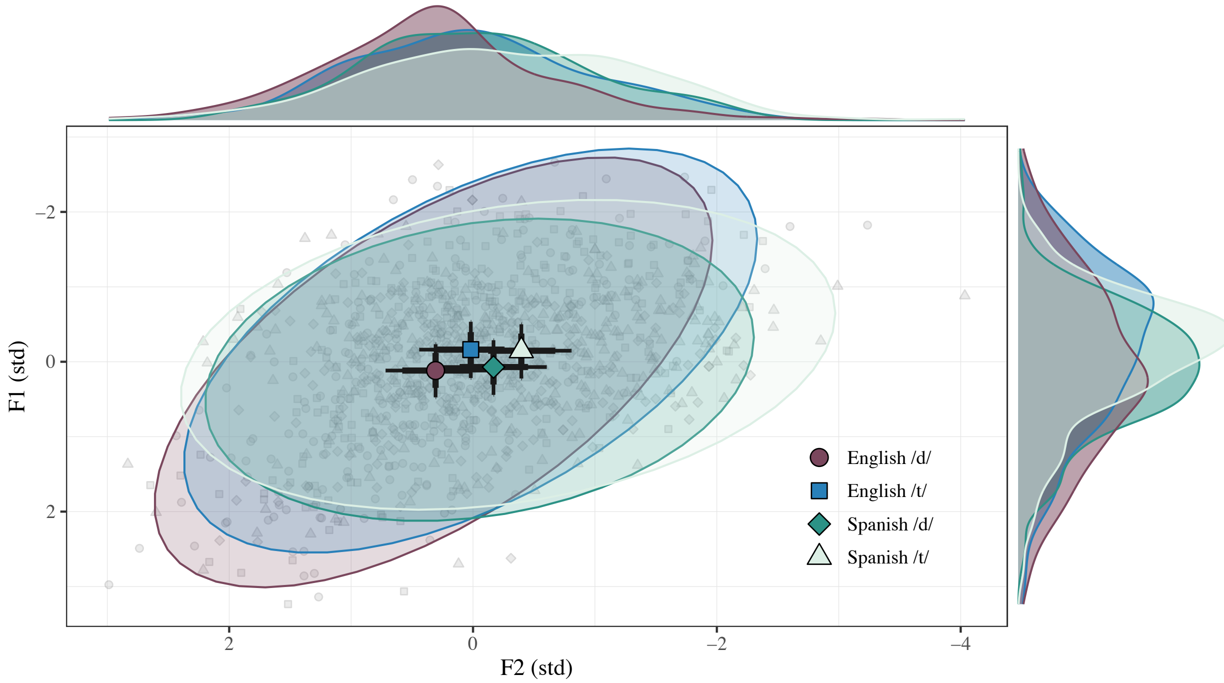

Figure 9: F1 × F2 of /a/ from monolingual speakers as a function of language (English, Spanish). Transparent points represent raw data. Solid points indicate posterior means ±95% and 80% credible intervals.

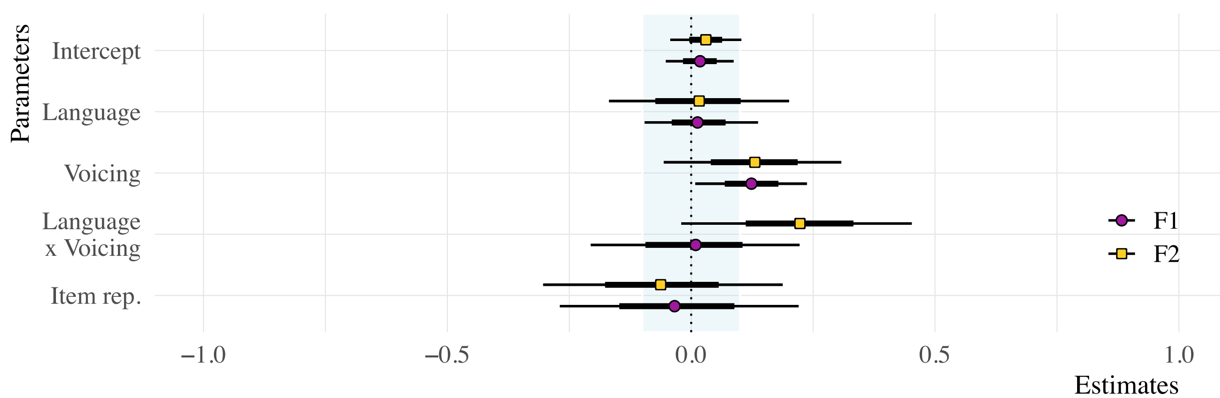

Figure 10: Posterior medians ±95% and 66% credible intervals for F1 and F2 from monolingual speaker data.

Table 1: Model summary for F1 and F2 as a function of language (English, Spanish), phoneme (/d/, /t/), and item repetition for monolingual data. The percentage of the HDI contained within the ROPE is based on an effect size of ±0.1.

Metric	Parameter	Estimate	HDI	ROPE	MPE
F1	Intercept	−0.031	[−0.283, 0.205]	0.587	0.606
	Language	0.008	[−0.221, 0.203]	0.693	0.536

Metric	Parameter	Estimate	HDI	ROPE	MPE
	Voicing	0.124	[0.007, 0.235]	0.328	0.981
	Item rep.	0.018	[−0.050, 0.089]	1.000	0.689
	Language:Voicing	0.015	[−0.097, 0.134]	0.951	0.595
F2	Intercept	−0.060	[−0.289, 0.198]	0.563	0.692
	Language	0.222	[−0.007, 0.462]	0.121	0.967
	Voicing	0.129	[−0.042, 0.318]	0.353	0.917
	Item rep.	0.030	[−0.045, 0.099]	1.000	0.802
	Language:Voicing	0.015	[−0.172, 0.194]	0.757	0.566

Appendix B: Supplementary analyses (monolinguals)

In this section we present additional information from the analyses of the monolingual data.

Table 2 includes the full output of the model summary from the primary multivariate analysis (Note: this table is equivalent to the forest plot in Figure 1).

Table 2: Model summary for VOT and burst metrics as a function of language (English, Spanish), phoneme (/d/, /t/), F1, F2, and item repetition for monolingual coronal stops. The percentage of the HDI contained within the ROPE is based on an effect size of ±0.1.

Metric	Parameter	Estimate	HDI	ROPE	MPE
VOT	Intercept	0.010	[−0.106, 0.133]	0.943	0.566
	Language	0.660	[0.556, 0.763]	0.000	1.000
	Voicing	−0.617	[−0.697, −0.544]	0.000	1.000
	F1	−0.010	[−0.041, 0.020]	1.000	0.752
	F2	0.010	[−0.023, 0.042]	1.000	0.730
	Item rep.	−0.014	[−0.045, 0.018]	1.000	0.806
	Language:Voicing	0.108	[0.033, 0.180]	0.406	0.995
COG	Intercept	0.048	[−0.104, 0.185]	0.775	0.746
	Language	0.687	[0.555, 0.810]	0.000	1.000
	Voicing	−0.298	[−0.354, −0.244]	0.000	1.000
	F1	0.005	[−0.033, 0.044]	1.000	0.590
	F2	0.008	[−0.028, 0.050]	1.000	0.653
	Item rep.	−0.031	[−0.071, 0.008]	1.000	0.934
	Language:Voicing	0.214	[0.157, 0.266]	0.000	1.000
KT	Intercept	−0.026	[−0.163, 0.109]	0.864	0.640
	Language	−0.646	[−0.755, −0.542]	0.000	1.000
	Voicing	0.290	[0.221, 0.361]	0.000	1.000
	F1	−0.027	[−0.067, 0.015]	1.000	0.903
	F2	0.001	[−0.038, 0.043]	1.000	0.509
	Item rep.	0.019	[−0.021, 0.062]	1.000	0.809
	Language:Voicing	−0.263	[−0.332, −0.195]	0.000	1.000
RI	Intercept	−0.111	[−0.347, 0.141]	0.446	0.820
	Language	0.026	[−0.171, 0.246]	0.666	0.602
	Voicing	−0.060	[−0.192, 0.055]	0.742	0.830
	F1	−0.172	[−0.252, −0.085]	0.026	1.000
	F2	−0.080	[−0.177, 0.015]	0.676	0.948
	Item rep.	0.051	[−0.023, 0.131]	0.916	0.912

Metric	Parameter	Estimate	HDI	ROPE	MPE
	Language:Voicing	0.025	[−0.094, 0.157]	0.906	0.657
SD	Intercept	0.039	[−0.148, 0.213]	0.717	0.661
	Language	0.490	[0.343, 0.648]	0.000	1.000
	Voicing	−0.282	[−0.355, −0.204]	0.000	1.000
	F1	−0.001	[−0.051, 0.051]	1.000	0.509
	F2	0.007	[−0.044, 0.061]	1.000	0.595
	Item rep.	−0.027	[−0.077, 0.027]	1.000	0.833
	Language:Voicing	0.210	[0.141, 0.293]	0.000	1.000
SK	Intercept	−0.052	[−0.187, 0.084]	0.777	0.780
	Language	−0.665	[−0.774, −0.561]	0.000	1.000
	Voicing	0.290	[0.222, 0.362]	0.000	1.000
	F1	−0.016	[−0.058, 0.022]	1.000	0.771
	F2	−0.011	[−0.051, 0.030]	1.000	0.708
	Item rep.	0.032	[−0.009, 0.075]	1.000	0.933
	Language:Voicing	−0.221	[−0.292, −0.155]	0.000	1.000

Figure 11 shows the comparison of English /d/ and Spanish /t/ for voice-onset time. One can observe that there is not a meaningful difference between the short-lag segments. Figure 12 shows the comparison of English /d/ and Spanish /t/ for relative intensity and spectral moments. There appears to be a meaningful difference between the short-lag segments for all metrics except relative intensity. Table 3 summarizes all of the short-lag stop comparisons.

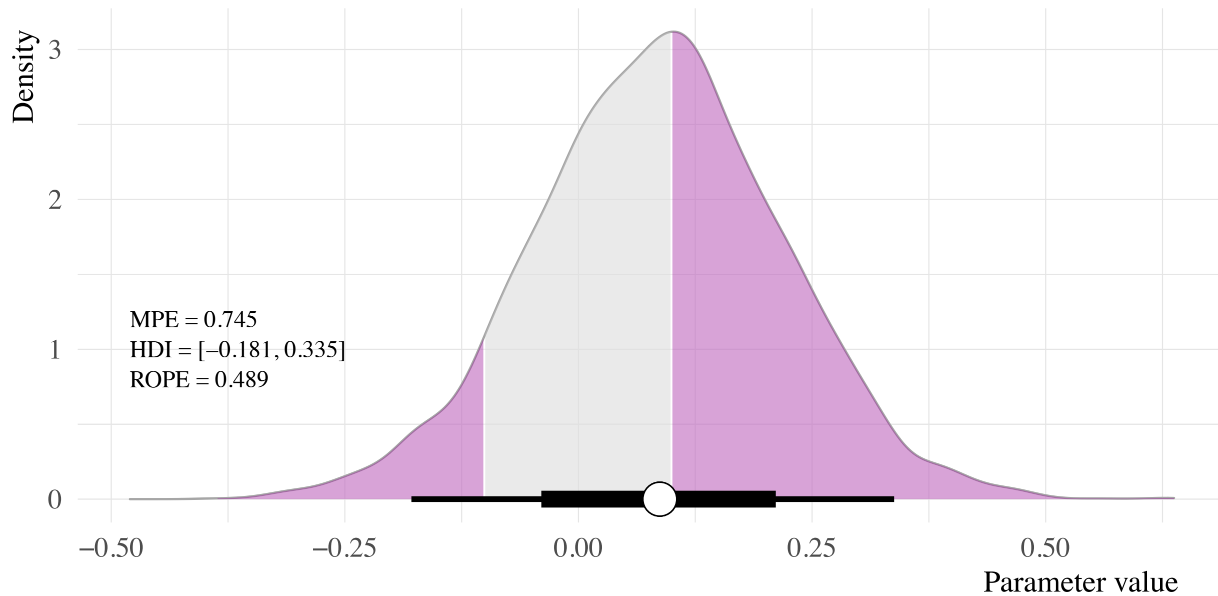

Figure 11: Posterior distribution comparing VOT of the short-lag stops, English /d/ and Spanish /t/. The white point represents the posterior mean ±95% HDI and the grey region represents the ROPE (±0.1).

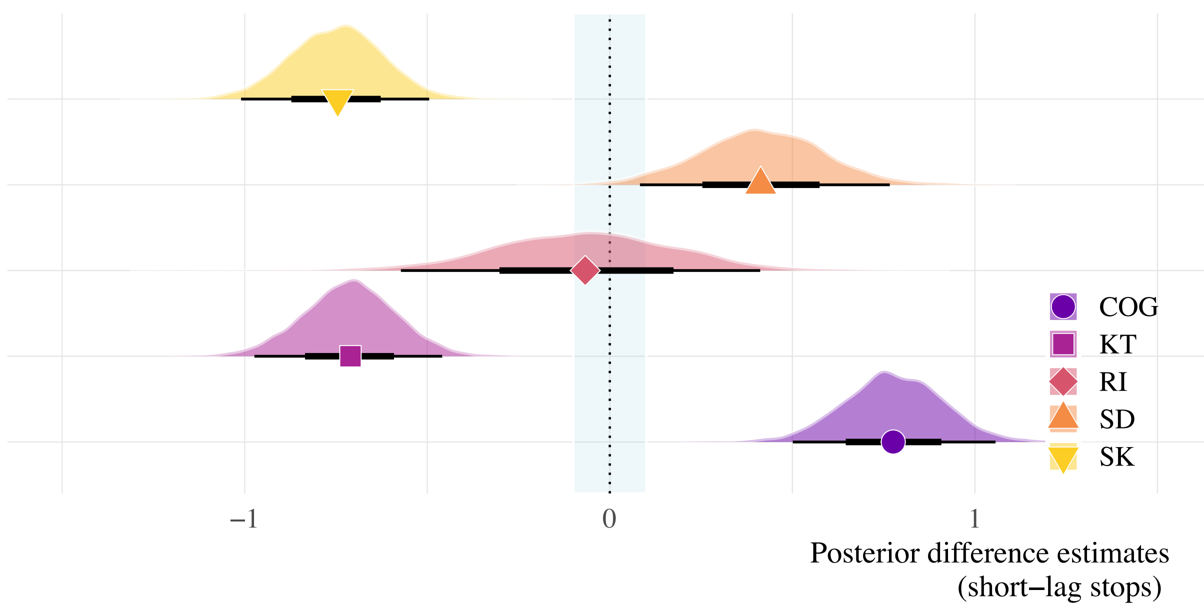

Figure 12: Posterior distribution comparing relative intensity and spectral moments of the short-lag stops, English /d/ and Spanish /t/. The points represents the posterior mean ±95% HDI and the grey region represents the ROPE (±0.1).

Table 3: Post-hoc comparisons of short-lag stops between English and Spanish for each burst metric. Each estimate is derived from the mean difference of the posterior of the Spanish segment subtracted from the posterior of the Englsh segment.

Metric	Estimate	HDI	ROPE	MPE
VOT	0.085	[−0.181, 0.335]	0.489	0.745
COG	0.778	[0.496, 1.048]	0.000	1.000
RI	−0.069	[−0.546, 0.431]	0.326	0.608
SD	0.417	[0.075, 0.753]	0.011	0.994
SK	−0.748	[−0.995, −0.484]	0.000	1.000
KT	−0.713	[−0.969, −0.456]	0.000	1.000

Figure 13 and Table 4 present the post-hoc analyses of within-language comparisons, that is, Spanish /d/ vs. Spanish /t/ and English /d/ vs. English /t/.

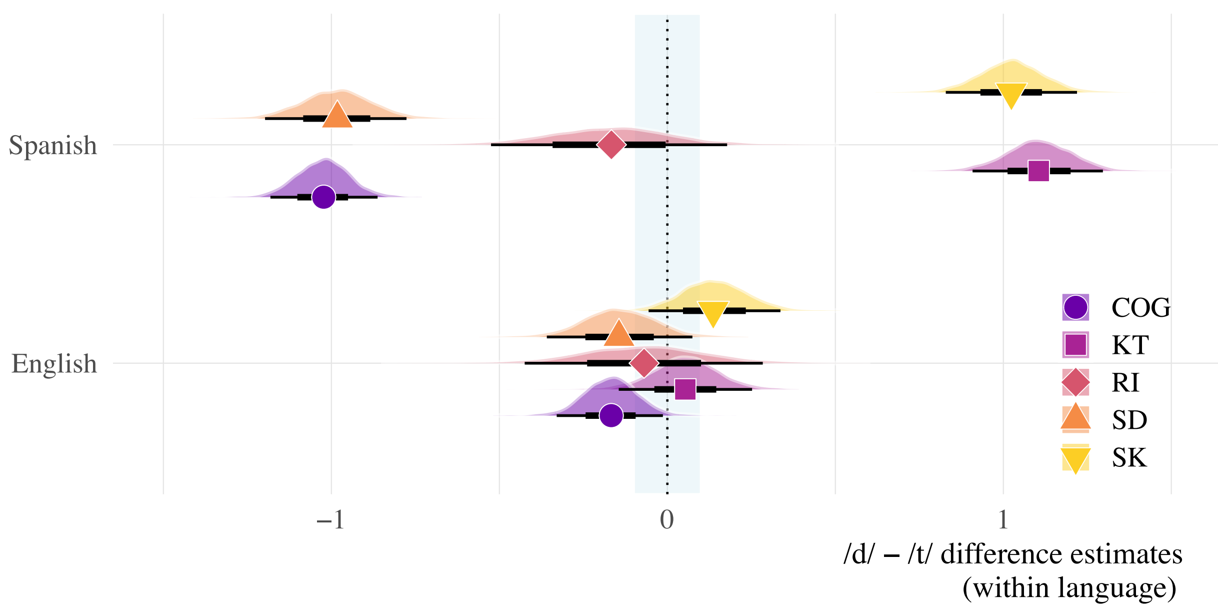

Figure 13: Posterior distributions of difference estimates between voiced and voiceless stops for Spanish and English. The point represents the posterior mean difference ±95% HDI of the posterior of the voiceless segment subtracted from the posterior of the voiced. The light blue region represents the ROPE (±0.1).

Table 4: Post-hoc comparisons of within-language voiced-voiceless pairs for each burst metric. Each point represents the posterior mean difference ±95% HDI of the posterior of the voiceless segment subtracted from the posterior of the voiced segment.

Language	Metric	Estimate	HDI	ROPE	MPE

Language	Metric	Estimate	HDI	ROPE	MPE
Spanish	COG	−1.024	[−1.184, −0.868]	0.000	1.000
	RI	−0.171	[−0.506, 0.197]	0.305	0.839
	SD	−0.985	[−1.193, −0.772]	0.000	1.000
	SK	1.023	[0.827, 1.214]	0.000	1.000
	KT	1.106	[0.905, 1.292]	0.000	1.000
English	COG	−0.169	[−0.327, −0.012]	0.171	0.983
	RI	−0.07	[−0.405, 0.296]	0.420	0.652
	SD	−0.143	[−0.365, 0.066]	0.338	0.910
	SK	0.139	[−0.049, 0.34]	0.343	0.920
	KT	0.054	[−0.149, 0.247]	0.659	0.712

Appendix C: Supplementary analyses (bilinguals)

In this section we present additional information from the analyses of the bilingual data. Table 5 includes the full output of the model summary from the primary multivariate analysis (Note: this table is equivalent to the forest plot in Figure 3).

Table 5: Model summary for VOT and burst metrics as a function of language (English, Spanish), phoneme (/d/, /t/), F1, F2, and item repetition for bilingual coronal stops. The percentage of the HDI contained within the ROPE is based on an effect size of ±0.1.

Metric	Parameter	Estimate	HDI	ROPE	MPE
VOT	Intercept	−0.075	[−0.209, 0.059]	0.647	0.863
	Language	0.523	[0.481, 0.566]	0.000	1.000
	Voicing	−0.633	[−0.738, −0.522]	0.000	1.000
	F1	0.016	[−0.013, 0.049]	1.000	0.847
	F2	−0.016	[−0.053, 0.020]	1.000	0.813
	Item rep.	−0.009	[−0.038, 0.019]	1.000	0.728
	Language:Voicing	0.025	[−0.017, 0.067]	1.000	0.883
COG	Intercept	−0.130	[−0.248, −0.003]	0.318	0.984
	Language	0.661	[0.631, 0.692]	0.000	1.000
	Voicing	−0.308	[−0.340, −0.270]	0.000	1.000
	F1	0.036	[−0.001, 0.074]	1.000	0.967
	F2	−0.015	[−0.053, 0.022]	1.000	0.783
	Item rep.	0.012	[−0.023, 0.049]	1.000	0.746
	Language:Voicing	0.096	[0.067, 0.126]	0.620	1.000
KT	Intercept	0.162	[0.047, 0.291]	0.128	0.994
	Language	−0.635	[−0.668, −0.603]	0.000	1.000
	Voicing	0.276	[0.237, 0.313]	0.000	1.000
	F1	−0.018	[−0.056, 0.022]	1.000	0.808
	F2	−0.009	[−0.048, 0.029]	1.000	0.666
	Item rep.	−0.028	[−0.064, 0.010]	1.000	0.936
	Language:Voicing	−0.173	[−0.204, −0.143]	0.000	1.000
RI	Intercept	−0.020	[−0.217, 0.170]	0.711	0.582
	Language	0.097	[0.030, 0.161]	0.543	0.997
	Voicing	−0.070	[−0.147, 0.011]	0.787	0.959
	F1	−0.224	[−0.303, −0.139]	0.000	1.000
	F2	−0.073	[−0.134, −0.014]	0.821	0.991
	Item rep.	0.005	[−0.039, 0.053]	1.000	0.587

Metric	Parameter	Estimate	HDI	ROPE	MPE
	Language:Voicing	0.002	[−0.066, 0.063]	1.000	0.512
SD	Intercept	−0.111	[−0.240, 0.022]	0.426	0.951
	Language	0.568	[0.534, 0.604]	0.000	1.000
	Voicing	−0.230	[−0.282, −0.178]	0.000	1.000
	F1	0.000	[−0.043, 0.046]	1.000	0.506
	F2	0.004	[−0.038, 0.050]	1.000	0.571
	Item rep.	0.014	[−0.024, 0.056]	1.000	0.750
	Language:Voicing	0.082	[0.046, 0.115]	0.866	1.000
SK	Intercept	0.116	[−0.014, 0.253]	0.403	0.959
	Language	−0.611	[−0.642, −0.576]	0.000	1.000
	Voicing	0.298	[0.264, 0.337]	0.000	1.000
	F1	−0.029	[−0.068, 0.012]	1.000	0.921
	F2	−0.008	[−0.047, 0.031]	1.000	0.649
	Item rep.	−0.008	[−0.046, 0.029]	1.000	0.664
	Language:Voicing	−0.140	[−0.169, −0.105]	0.000	1.000

Figure 14 shows the comparison of English /d/ and Spanish /t/ for voice-onset time. One can observe that there is not a meaningful difference between the short-lag segments. Figure 15 shows the comparison of English /d/ and Spanish /t/ for relative intensity and spectral moments. There appears to be a meaningful difference between the short-lag segments for all metrics except relative intensity. Table 6 summarizes all of the short-lag stop comparisons.

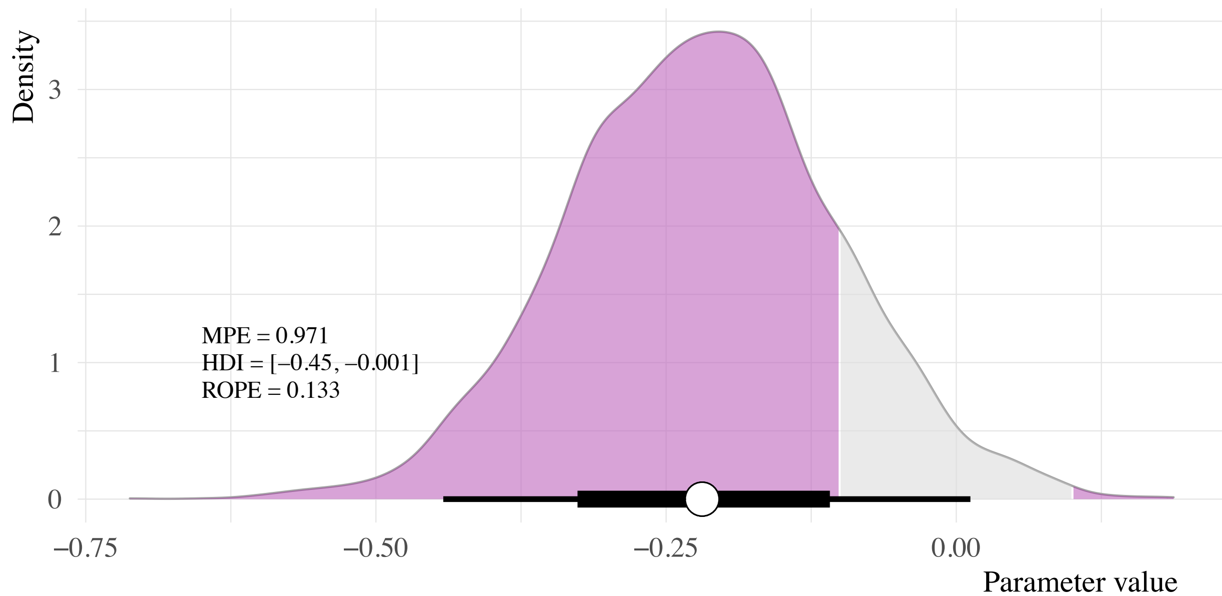

Figure 14: Posterior distribution comparing the short-lag stops, English /d/ and Spanish /t/. The white point represents the posterior mean ±95% HDI and the grey region represents the ROPE (±0.1).

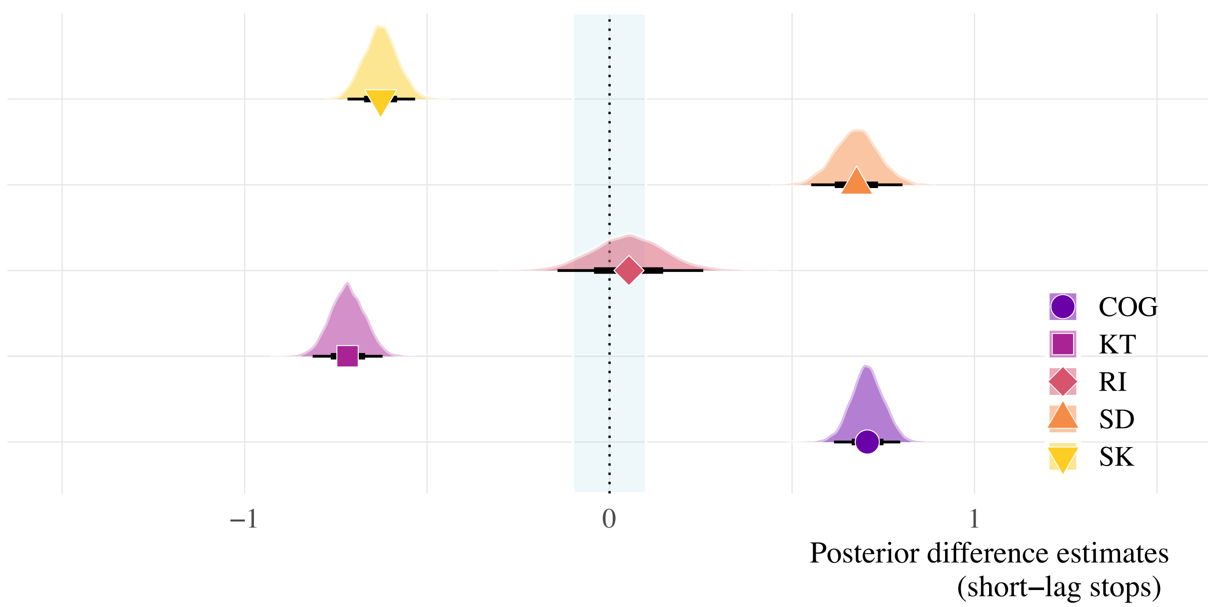

Figure 15: Posterior distribution comparing relative intensity and spectral moments of the short-lag stops, English /d/ and Spanish /t/. The points represents the posterior mean ±95% HDI and the grey region represents the ROPE (±0.1).

Table 6: Post-hoc comparisons of short-lag stops between English and Spanish for each burst metric. Each estimate is derived from the mean difference of the posterior of the Spanish segment subtracted from the posterior of the Englsh segment.

Metric	Estimate	HDI	ROPE	MPE
VOT	−0.219	[−0.45, −0.001]	0.133	0.971
COG	0.707	[0.62, 0.799]	0.000	1.000
RI	0.053	[−0.141, 0.257]	0.652	0.702
SD	0.677	[0.552, 0.802]	0.000	1.000
SK	−0.627	[−0.72, −0.536]	0.000	1.000
KT	−0.718	[−0.812, −0.621]	0.000	1.000

Figure 16 and Table 7 presents the post-hoc analyses of within-language comparisons, that is, Spanish /d/ vs. Spanish /t/ and English /d/ vs. English /t/. This table presents the same data as the forest plot in Figure 5.

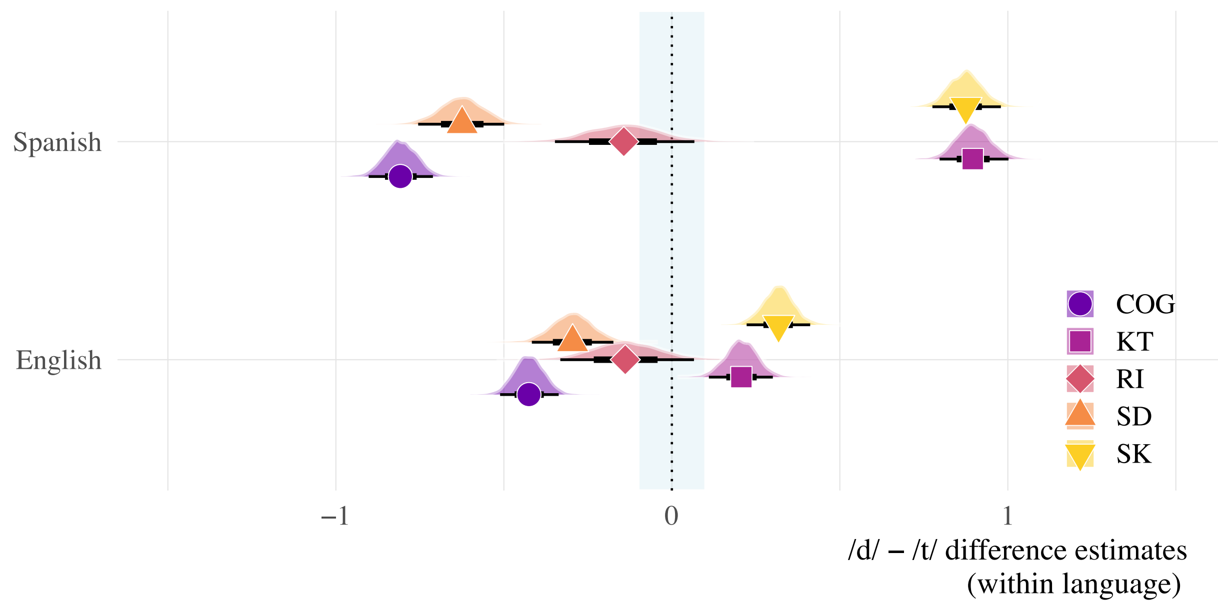

Figure 16: Posterior distributions of difference estimates between voiced and voiceless stops for Spanish and English. The point represents the posterior mean difference ±95% HDI of the posterior of the voiceless segment subtracted from the posterior of the voiced. The light blue region represents the ROPE (±0.1).

Table 7: Post-hoc comparisons of within-language voiced-voiceless pairs for each burst metric. Each point represents the posterior mean difference ±95% HDI of the posterior of the voiceless segment subtracted from the posterior of the voiced segment.

Language	Metric	Estimate	HDI	ROPE	MPE
Spanish	COG	−0.807	[−0.9, −0.71]	0.000	1.000
	RI	−0.143	[−0.346, 0.069]	0.331	0.917
	SD	−0.625	[−0.754, −0.499]	0.000	1.000
	SK	0.875	[0.774, 0.978]	0.000	1.000
	KT	0.897	[0.798, 1.003]	0.000	1.000
English	COG	−0.425	[−0.509, −0.335]	0.000	1.000
	RI	−0.137	[−0.327, 0.069]	0.344	0.914
	SD	−0.296	[−0.417, −0.174]	0.000	1.000
	SK	0.317	[0.221, 0.41]	0.000	1.000
	KT	0.207	[0.11, 0.3]	0.000	1.000

Appendix D: Supplementary analyses (place of articulation)

In this section we present additional information from the analyses of the bilingual place of articulation data. Table 8 includes the full output of the model summary from the primary multivariate analysis (Note: this table is equivalent to the forest plot in Figure 6).

Table 8: Model summary for VOT and burst metrics as a function of language (English, Spanish), place of articulation (bilabial, coronal), F1, F2, and item repetition for bilingual voiceless stops. The percentage of the HDI contained within the ROPE is based on an effect size of ±0.1.

Metric	Parameter	Estimate	HDI	ROPE	MPE
VOT	Intercept	−0.086	[−0.200, 0.027]	0.601	0.928
	Language	0.802	[0.750, 0.852]	0.000	1.000
	Place	0.134	[0.029, 0.236]	0.248	0.992
	F1	0.013	[−0.025, 0.050]	1.000	0.751
	F2	0.000	[−0.046, 0.043]	1.000	0.505
	Item rep.	−0.003	[−0.042, 0.034]	1.000	0.567
	Language:Place	0.055	[0.004, 0.105]	0.983	0.984
COG	Intercept	−0.569	[−0.665, −0.475]	0.000	1.000
	Language	0.206	[0.177, 0.236]	0.000	1.000
	Place	0.953	[0.862, 1.041]	0.000	1.000
	F1	0.019	[−0.008, 0.049]	1.000	0.905
	F2	−0.010	[−0.041, 0.020]	1.000	0.743
	Item rep.	0.007	[−0.026, 0.039]	1.000	0.659
	Language:Place	0.211	[0.181, 0.239]	0.000	1.000
KT	Intercept	0.571	[0.428, 0.690]	0.000	1.000
	Language	−0.204	[−0.237, −0.173]	0.000	1.000
	Place	−0.888	[−1.006, −0.774]	0.000	1.000
	F1	0.015	[−0.016, 0.047]	1.000	0.831
	F2	−0.015	[−0.047, 0.017]	1.000	0.822
	Item rep.	−0.028	[−0.065, 0.011]	1.000	0.930
	Language:Place	−0.169	[−0.200, −0.137]	0.000	1.000
RI	Intercept	0.232	[0.013, 0.434]	0.091	0.983
	Language	0.099	[0.042, 0.155]	0.518	1.000
	Place	−0.475	[−0.680, −0.269]	0.000	1.000
	F1	−0.222	[−0.304, −0.136]	0.000	1.000
	F2	−0.059	[−0.116, −0.001]	0.945	0.978

Metric	Parameter	Estimate	HDI	ROPE	MPE
	Item rep.	0.028	[−0.049, 0.103]	0.995	0.780
	Language:Place	−0.003	[−0.056, 0.057]	1.000	0.542
SD	Intercept	−0.476	[−0.594, −0.356]	0.000	1.000
	Language	0.230	[0.186, 0.271]	0.000	1.000
	Place	0.766	[0.671, 0.870]	0.000	1.000
	F1	−0.012	[−0.050, 0.030]	1.000	0.725
	F2	0.008	[−0.035, 0.049]	1.000	0.662
	Item rep.	0.016	[−0.033, 0.061]	1.000	0.748
	Language:Place	0.209	[0.169, 0.250]	0.000	1.000
SK	Intercept	0.527	[0.410, 0.656]	0.000	1.000
	Language	−0.222	[−0.255, −0.188]	0.000	1.000
	Place	−0.888	[−0.994, −0.781]	0.000	1.000
	F1	−0.024	[−0.058, 0.009]	1.000	0.924
	F2	0.009	[−0.025, 0.042]	1.000	0.696
	Item rep.	0.000	[−0.037, 0.039]	1.000	0.518
	Language:Place	−0.192	[−0.226, −0.159]	0.000	1.000

Table 9 presents the post-hoc analyses of between-language place comparisons, that is, bilabials vs. coronals. This table presents the same information available in the forest plot in Figure 8.

Table 9: Post-hoc between-language comparisons of bilabial and coronal stops. Each point represents the posterior mean difference ±95% HDI of the posterior of the Spanish values subtracted from the posterior of the English values.

Place	Metric	Estimate	HDI	ROPE	MPE
Coronal	VOT	1.713	[1.582, 1.856]	0.000	1.000
	COG	0.835	[0.781, 0.888]	0.000	1.000
	RI	0.19	[0.068, 0.314]	0.047	0.997
	SD	0.876	[0.802, 0.951]	0.000	1.000
	SK	−0.829	[−0.889, −0.765]	0.000	1.000
	KT	−0.747	[−0.809, −0.692]	0.000	1.000
Bilabial	VOT	1.493	[1.341, 1.64]	0.000	1.000
	COG	−0.01	[−0.112, 0.093]	0.975	0.574
	RI	0.203	[0.012, 0.4]	0.139	0.982
	SD	0.042	[−0.107, 0.187]	0.784	0.722

Place	Metric	Estimate	HDI	ROPE	MPE
	SK	−0.061	[−0.176, 0.06]	0.747	0.845
	KT	−0.071	[−0.186, 0.039]	0.701	0.891

Appendix E: Bayesian data analysis

This study employs Bayesian Data Analysis for quantitative inferential statistics. Specifically, this implies that we use Bayesian *credible intervals*—and other metrics—to draw statistical inferences. A Bayesian model calculates a posterior distribution, i.e., a distribution of plausible parameter values, given the data, a data-generating model, and any prior assumptions we have about those parameter values. Posterior distributions are computationally costly. For this reason, we use the Hamiltonian Markov Chain Monte Carlo algorithm to obtain a sample that includes thousands of values from the posterior distribution. In practical terms, what this means is that we do not calculate a single point estimate for an effect β, but rather we draw a sample of 4,000 plausible values for β. This allows us to quantify our uncertainty regarding β by summarizing the distribution of those values. We will use 4 statistics to describe the posterior distribution: (1) the posterior mean, (2) the highest density credible interval (HDI), (3) the proportion of the HDI that falls within a Region of Practical Equivalence (ROPE), and (4) the Maximum Probability of Effect (MPE). The posterior mean provides a point estimate for the distribution. The 95% highest density credible interval provides bounds for the effect. The ROPE designates a region of practical equivalence for a negligible effect and calculates the proportion of the HDI that falls within this interval.[2] The MPE calculates the proportion of the posterior distribution that is of the median's sign (or the probability that the effect is positive or negative).

$$es = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}} \qquad (1)$$

If, for instance, a hypothesis states that β > 0, we judge there to be *compelling evidence* for this hypothesis if the mean point estimate is a positive number, if the 95% credible interval of β does

---

[2] We utilize a ROPE of ±1 for standardized values. For non-standardized values Kruschke (2018) recommends using the formula in Equation 1.

not contain 0 and is outside the ROPE by a reasonably clear margin, and the posterior $P(\beta > 0)$ is close to one. Together these four statistics allow us to quantify our uncertainty and provide an intuitive interpretation of any given effect. Consider a case in which the posterior mean of β is 100 and the 95% credible interval is [40, 160]. The interval tells us that we can be 95% certain the *true* value of β is between 40 and 160, given the data, our model, and our prior information. Furthermore, the interval allows us to specify areas of uncertainty. In this example, we can conclude that the effect is almost certain to be positive. The lower interval value of 40 tells us that 95% of the plausible values are greater than 40. We also note that the interval covers a wide range of values, thus we also conclude that we are not very certain about the size of the effect. This type of interpretation is not possible under a frequentist paradigm.

Appendix F: Reproducibility information

About this document

This document was written in Markdown using `quarto`.

Session info

```
>   setting  value
>   version  R version 4.2.1 (2022-06-23)
>   os       macOS Big Sur ... 10.16
>   system   x86_64, darwin17.0
>   ui       X11
>   language (EN)
>   collate  en_US.UTF-8
>   ctype    en_US.UTF-8
>   tz       America/New_York
>   date     2022-09-07
>   pandoc   2.18 @
/Applications/RStudio.app/Contents/MacOS/quarto/bin/tools/ (via rmarkdown)

>                  loadedversion       date
>  academicWriteR          0.4.1 2022-09-05
>  bayesplot               1.9.0 2022-03-10
>  bayestestR             0.12.1 2022-05-02
>  brms                   2.17.0 2022-04-13
>  broom                   1.0.0 2022-07-01
>  dplyr                   1.0.9 2022-04-28
>  flextable               0.7.3 2022-08-09
>  forcats                 0.5.2 2022-08-19
>  future                 1.27.0 2022-07-22
>  ggbeeswarm              0.6.0 2017-08-07
>  ggExtra                0.10.0 2022-03-23
>  ggplot2                 3.3.6 2022-05-03
>  ggstance                0.3.5 2020-12-17
>  glue                    1.6.2 2022-02-24
>  here                    1.0.1 2020-12-13
>  knitr                    1.40 2022-08-24
>  modelr                  0.1.9 2022-08-19
>  purrr                   0.3.4 2020-04-17
>  Rcpp                    1.0.9 2022-07-08
>  readr                   2.1.2 2022-01-30
>  stringr                 1.4.1 2022-08-20
>  tidybayes               3.0.2 2022-01-05
>  tidyr                   1.2.0 2022-02-01
```