# High Vowels in Unstressed Final Position

## in Fortalezenses' informal speech

Alerrandro da Silva Araujo [1] Ronaldo Lima Jr. [2]

[1] Universidade Federal do Ceará – Unicamp
*alesilvaraujo@gmail.com*

[2] Universidade Federal do Ceará
*ronaldojr@ufc.br*

# Outline

1. Introduction
2. Method
3. Results
4. Discussion
5. Final remarks

# Introduction

# Context

Behavior of unstressed vowels:

- ’espera’ [ˈspɛrɐ] (Gomes 2019, Silva 2019)
- ’potes’ [pɔts] (Leite 2006, Nascimento 2016)
- ’hoje’ [oʒ], ’bosque’ [bɔsk], ’peixe’ [peʃ], ’doce’ [dos] (Dubiela 2013)
- ’árduo’ [ahd], ’cárie’ [kaɾ] (Cristófaro Silva & Faria 2014)
- ’chave’ [ʃav] (Cristófaro Silva & Vieira 2015)
- ’sapato’ [saˈpat], ’casaco’ [kaˈzak] (Dias & Seara 2013)

# Context

Behavior of unstressed vowels:

- 'espera' ['spɛrɐ] (Gomes 2019, Silva 2019)
- 'potes' [pɔts] (Leite 2006, Nascimento 2016)
- 'hoje' [oʒ], 'bosque' [bɔsk], 'peixe' [peʃ], 'doce' [dos] (Dubiela 2013)
- 'árduo' [ahd], 'cárie' [kaɾ] (Cristófaro Silva & Faria 2014)
- 'chave' [ʃav] (Cristófaro Silva & Vieira 2015)
- 'sapato' [sa'pat], 'casaco' [ka'zak] (Dias & Seara 2013)

→ Gradient (Albano 1999, Meneses 2012, Guzzo & Garcia 2021, Silva & Lima Jr. 2021)

# Gaps and Goals

Gaps:

- Dialects of the northeast $\rightarrow$ Ceará $\rightarrow$ Fortaleza
- (Semi-)spontaneous speech

# Gaps and Goals

Gaps:

- Dialects of the northeast $\rightarrow$ Ceará $\rightarrow$ Fortaleza
- (Semi-)spontaneous speech

Goals:

- Investigate the deletion/reduction of high vowels in unstressed final position in informal (semi-)spontaneous speech of Fortalezenses, considering the following variables:
  - speakers' age, formal education and sex
  - word frequency, number of syllables in the word, stressed vowel, preceding segment (place, manner, voicing)

# Gaps and Goals

Gaps:

- Dialects of the northeast → Ceará → Fortaleza
- (Semi-)spontaneous speech

Goals:

- Investigate the deletion/reduction of high vowels in unstressed final position in informal (semi-)spontaneous speech of Fortalezenses, considering the following variables:
  - speakers' age, formal education and sex ← controlled for
  - word frequency, number of syllables in the word, stressed vowel, preceding segment (place, manner, voicing) ← not controlled for

# Method

# Corpus

NORPOFOR

*Norma Oral do Português Popular de Fortaleza-CE*

(Araújo, Viana & Pereira 2018)

- Interviews ($\approx$ 1 hr) collected between 2003–2006
- for sociolinguistics
- 198 speakers
  - raised, living, never left Fortaleza (+ 2 yrs), Fortalezense parents
  - male / female
  - age group (I: 15–25, II: 26–49, III: +50)
  - formal education (A: 0–4, B: 5–8, C: 9–11)

# Corpus

NORPOFOR

*Norma Oral do Português Popular de Fortaleza-CE*

(Araújo, Viana & Pereira 2018)

- Interviews ($\approx$ 1 hr) collected between 2003–2006
- for sociolinguistics
- ~~198~~ 16 speakers
  - raised, living, never left Fortaleza (+ 2 yrs), Fortalezense parents
  - 8 male / 8 female
  - age group 8 I: 15–25, ~~II: 26–49~~, 8 III: +50
  - formal education 8 A: 0–4, ~~B: 5–8~~, 8 C: 9–11

# Corpus

Biggest challenge


# Corpus

Biggest challenge



→ extreme noise reduction in Audacity, prevented spectral analyses

Noise Reduction (dB): 10; Sensitivity: 4; Frequency smoothing (bands): 2

# Procedure

Pilot analysis to come to these decisions:

- First ten minutes of recording ignored
- content words (nouns, adjectives and verbs)
- penultimate stress
- CV syllable
- 25 words ending in unstressed [i] per recording ($\times 16 = 400$)
- 25 words ending in unstressed [u] per recording ($\times 16 = 400$)

# Procedure

1. Potential words identified in orthographic transcription of interview

# Procedure

1. Potential words identified in orthographic transcription of interview
2. Best 25 of each vowel (noise-wise) selected and segmented in Praat
   - 2.1 word
   - 2.2 stressed vowel
   - 2.3 preceding consonant
   - 2.4 unstressed vowel

     → vowel produced or deleted?

# Procedure

1. Potential words identified in orthographic transcription of interview
2. Best 25 of each vowel (noise-wise) selected and segmented in Praat
   - 2.1 word
   - 2.2 stressed vowel
   - 2.3 preceding consonant
   - 2.4 unstressed vowel

     → vowel produced or deleted?

     conservative decision

# Procedure

1. Potential words identified in orthographic transcription of interview
2. Best 25 of each vowel (noise-wise) selected and segmented in Praat
   - 2.1 word
   - 2.2 stressed vowel
   - 2.3 preceding consonant
   - 2.4 unstressed vowel

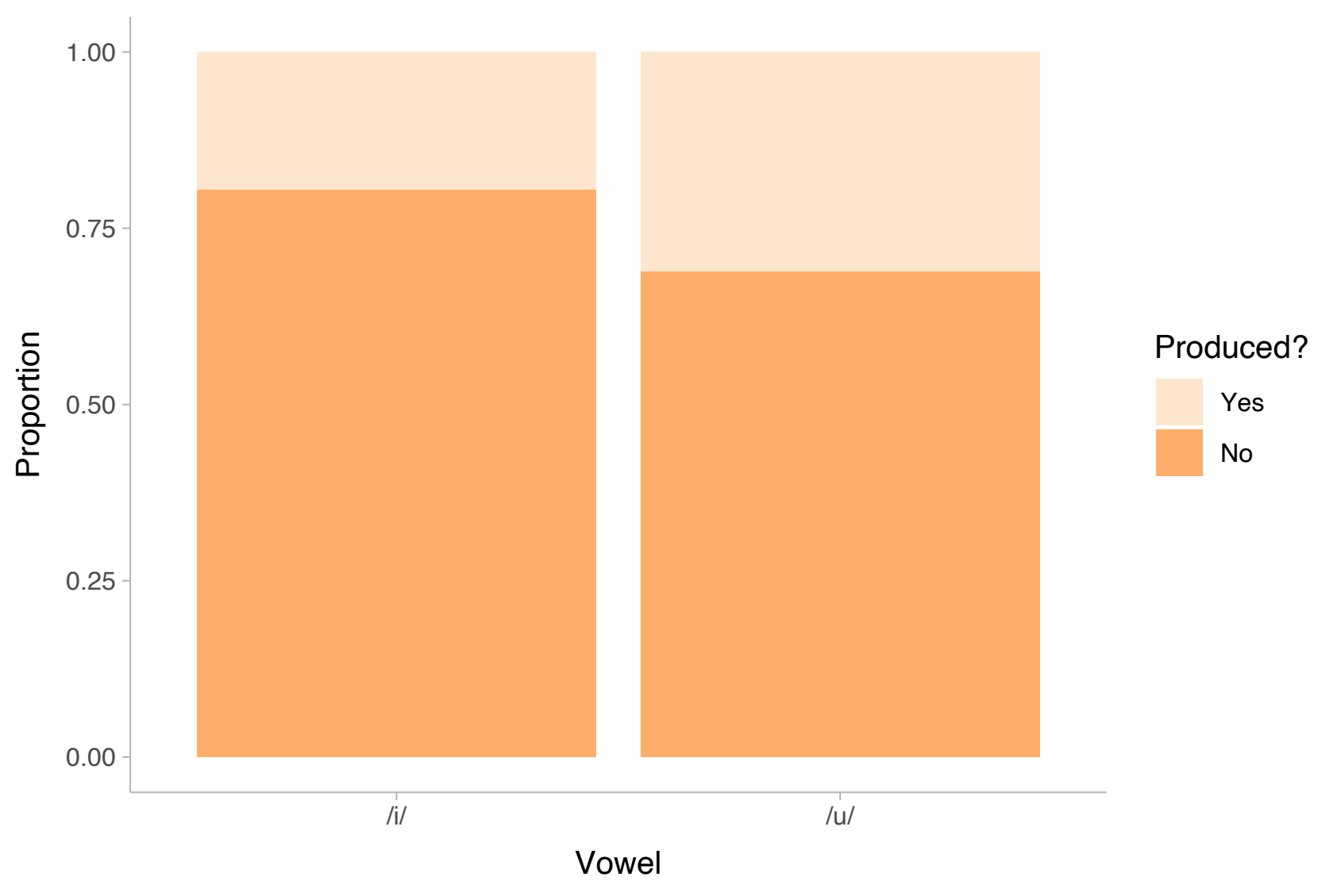
     → vowel produced or deleted?

     conservative decision
3. Word frequency checked in *Corpus Brasileiro* (PUCSP)

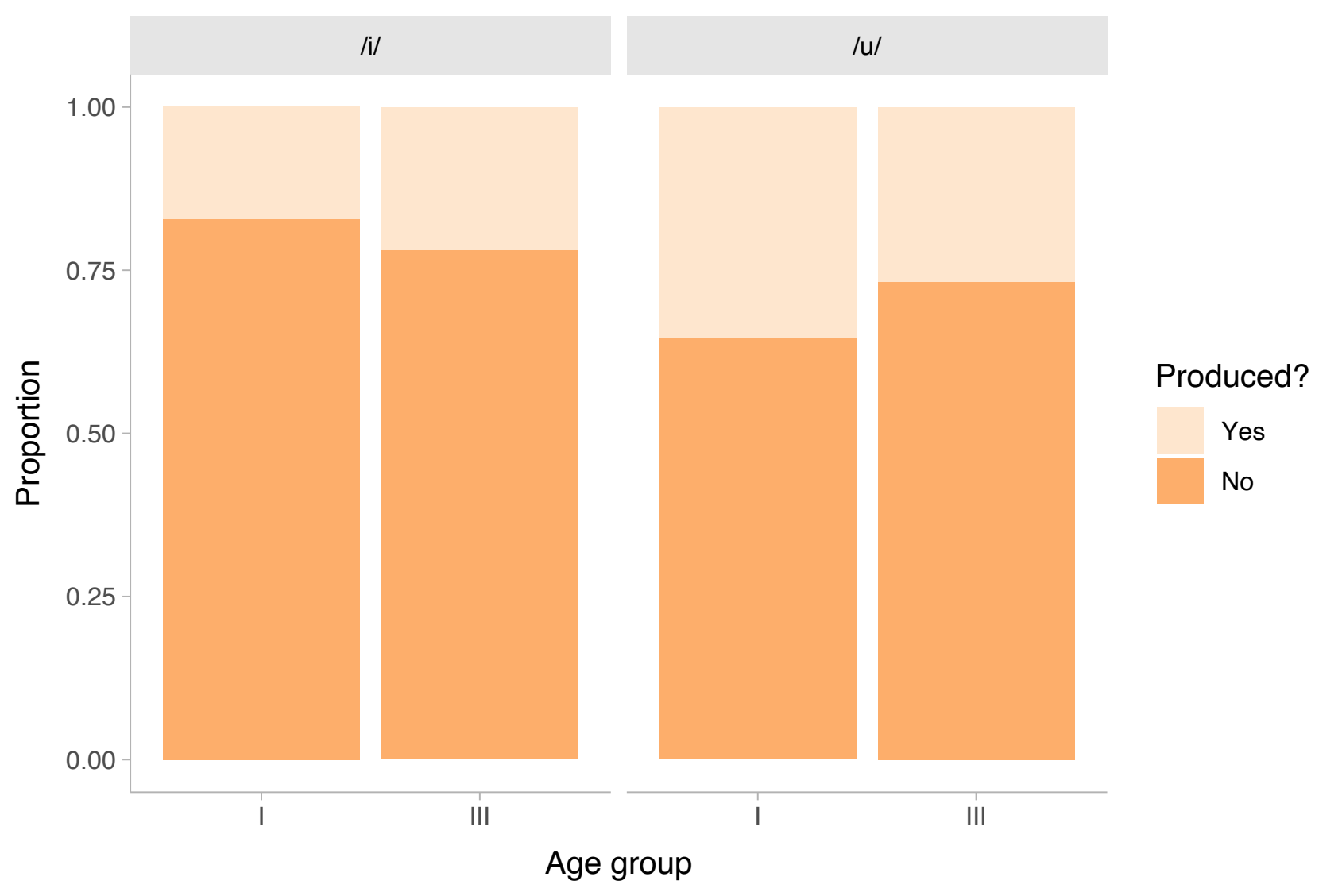
   http://corpusbrasileiro.pucsp.br & https://www.linguateca.pt
4. Bayesian hierarchical model

# Results

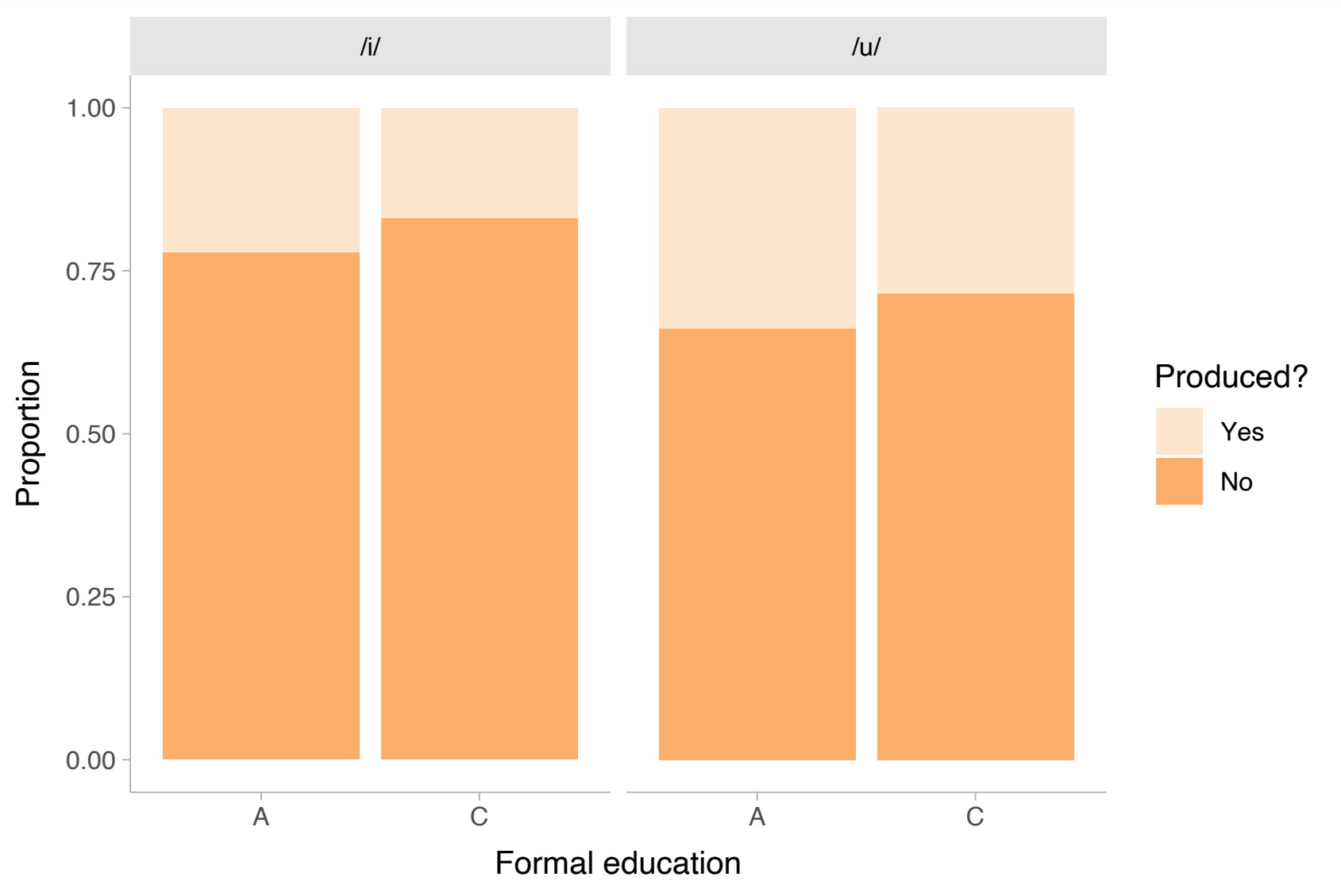
# Descriptive statistics



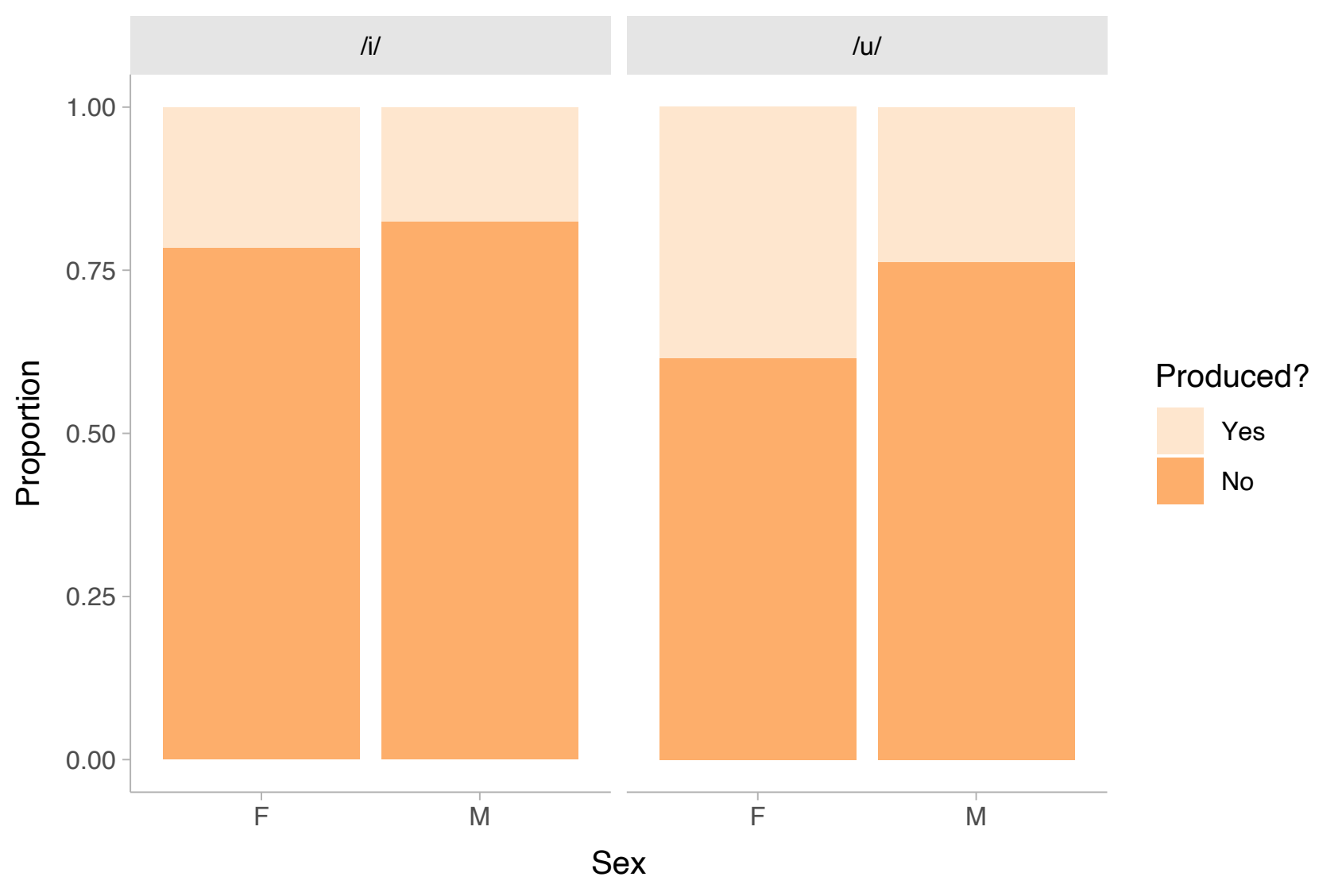
# Descriptive statistics – age



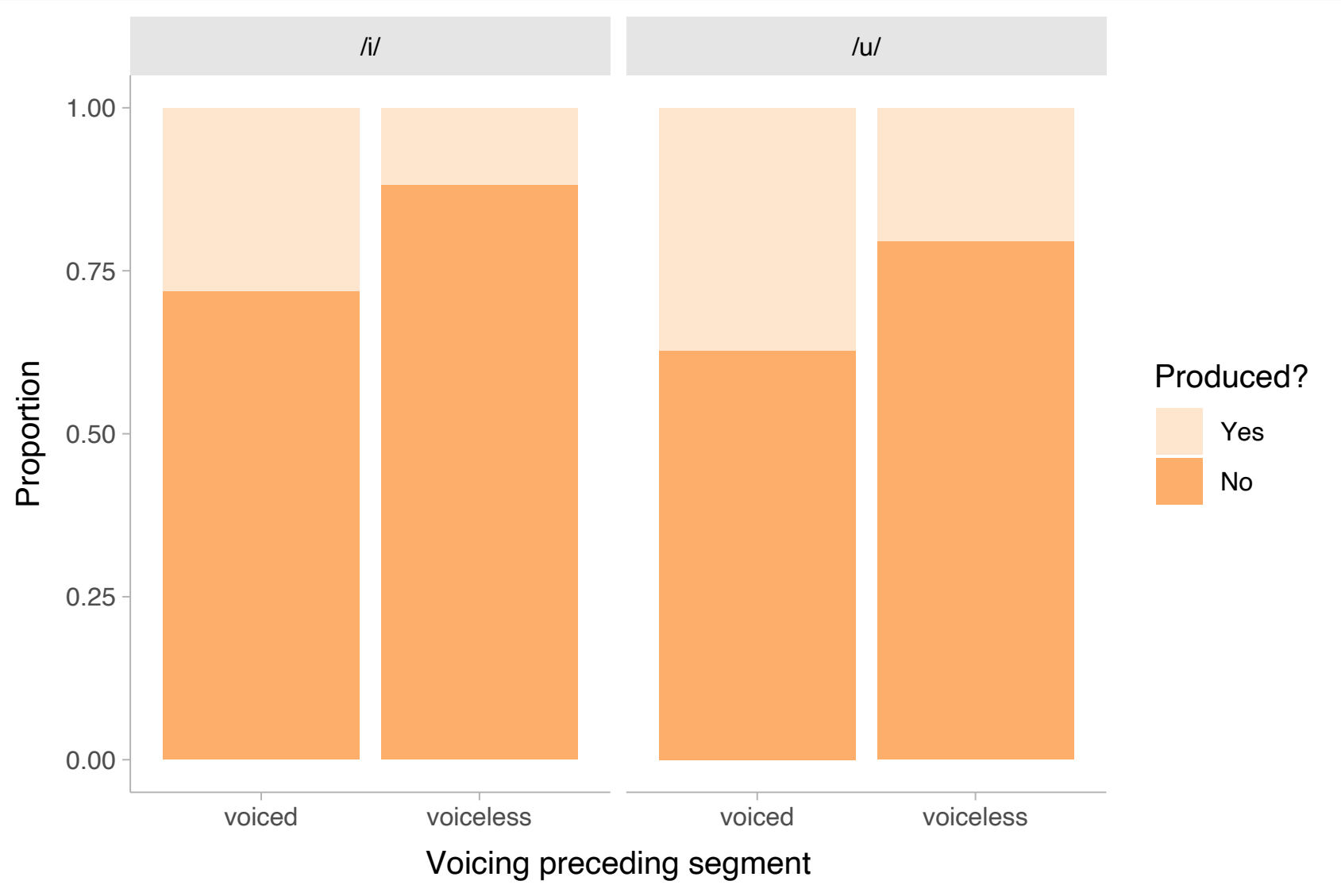
# Descriptive statistics – education



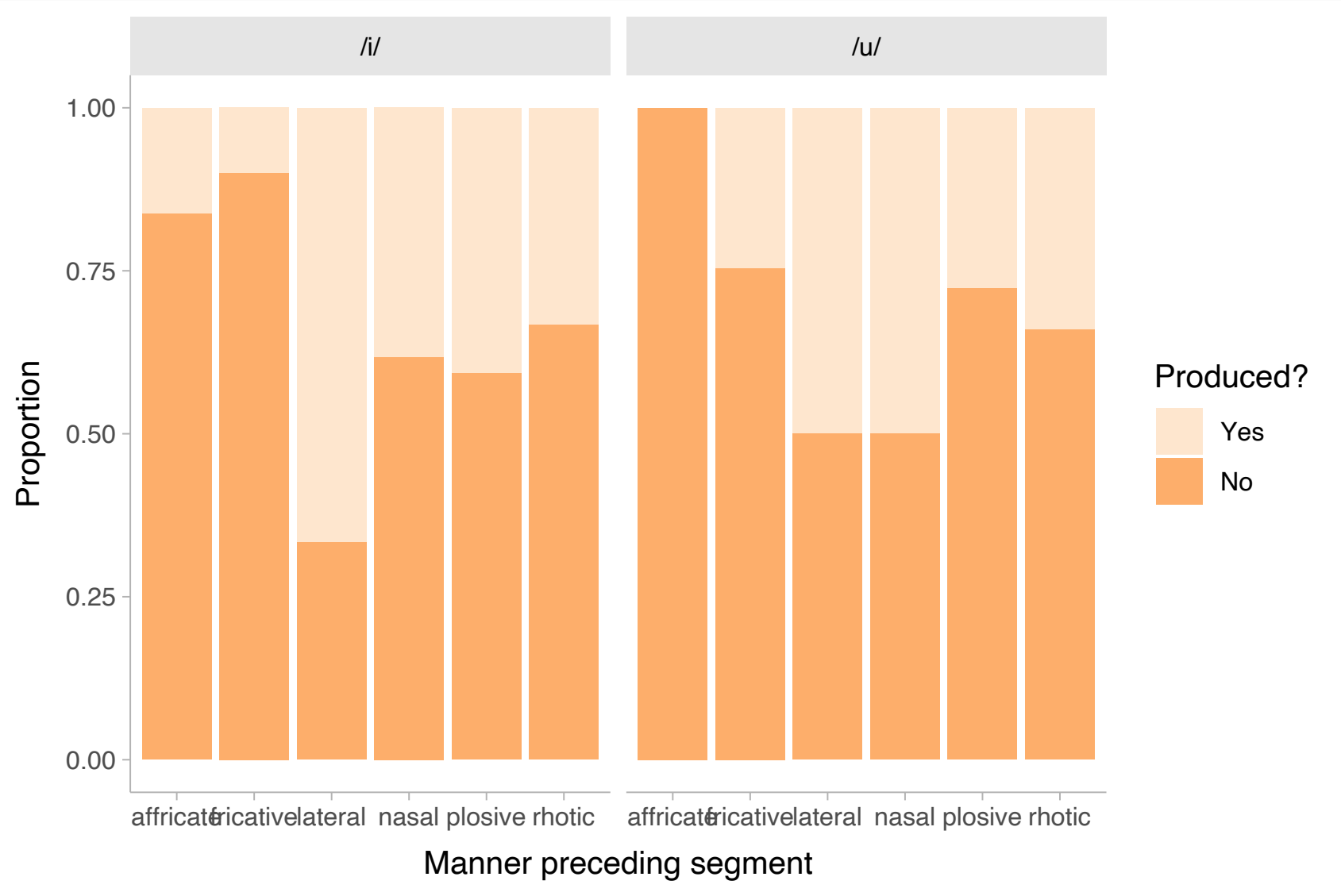
# Descriptive statistics – sex



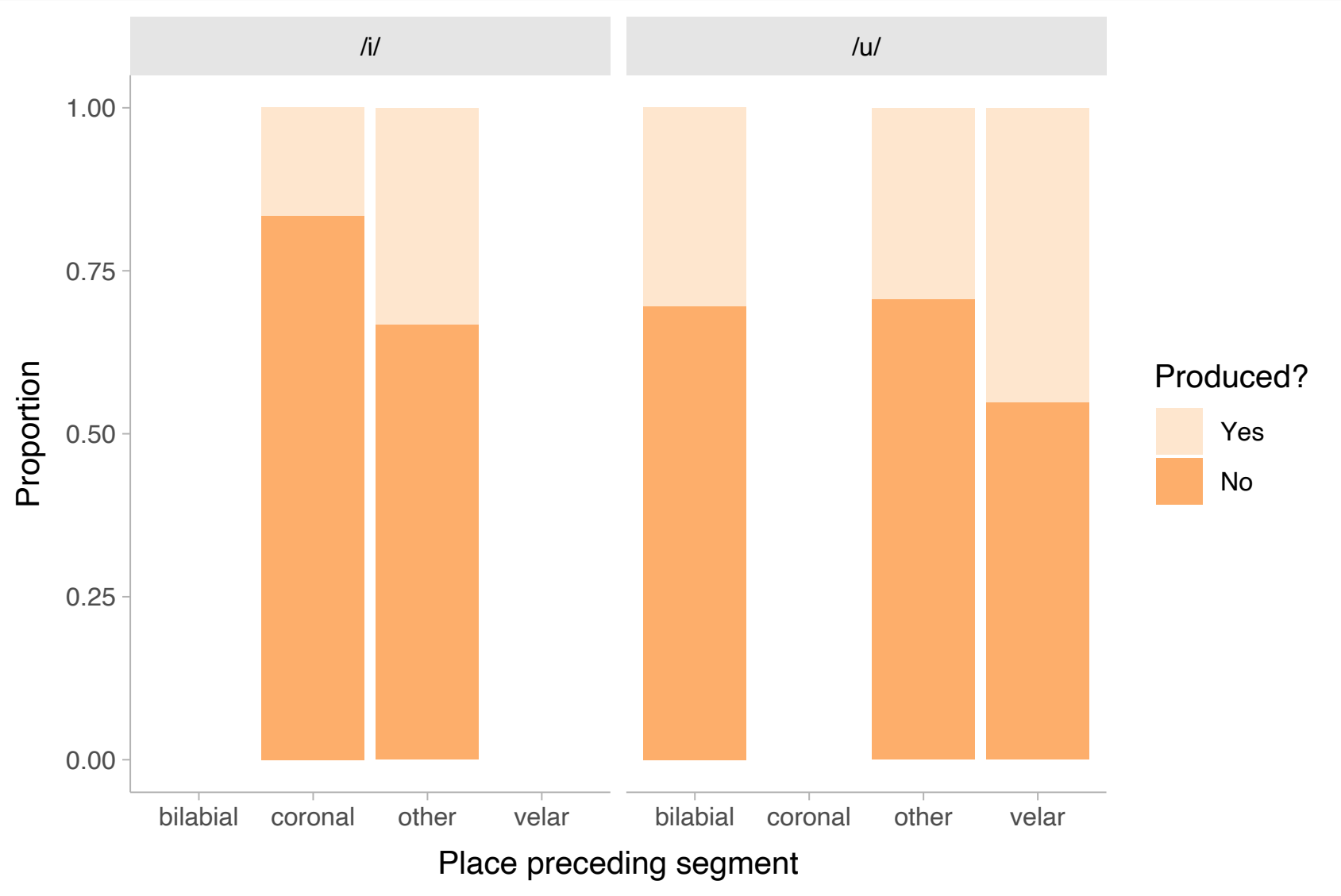
# Descriptive statistics – preceding segment



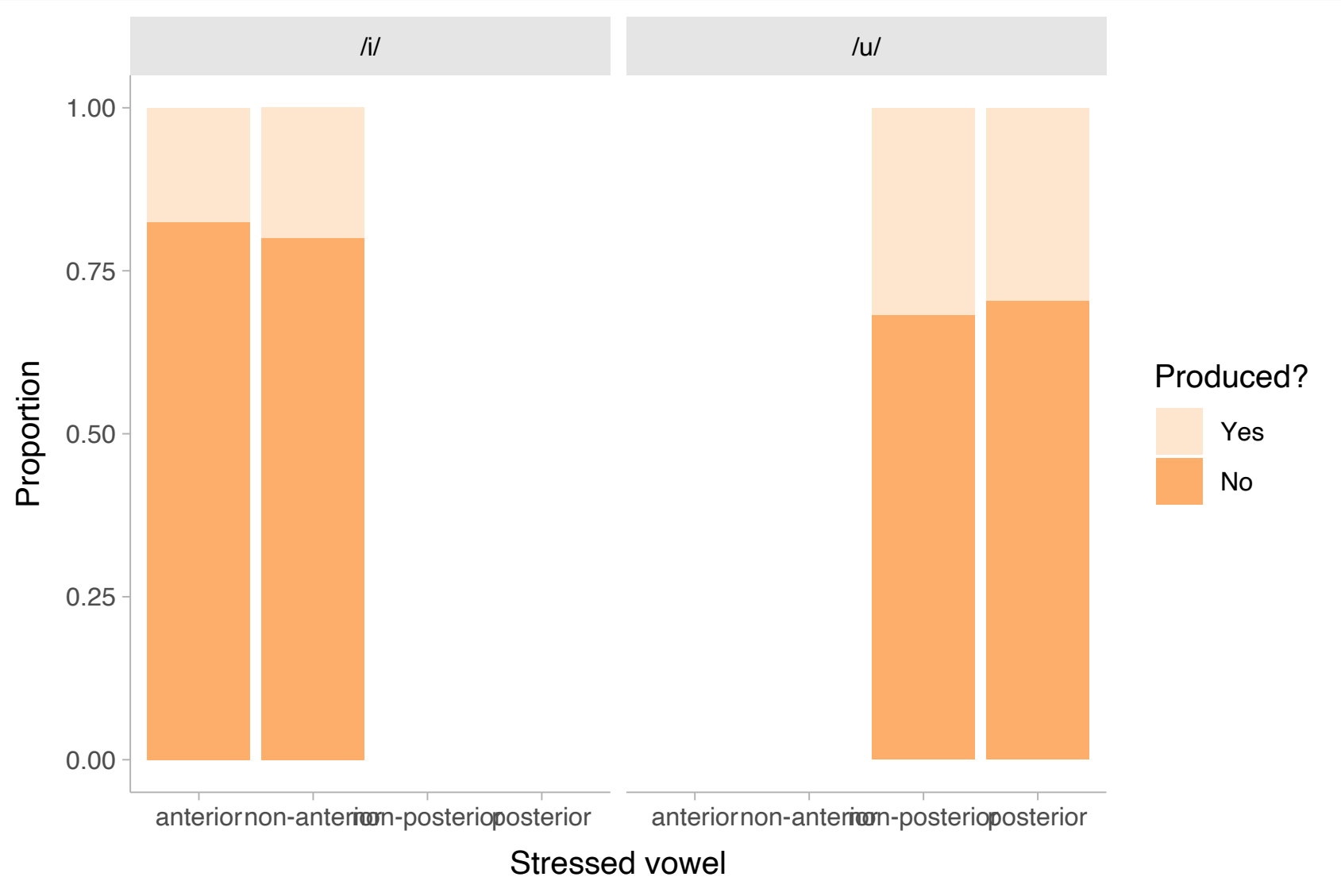
# Descriptive statistics – preceding segment



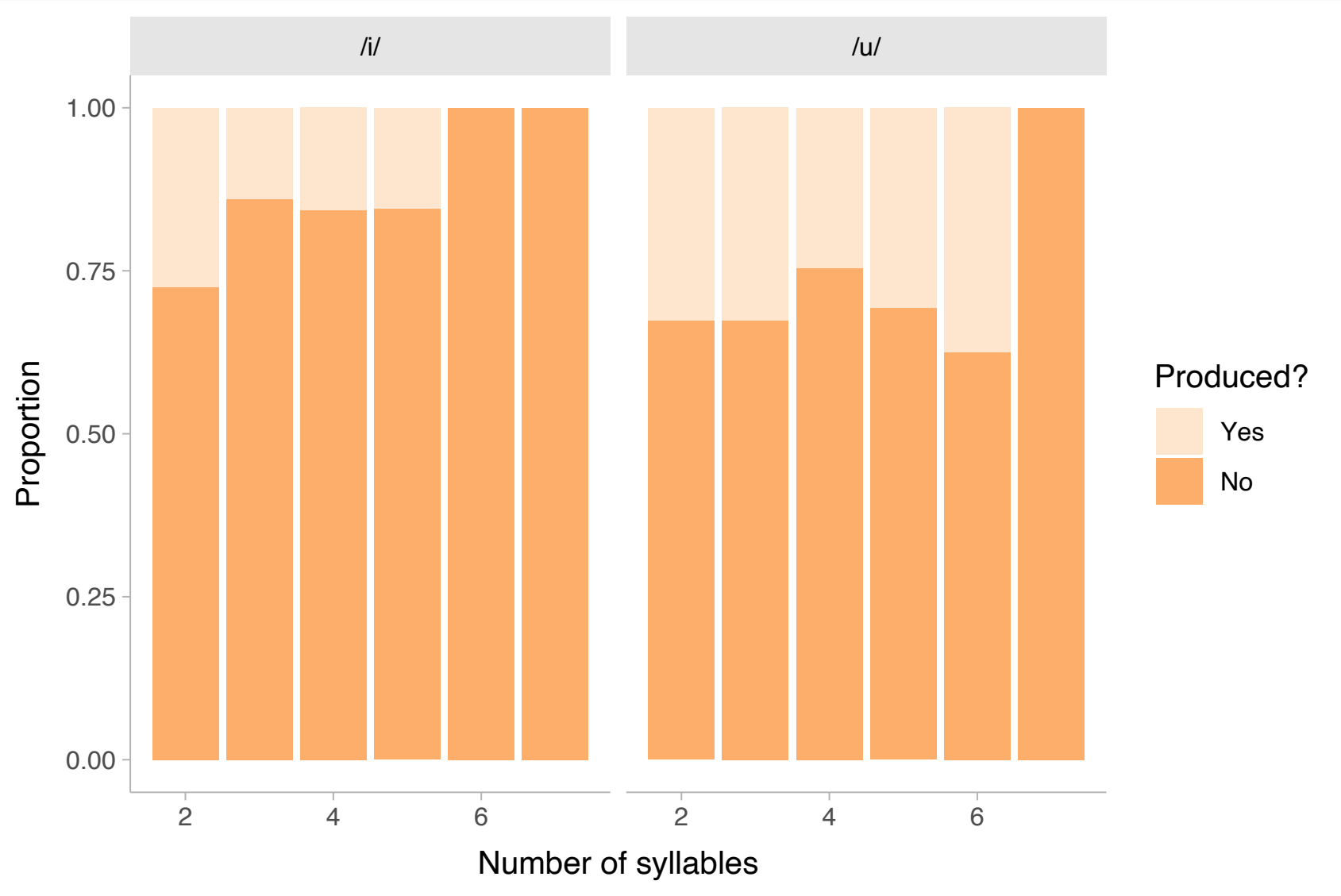
# Descriptive statistics – preceding segment



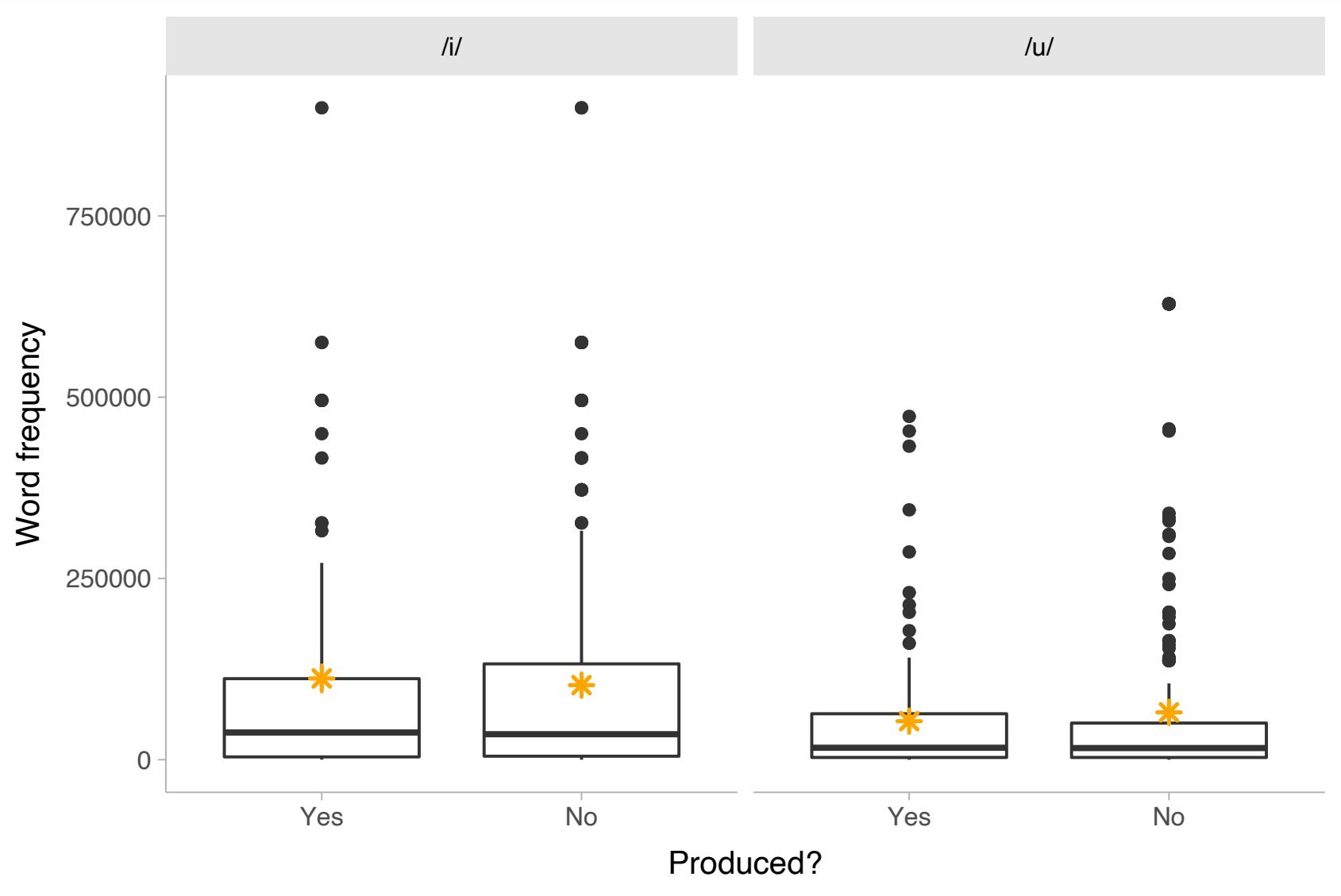
# Descriptive statistics – vowel of stressed syllable



# Descriptive statistics – number of syllables

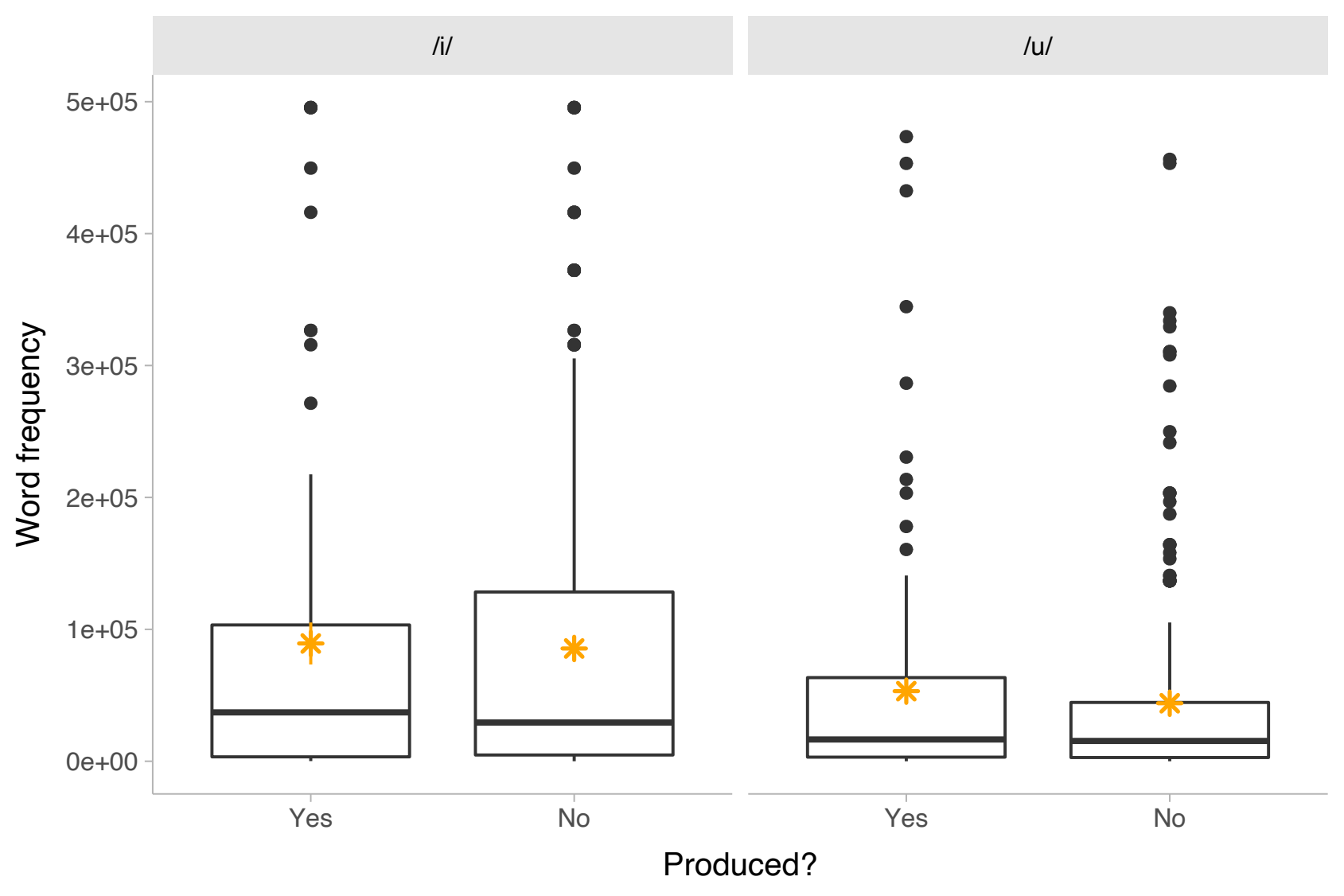


# Descriptive statistics – word frequency



# Descriptive statistics – word frequency

- Removing frequencies above 500k



# Inferential statistics – Bayesian hierarchical model

Model:

- Logistic regression model, flat priors, random intercepts for participants
- Model comparison with `LOO()` (Leave-One-Out Cross-validation)
  - w/o interactions, w/o random effects

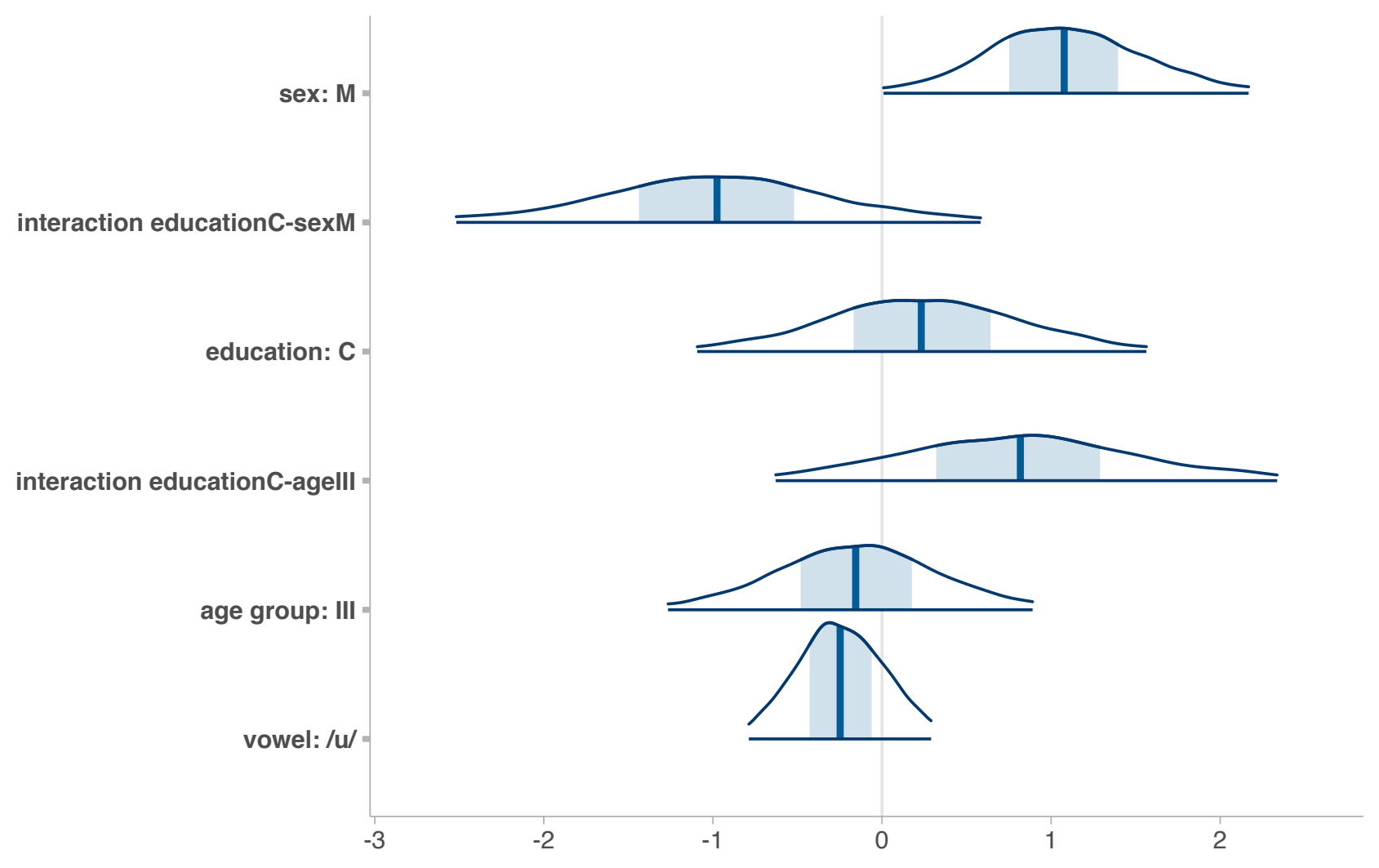
```
bm1 = brm(deletion ~ vowel + age.group + education + sex +
            preceding.manner + preceding.voicing + n.syllables
          + age.group:education + education:sex
          + (1|participant),
          family = bernoulli(link = "logit"))
```

# Inferential statistics – Bayesian hierarchical model

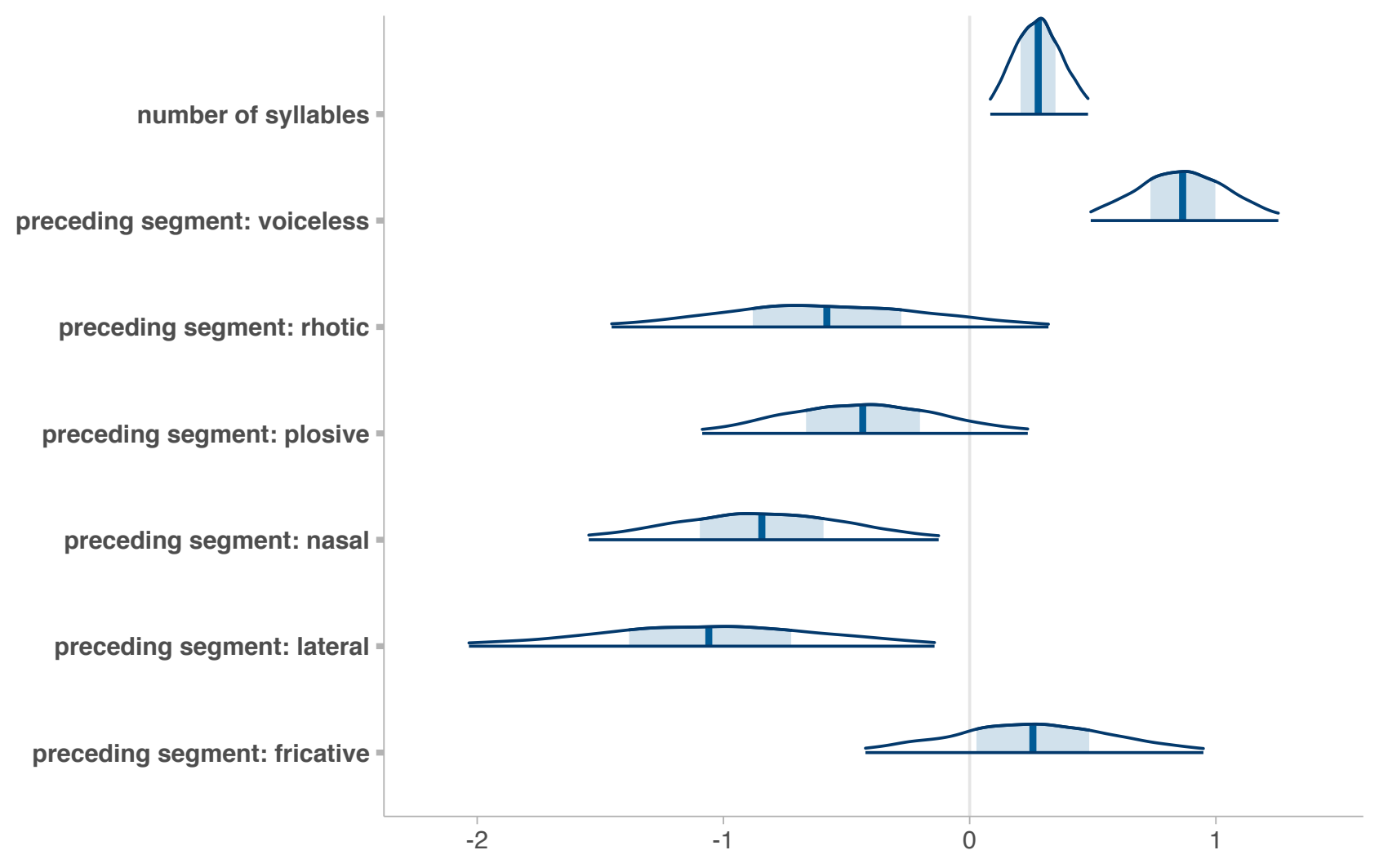
```
Group-Level Effects:
~participant (Number of levels: 16)
               Estimate Est.Error l-95% CI u-95% CI
sd(Intercept)      0.62      0.24     0.26     1.22

Population-Level Effects:
                            Estimate Est.Error l-95% CI u-95% CI
Intercept                      -0.12      0.61    -1.37     1.06
vowel/u/                       -0.25      0.28    -0.79     0.29
age.groupIII                   -0.16      0.54    -1.27     0.89
educationC                      0.23      0.66    -1.09     1.56
sexM                            1.08      0.53     0.01     2.17
prev.mannerfricative            0.26      0.35    -0.42     0.95
prev.mannerlateral             -1.06      0.49    -2.03    -0.14
prev.mannernasal               -0.84      0.37    -1.55    -0.13
prev.mannerplosive             -0.43      0.34    -1.09     0.24
modo.art.anteriorrhotic        -0.58      0.45    -1.45     0.32
prev.voicingvoiceless           0.86      0.20     0.49     1.25
n.syllables                     0.28      0.10     0.08     0.48
age.groupIII:educationC         0.82      0.76    -0.63     2.34
educationC:sexM                -0.98      0.77    -2.52     0.58
```

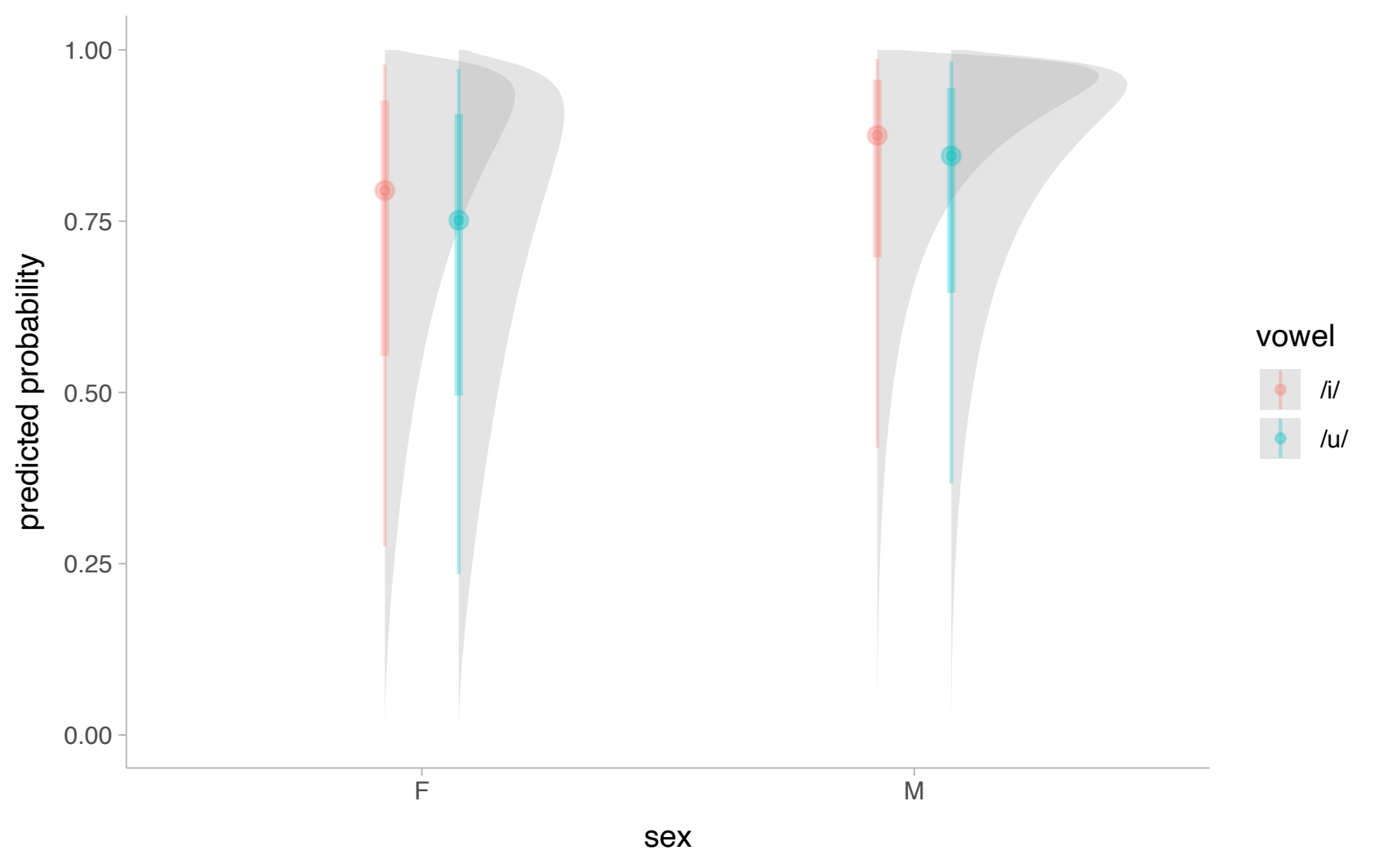
# Inferential statistics – posterior distributions



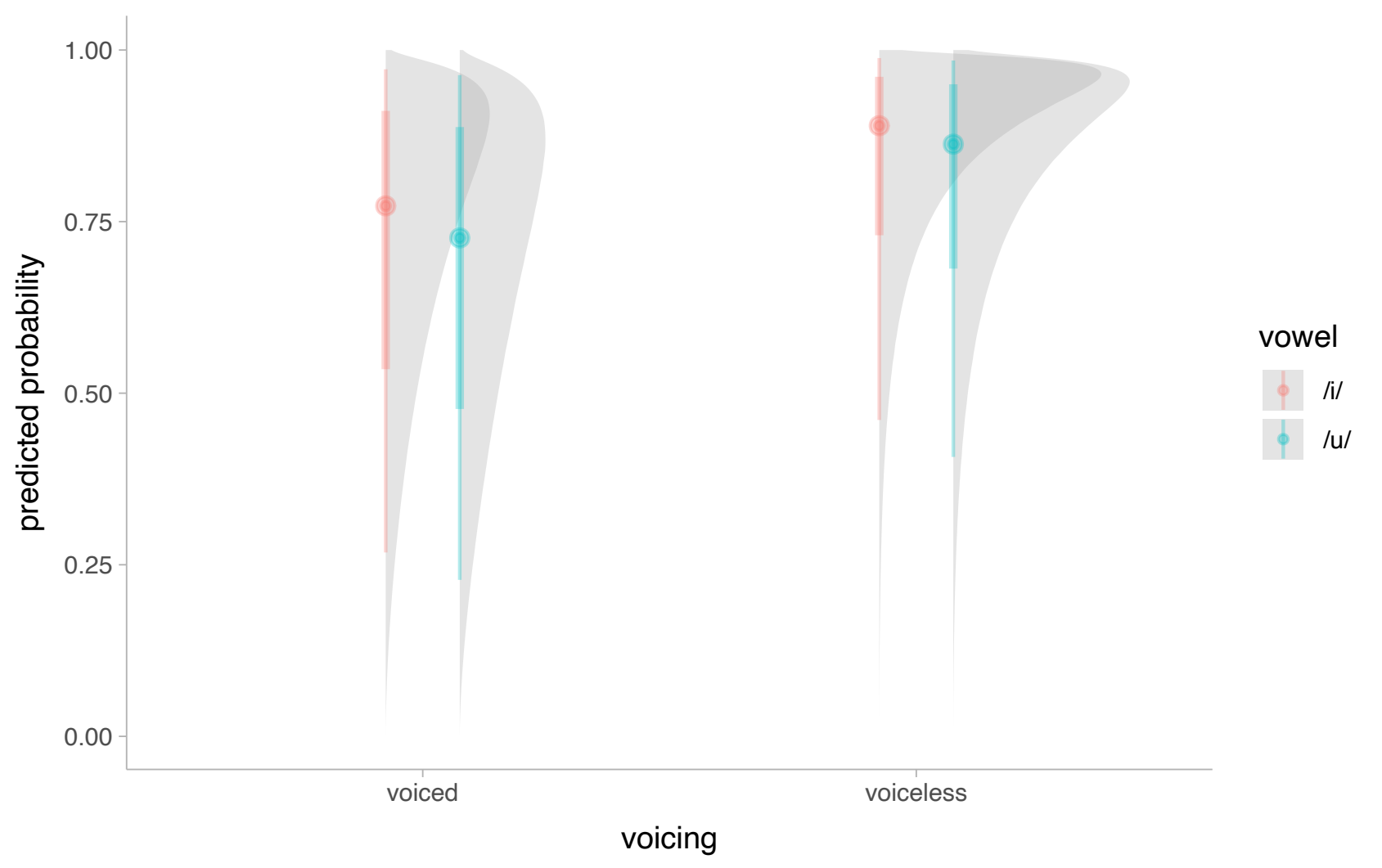
# Inferential statistics – posterior distributions



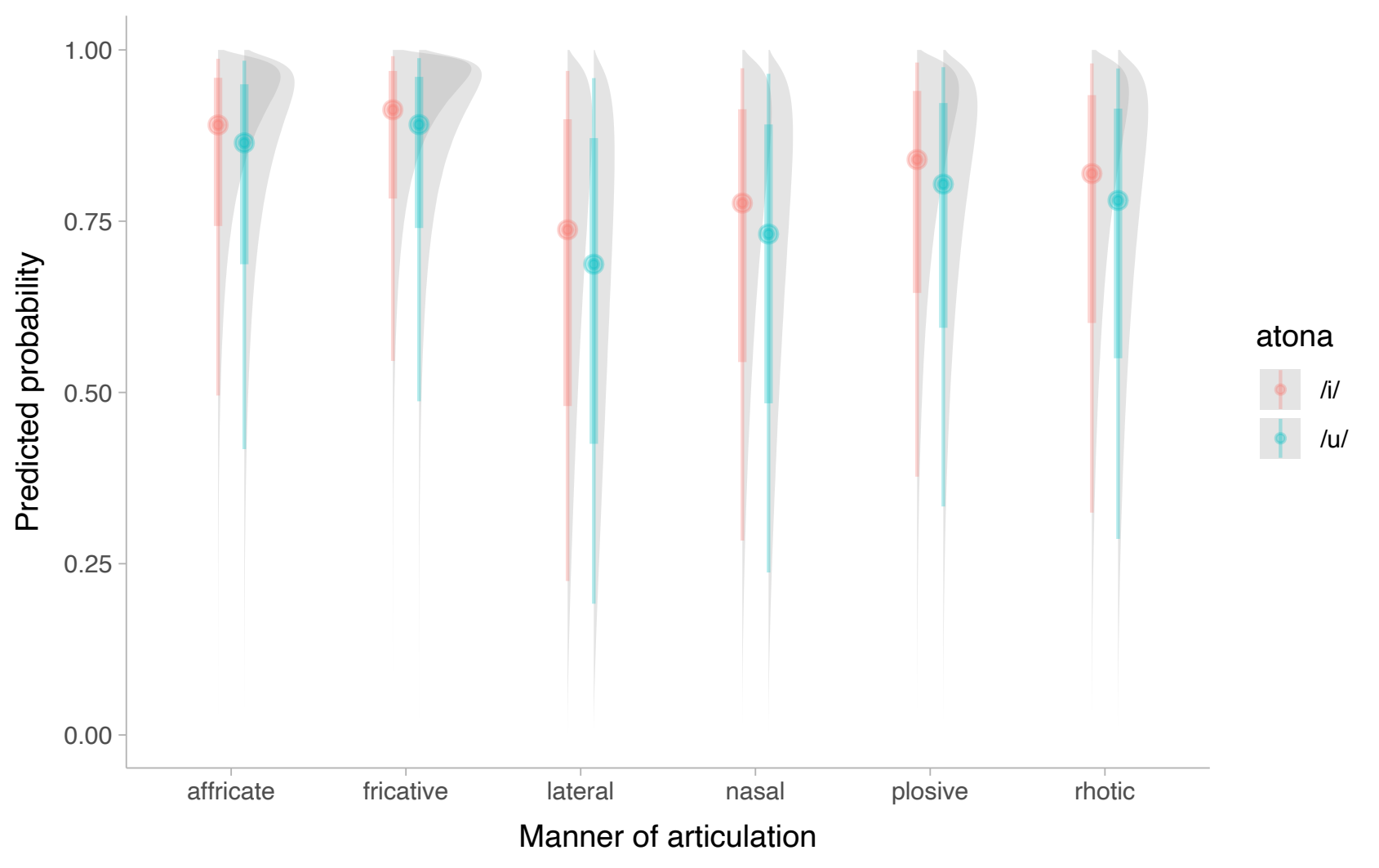
# Inferential statistics – predicted probabilities of reduction



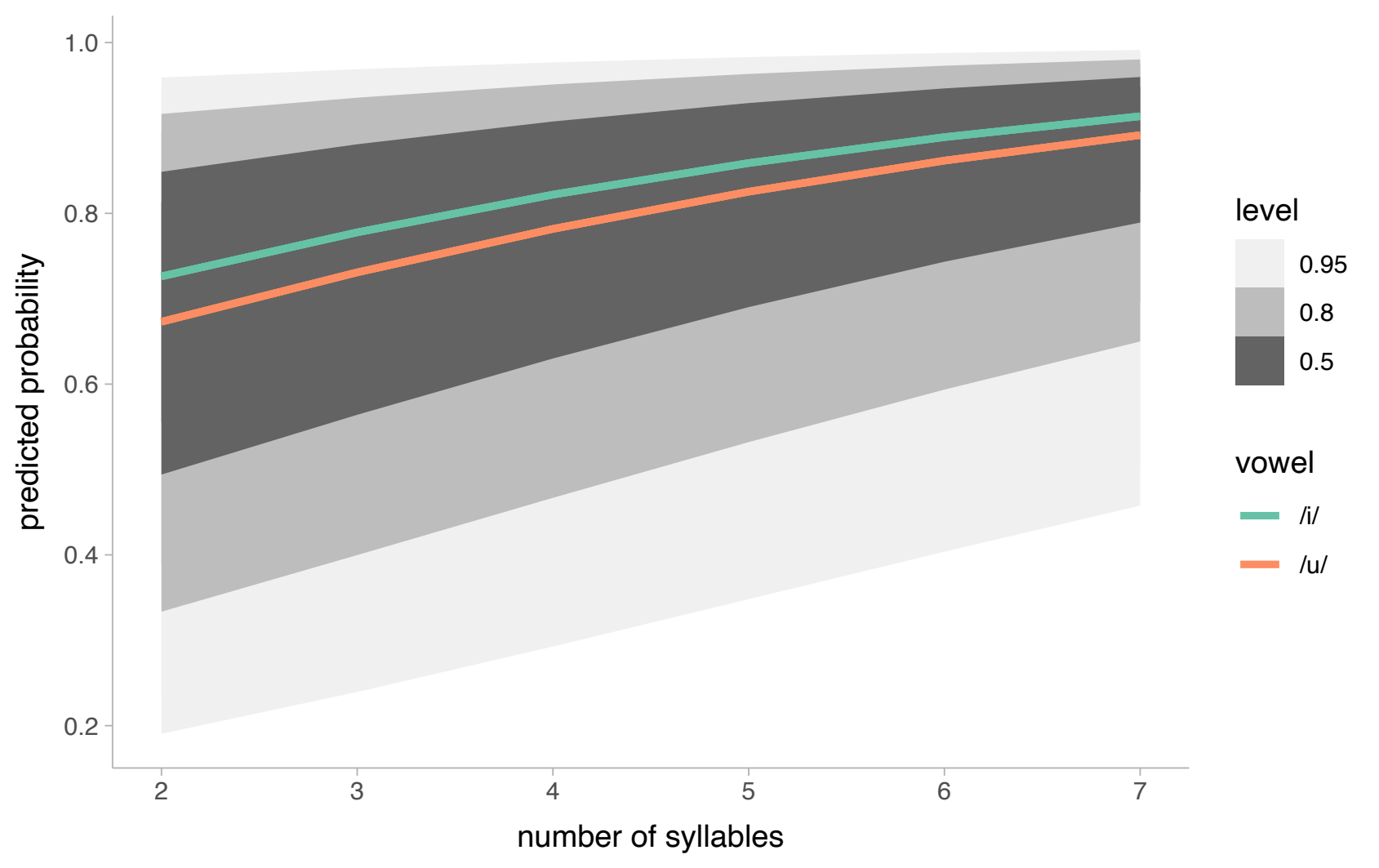
# Inferential statistics – predicted probabilities of reduction



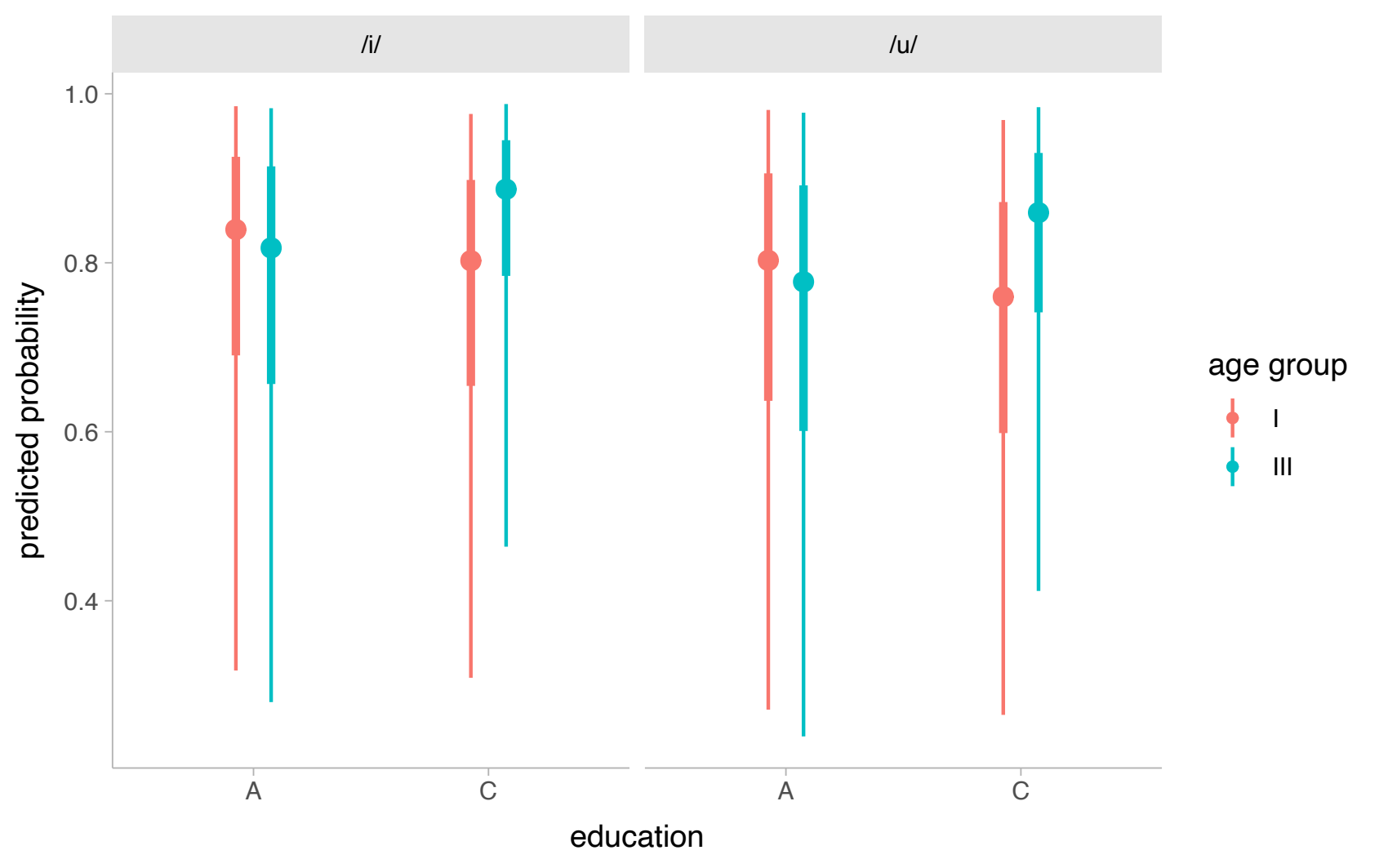
# Inferential statistics – predicted probabilities of reduction



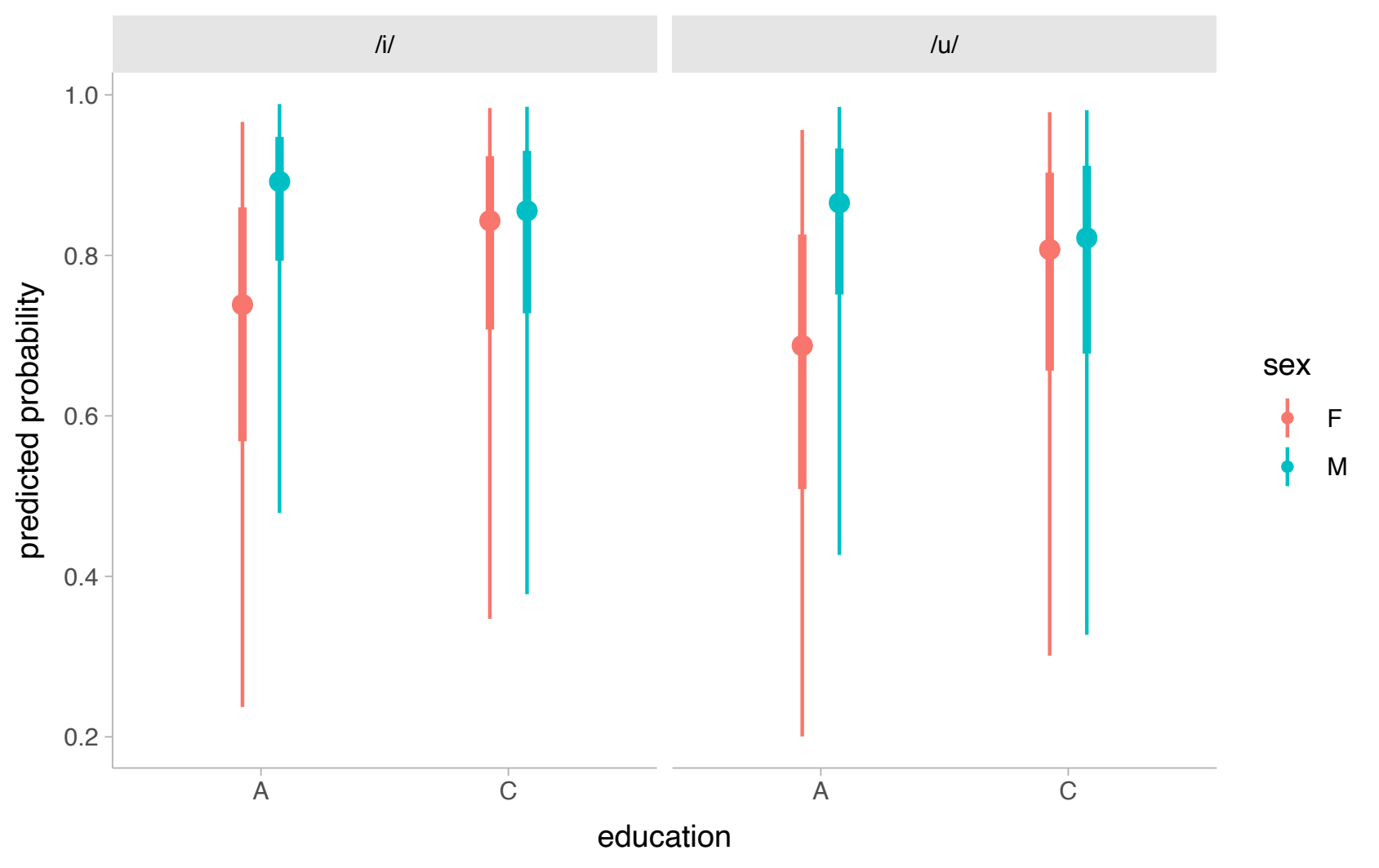
# Inferential statistics – predicted probabilities of reduction



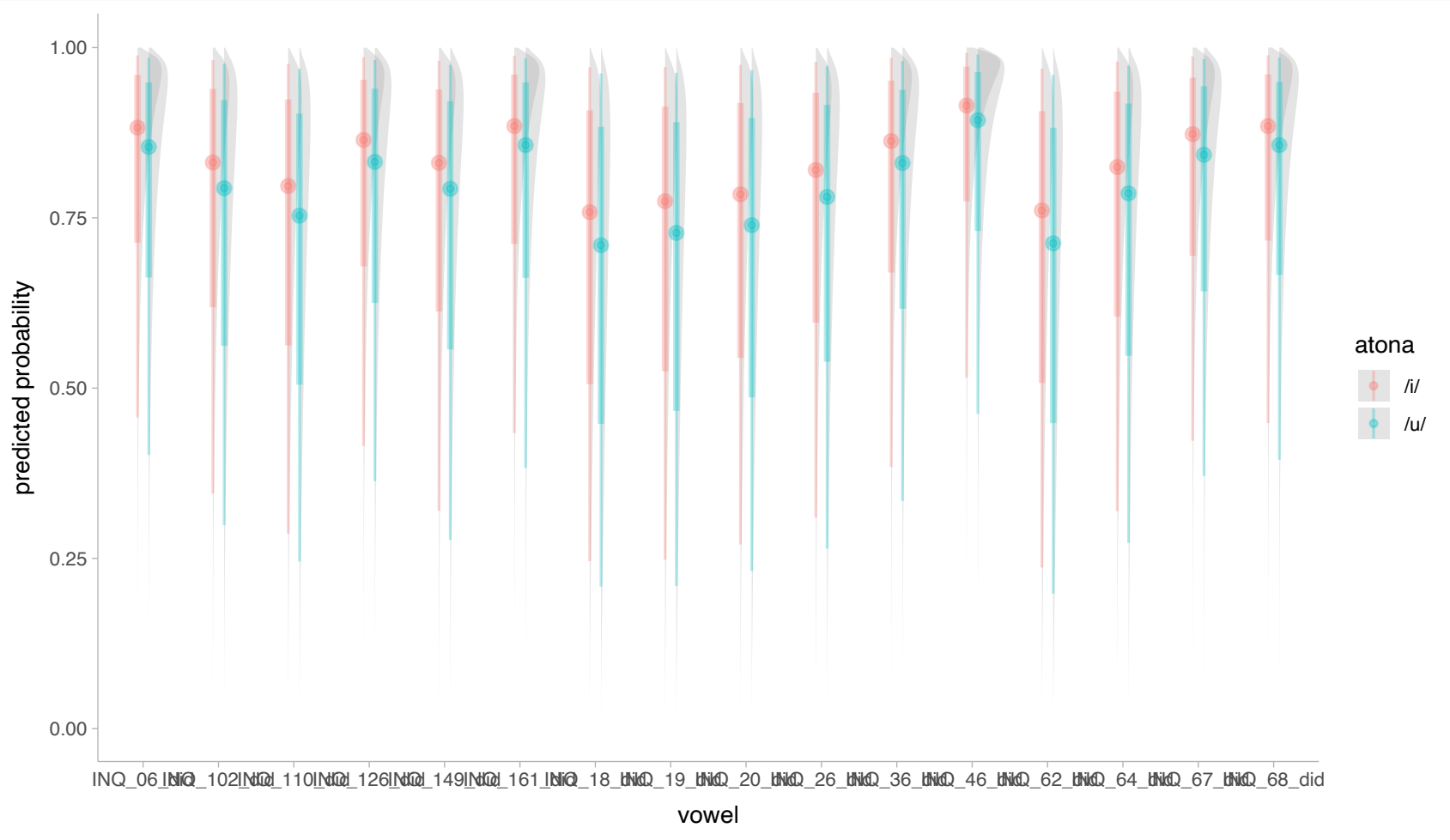
# Inferential statistics – Interactions



# Inferential statistics – Interactions



# Inferential statistics – Random effects



# Discussion

# Discussion

| YES | NO |
|---|---|
| male | education |
| | sex |
| number of syllables | [i] [u] |
| preceding voiceless | preceding place |
| preceding affricates | stressed vowel |
| preceding fricatives | word frequency |

# Discussion

| YES | NO |
|---|---|
| male | education |
| | sex |
| number of syllables | [i] [u] |
| preceding voiceless | preceding place |
| preceding affricates | stressed vowel |
| preceding fricatives | word frequency |

→ nothing relevant with social variables

→ linguistic variables as expected from previous studies

  → word frequency was a surprise (confounding effects?)

  → number of syllables not entirely credible

# Discussion

Contributions:

- Description of the phenomenon in the dialect of Fortalezenses
- Use of informal (semi-)spontaneous speech
- Less deterministic inferential model (soon on github and OSF)

# Final remarks

# Limitations — next steps

- Check speech rate
- Look into relative durational patterns
- Investigate speech of individuals who delete/reduce more/less
- Include other positions of unstressed vowels
- Revisit frequency (include type frequency)
- Laboratory data collection

Questions?

Suggestions?