The categorization of L3 sounds in Spanish/English bilinguals during the initial stage of L3 acquisition

Kyle Parrish1

1 Rutgers University

Author note

Correspondence concerning this article should be addressed to Kyle Parrish, 15 Seminary Place, New Brunswick, NJ 08904. E-mail: [kcp74@scarletmail.rutgers.edu](mailto:kcp74@scarletmail.rutgers.edu)

The categorization of L3 sounds in Spanish/English bilinguals during the initial stage of L3 acquisition

# 1 Methods

## 1.1 Participants

A total of 100, NA, NA, and NA participants took part in at least one experiment of the study and were recruited on the online platform Prolific. A total of 16 participants did not complete all blocks of the experiment, and their data was removed from the final analysis, resulting in a total of 84 participants who completed all portions of the experiment. In order to tease apart L2 status effects from CLI brought about by typological similarity across languages, participants were recruited with both orders of acquisition. That is, one group of participants spoke English as a first language and Spanish as a second language (the ES group, n = 36, and another group spoke Spanish as their first language and English as their second (the SE group, n = 55). Participants were screened prior to taking part in the study such that their age of onset was greater than 12 years old. Then, in order to examine how the same groups of bilinguals categorize different L3s, the SE and ES group were then randomly assigned either Hungarian (in the ES group, n = 17 and in the SE group, n = 31) or French (in the ES group, n = 17 and in the SE group, n = 18). In order to assess whether participants were eligible to take part in the study, a screening process was carried out in prolific which involved the Language History Questionnaire (Li, Zhang, Yu, & Zhao, 2020), and a proficiency test in the participant’s L2. Speakers who reported knowing a third language or who scored below 60% on the lexTALE proficiency test were deemed ineligible for the study, and were not permitted to continue.

Additionally, participants who failed attention checks, chose the same answer for every trial, or who answered over 80% /p/ or /b/ in the stops trials were removed from the data analysis. The data of 5 participants were removed for English stops, 7 for Spanish stops, 4 for French stops, and 1 for Hungarian. The same procedure was carried out for vowels and a total of 5 participants’ data was removed under the same criteria.

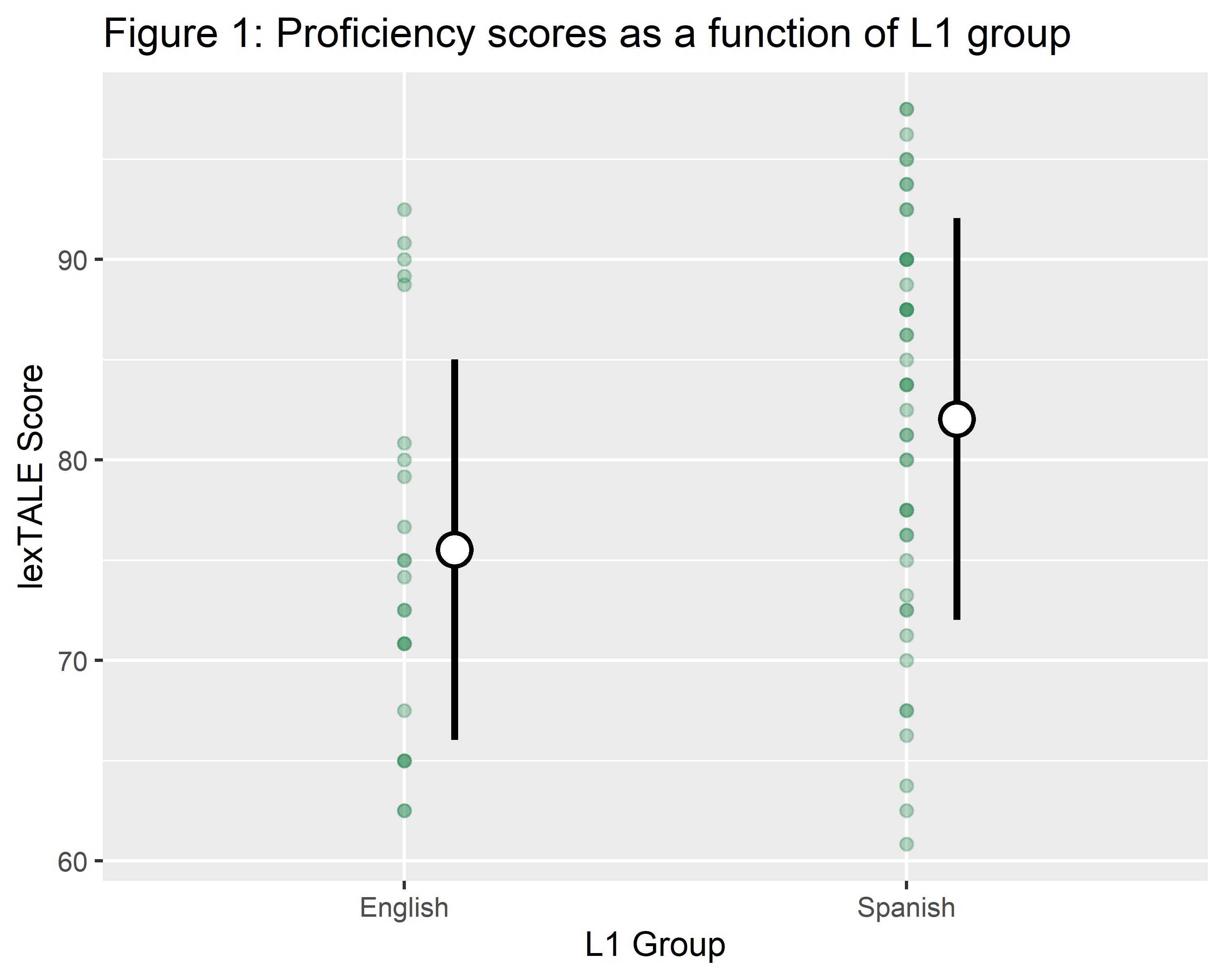
## 1.2 Materials

The experiment was composed of 5 total blocks. An itemized version of the Language History Questionnaire (Li et al., 2020) was given to participants and asked their first language, second language, and the age at which they began learning their L2. The measure of L2 proficiency, the lexTALE (Izura, Cuetos, & Brysbaert, 2016; Lemhöfer & Broersma, 2016), was given in Spanish to the ES group and English to the SE group. The lexTALE is a lexical decision task in which the participants are presented with either a word or a non-word, and they must decide whether the word presented exists in that language. Both the Spanish and English versions are scored in the same manner, and allow for comparability of proficiency across groups. Both versions contain 60 words total, 40 of which are real words and 20 of which are non-words. A formula then calculates a score based on the percentage of real words correct plus the percentage of non-words correct divided by 2. Table 1 provides descriptive statistics which shows the mean lexTALE score per L1 group. As the table shows, the Spanish L1 group was more proficient in their L2 on average than the English L1 group. Specifically, the mean lexTALE score of the Spanish L1 group was 81.58, (SD = 10.43) and the mean score of the English L1 group was 73.85, (SD = 10.78). Figure 1 shows a graphical relationship of the distribution of proficiency scores per L1 group. A test of equivalence of independent samples revealed that the Spanish L1 group and the English L1 group were not equivalent (*t*(50.52) = 0.957, p = 0.828), with equivalence bounds of -5.227 and 5.227 (on a raw scale) and an alpha of 0.05. A null hypothesis test of the same samples suggests that the null hypothesis is also rejected (*t*(50.52) = 2.958, p = 0.0047, given an alpha of 0.05). Concretely, these results suggest that the difference in proficiency between groups was 7.73 percentage points (90% CI 3.35-12.10) on the lexTALE vocabulary task, with the Spanish group scoring slightly higher.

Table 1:

*Descriptive Statistics of proficiency per L1 group*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X | l1 | Mean.lexTALE.Score | SD | n |
| 1 | English | 73.85 | 10.78 | 25 |
| 2 | Spanish | 81.58 | 10.43 | 47 |



*Figure* *1.*  Add figure caption here.

Following these tasks, participants then took a series of three two-alternative forced choice tasks (2afc). The tasks were programmed in Psychopy (Peirce et al., 2019) and were hosted on Pavlovia.org. Participants received written instructions prior to taking part in each phase of the experiment. The instructions for the “English” task were in English, the instructions for the “Spanish” task were in Spanish, and the instructions for the L3 tasks were bilingual (both Spanish and English). In each block, the instructions on profilic were in the language mode that the participants were told that they would hear, and the instructions within the experiment itself were also language specific, with the exception of all L3 prolific instructions being in English. These instructions read “Final step! You have been assigned to the Hungarian/French group. The task is the same as the first two sessions, but this time, the words are from the Hungarian/French language.” In the L3 experimental session, written instructions were more scarce, since the participants had already completed the same procedure twice, but were always presented in both English and Spanish at the same time. Instructions appeared at the beginning of the experiment, after the initial practice trials, and between the stop and vowel sessions, and for the attention checks. In each case, and English instruction was given and its Spanish translation appeared either to the right of the English instruction or immediately below it. As in previous studies, the instructions contained no instances of the experimental stimuli, in this case “pafri,” “bafri,”ifri, or “ufri,” or any of their language specific forms appeared during the instructions.

## 1.3 Stimuli

The 2afc tested the perception of the same two continua (a /p/-/b/ VOT continuum and a /i/-/u/ F2 continuum) in three different language modes: English, Spanish and the L3 (either Hungarian or French). The VOT continuum has been used in previous studies (Gonzales, Byers-Heinlein, & Lotto, 2019; Lozano-Argüelles et al., 2020) and were created from a recording by a female Spanish-English bilingual. The original VOT continuum contained the word “pafri” and “bafri.” The recording was resynthesized and ranged in VOT from -35ms to 35ms, excluding a 0 VOT step. In the present study, three additional steps were added on the positive end of the scale in order to increase the chances of /p/ responses at the positive extreme of the continuum, such that steps of 40ms, 45ms, and 50ms were included and made a total of 17 total steps. The extra steps were created manually in PRAAT by extracting 5ms of the voiceless part of the previous segments and adding these portions on to the beginning of the 35ms stimulus up to three times. Just as in Lozano-Argüelles et al. (2020), the present study conceptually cued participants, and did not utilize the language specific rhotics used in Gonzales and Lotto (2013) and Casillas and Simonet (2018).

The vowel continua were created using productions of the pseudowords “ifri” and “ufri” by a male American English speaker. These natural productions were then resynthesized using a PRAAT script which manipulated the second formant while holding the other formants constant. Just as was done with the stop continuum, only the first syllable together with [f] were produced in each stimulus to avoid perceptual cueing. As a result, the stimuli, ranging from “if” to “uf” varied only by the second formant (F2: a correlate for frontedness in the vowel space) in a total of 11 steps.

During the 2afc task, participants saw both the words *pafri* and *bafri* or language specific versions of *ifri* and *ufri* on the screen and just the first syllable of these words in which the initial consonant was a member of the continuum (“paf” or “baf” for vowels, “if” or “uf” for vowels). Each step of the continuum was played 5 times total, leading to a total of 85 trials in the VOT continuum per participant and 55 total trials in the vowel continuum. Each trial was drawn randomly from the continuum in a single block, so no participant heard the same stimulus order. Due to language specific orthographic differences, the non-words on screen were distinct orthographically in each language, with the exception of French and Spanish. In Spanish and French, the participants saw *ifri* and *ufri* on the screen, in English they saw *eefri* and *oofri*, and in Hungarian, they saw *ifrelo* and *ufrelo*. This spelling difference was necessary because of the orthographic conventions in the realizations of similar sounds between English and Spanish. For example, the Spanish grapheme “u” and the English graphemes “oo” are similar in their acoustic realizations. On the other hand, the choice of spelling in the French and Hungarian stimuli were made to reflect that high degree of lexical overlap between Spanish and French, and to reinforce the ambiguous relationship between Hungraian and English or Spanish.

## 1.4 Procedure

All participants completed a total of five tasks, the Language History Questionnaire, the lexTALE proficiency test in their L2, and three 2afc tasks. First, the participants completed the Language History Questionnaire online (Li et al., 2020), followed by the lexTALE proficiency test. Following these two tasks used for screening, only those participants who did have knowledge of a third language and scored at least 60% on the lexTALE were deemed eligible and continued the experiment. This cutoff was chosen to focus the analysis on more proficient bilinguals, rather than demonstrate that increasing proficiency is correlated to a larger double phonemic boundary effect size, as has been found in previous studies (Casillas & Simonet, 2018; Garcia-Sierra, Diehl, & Champlin, 2009; Gonzales & Lotto, 2013).

Following these screening tasks, and in order to control for language mode, the participants completed the 2afc tasks in a semi-longitudinal design. Participants were invited to take part in the second portion of the experiment at least 30 minutes after their completion of the previous step. In many cases, however, participants completed different blocks of the experiment hours or days apart. All participants began with the “English” task, followed by “Spanish,” and finished with either “French” or “Hungarian.” Unfortunately, the lack of counterbalancing across participants is a limitation in the design of the current study. In the English and Spanish tasks, the instructions were given in that respective language, and it was explained that the participants were going to hear rare words in the language in question (English, Spanish or the L3). In the L3 session, the same instructions were given in both English and Spanish simultaneously and on the same screen in order to avoid biasing the participants to either English or Spanish mode.

## 1.5 Statistical analyses

In order to determine a crossover boundary for analysis in a general linear model, two logistic regression models were fit to each participant with their responses of ‘bafri’/‘pafri,’ in the VOT continuum, and ‘ifri/ufri,’ in the F2 continuum as a function of the standardized continuum step for each language. The continuum step standardization was done in order to allow for the comparability of the stop and vowel crossover boundaries, which had a distinct number of steps. As a result, 6 crossover boundaries total were calculated per participant that provided which step in the continuum the probability of the participant choosing either “pafri” or “bafri” was 50%. These boundaries were used as continuous variables in subsequent general linear regression models and t-tests. All analyses were carried out in R (R Core Team, 2020) and used either the stats package or the lmer package (Bates, Mächler, Bolker, & Walker, 2015). In order to determine the independence of the crossover samples, paired t-tests were carried out in each language combination in both stops and vowels. A significant p-value on a paired t-test would indicate that there is non-zero difference between individual participants’ crossover boundaries between the two languages.

# 2 Results

## 2.1 Descriptive statstics

Tables 2 and 3 provide descriptive statistics of the crossover boundaries per group per language. The crossover boundary is the continuum step at which the probability of choosing both options (/p/ or /b/, or /i/ or /u/) was 50% and was calculated by fitting a logistic regression model to each participant in each language mode session. The crossover boundaries were standardized to z-scores in order to allow for comparability across all language modes and over features (between stops and vowels).

Table 2:

*Descriptive statistics of vowel crossovers by group*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| group | Mean English Crossover | SD English | Mean Spanish Crossover | SD Spanish | Mean L3 Crossover | SD L3 | n |
| English\_french | -0.14 | 0.03 | -0.13 | 0.04 | -0.14 | 0.06 | 17 |
| English\_hungarian | -0.14 | 0.07 | -0.13 | 0.05 | -0.13 | 0.05 | 17 |
| Spanish\_french | -0.07 | 0.05 | -0.11 | 0.06 | -0.10 | 0.06 | 17 |
| Spanish\_hungarian | -0.08 | 0.06 | -0.10 | 0.05 | -0.10 | 0.05 | 28 |

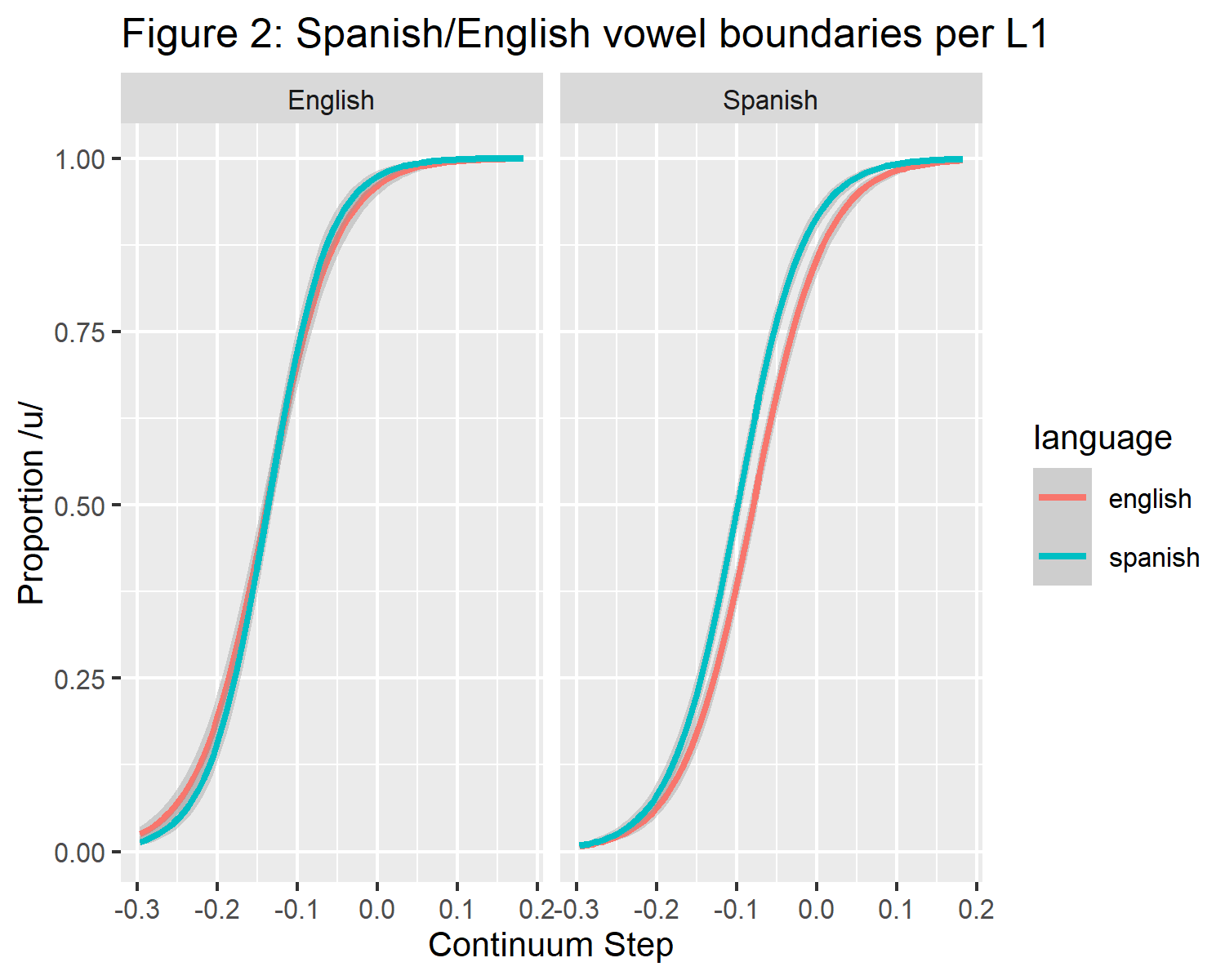
Table 3:

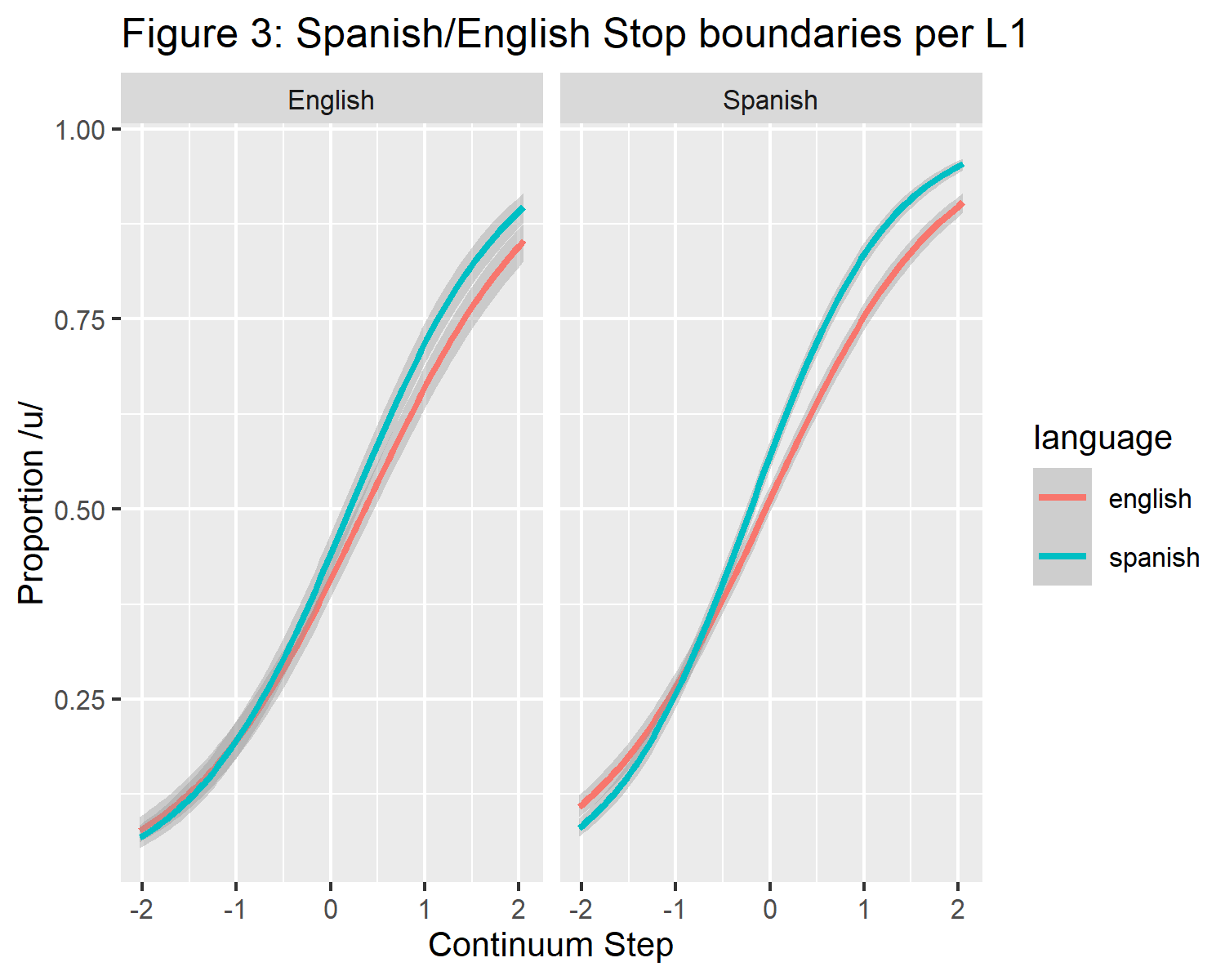
*Descriptive statistics of stops crossovers by group*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| group | Mean English Crossover | SD English | Mean Spanish Crossover | SD Spanish | Mean L3 Crossover | SD L3 | n |
| English\_french | 0.37 | 0.60 | 0.07 | 0.73 | 0.14 | 0.60 | 9 |
| English\_hungarian | 0.51 | 0.47 | 0.39 | 0.37 | 0.20 | 0.32 | 14 |
| Spanish\_french | -0.14 | 0.74 | -0.30 | 0.50 | -0.27 | 0.46 | 17 |
| Spanish\_hungarian | 0.06 | 0.57 | -0.19 | 0.47 | -0.19 | 0.53 | 29 |

## 2.2 Question 1: Was the double phonemic boundary effect replicted in stops and extended to vowels?

In order to answer the research questions of the present study, it was vital to replicate previous findings of double phonemic boundary effect in Spanish-English bilinguals. Thus, an implicit research question in this study was whether Spanish-English bilinguals would show evidence of a double phonemic boundary in stops, as in previous studies, and, newly, whether this effect could be extended to vowels. Of course, without finding two language specific language mode-driven categorizations of English and Spanish, conclusions cannot be drawn as to whether bilinguals approach a third language using their first or second language perceptual routines. The double phonemic boundaries were assessed per language using a series of paired t-tests. The results showed that the English L1 group showed a double phonemic boundary effect in stops (, ), but not in vowels (, ). The Spanish L1 group showed evidence of the double phonemic boundary in both stops (, ) and vowels (, ). Figure 2 shows a graphical representations of the categorizations of vowels (figure 2) and stops (figure 3) in both languages in both L1 groups.





## 2.3 RQ2: which language-specific phonemic boundaries will participants perceive when they are introduced to the same VOT/F2 continuum in a “third language?”

Due to the results of the initial series of paired t-tests, only the results of the stops in both L1 groups and the vowels in the Spanish groups were useful in answering this question, since English vowels violated the necessary assumption that all groups would show double phonemic boundary effects. A series of three paired t-tests are reported per each of the four participant groups. In each group, the paired t-tests serve to provide evidence that language categorizations are independent samples (if the p-value falls below .05) or not independent samples (if the p-value is greater that .05). The three t-tests in each group test the independence of their L1 and L2 categorization (in all 4 groups), the categorization of their L1 and their L3 (English or Spanish with Hungarian or French), and their L2 and their L3 (also English or Spanish with Hungarian or French). By having all three of the t-tests together, it can more easily be argued that one specific language has greater influence on L3 categorization if, first the Spanish and English paired t-test is significant, and one of the two remaining t-tests is significant and the other is not (i.e. if the L2 to L3 t-test is significant but the L1 to L3 t-test is not, that is taken as evidence that the L2 and the L3 are independent samples, but that the L1 and the L3 are not).

### 2.3.1 Stops English L1 French L3 group.

The English L1 group appeared to use their L1 boundary in all three continua in both L1 groups. English L1 vowels did not show a group wide L1-L2 double phonemic boundary effect, and subsequent analyses in the English L1 groups focuses on stops.Despite the group-wide differences in the English L1 group, double phonemic boundary effects were not found in the individual English L1 groups. In stops, paired t-tests found non-significant results, interpreted as similar categorizations in English-Spanish categorization (, ), Spanish-French categorization (L2 to L3) (, ) and English-French categorization (L1 to L3) (, ).

### 2.3.2 Stops English L1 Hungarian L3 group.

The English-Hungarian stops were found to be distinct. English-Hungarian categorization (L1 to L3) (, ), but not the English-Spanish categorization (, ) nor Spanish-Hungarian categorization (L2 to L3) (, ).

### 2.3.3 Stops Spanish L1 French L3 group.

No categorization differences were found in Spanish L1 speakers assigned to the “French” group in stops. There was no significant English-Spanish categorization difference (, ), no significant English-French categorizations difference (, ) and no significant Spanish-French categorization difference (, ).

### 2.3.4 Vowels Spanish L1 French L3 group.

Evidence of a double phonemic Spanish-English boundary was not found in this group (, ). Likewise, Spanish-French categorizations had no evidence of being distinct (, ). There was, however, evidence of a distinction in English-French categorizations (, ). The lack of evidence for the double phonemic boundary effect in this group alone does not allow for definitive evidence that can inform which language boundary this group used in the perception of what they believed were French vowels.

### 2.3.5 Stops Spanish L1 Hungarian L3 group.

In this group, there was evidence that there was a English-Spanish categorization difference (, ), implying that this group showed evidence of the double phonemic boundary. The L2 and L3 English-Hungarian categorizations were also found to be categorized differently (, ), whereas the L1 and the L3 Spanish-Hungarian categorizations (, ) were not categorized differently. These results suggest that the Spanish L1 group used their L1 boundaries to categorize what they believed were Hungarian stops.

### 2.3.6 Vowels Spanish L1 Hungarian L3 group.

The same trend is present in vowels as in stops for the Spanish L1 Hungarian L3 group. A paired t-test examining the cross over boundaries of English and Spanish (, ) reveals a double phonemic boundary effect. There was evidence that there is a difference in L2-L3 English-Hungarian categorizations (, ), but another paired t-test suggests that L1-L3 Spanish-Hungarian (, ) are not categorized distinctly. Like stops, taken together, these results suggest that the Spanish L1 group used their L1 boundaries to categorize what they believed were Hungarian vowels.

## 2.4 RQ3: Did the L3 groups categorize the continua differently?

T-tests of independent samples revealed that L3 groups did not categorize their respective continua differently if they were assigned to the French group or the Hungarian group. The L3 French group with Spanish L1 and the L3 Hungarian group with a Spanish L1 did not categorize their respective L3s differently in vowels (, ) or stops (, ). The same evidence was found the the English L1 groups; they categorized L3 sounds similarly, whether they were in the L3 French group or Hungarian group in both vowels (, ) and stops (, ).

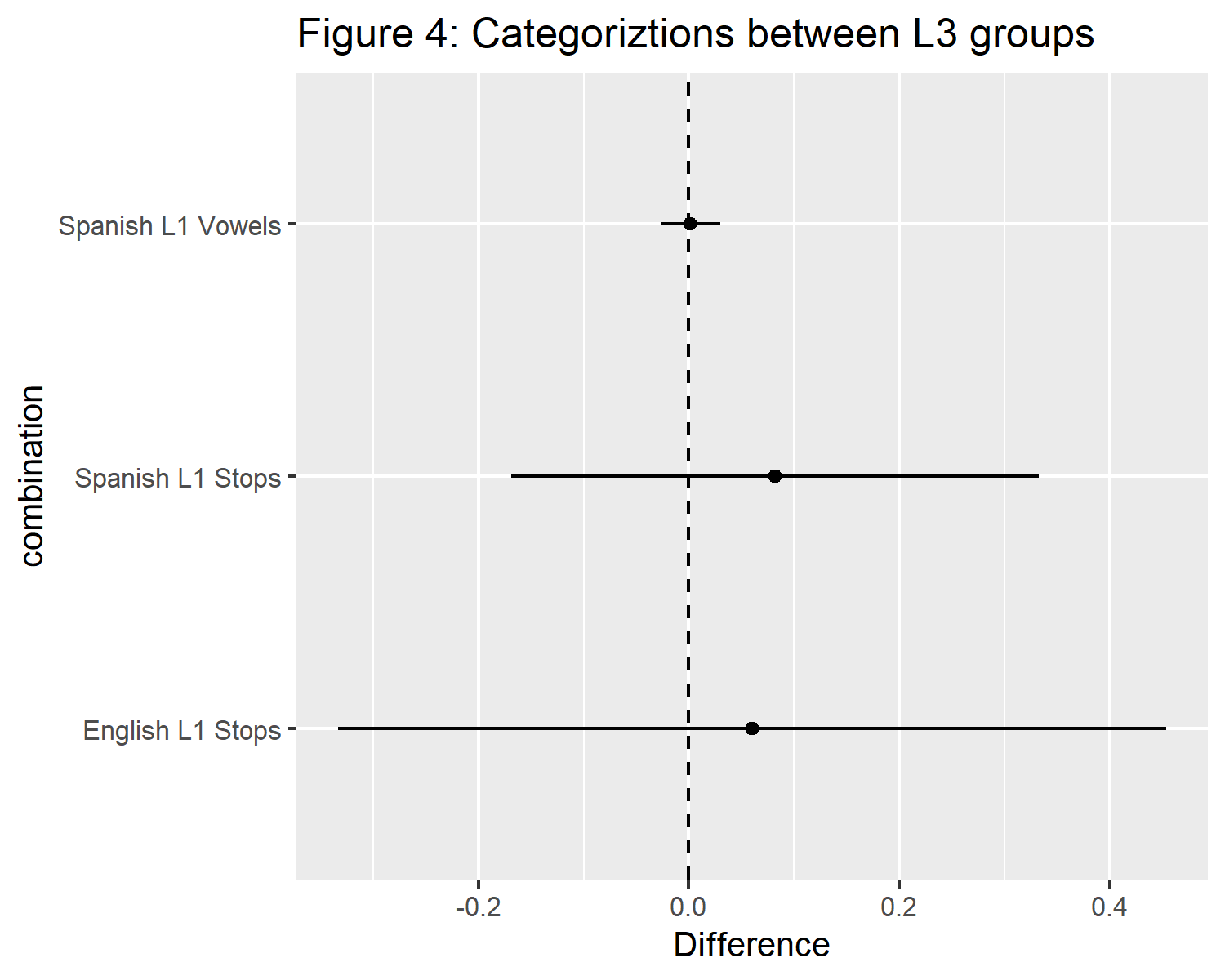
## 2.5 Post-hoc analysis

Based on the findings that the suggested that L3 continua were not categorized differently by the same L1 group, a post-hoc series of paired tests of equivalence were carried out in which the L3 groups were pooled together. First, tests of equivalence were carried out to determine whether L3 groups categorized the continua equivalently whether they believed the were hearing Hungarian or French. Importantly, the same participants did not hear both French and Hungarian, but were randomly assigned to one of them. As a result, a non-paired test of equivalence was carried out. Cohen’s D was set to .5 in these tests, which has been reported as a medium effect size in the literature (Cohen, 2013). All tests were carried out in R (R Core Team, 2020), using the TOSTER function (Lakens, 2017).

Figure 4 shows a graphical summary of three non-paired tests of equivalence. The first was a comparison between categorization of vowels between the French L3 and Hungarian L3 groups with Spanish as their L1. Surprisingly, the equivalence test was did not quite show that the categorizations of French and Hungarian vowels by the Spanish L1 groups were equivalent (t(31.25) = -1.510, p = 0.0705, given equivalence bounds of -0.0266 and 0.0266 (on a raw scale) and an alpha of 0.05). It is likely that, it this case, the lack of equivalence found is due to a low sample size.

The second test of equivalence measured the categorizations of stops between the French and Hungarian L3 groups who speak L1 Spanish. These results are similar to those reported in vowels in that they do not provide enough evidence that the L3 groups categorized what they believed to be L3 sounds the same (t(37.81) = -1.118, p = 0.135, given equivalence bounds of -0.248 and 0.248 (on a raw scale) and an alpha of 0.05). Again, this is suggested to be due to low sample size.

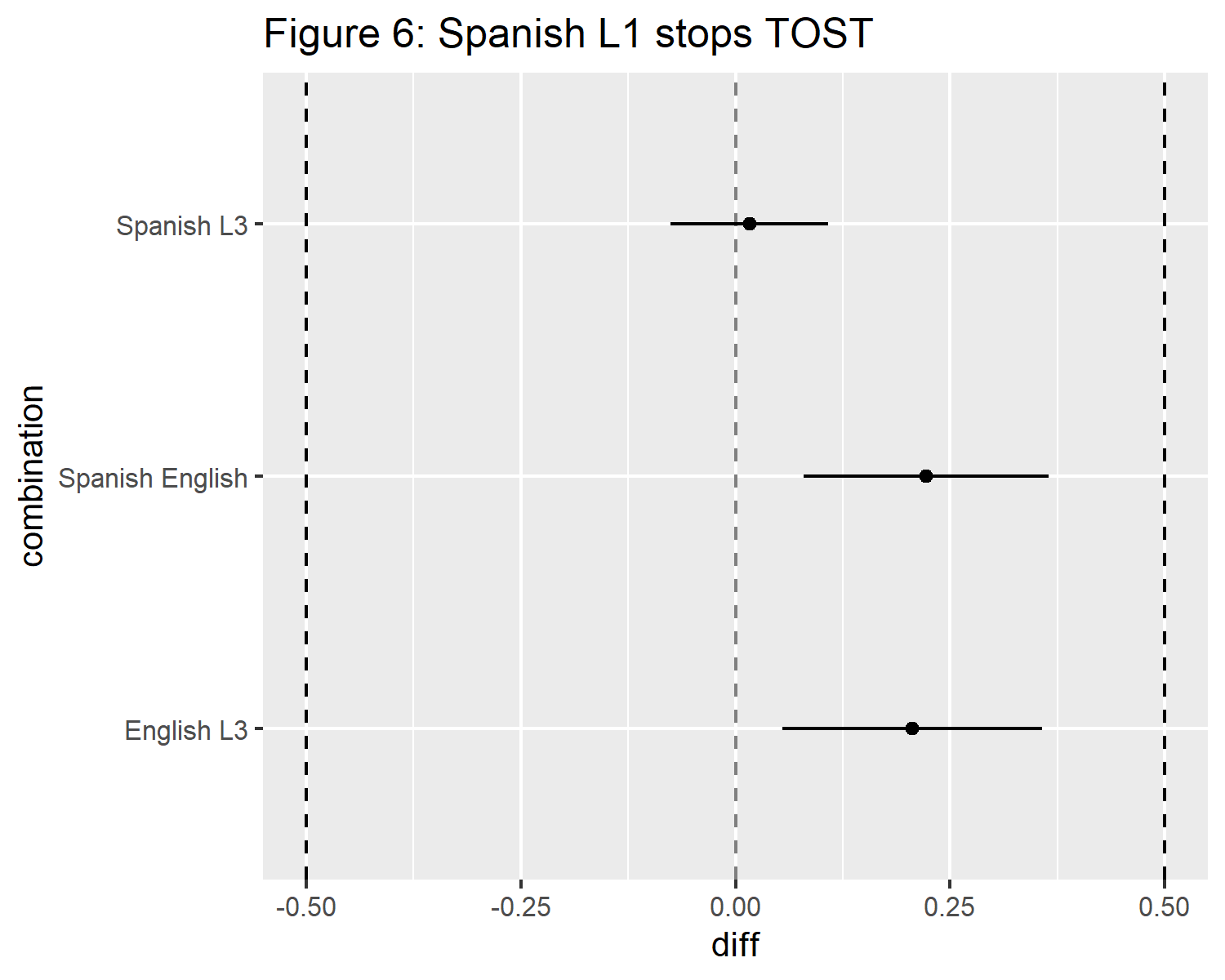
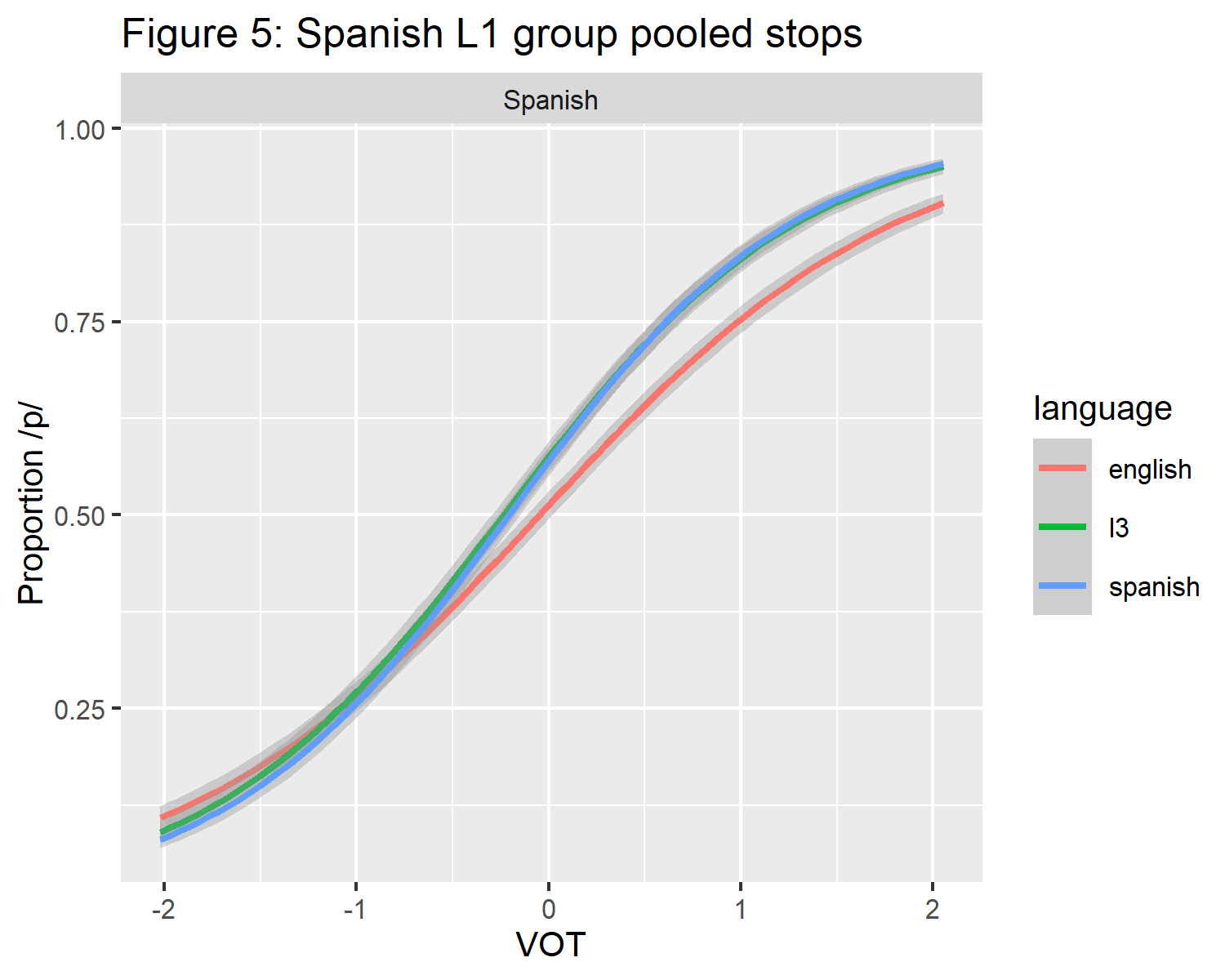
The third test of equivalence measured the categorizations of stops between the French and Hungarian L3 groups who speak L1 English. A similar trend was observed in this case, in which equivalence was not detected, likely due to low sample size (t(10.89) = -0.829, p = 0.212, given equivalence bounds of -0.242 and 0.242 (on a raw scale) and an alpha of 0.05.).



Inspired by the idea that low sample size may be the cause of the lack of equivalence in the findings, an additional series of paired equivalence tests were carried out in Spanish L1 vowels, Spanish L1 stops, and English L1 stops. English L1 vowels were omitted due to the lack of any recognizable trend towards a meaningful difference in the initial analysis.

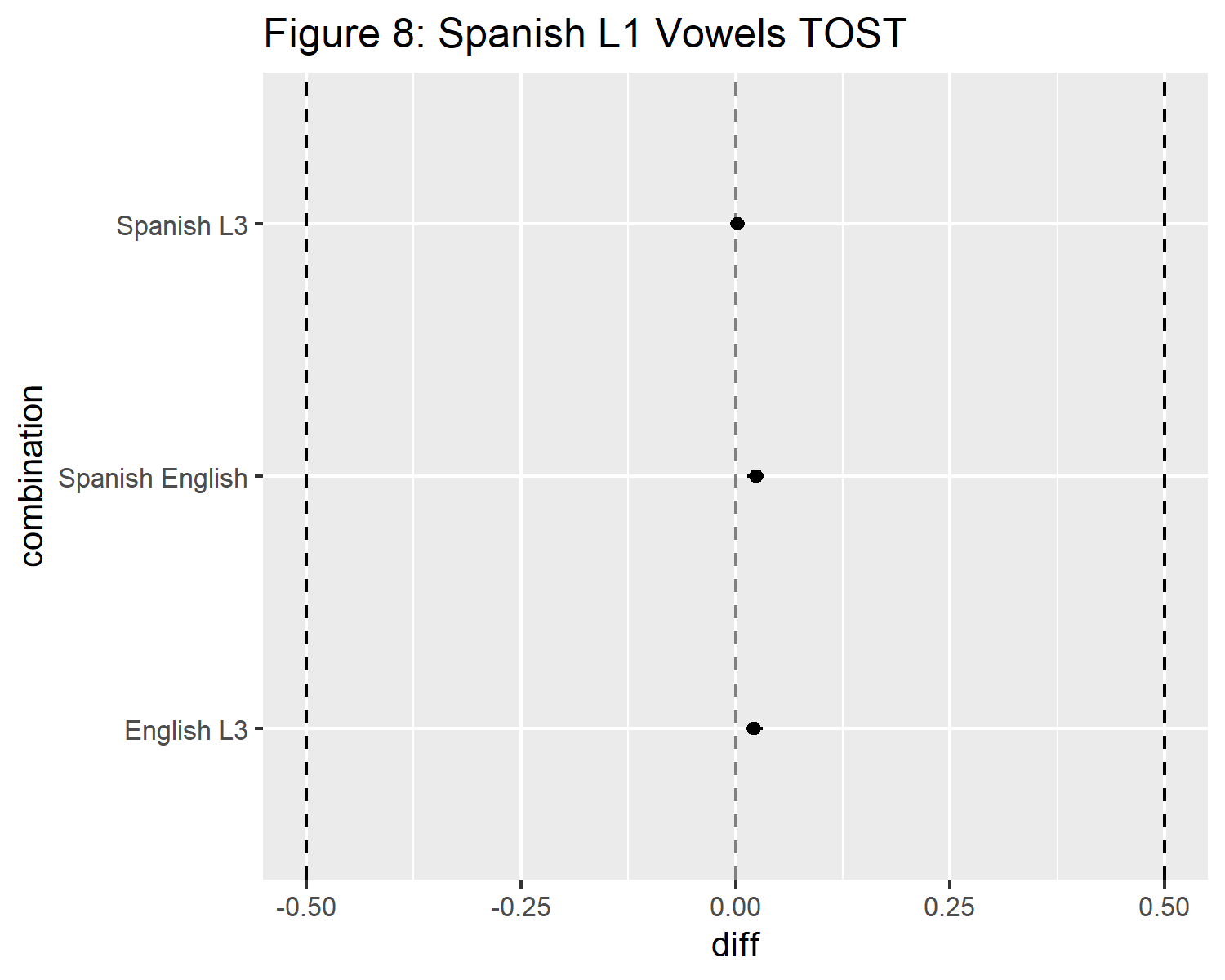
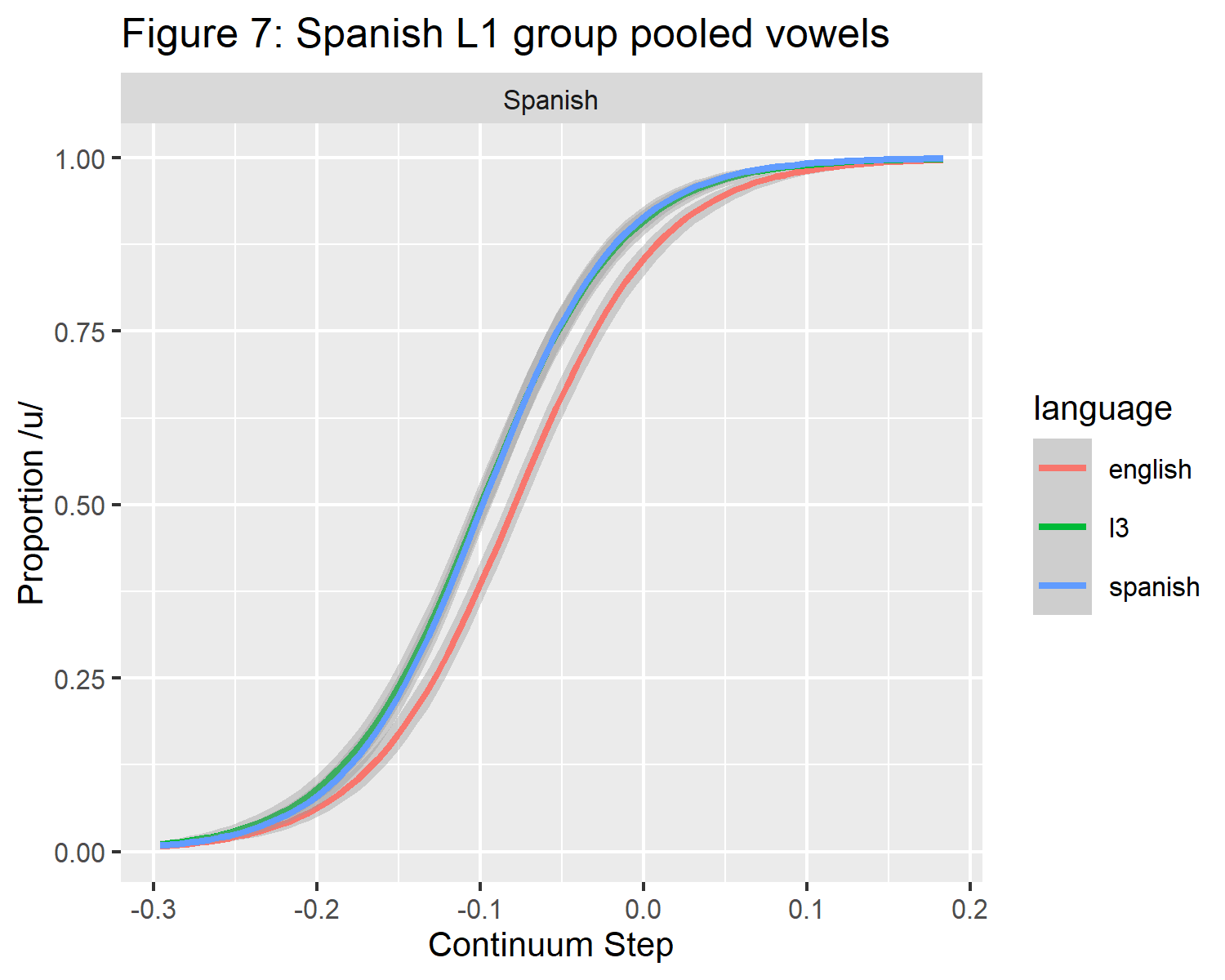
### 2.5.1 Spanish L1 group pooled stops.

Firstly, this group showed evidence of the double phonemic boundary effect when they were pooled (, ). Next, of interest in this group was whether the Spanish L1 group categorized the L3 stop continua using their L1 boundaries. The test of equivalence suggest that this is the case (t(45) = -3.094, p = 0.00169). On the other hand, English and the L3 were not found to be equivalent (t(45) = -1.103, p = 0.138), and nor were Spanish and English (t(45) = -0.770, p = 0.223). Figures 5 is a graphical representation of the crossover boundaries of stops in the pooled data of both Spanish L1 groups. Figure 6 shows the effect sizes derived from each of the three tests of equivalence. As can be seen in figure 6, the smallest effect size is the difference between the L1 and the L3, which is rather close to zero.



### 2.5.2 Spanish L1 group pooled vowels.

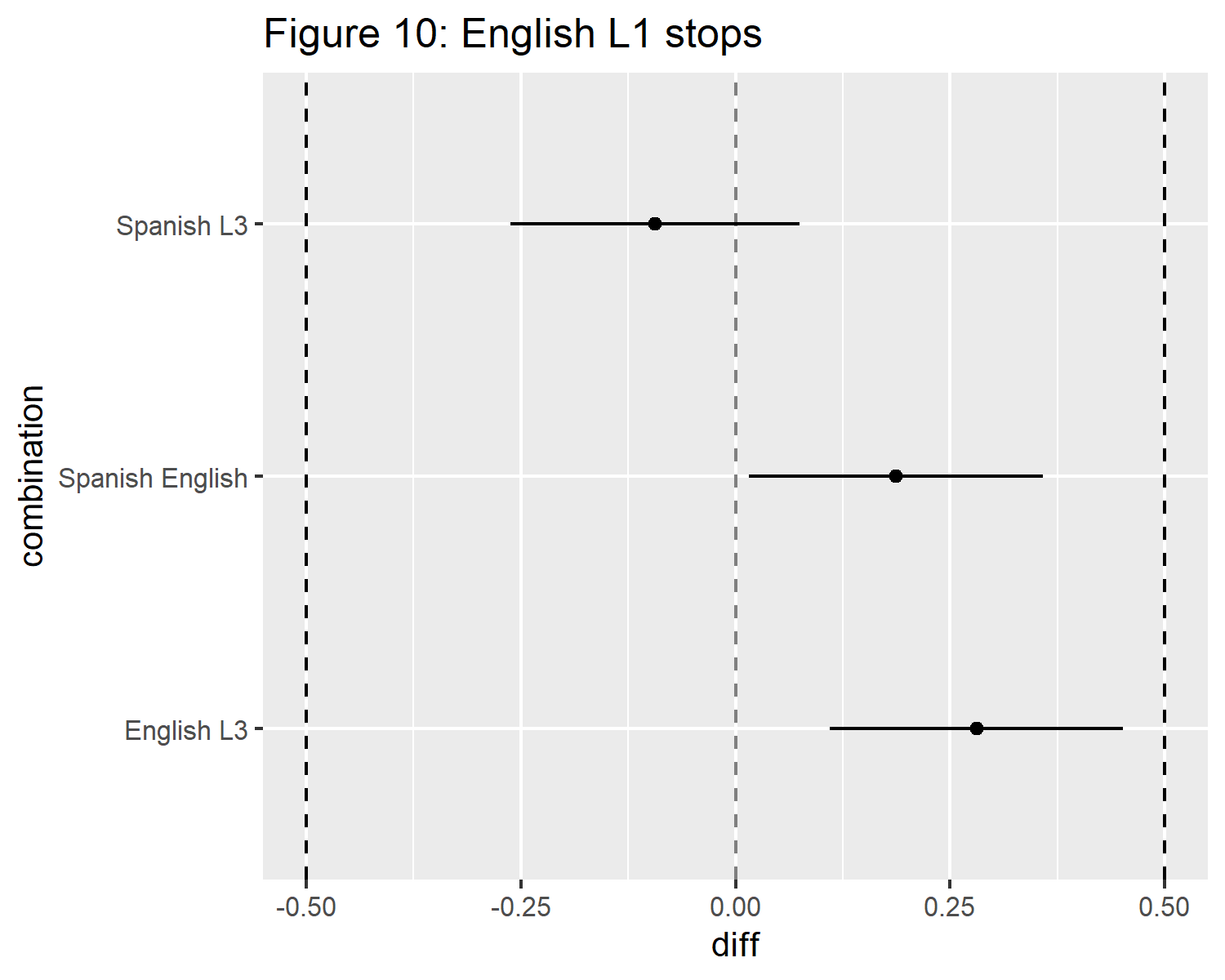
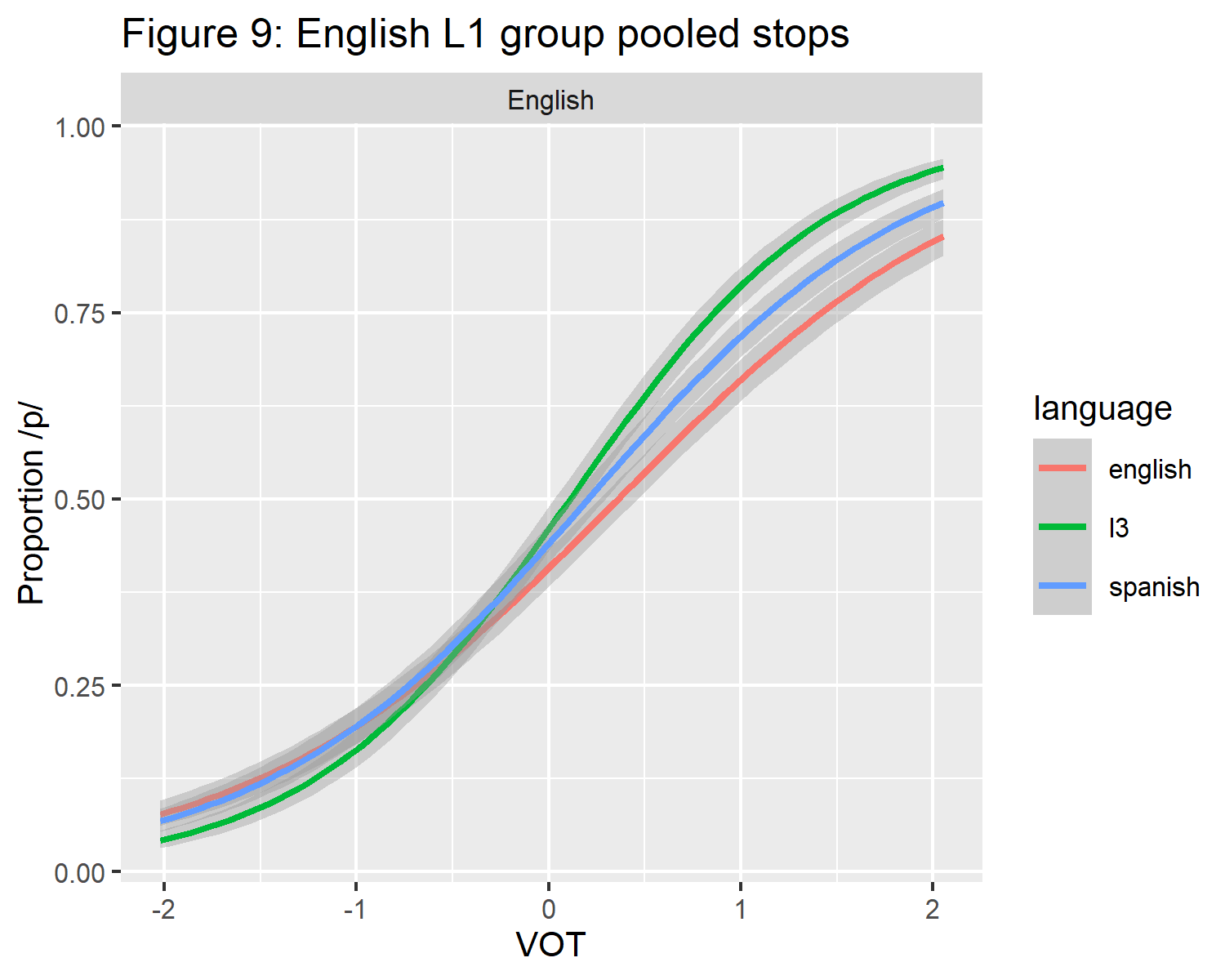
A similar results was found in the Spanish L1 group in vowels. First, there was evidence of a double phonemic boundary (, ) between English and Spanish. Next, there is evidence that the L1 and the L3 were categorized equivalently (t(45) = -2.949, p = 0.00252), where the L2 and L3 (t(45) = 0.198, p = 0.578) and the L1 and the L2 (t(45) = 0.536, p = 0.703) were not. Just as in stops, Figure 7 is a graphical representation of the crossover boundaries of vowels in the pooled data of both Spanish L1 groups. Figure 8 shows the effect sizes derived from each of the three tests of equivalence. Again, similarly to stops, in figure 8, the smallest effect size is the difference between the L1 and the L3, which is rather close to zero.



### 2.5.3 English L1 group pooled stops.

Overall, clear conclusions cannot be drawn regarding the English L1 group’s categorization of stops. In the fist analysis, only the stops showed evidence of a double phonemic boundary modulated by language mode, however, this Spanish-English double phonemic boundary effect does not quite persist in those who completed all phases of the experiment (, ).

The data also suggested that this group showed evidence of their L2 influencing L3 comprehension, rather than their L1. The equivalence test between L2 and L3 was the closest to equivalent, t(22) = 1.441, p = 0.0818), but did not provide sufficient evidence to classify the data as evidence for an L2 status effect. Importantly, this groups L1 and L3 categorizations were not equivalent (t(22) = 0.427, p = 0.663), and neither were their L1 and L2 categorizations (t(22) = -0.527, p = 0.302) Just as in the previously reported tests of equivalence, figures 9 shows the graphical representation of the crossover boundaries of stops in the pooled data of both English L1 groups. Figure 10 shows the relative effect sizes of each language combination. Taken together, it is difficult to draw clear conclusions in regard to the English L1 group.



# 3 References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>

Casillas, J. V., & Simonet, M. (2018). Perceptual categorization and bilingual language modes: Assessing the double phonemic boundary in early and late bilinguals. *Journal of Phonetics*, *71*, 51–64. <https://doi.org/10.1016/j.wocn.2018.07.002>

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences.* Burlington: Elsevier Science. Retrieved from <https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=923159>

Garcia-Sierra, A., Diehl, R. L., & Champlin, C. (2009). Testing the double phonemic boundary in bilinguals. *Speech Communication*, *51*(4), 369–378. <https://doi.org/10.1016/j.specom.2008.11.005>

Gonzales, K., Byers-Heinlein, K., & Lotto, A. J. (2019). How bilinguals perceive speech depends on which language they think they’re hearing. *Cognition*, *182*, 318–330.

Gonzales, K., & Lotto, A. J. (2013). A *bafri* , un *pafri*: Bilinguals’ pseudoword identifications support language-specific phonetic systems. *Psychological Science*, *24*(11), 2135–2142. <https://doi.org/10.1177/0956797613486485>

Izura, C., Cuetos, F., & Brysbaert, M. (2016). *Lexical test for advanced learners of spanish*. American Psychological Association. <https://doi.org/10.1037/t47086-000>

Lakens, D. (2017). Equivalence tests: A practical primer for t-tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *1*, 1–8. <https://doi.org/10.1177/1948550617697177>

Lemhöfer, K., & Broersma, M. (2016). *Lexical test for advanced learners of english*. American Psychological Association. <https://doi.org/10.1037/t47085-000>

Li, P., Zhang, F., Yu, A., & Zhao, X. (2020). Language history questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, *23*(5), 938–944. <https://doi.org/10.1017/S1366728918001153>

Lozano-Argüelles, C., Fernández Arroyo, L., Rodríguez, N., Durand López, E. M., Garrido Pozú, J. J., Markovits, J., … Casillas, J. V. (2020). CONCEPTUALLY CUED PERCEPTUAL CATEGORIZATION IN ADULT L2 LEARNERS. *Studies in Second Language Acquisition*, 1–16. <https://doi.org/10.1017/S0272263120000273>

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>