Glide affiliation

# 1 Methods

## 1.1 Statistical analyses

We report four primary statistical analyses in order to address the aforementioned research questions. To this end, we fit a series of Bayesian regression models, which are described in detail in the corresponding sections below and in the supplementary materials (LINK). All analyses were conducted in R (R Core Team, 2018, version 4.1.0). The models were fit using stan (Stan Development Team, 2018) via the R package brms (Bürkner, 2017). All models included a maximal grouping-effects structure (CITE), as well as regularizing, weakly informative priors (Gelman, Simpson, & Betancourt, 2017). Additionally, models were fit with 2000 iterations (1000 warm-up) and Hamiltonian Monte-Carlo sampling was carried out with 4 chains distributed between 4 processing cores. We report point estimates (posterior medians) for each parameter of interest, along with the 95% highest density interval (HDI), and the maximum probability of effect (MPE). A complete description of the models and output summaries in table format are available in the supplementary materials (LINK).

# 2 Results

## 2.1 Syllable division task

The first model analyzed the Syllable division task data using hierarchical multinomial logistic regression. The participants responses to the critical CGVG sequences were classified as triphthong, hiatus, or simplification, i.e., [la.ka.ˈpi̯ai̯s.to], [la.ka.pi.ˈai̯s.to], or [la.ka.ˈpai̯s.to], respectively. The responses were modeled in a simple, intecept-only model, and as a function of the post-consonantal glide ([j], [w]).

The intercept was something (β = -0.30 [-0.96, 0.29]; MPE = 0.84).

Figure 2.1 shows stuff.

* Questions:
  + How do participants respond? Is GVG possible?
  + Does production depend on the glide type
* Coding
  + Triphthong: critical sequence produced in a single syllable
    - i.e. “lakapiaisto” ⇾ [la.ka.ˈpi̯ai̯s.to]
  + Hiatus: vowel + diphthong (CV + VGC),
    - i.e., “lakapiaisto” ⇾ [la.ka.pi.ˈai̯s.to]
  + Simplification: a segment was elided (typically the pre-vocalic glide)
    - i.e., “lakapiaisto” ⇾ [la.ka.ˈpai̯s.to]
* Triphthongs were produced in approximately 45% of the targets.
* A production containing a hiatus made up roughly 30% of the data, followed by a simplification of some sort (~25% of the time)
* Overall, the task provides evidence supporting the hypothesis that pre-vocalic glides can be part of the onset in this variety of Spanish.
* Main finding: we have evidence supporting the hypothesis that pre-vocalic glides can be part of the onset
* Why: because the participants produced triphthongs at least some of the time. Responses were variable and we cannot account for this variability with glide type nor preceding consonant

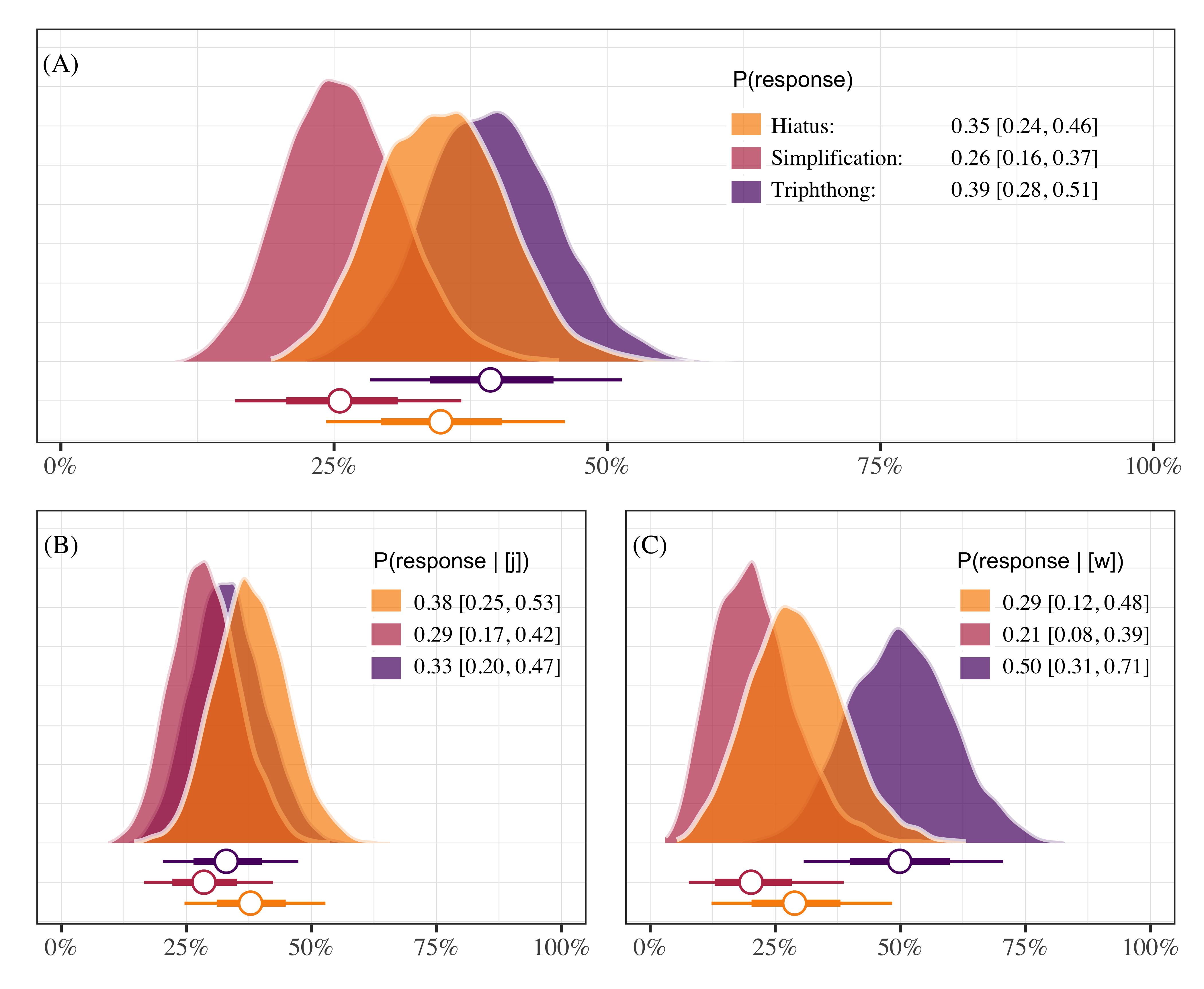


Figure 2.1: This is a figure caption.

## 2.2 Phrase reading task

For the carrier phrase production task data, we fit three models: (1) a multi-level linear regression analyzing segment duration as a function of whether the onset included a palatal consonant or not (sum coded as 1, -1), and two Generalized Additive Mixed Models (GAMMs) analyzing the time course of (2) F1 (Hz) as a function of palatal status, and (2) intensity (dB) as a function of palatal status. In both cases, palatal status (palatal, other) was set as an ordinal factor and included a difference smooth term.

* Hypothesis: Pre-vocalic glides will be disallowed if preceded by a palatal consonant
* Measure duration, F1, and intensity of the pre-vocalic glide in two environments: after a palatal consonant, after any other consonant
* If pre-vocalic glides are blocked after palatals (i.e., “lliape”), when compared with pre-vocalic glides that are not preceded by a palatal segment (i.e., “piano”) we expect to observe…
  + differences in overall duration
  + formant trajectory differences related to height (F1)
* Analysis 1 (duration)
  + Linear mixed effects model
  + Model: duration ~ environment
  + Random effect: by-subject/by-item intercepts with random slopes for preceding consonant

Differences in duration: Why? Because something is theoretically elided Differences in F1: Why? Because in one case (piano) a glide should be produced without problems and in the other case (lliape) it shouldn’t thus there should be differences in formant movement between the two conditions: (F1 correlates with height) PIANO: formant movement the high vocoid [j] to the low vowel [a] LLIAPE: if pre-vocalic glide is blocked then we shouldn’t see the same formant movement

* If the participants are able to produce a glide in some contexts (piano) but not others (lliape) we should also see differences in F1 as a function of the preceding consonant.
* Analysis 2
  + Generalized Additive Mixed Model
  + Model: F1 ~ preceding\_consonant + time
  + Reference smooth for time, participant
  + Difference smooth for preceding\_consonant
* Participants might be producing a longer segment in the palatal condition because of the fact that they cannot naturally produce both.
* In other words, they might be trying to produce something, but because the target is illicit, they resort to lengthening.
* Prediction: intensity of the lengthened segment should be lower than that of a glide.
* Why? More consonant-like productions have lower intensity than more vowel-like productions.
* Analysis 3
  + Generalized Additive Mixed Model
  + Model: Intensity ~ preceding\_consonant + time

Summary - Sonoran speakers used variable strategies when producing the CGVGC sequences - Importantly, they produced the GVG sequence at least some of the time (thus it is possible) - Acoustic analysis shows that pre-vocalic segments are longer after palatal consonants - Not shorter (i.e., elided because they are blocked) - Analysis of the time course suggests the duration increase could be due to lengthening of onset (Strategy to avoid illicit sequence)

The intercept was something (β = -0.07 [-0.41, 0.27]; MPE = 0.65).

Table 2.2 shows a picture table.

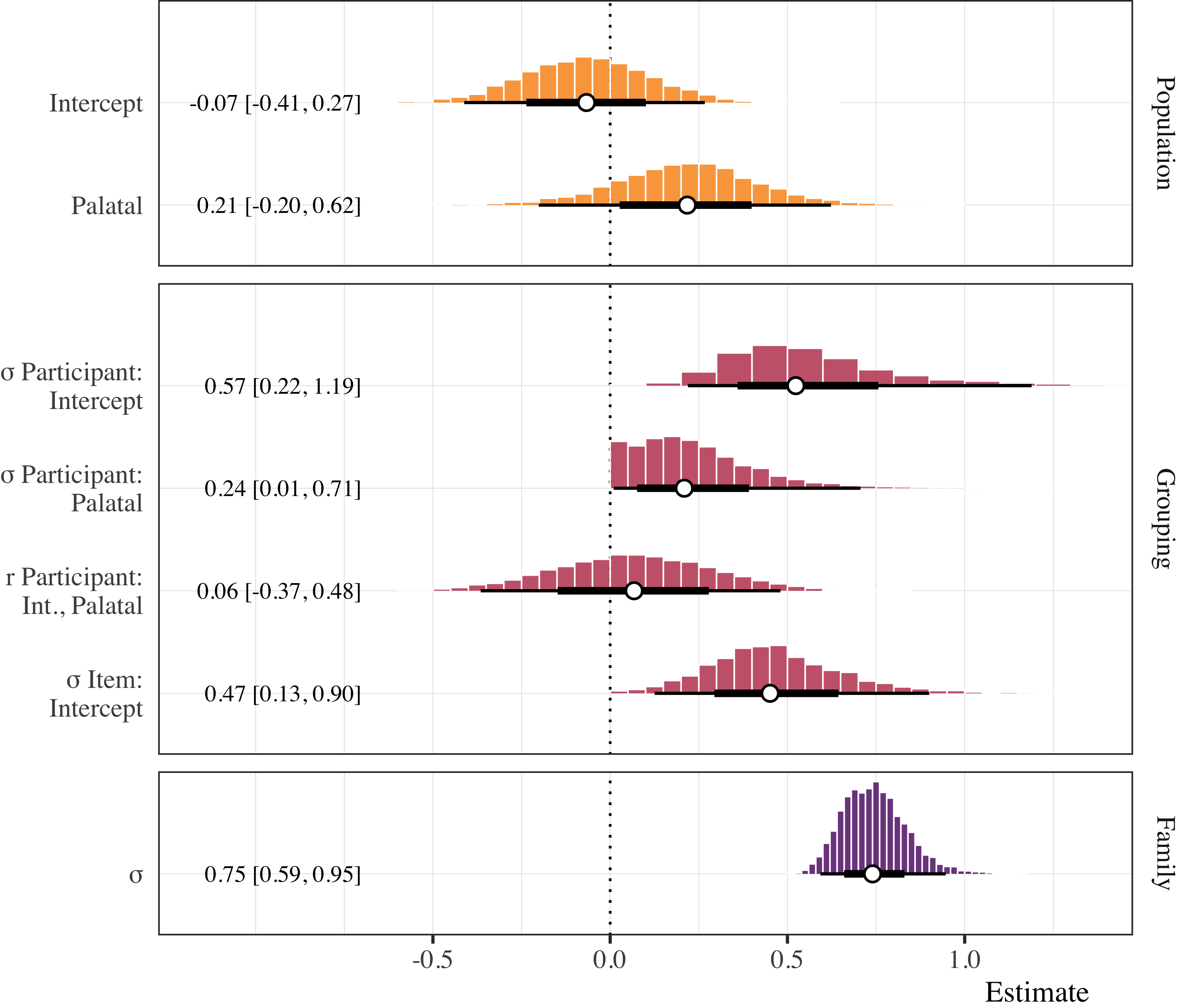


Figure 2.2: This is a table caption.

Figure 2.3 shows the F1 GAM stuff.

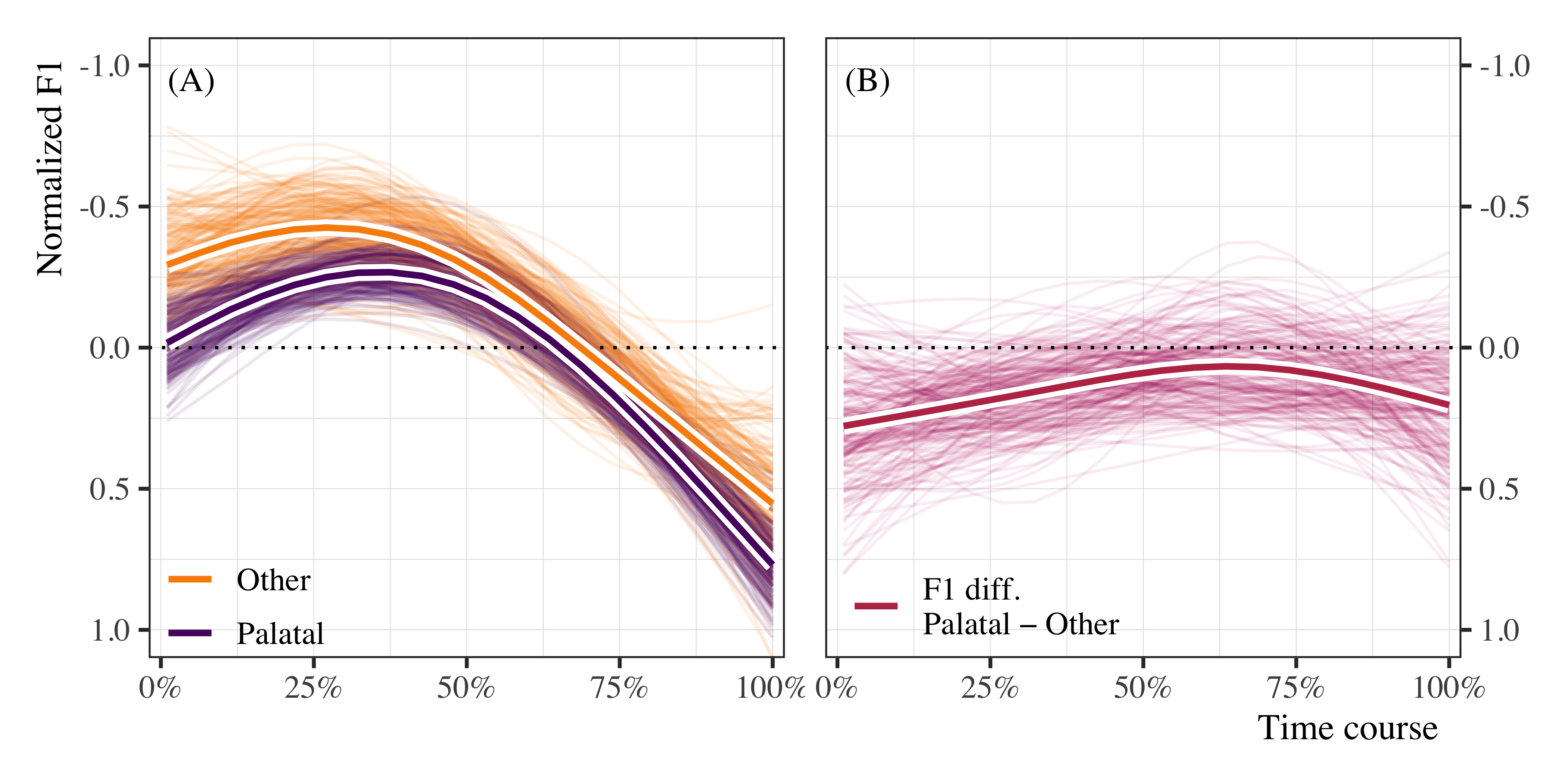


Figure 2.3: This is a figure caption for the F1 GAM.

Figure 2.4 shows the Intensity GAM stuff.

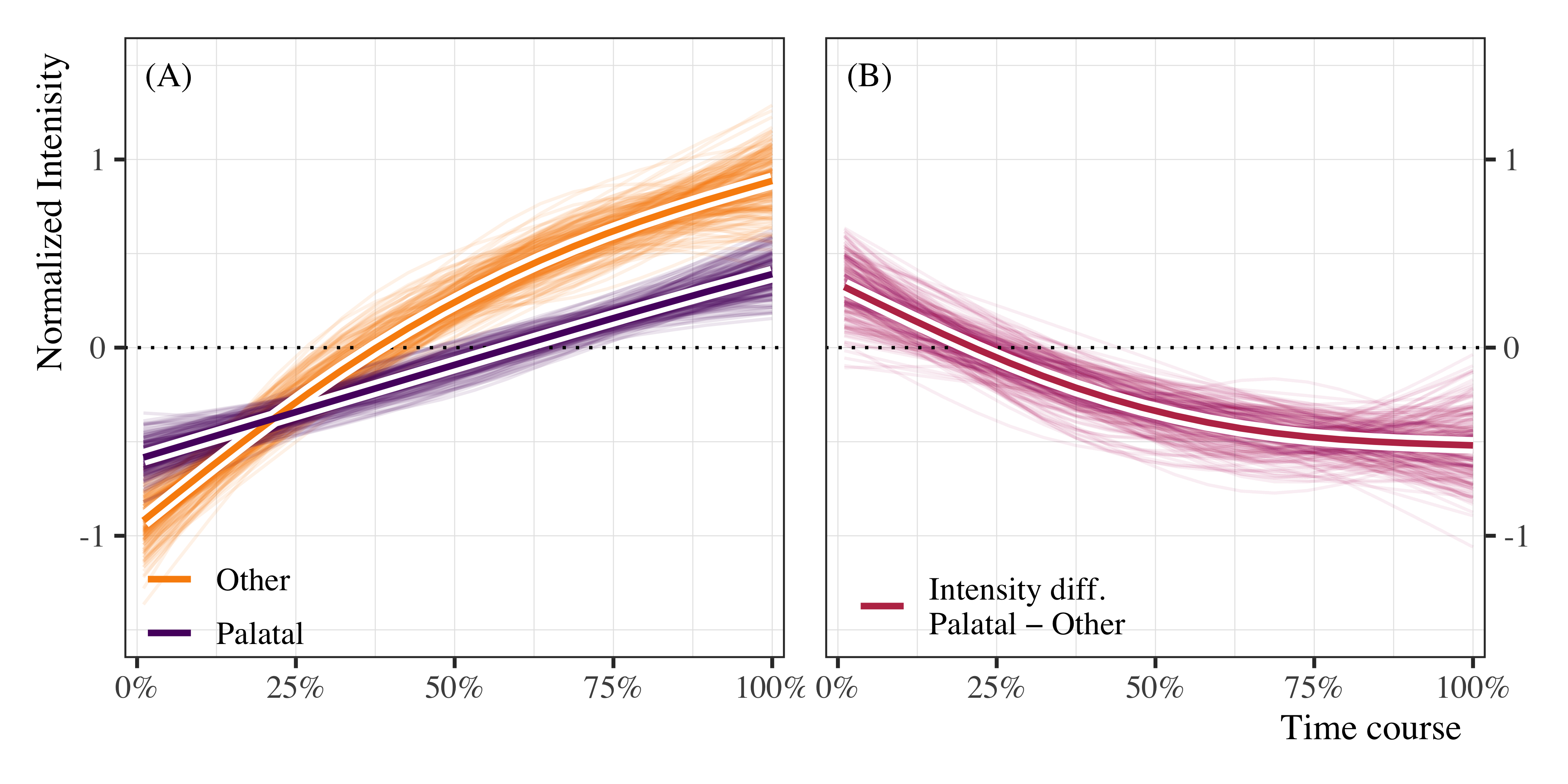


Figure 2.4: This is a figure caption for the intensity GAM.

Figure 2.5 shows the GAM forest plot

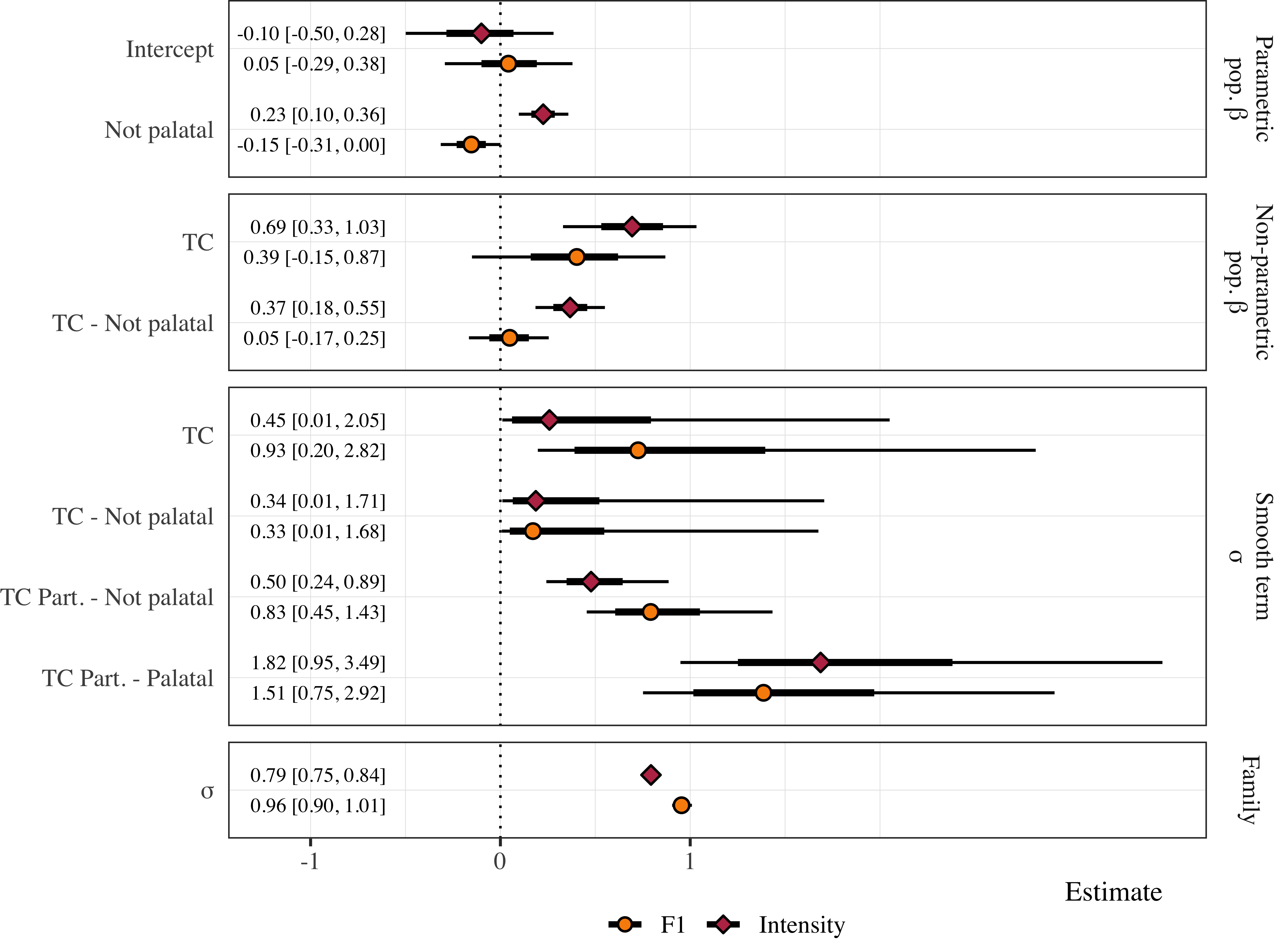


Figure 2.5: This is a figure caption for the forest plot of the gams.

Answers to Research questions

* RQ1: Is a postconsonantal, prevocalic glide parsed as the second segment of a complex onset in Sonoran Spanish?
  + A1:Yes, at least some of the time.
* RQ2: In more general terms, the question is: are Spanish prevocalic glides always part of a complex nucleus (preceding a full vowel, in a diphthong) or can the glide be parsed in the onset in some dialects?
  + A2:The glide can be parsed in the onset in dialects like Sonoran Spanish, but it is not categorical

# 3 Supplementary materials

## 3.1 Bayesian Data Analysis

In all cases we used Bayesian Data Analysis (BDA, CITE) for statistical inferences.

## 3.2 Syllabification task

The first model analyzed the syllabification task data using multinomial logistic regression. The participants responses (triphthong, hiatus, simplification) were modeled in a simple, intecept-only model, and as a function of the post-consonantal glide ([j], [w]).

Table 3.1 below.

Table 3.1: Model summary for the syllabification task. The table reports posterior medians, 95% credible intervals, and probability of direction to assess estimates, along with Rhat and Effective sample size to assess model fit.

| Model | Parameter | Estimate | P(direction) | Rhat | ESS | Prior |
| --- | --- | --- | --- | --- | --- | --- |
| Main | μ Simplification: Intercept | -0.30 [-0.96, 0.29] | 0.84 | 1.00 | 1724.03 | Normal(0, 20) |
| Main | μ Triphthong: Intercept | 0.12 [-0.45, 0.66] | 0.66 | 1.00 | 1974.05 | Normal(0, 20) |
| [j] | μ Simplification: Intercept | -0.29 [-1.00, 0.47] | 0.78 | 1.00 | 1576.48 | Normal(0, 20) |
| [j] | μ Triphthong: Intercept | -0.14 [-0.89, 0.52] | 0.65 | 1.00 | 1604.93 | Normal(0, 20) |
| [w] | μ Simplification: Intercept | -0.37 [-1.61, 0.76] | 0.73 | 1.00 | 1480.26 | Normal(0, 20) |
| [w] | μ Triphthong: Intercept | 0.54 [-0.42, 1.55] | 0.88 | 1.00 | 1238.82 | Normal(0, 20) |

## 3.3 Carrier task

Table 3.2 below.

Table 3.2: Model summary for the duration analysis. The table reports posterior medians, 95% credible intervals, and probability of direction to assess estimates, along with Rhat and Effective sample size to assess model fit.

| Parameter | Estimate | P(direction) | Rhat | ESS | Prior |
| --- | --- | --- | --- | --- | --- |
| Intercept | -0.07 [-0.41, 0.27] | 0.65 | 1.00 | 2540.31 | Normal(0, 0.2) |
| Palatal | 0.22 [-0.20, 0.62] | 0.86 | 1.00 | 2091.63 | Normal(0, 0.5) |

Table 3.3 below.

Table 3.3: Model summary for the F1 GAMM. The table reports posterior medians, 95% credible intervals, and probability of direction to assess estimates, along with Rhat and Effective sample size to assess model fit.

| Parameter | Function | Estimate | P(direction) | Rhat | ESS | Prior |
| --- | --- | --- | --- | --- | --- | --- |
| Intercept |  | 0.04 [-0.30, 0.37] | 0.62 | 1.00 | 1241.48 | Normal(0, 0.5) |
| Not palatal |  | -0.15 [-0.31, 0.00] | 0.98 | 1.00 | 5632.28 | Normal(0, 0.5) |
| Time course | Smooth | 0.40 [-0.15, 0.86] | 0.93 | 1.00 | 1617.80 | student\_t(3, 0, 1) |
| Time course: Not palatal | Smooth | 0.05 [-0.17, 0.24] | 0.67 | 1.00 | 5176.51 | student\_t(3, 0, 1) |

Table 3.4 below.

Table 3.4: Model summary for the F1 GAMM. The table reports posterior medians, 95% credible intervals, and probability of direction to assess estimates, along with Rhat and Effective sample size to assess model fit.

| Parameter | Function | Estimate | P(direction) | Rhat | ESS | Prior |
| --- | --- | --- | --- | --- | --- | --- |
| Intercept |  | -0.10 [-0.48, 0.29] | 0.72 | 1.00 | 1178.24 | Normal(0, 0.5) |
| Not palatal |  | 0.23 [0.10, 0.36] | 1.00 | 1.00 | 5471.15 | Normal(0, 0.5) |
| Time course | Smooth | 0.69 [0.32, 1.02] | 1.00 | 1.00 | 1761.32 | student\_t(3, 0, 1) |
| Time course: Not palatal | Smooth | 0.37 [0.20, 0.56] | 1.00 | 1.00 | 4063.34 | student\_t(3, 0, 1) |

# References

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>

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>

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

Stan Development Team. (2018). Stan modeling language users guide and reference manual (Version 2.18.0). Stan Development Team. Retrieved from <http://mc-stan.org>