The title

Lena V. Kremin1,2, Andrea Sander-Montant1,2, & Krista Byers-Heinlein1,2

1 Concordia University

2 Centre for Research on Brain, Language and Music

Author note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Enter author note here.

Correspondence concerning this article should be addressed to Lena V. Kremin, 7141 Sherbrooke St. West, PY-033, Montréal, QC, H4B 1R6. E-mail: [lena.kremin@mail.concordia.ca](mailto:lena.kremin@mail.concordia.ca)

Abstract

One or two sentences providing a **basic introduction** to the field, comprehensible to a scientist in any discipline.

Two to three sentences of **more detailed background**, comprehensible to scientists in related disciplines.

One sentence clearly stating the **general problem** being addressed by this particular study.

One sentence summarizing the main result (with the words “**here we show**” or their equivalent).

Two or three sentences explaining what the **main result** reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge.

One or two sentences to put the results into a more **general context**.

Two or three sentences to provide a **broader perspective**, readily comprehensible to a scientist in any discipline.

*Keywords:* keywords

*Word count:* X

The title

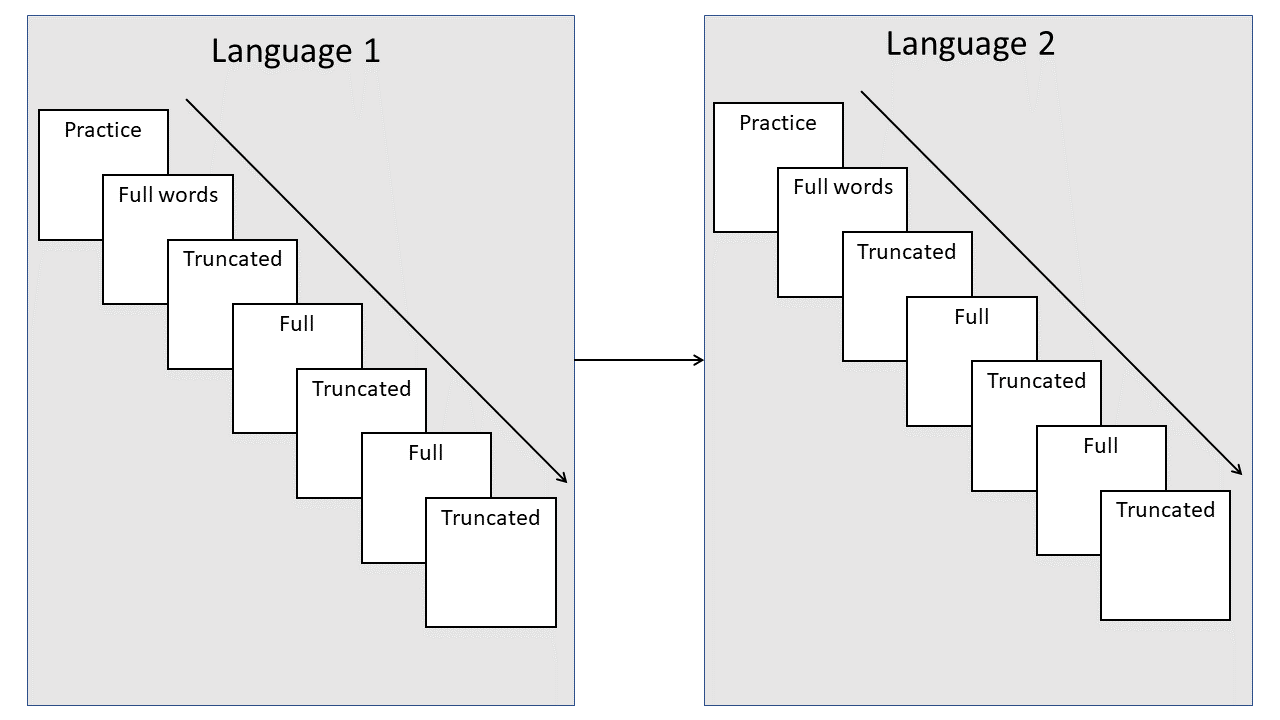
# Method

## Participants

Participants were 75 bilingual adult speakers of French and English from Quebec, Canada (mean age = 24.76, age range = 19 - 44). Participants were recruited in one of two primary ways: 1) from the Concordia University Participant Pool or 2) from the International Laboratory for Brain, Music and Sound Research (BRAMS) online testing platform. Participants from Concordia University received course credit, and participants recruited from BRAMS received a $10 gift card for their participation. An additional 69 participants completed the study but were removed from the analysis for not meeting the pre-registered language criteria (see below; 49), not following instructions to complete the task while wearing headphones (18), or not meeting our pre-registered attention criteria (see below; 2).

Language background was assessed using the Language History Questionnaire (LHQ; REF). Participants were required to be a native speaker of French (N = 14) or English (N = 34) or be a native French-English bilingual (N = 27). On average, participants reported learning English from the age of 1.27 years (SD = 2.94, Range = 0 – 11) and French from the age of 2.73 years (SD = 3.92, Range = 0 – 18). Participants rated their proficiency in each language for listening, speaking, reading, and writing separately on a scale of 1-7. We calculated the mean of these four values to calculate participants’ proficiency in both English and French. Participants’ mean proficiency was 6.69 (SD = 0.63, Range = 3.75 - 7) in English and 5.45 (SD = 1.39, Range = 1.50 - 7) in French. We did not require participants to have minimum proficiency in either French or English, as one of the main goals of this study was to investigate how various levels of proficiency affect phoneme perception. We then calculated a single proficiency difference score by subtracting French proficiency from English proficiency. Thus, a negative score indicated higher proficiency in French, a score close to 0 indicated relatively equal proficiency in each language, and a positive score indicated higher proficiency in English. Proficiency difference scores ranged from -3.25 - 5.50 (*Mean* = 1.24, *SD* = 1.59), indicating that on average, participants were more proficient in English.

Due to high rates of multilingualism in Montreal and Quebec, and following similar criteria that have been used to investigate speech sound perception in this population (Gonzales et al., 2019), we included participants with knowledge of additional languages, as long as they were not learned from birth (i.e., they were learned after English and/or French). Sixty-four participants reported knowing additional languages, including Spanish (43), Italian (11), Hebrew (6), German (4), Mandarin (3), Russian (3), Dutch (2), Arabic (1), Indonesian (1), Japanese (1), Korean (1), Mandarin Chinese (1), Norwegian (1), Portuguese (1), Punjabi (1), and Romanian (1). Participants started learning these additional languages between 1 - 43 (mean = 13.36, SD = 8.10) years of age.



*Figure* *1.*  Illustration of the design of the phonological categorization task.

# Results

## General Analytic Strategy

The dependent variable for all analyses was the proportion of /p/ responses. Predictor variables included language (French versus English, contrast coded), VOT (a continuous variable ranging from -30ms to 30ms), target onset (the onset of the original word, contrast coded; e.g., the word “book” was always coded as /b/ even for trials when its onset was phonetically realized as /p/). In our pre-registration, we had planned to investigate English and French proficiency as separate predictors, but as most participants ended up being quite proficient in English (*M* = 6.69/7), we opted to use the proficiency difference score (i.e., English proficiency - French proficiency) to capture both measures of proficiency in a single variable. Pre-registered confirmatory analyses were conducted by fitting logistic mixed effects models to the full dataset. Models were fit separately for full words and truncated words. The final models are reported in the main text, and intermediary models are reported in the supplementary materials. We conducted additional exploratory analyses following Gonzales et al. (2019), by first fitting logistic models to individuals’ data, and then comparing the resulting individual VOT boundaries across relevant predictor variables.

## Logistic Mixed-Effects Model

Our pre-registered analytic approach began with a base logistic mixed-effects model developed through pilot data (report of the pilot data available at <https://osf.io/bx4ad/>). We used mixed-effects models, because this technique allows for individual-level and group-level data to vary independently, thus accounting for the dependence of error terms within each level (Sonderegger, 2021). We used a logistic model because our response variable in the forced-choice task was binary (either /b/ or /p/). Our main research question – whether bilinguals perceive phonemes in a language-specific manner – was tested via effects of and interactions with the language predictor. Significant effects of language would indicate that participants had language-specific phoneme perception based on the phonetic cue of VOT. We predicted a significant effect of language for the truncated words, but not for the full words. We also expected to find significant main effects of VOT, whereby participants should report hearing /p/ more often at higher VOTs for both full and truncated words. For the full words only, under the Ganong effect, we expected a main effect of target onset, whereby known words that begin with /p/ (e.g., “puppy”) would be more often heard with /p/ than known words that begin with /b/. When looking at the effect of proficiency, we expected bilinguals with higher proficiency in each language to exhibit a more pronounced effect of language context than those with lower proficiency in their second language.

### Model Selection.

The base model included main effects for VOT, language, and target onset, and two way-interactions between VOT and language as well as VOT and target onset. The base model also included random intercepts for participants and items (i.e., individual target words; base model equation: response ~ VOT + language + target onset + VOT:language + VOT:target onset + (1 | participant) + (1 | word)). We then added the proficiency difference score, which was retained in the subsequent model iterations if the associated regression coefficient had an associated z value larger than 1.96 or less than -1.96. If proficiency was retained, we then entered in 2-way interactions between proficiency and language then target onset, and these were retained if the interaction regression coefficient had an associated z value larger than 1.96 or less than -1.96.

We followed a similar procedure for determining the random effects structure. We started with an intercepts-only model (Sonderegger, 2021) with subject and item as random intercepts. From there, we added random slopes in the following order: VOT, language, target onset, proficiency (if fixed effects included in model). Random slopes were retained if they statistically significantly reduced the deviance based on a Likelihood ratio test and pruned if they led to a convergence error.

We had also pre-registered conducting exploratory analyses to investigate the potential effect of age of acquisition on phoneme perception. While inspecting this variable, we discovered that most of our participants began learning both languages early in life (at or before 5 years of age). Thus, our data did not have enough variability to properly address this question, as the magnitude of an effect is reduced when the data have a restricted range, and thus we did not perform this analysis.

The final model for full words built upon the base model and additionally included the main effect of proficiency, the interaction between proficiency and target onset, random slopes for VOT and language by participant, and a random slope for VOT by word (full words model: response ~ VOT + language + target onset + proficiency + VOT:language + VOT:target onset + proficiency:target onset + (1 + VOT + language || participant) + (1 + VOT || word)). The final model for the truncated words, compared to the baseline model, included the additional main effect of proficiency (truncated words model: response ~ VOT + language + target onset + VOT:language + VOT:target onset + proficiency + (1 | participant) + (1 | word)). For these final models, we obtained several model diagnostic measurements. The Brier Score (.12 for both the full words and truncated words models) indicated a small mean difference between the fitted and actual values of the dependent variable, which indicates a good model fit (Brier, 1950). The Expected Percentage of Correct Predictions (76.38% for the full words model and 74.18% for the truncated words model; Herron, 1999) and the Classification Accuracy Measures (88.6 % for the full words model and 82.6% for the truncated words model) also indicated a good model fit. We report the outcomes of each model below.

### Full Words.

Addressing our main research question, we found no main effect of language nor an interaction between language and VOT (See Table 1). , suggesting that bilinguals did not perceive the consonants in a language-specific way as in full word contexts, described in previous literature for syllables and non-words. Further examining the fixed effects, as anticipated, we observed main effects of VOT and target onset: participants reported hearing /p/ more often with increased VOT and for real words that begin with /p/. The interaction between VOT and the target onset approached statistical significance, tentatively suggesting that the effect of VOT on participants’ perception depended on the onset of the target word (See Figure 2). This supports the prediction stemming from the Ganong effect that participants would be more likely to report hearing the sound that resulted in a real word. Finally, there was a significant effect of proficiency. As participants’ relative proficiency in English increased, they gave fewer /p/ responses, indicating they perceived phonemes differently than participants who were relatively more proficient in French . This is consistent with what would be expected from monolinguals in each language (Abramson & Lisker, 1973; Caramazza et al., 1973).

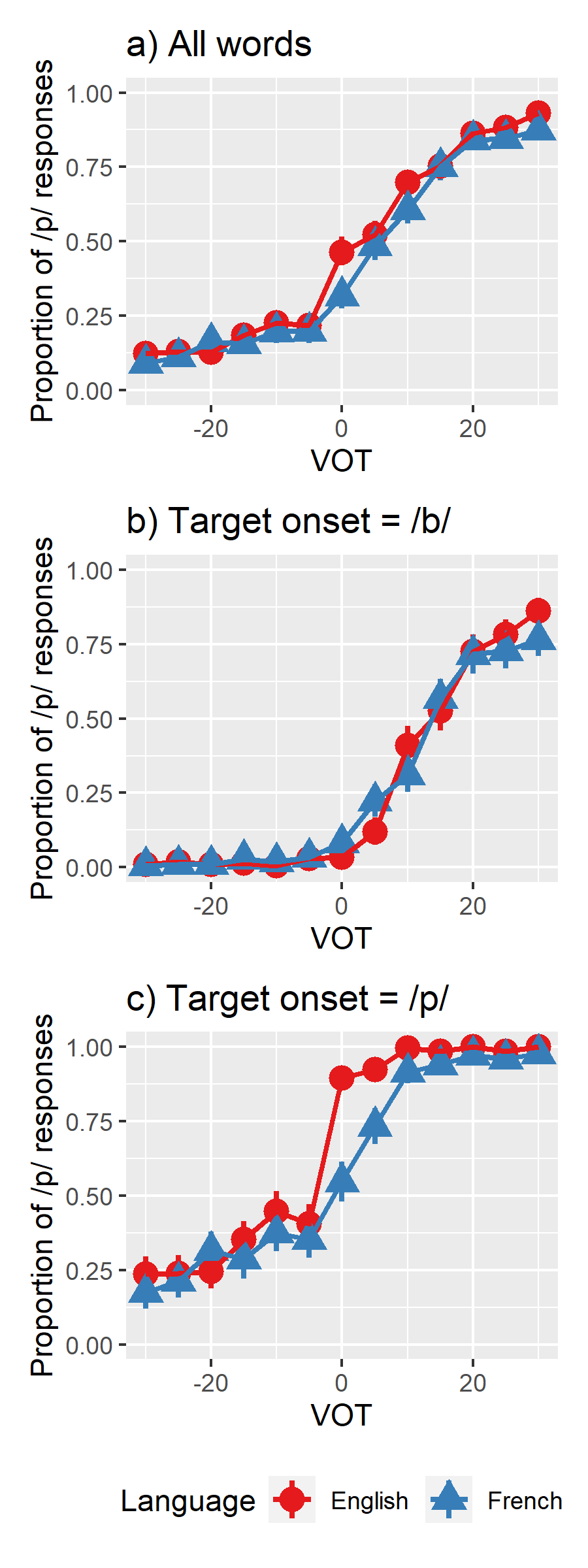
Examining the random effects, there was little variance in the random slopes. The random intercept by participant was also small, suggesting that the effect of VOT did not vary strongly across participants. However, there was high variance on the random slope of item, indicating that VOT boundaries strongly depended on the word tested. Individual items are visualized in Figure 3.

Table 1:

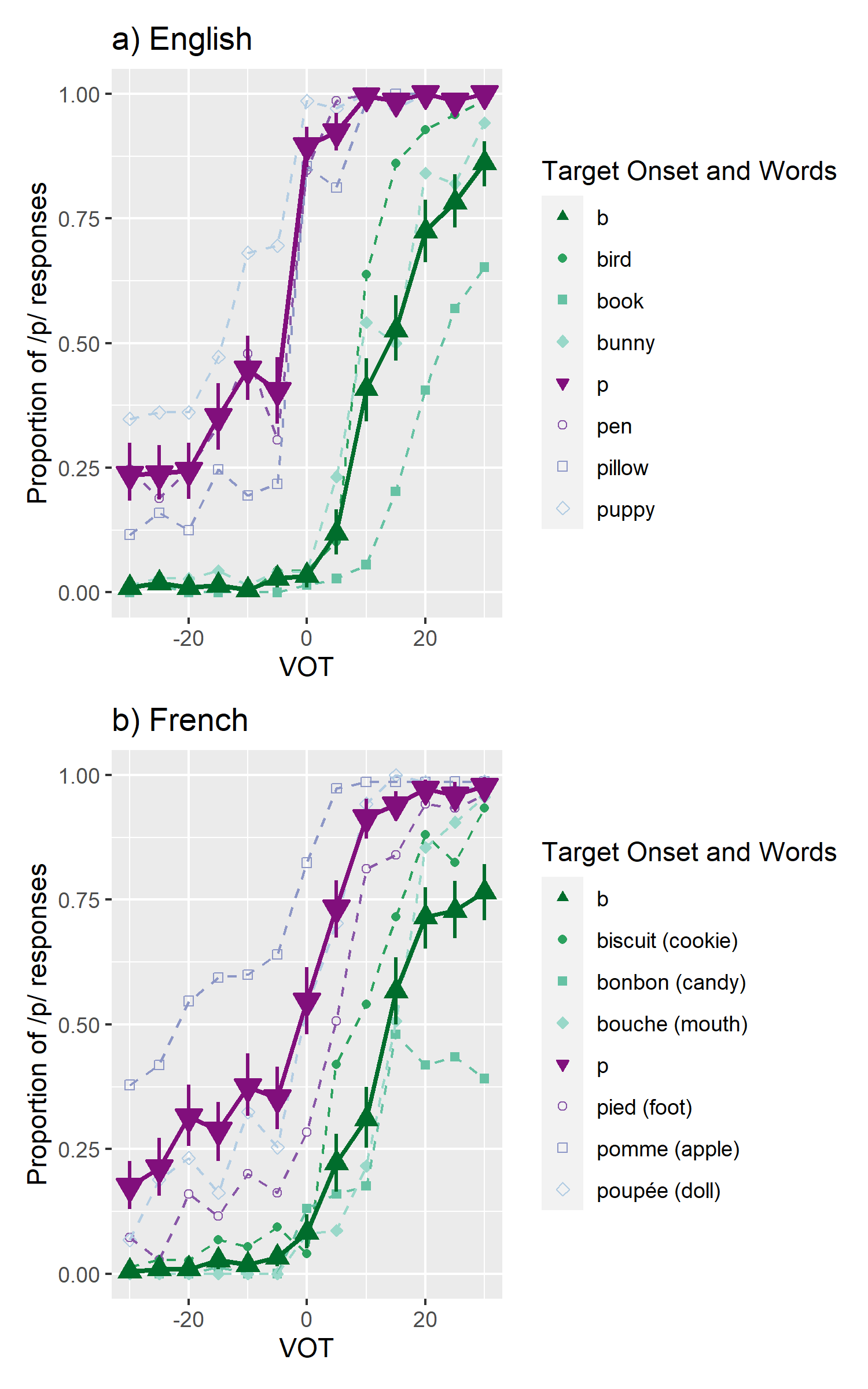
*Results from logistic mixed effect model for full words*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Fixed effects | Estimate | Standardized estimate | Standard error | *z*-statistic | *p*-value |  |
| Intercept | -0.569 | -0.829 | 0.295 | -1.928 | 0.054 | + |
| VOT | 0.174 | 3.246 | 0.015 | 11.949 | < 0.001 | \*\*\* |
| Language | -0.24 | -0.12 | 0.564 | -0.426 | 0.67 |  |
| Target onset | 4.131 | 2.032 | 0.564 | 7.33 | < 0.001 | \*\*\* |
| Proficiency difference | -0.219 | -0.345 | 0.05 | -4.348 | < 0.001 | \*\*\* |
| VOT x Language | -0.029 | -0.276 | 0.025 | -1.164 | 0.244 |  |
| VOT x Target onset | -0.049 | -0.458 | 0.025 | -1.933 | 0.053 | + |
| Target onset x Proficiency difference | -0.056 | -0.044 | 0.053 | -1.064 | 0.287 |  |
| **Random effects** |  | **Variance** |  |  |  |  |
| Participant | Intercept | 0.38 |  |  |  |  |
| Participant | VOT | 0.003 |  |  |  |  |
| Target word | Intercept | 0.906 |  |  |  |  |
| Target word | VOT | 0.002 |  |  |  |  |

*Note.* Equation: VOT + Language + Target onset + VOT x Language + VOT x Target onset + Proficiency difference + Target onset x Proficiency difference + (1 + VOT || participant) + (1 + VOT || Target word). + *p* < .1, \* *p* < .05, \*\* *p* < .01, \*\*\* *p* < .001



*Figure* *2.*  Proportion of /p/ responses as a function of VOT and language for both target onsets combined (a) and each target onset separately (b and c) for full words.



*Figure* *3.*  Proportion of /p/ responses as a function of VOT and Target Onset. The solid lines represent the average values for each Target Onset across all Target Words. The dotted lines represent the average values for each Target Word.

### Truncated Words.

Our second model analyzed participants’ responses to the truncated words, and the final set of predictors included main effects of VOT, language, target onset, and the proficiency difference score, and interactions between VOT and language and VOT and target onset (See Table 2). The model also included random intercepts for participants and target words and random slopes for VOT and language.

Like the model for full words, the main effects of VOT and target onset, and the interaction between VOT and target onset were statistically significant, suggesting that while VOT is a large contributor to participants’ perception, they were still influenced by the onset of the target word. There was no main effect for language, but unexpectedly there was a significant interaction between VOT and language, indicating that participants were responding to the cue of VOT differently across languages. Specifically, participants were more likely to report hearing a /p/ during English blocks compared to French blocks. Thus, we did observe some evidence for language-specific phoneme perception in the truncated words condition (See Figure 4). Lastly, there was a significant effect of proficiency. Like in the model for the full words, participants who were relatively more proficient in English gave fewer /p/ responses, indicating a higher VOT boundary than participants relatively more proficient in French.

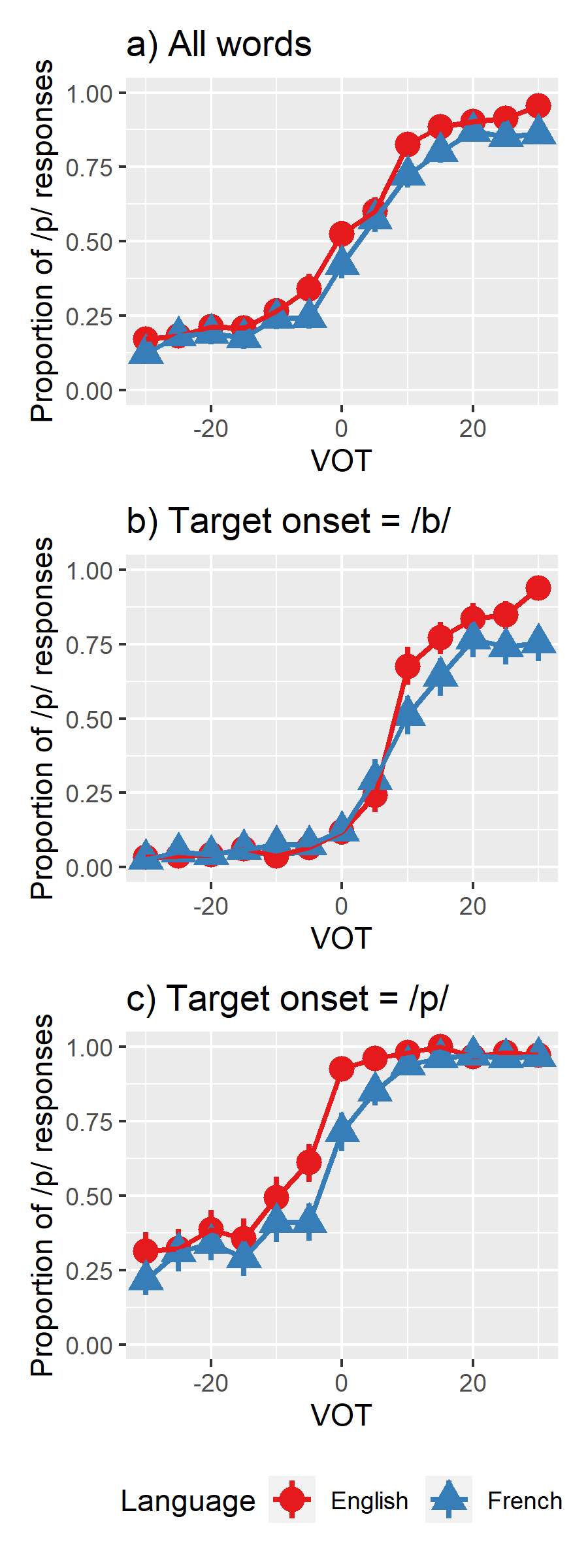
As in the model for the full words, the random slopes of target word accounted for more variance in this model than the random slopes of participants, meaning that differences across items accounted for more variation than the differences between participants. This suggests that the outcomes varied a bit for each individual but they varied more depending on the truncated word presented (see Figure 5).

Table 2:

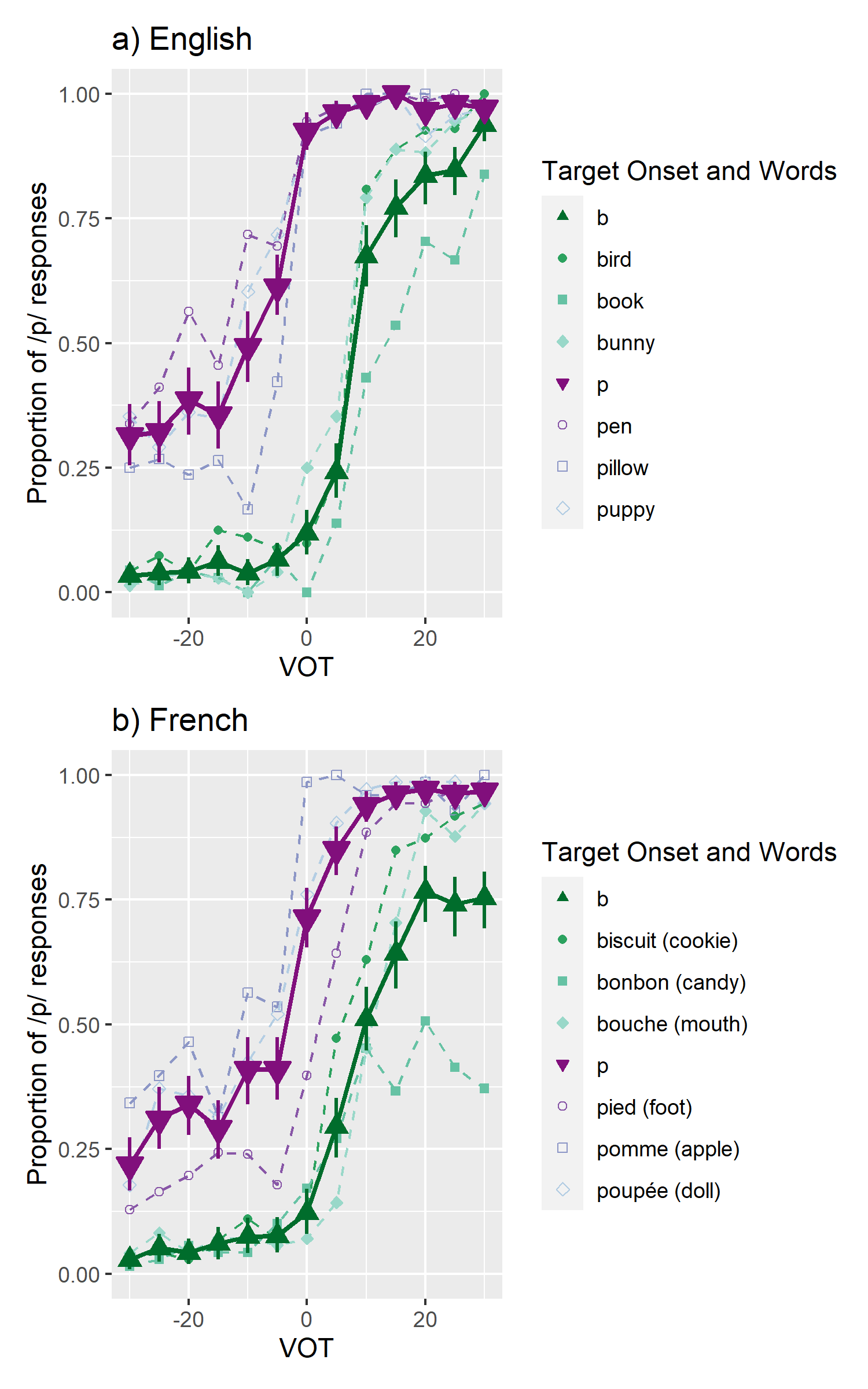
*Results from logistic mixed effect model for truncated words*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Fixed effects | Estimate | Standardized estimate | Standard error | *z*-statistic | *p*-value |  |
| Intercept | 0.107 | 0.013 | 0.183 | 0.581 | 0.561 |  |
| VOT | 0.115 | 2.151 | 0.002 | 51.36 | < 0.001 | \*\*\* |
| Language | -0.476 | -0.238 | 0.342 | -1.393 | 0.164 |  |
| Target onset | 2.722 | 1.361 | 0.344 | 7.913 | < 0.001 | \*\*\* |
| Proficiency difference | -0.081 | -0.126 | 0.035 | -2.305 | 0.021 | \* |
| VOT x Language | -0.02 | -0.187 | 0.004 | -4.65 | < 0.001 | \*\*\* |
| VOT x Target onset | -0.02 | -0.187 | 0.004 | -4.642 | < 0.001 | \*\*\* |
| **Random effects** |  | **Variance** |  |  |  |  |
| Participant | Intercept | 0.169 |  |  |  |  |
| Target word | Intercept | 0.341 |  |  |  |  |

*Note.* Equation: VOT + Language + Target onset + Proficiency difference + VOT x Language + VOT x Target onset + (1 | participant) + (1 | Target Word). + *p* < .1, \* *p* < .05, \*\* *p* < .01, \*\*\* *p* < .001



*Figure* *4.*  Proportion of /p/ responses as a function of VOT and language for both target onsets combined (a) and each target onset separately (b and c) for truncated words.



*Figure* *5.*  Proportion of /p/ responses as a function of VOT and Target Onset for truncated words. The solid lines represent the average values for each Target Onset across all Target Words. The dotted lines represent the average values for each Target Word. Note that the full word is included in the legend, but participants only heard the first 200ms of the word.

## Individual Curve Analysis

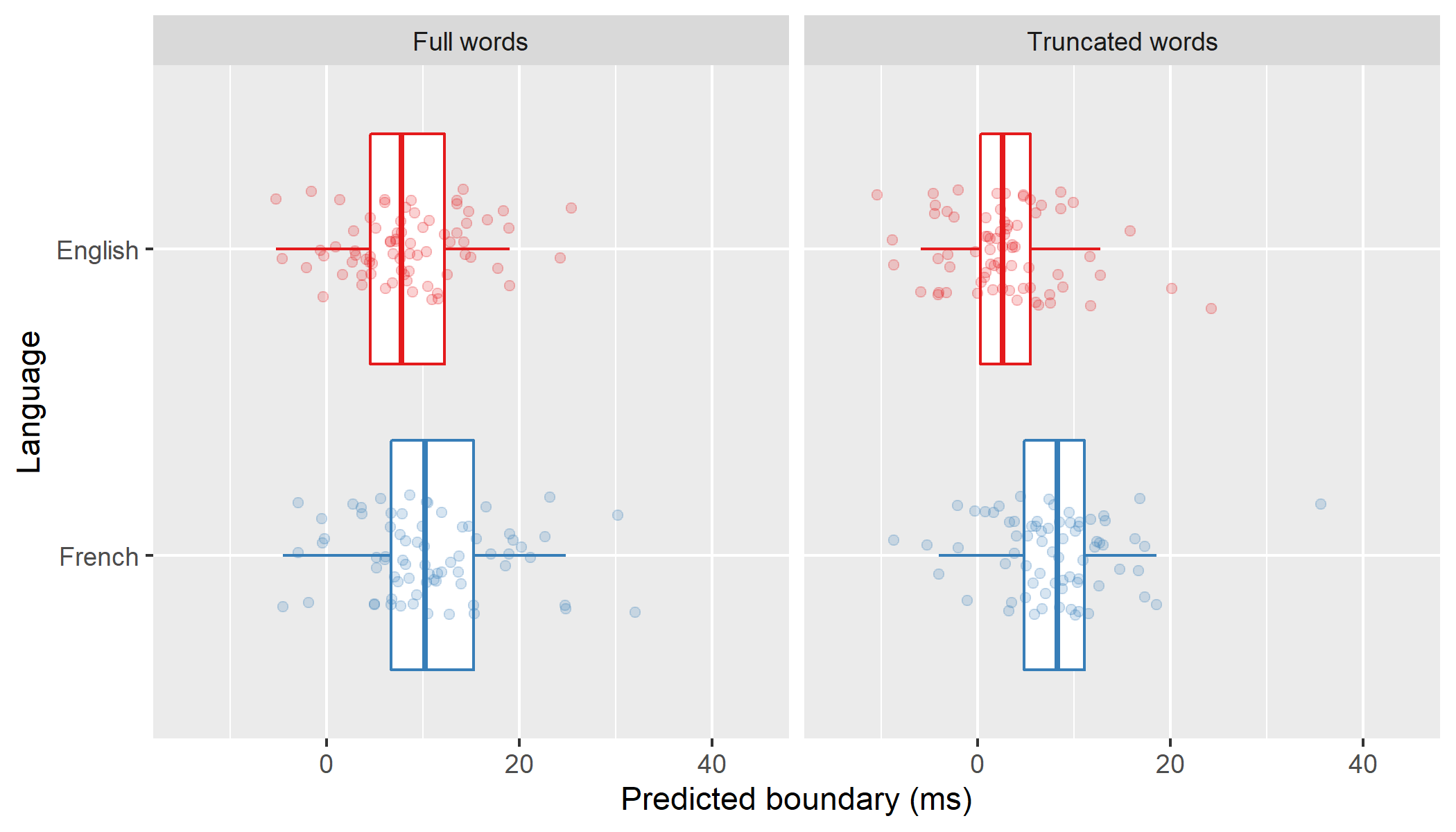
In addition to our pre-registered analyses, we performed a complementary analysis following analyses performed by Gonzales and colleagues (2019), by computing phoneme boundaries for each participant for both full and truncated words using a logistic regression model fitted to their individual data. Computing individual boundaries is a way to examine how phoneme perception varies between individuals based on a specified set of variables. First, we computed individual boundaries based on VOT, language, and their interaction. Second, we computed individual boundaries when adding target onset and the interaction between target onset and VOT. Using each model’s intercept and coefficient terms, we calculated the VOT value where a /b/ and a /p/ response were equally likely. As the VOT values of the individual boundaries were skewed, here, we report median values and compare boundaries across conditions with the non-parametric paired samples Wilcoxon test.

### Full Words.

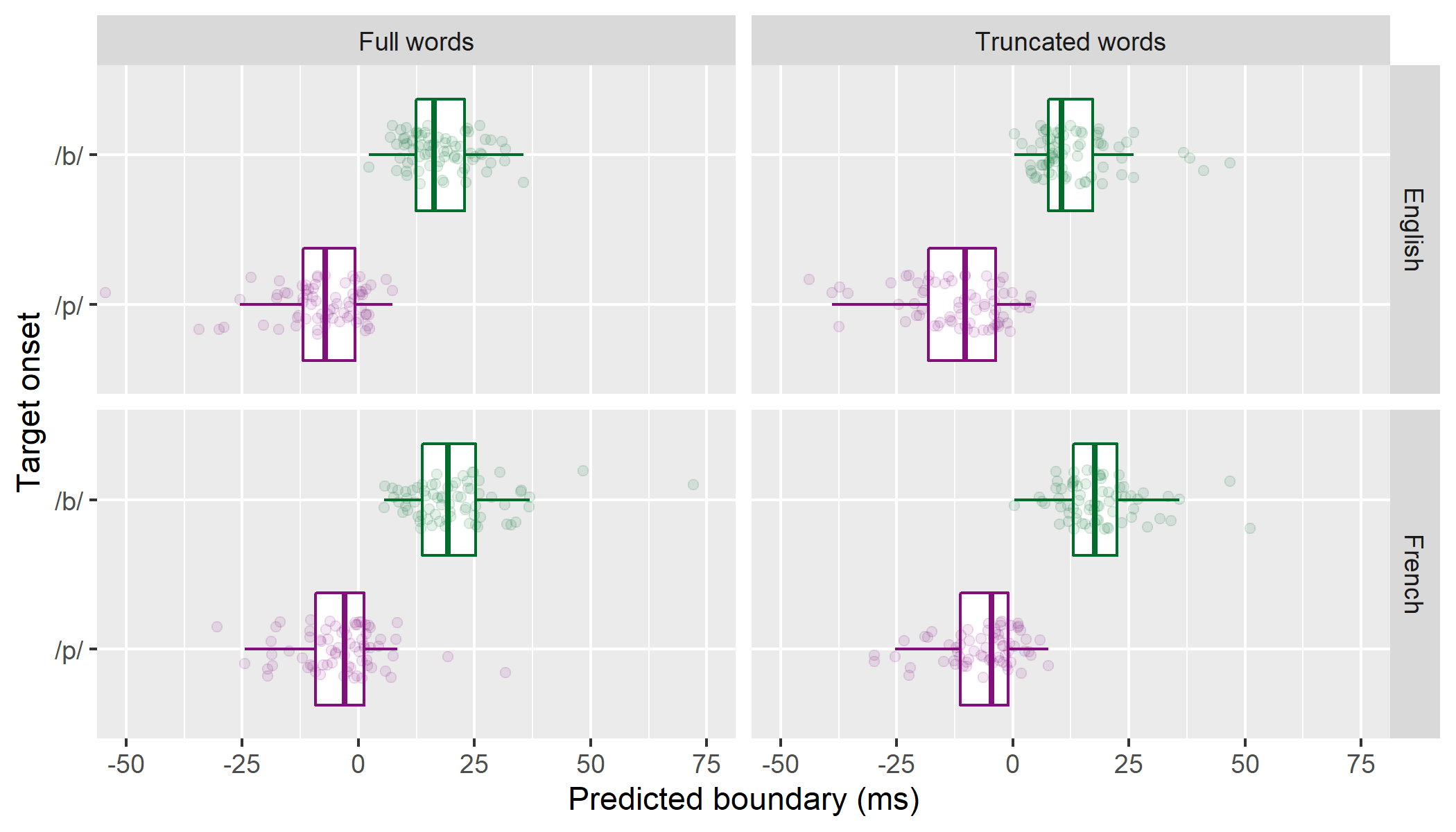
When considering only language, participants’ individual VOT boundaries were lower in English (*Median* = 7.79, *SD* = 6.10) than in French (*Median* = 10.23, *SD* = 12.10), , (See Figure 6, left). When considering both language and target onset, in English, participants’ individual VOT boundaries were lower for words beginning with /p/ (*Median* = -7.15, *SD* = 10.11) than words beginning with /b/ (*Median* = 16.23, *SD* = 7.01), , . Similarly, in French, participants’ individual VOT boundaries were lower for words beginning with /p/ (*Median* = -2.93, *SD* = 11.67) than words beginning with /b/ (*Median* = 19.22, *SD* = 13.59), , (See Figure 7, left). While we saw statistically significant patterns in individual boundaries across language and target onset, the spread of these boundaries suggests that individuals weigh cues differently for phoneme perception. Trying to identify a single boundary value for all listeners may neglect this variation and obscure the importance of other cues.

### Truncated Words.

When considering only language, participants’ individual VOT boundaries were lower in English (*Median* = 2.59, *SD* = 6.02) than in French (*Median* = 8.23, *SD* = 8.36), , (See Figure 6, right). When considering both language and target onset, in English, participants’ individual VOT boundaries were lower for words beginning with /p/ (*Median* = -10.25, *SD* = 34.23) than words beginning with /b/ (*Median* = 10.49, *SD* = 8.98), , . Similarly, in French, participants’ individual VOT boundaries were lower for words beginning with /p/ (*Median* = -4.52, *SD* = 39.51) than words beginning with /b/ (*Median* = 17.58, *SD* = 16.39), , (See Figure 7, right). Again, we saw a large spread in boundary values, suggesting that individuals weigh cues differently.



*Figure* *6.*  Individual VOT boundaries based on language. Note that for visualization purposes, outlier boundaries greater than 45ms have been removed.



*Figure* *7.*  Individual VOT boundaries based on language and target onset. Note that for visualization purposes, outlier boundaries less than -50ms or greater than 75ms have been removed.

### Proficiency Difference Scores.

Because proficiency was a significant predictor in our logistic mixed-effect models, we calculated correlations between participants’ proficiency difference scores and their individual boundaries (See Table 3 for language and Table 4 for language and onset). To account for multiple comparisons, we reduced to .013 for boundaries based on language and .006 for boundaries based on language and onset. With this correction, the only correlation that was statistically significant was between proficiency and French words beginning with /p/ in full blocks. Unlike in the previous analyses, these results suggest that overall, the difference between an individual’s proficiency in English and French did not affect how they weighed the cues of language and target onset, and that proficiency is not sufficient to explain the range of individual boundaries we observed across all testing conditions.

Table 3:

*Correlations between proficiency difference scores and individual boundaries calculated based on language.*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Block | Language |  | 95% CI |  |  |  |  |
| Full words | English | .26 | [.03, .46] | 2.24 | 71 | 0.028 |  |
| Full words | French | .19 | [-.04, .41] | 1.67 | 71 | 0.099 |  |
| Truncated words | English | .13 | [-.10, .35] | 1.12 | 70 | 0.269 |  |
| Truncated words | French | .01 | [-.22, .24] | 0.09 | 70 | 0.928 |  |

*Note.* Asterisk denotes *p*-value < 0.013

Table 4:

*Correlations between proficiency difference scores and individual boundaries calculated based on language and target onset.*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Block | Language | Target Onset |  | 95% CI |  |  |  |  |
| Full | English | /b/ | .10 | [-.14, .32] | 0.81 | 71 | 0.422 |  |
| Full | English | /p/ | .22 | [-.01, .43] | 1.88 | 71 | 0.064 |  |
| Full | French | /b/ | .04 | [-.19, .27] | 0.33 | 71 | 0.744 |  |
| Full | French | /p/ | .34 | [.12, .53] | 3.09 | 71 | 0.003 | \* |
| Truncated | English | /b/ | .06 | [-.17, .29] | 0.52 | 70 | 0.603 |  |
| Truncated | English | /p/ | .13 | [-.11, .35] | 1.07 | 70 | 0.288 |  |
| Truncated | French | /b/ | -.06 | [-.29, .17] | -0.53 | 70 | 0.601 |  |
| Truncated | French | /p/ | .03 | [-.21, .26] | 0.23 | 70 | 0.821 |  |

*Note.* Asterisk denotes *p*-value < 0.006