Analyzing a sound change from multiple sources and angles

Ander Egurtzegi (CNRS-IKER)

Advanced course in Sound Change with a focus on Basque

University of Chicago, 2025/04/21

Studying sound change with historical data

Sibilant mergers

Mergers in historical data

- How to interpret written sources?
- How to know a merger is complete?
- What methods can we use?

Introduction

Studying sound change with historical data

- Basque attestations preceding the 19th century are limited in size.

- "Lubieta's dictionary", dated 1728.
  - Gipuzkoan (Donostia-San Sebastián).

- Two different mergers in progress in this variety in the 18th century.

Krajewska, Zuloaga & Egurtzegi 2022: "Sibilant mergers in 18th-century Basque: A quantitative study". *Phonological Data & Analysis* 4.5, 1–31.

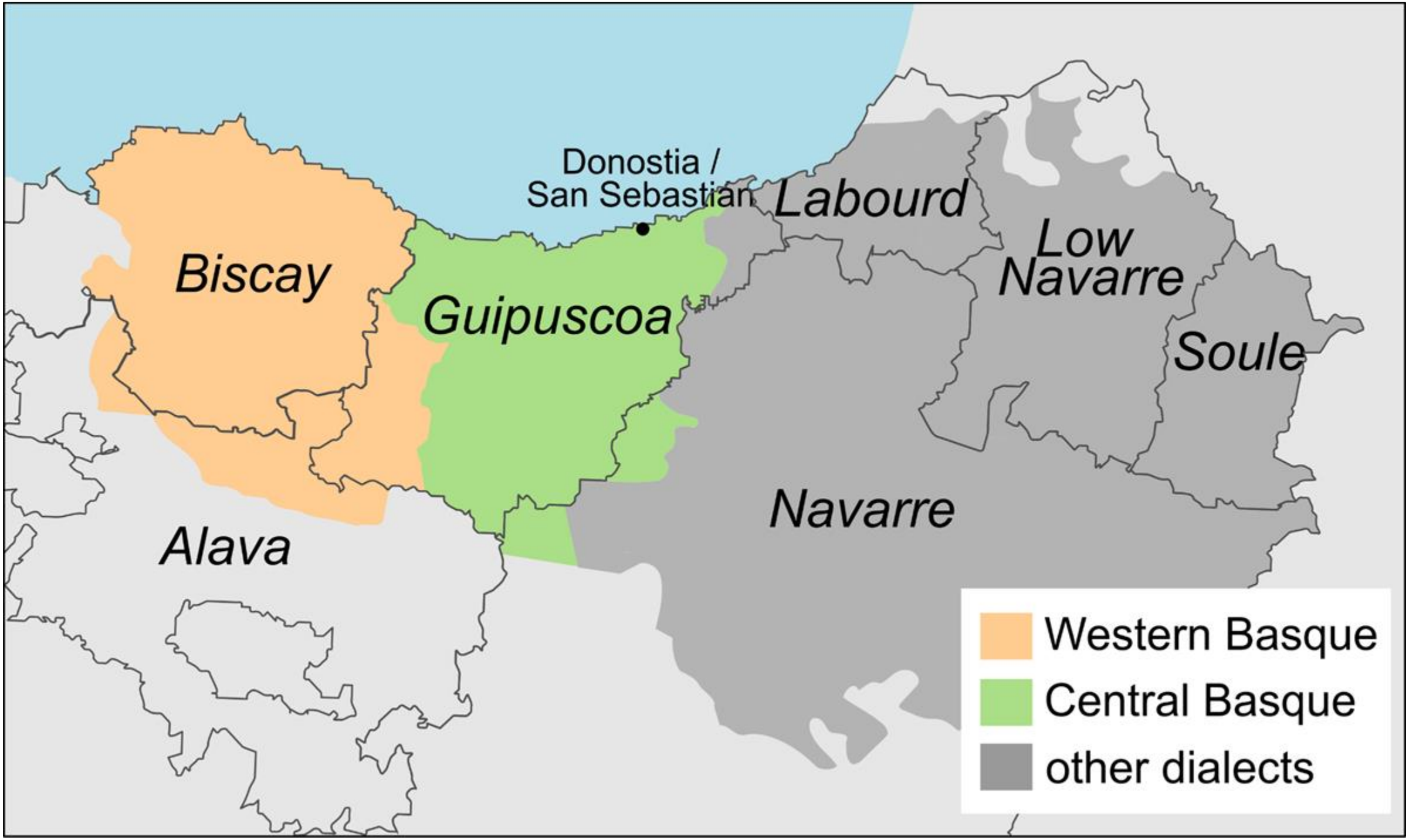
Place of articulation	Manner	IPA	Modern spelling	Old spelling
lamino-alveolar	fricative	/s̻/	z	z, c
	affricate	/ts̻/	tz	(t)z, (t)c
apico-alveolar	fricative	/s̺/	s	s
	affricate	/ts̺/	ts	(t)s
postalveolar	fricative	/ʃ/	x	
	affricate	/tʃ/	tx	

Historical mergers in written sources

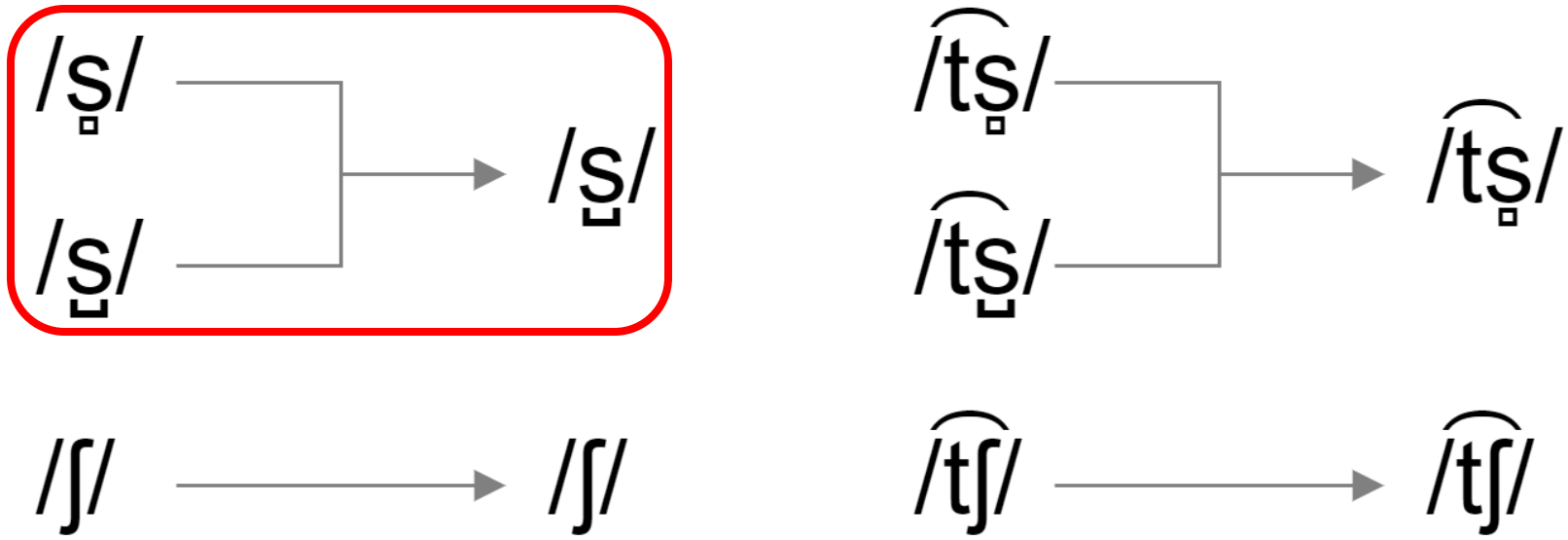
- **The Western merger**:
  - Western varieties, from the 17th century.
  - Most extended pattern (Bizkaia, parts of Gipuzkoa).

- **The Central merger**:
  - Central varieties, 17th-19th century.
  - Always documented alongside the Western merger.
  - The prevailing pattern in the Guipuzkoan coastal areas and Urola (19th c.).

Krajewska, Zuloaga & Egurtzegi 2022: "Sibilant mergers in 18th-century Basque: A quantitative study". *Phonological Data & Analysis* 4.5, 1–31.

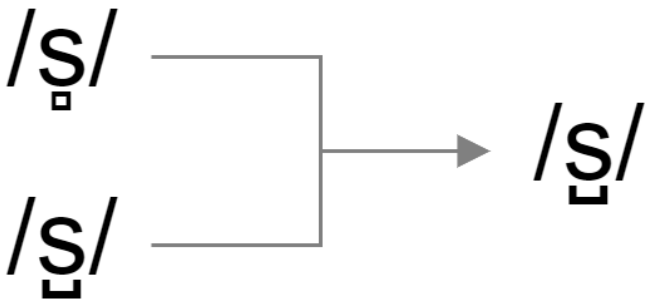

Donostia /
San Sebastian
Labourd
Low
Navarre
Soule
Biscay
Guipuscoa
Navarre
Alava
Western Basque
Central Basque
other dialects

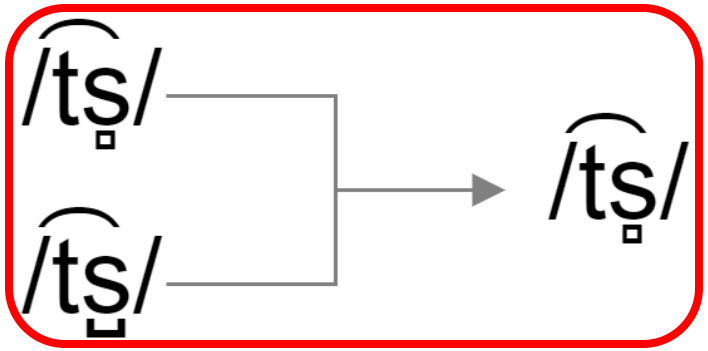
Western Merger



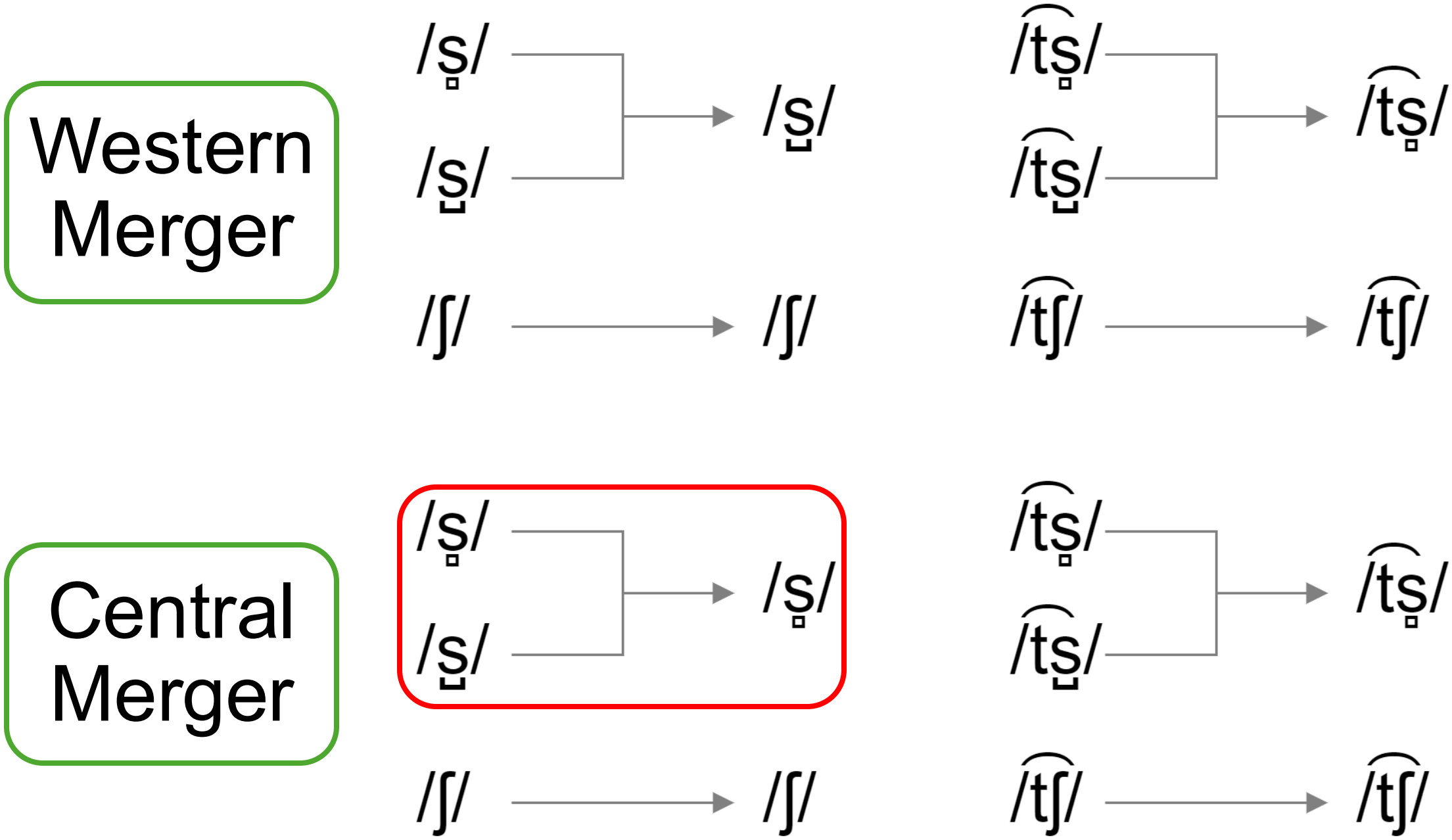
Krajewska, Zuloaga & Egurtzegi 2022: “Sibilant mergers in 18th-century Basque: A quantitative study”. *Phonological Data & Analysis* 4.5, 1–31.

Western Merger

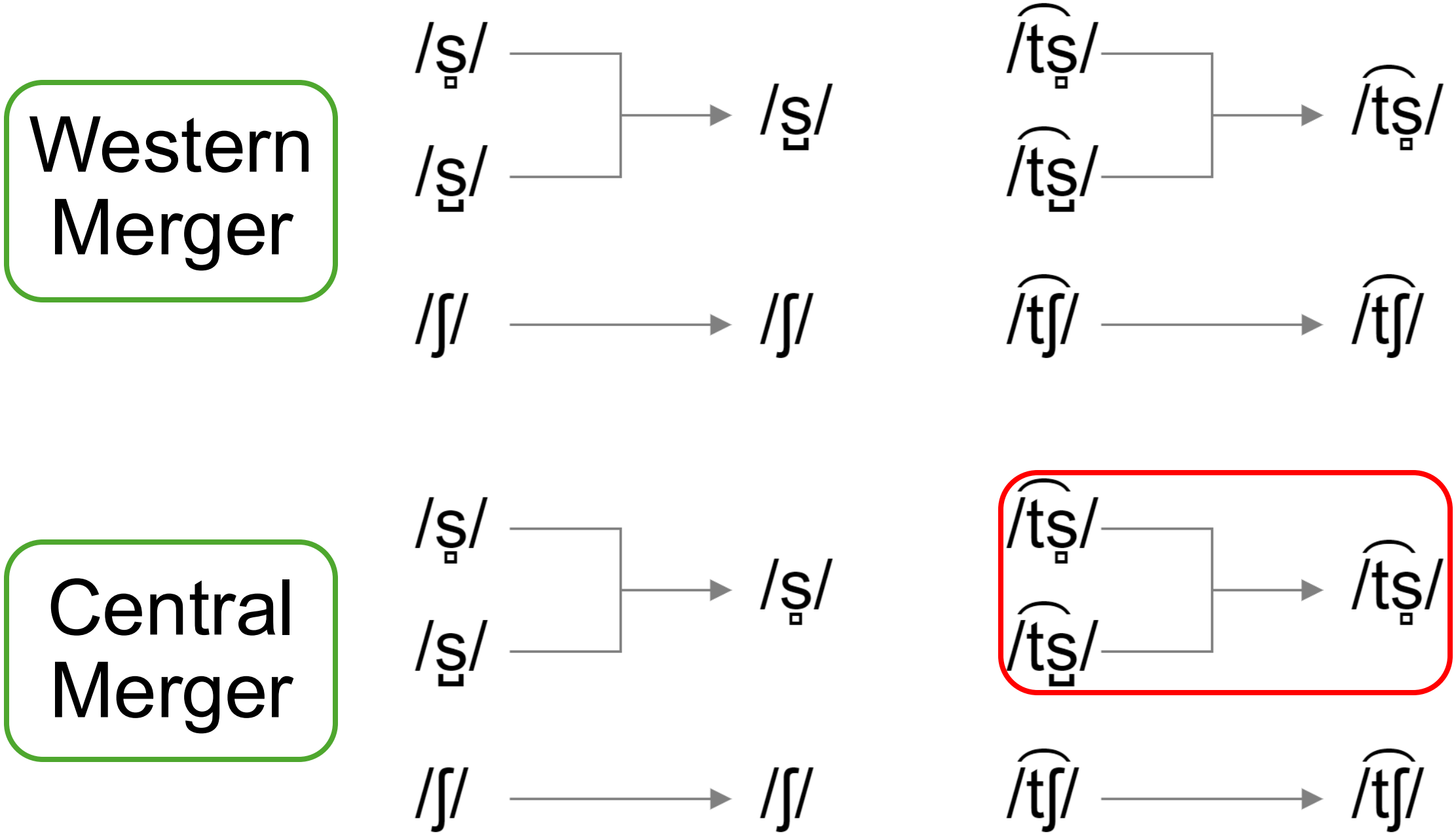




Krajewska, Zuloaga & Egurtzegi 2022: “Sibilant mergers in 18th-century Basque: A quantitative study”. *Phonological Data & Analysis* 4.5, 1–31.



Krajewska, Zuloaga & Egurtzegi 2022: "Sibilant mergers in 18th-century Basque: A quantitative study". *Phonological Data & Analysis* 4.5, 1–31.



Krajewska, Zuloaga & Egurtzegi 2022: "Sibilant mergers in 18th-century Basque: A quantitative study". *Phonological Data & Analysis* 4.5, 1–31.

Mergers of different nature

- Contexts of the Western merger:

_C & _# > #_ & V_V

- The Central merger:
  - It always co-occurs with the Western merger in older sources.
  - The Central merger as hypercorrection (Zuloaga 2020).

Objectives

- To study the changes in the sibilant system in the Central variety of Basque (in the city of San Sebastián) in the 18th century.

- Differences between Western and Central mergers:

  Articulatory vs. external factors (hypercorrection).

- To approach the problem with a quantitative analysis.

Lubieta’s dictionary

Lubieta’s dictionary

- From 1728.
- A bilingual manuscript.
- Variety: Donostia/San Sebastián Basque (Gipuzkoan).
- Original objective: teach/learn Basque.
- Content: dictionary, verses, dialogues, parts of the Catholic Christian doctrine…
- Linguistically: it shows both patterns of sibilant merger.
- Dataset: 6316 tokens of sibilants.

Aquel aurá amado - - - - Arg amatu alzuen =

Nos otros auremos amado - - Gug amatu alzenduen =

Vos otros aureis amado - - - Sueg amatu alzenuten =

Aquellos aurán amado - - - Areg amatu alzuten =

Ama tu - - - - - - - - - Eug amazazu =

Ame aquel - - - - - - - - Arg amaueza =

Amad vos otros - - - - - - Sueg amazute =

Amen aquellos - - - - - - Areg amauezate =

Yo ame - - - - - - - - - Nig amatunuen =

tu amaste - - - - - - - - Sug amatazenuen =

Aquel amó - - - - - - - - Arg amatuzuen =

Nos otros amemos - - - - - Gug amatu guenuen =

Vos otros amasteis - - - - Sueg amatuzenuten =

Aquellos amaron - - - - - - Areg amatuzuten =

Yo amara - - - - - - - - Nig amatuconuque =

tu amaras - - - - - - - - Sug amatuco cenuque =

Aquel amara - - - - - - - Arg amatuco luque =

Nos otros amaramos - - - Gug amatuco guenuque =

Vos otros amarais - - - - Sueg amatuco zenuteque =

Aquellos amaran - - - - - Areg amatuco luteque =

Yo haia amado - - - - - - Nig amatu izandunuala =

tu haias amado - - - - - - Sug amatu izanducenuala =

nos otros haiamos amado - - Gug amatu izandugenuela =

Vos otros haiais amado - - - Sueg amatu izanduzenutela =

Aquellos haian amado - - - Areg amatu izanduzutela =

Amar - - - - - - - - - - Amatu =

hauer amado - - - - - - - Amatu izanduzuen =

hauer de amar - - - - - - Amatu vear =

que amara - - - - - - - - Amatuco duela =

Place of articulation	Manner	IPA	Modern spelling	Conservative spelling	Spelling in Lubieta
lamino-alveolar	fricative	/s̻/	z	z, c	z, c, **s**
	affricate	/ts̻/	tz	z, c	z, c, **s**
apico-alveolar	fricative	/s̺/	s	s	**z**, **c**, s
	affricate	/ts̺/	ts	s	**z**, **c**, s

Questions

- Q1. Was the **opposition between apical and laminal alveolar sibilants** already lost in the 18th century Basque of San Sebastián?

- Q2. What was the role of **phonological context**, on the one hand, and **lexical and frequency effects**, on the other hand, in the observed changes?

- Q3. How can we account for the observed distribution of **spellings**?

Methodology

Spelling

Sound	IPA	Conservative spelling	Innovative spelling
laminal fricative	/s̻/	z, c	s
apical fricative	/s̺/	s	z, c
laminal affricate	/ts̻/	z, c	s
apical affricate	/ts̺/	s	z, c

Generalized linear mixed-effects model

Dependent variable/Response: Spelling

- ConSpell (conservative spelling)
- InnSpell (innovative spelling)

Predictors/Independent variables

Place_etym etymological place of articulation	• Apical • Laminal
Manner	• Fricative • Affricate
Context	• _V (prevocalic) • #_ (word-initial) • _# (word-final) • _C (preconsonantal)
Loan	• Yes (unadapted loanwords) • Old (adapted loanwords) • No
Frequency	• Low • Mid • High
Category	• Content • Grammatical

Interactions

- place_etym – manner
- place_etym – frequency_class
- place_etym – context
- place_etym – loan
- place_etym – category
- frequency_class – category
- category – manner

Random effects

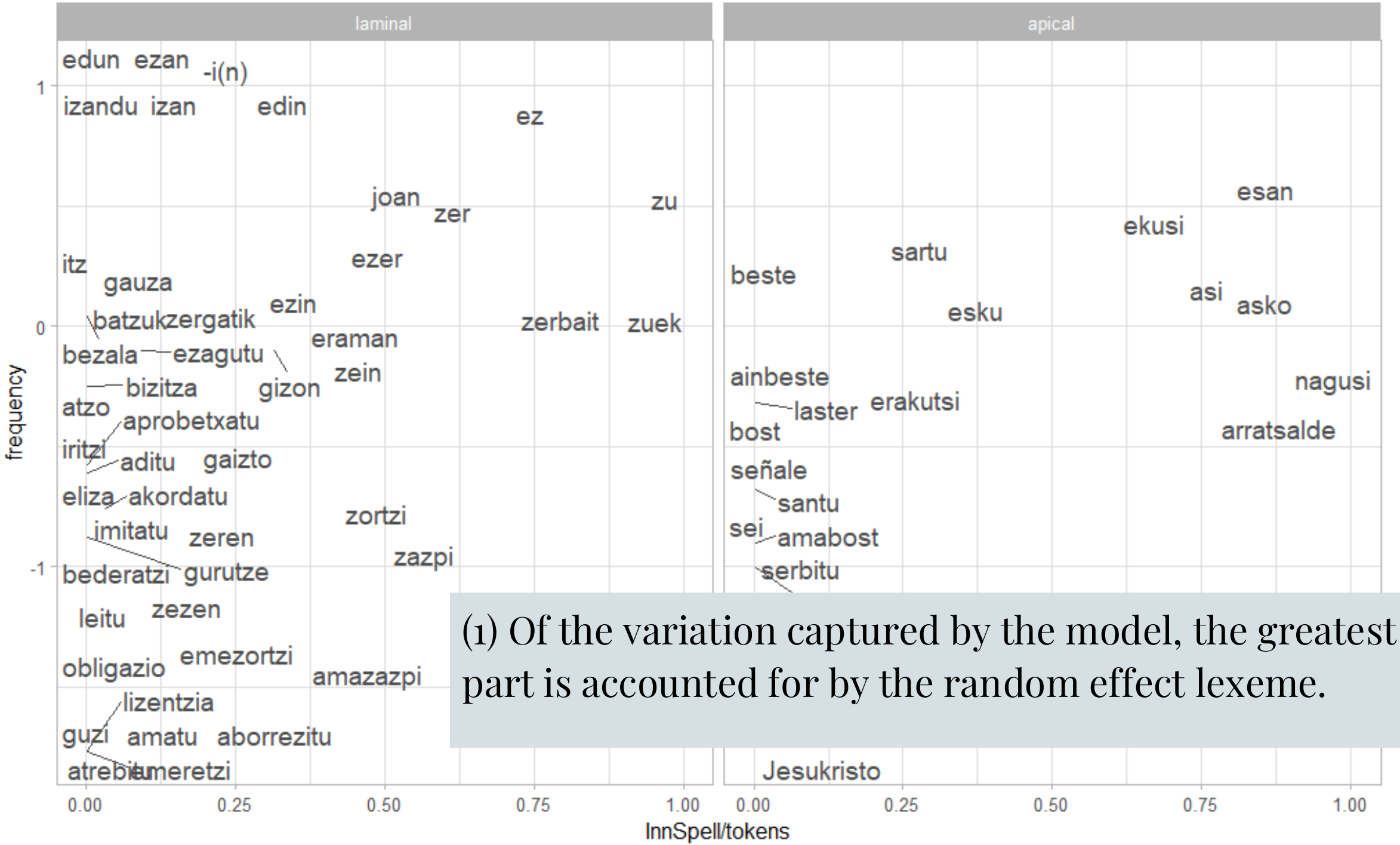
- **Lexeme**: Different lexemes might show different tendencies.

- **Page**: Page in which the example is found. Lubieta's text is a manuscript, and the effect of page is treated as a proxy of (writing) session (or trial in experimental studies).

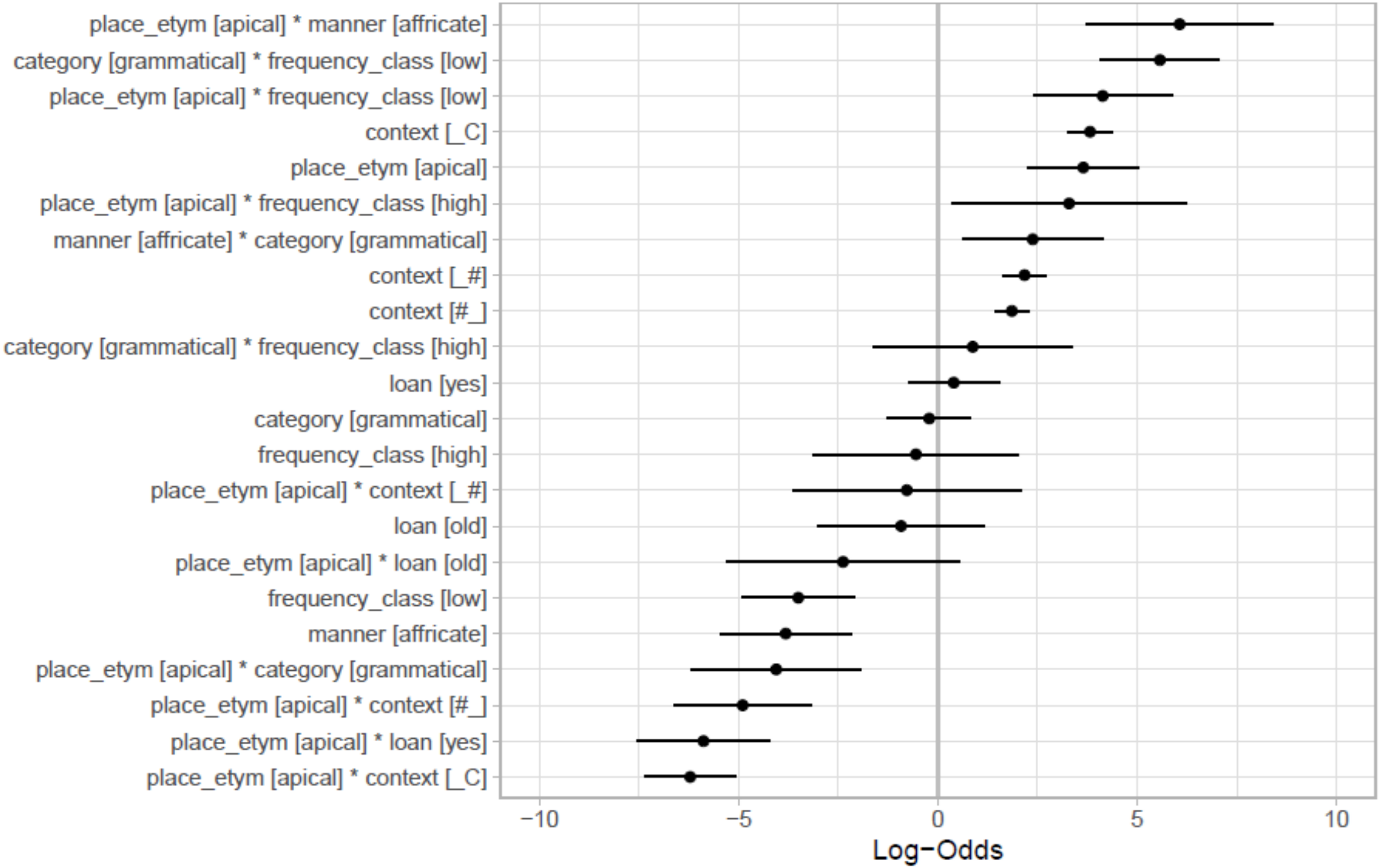
Statistical model

- spelling ~
  place_etym + manner + context + loan + category +
  frequency_class + place_etym:manner +
  frequency_class:place_etym + frequency_class:category +
  context:place_etym + loan:place_etym + category:place_etym +
  category:manner + (1 | lexeme) + (1 | page)

Results

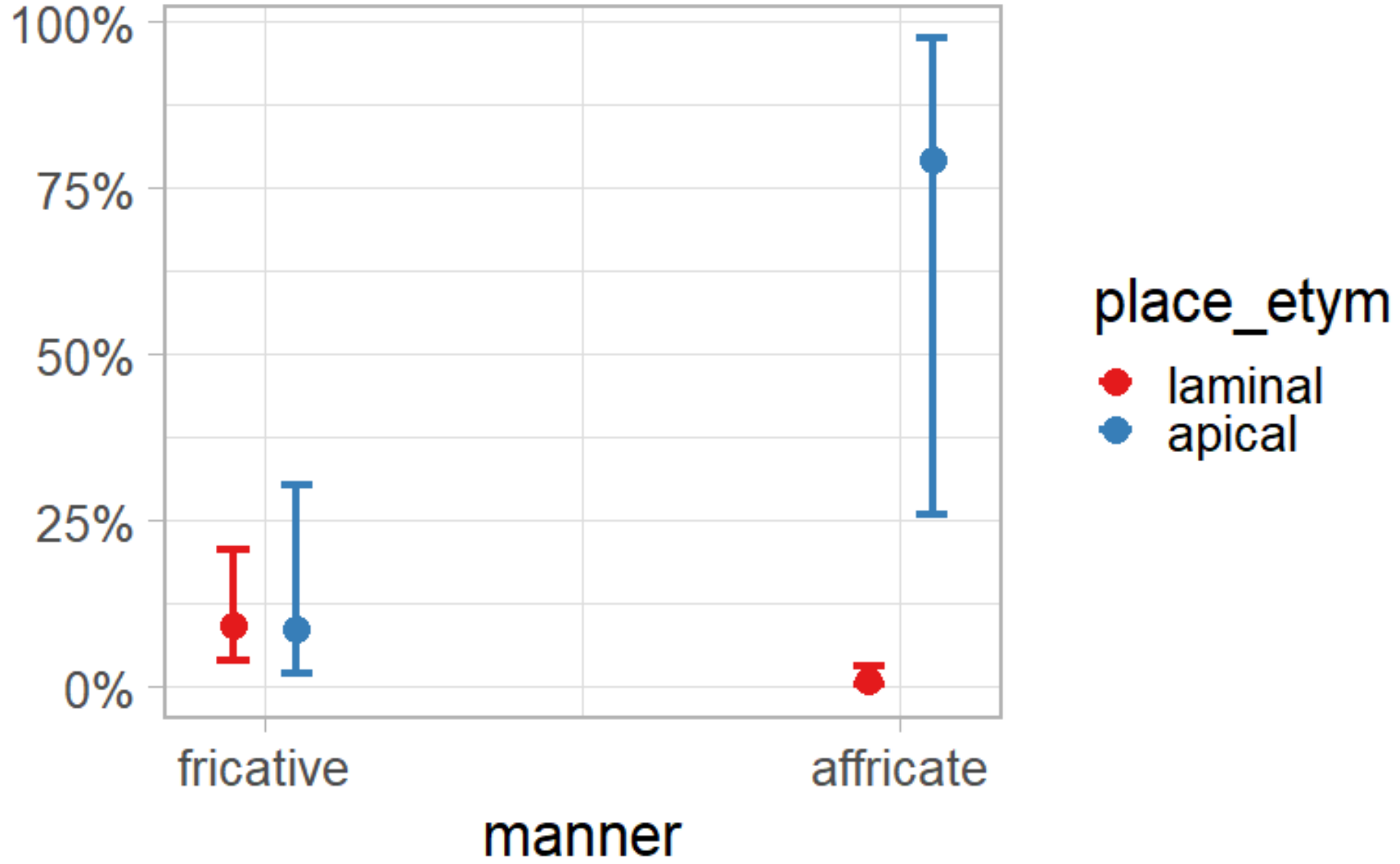


(1) Of the variation captured by the model, the greatest part is accounted for by the random effect lexeme.


place_etym [apical] * manner [affricate]
category [grammatical] * frequency_class [low]
place_etym [apical] * frequency_class [low]
context [_C]
place_etym [apical]
place_etym [apical] * frequency_class [high]
manner [affricate] * category [grammatical]
context [_#]
context [#_]
category [grammatical] * frequency_class [high]
loan [yes]
category [grammatical]
frequency_class [high]
place_etym [apical] * context [_#]
loan [old]
place_etym [apical] * loan [old]
frequency_class [low]
manner [affricate]
place_etym [apical] * category [grammatical]
place_etym [apical] * context [#_]
place_etym [apical] * loan [yes]
place_etym [apical] * context [_C]
−10
−5
0
5
10
Log−Odds

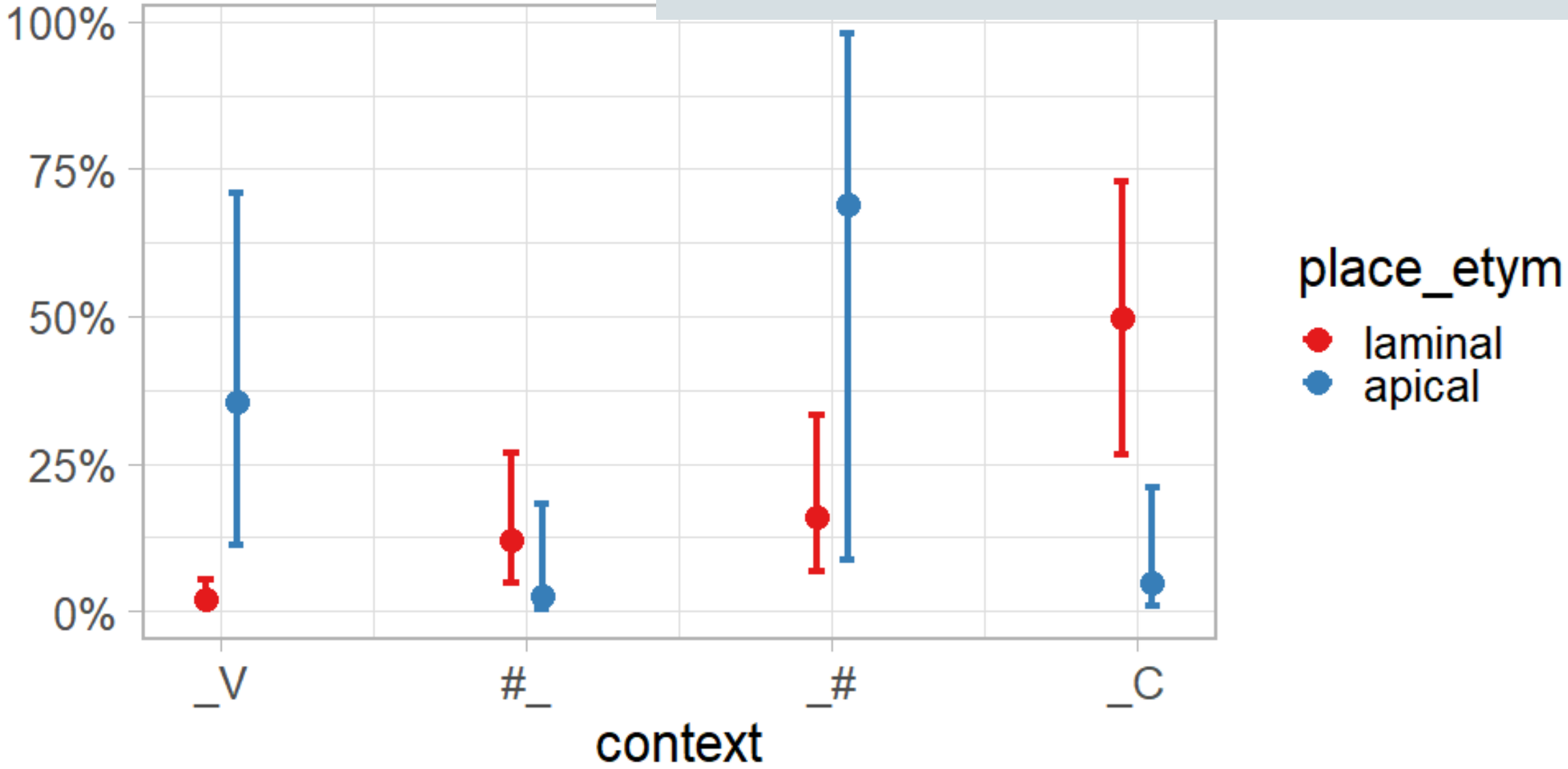
Manner

(2) In affricates the merger is more advanced than in fricatives.

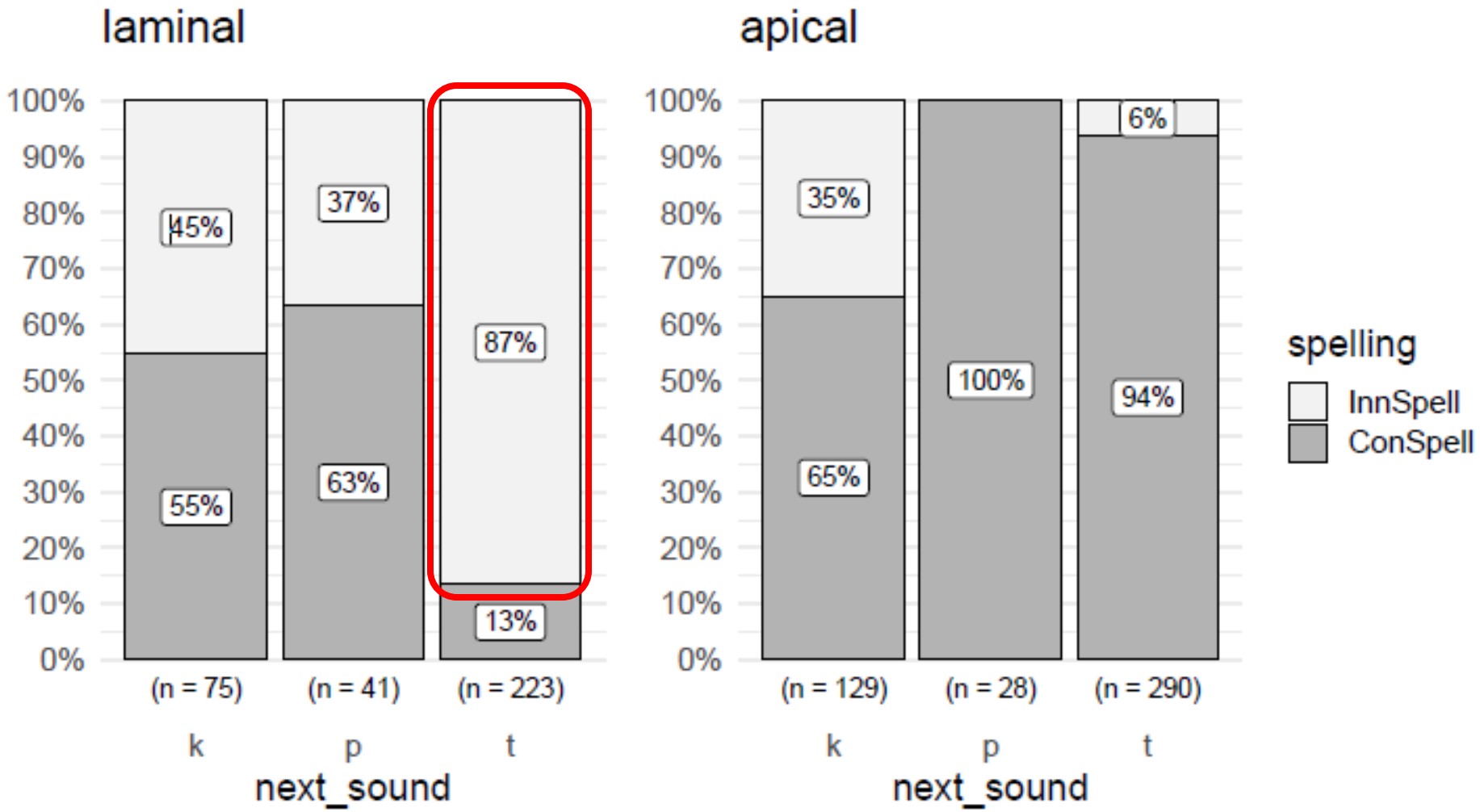


(3) The variable context is influential for both the Central and Western mergers (but differently!)

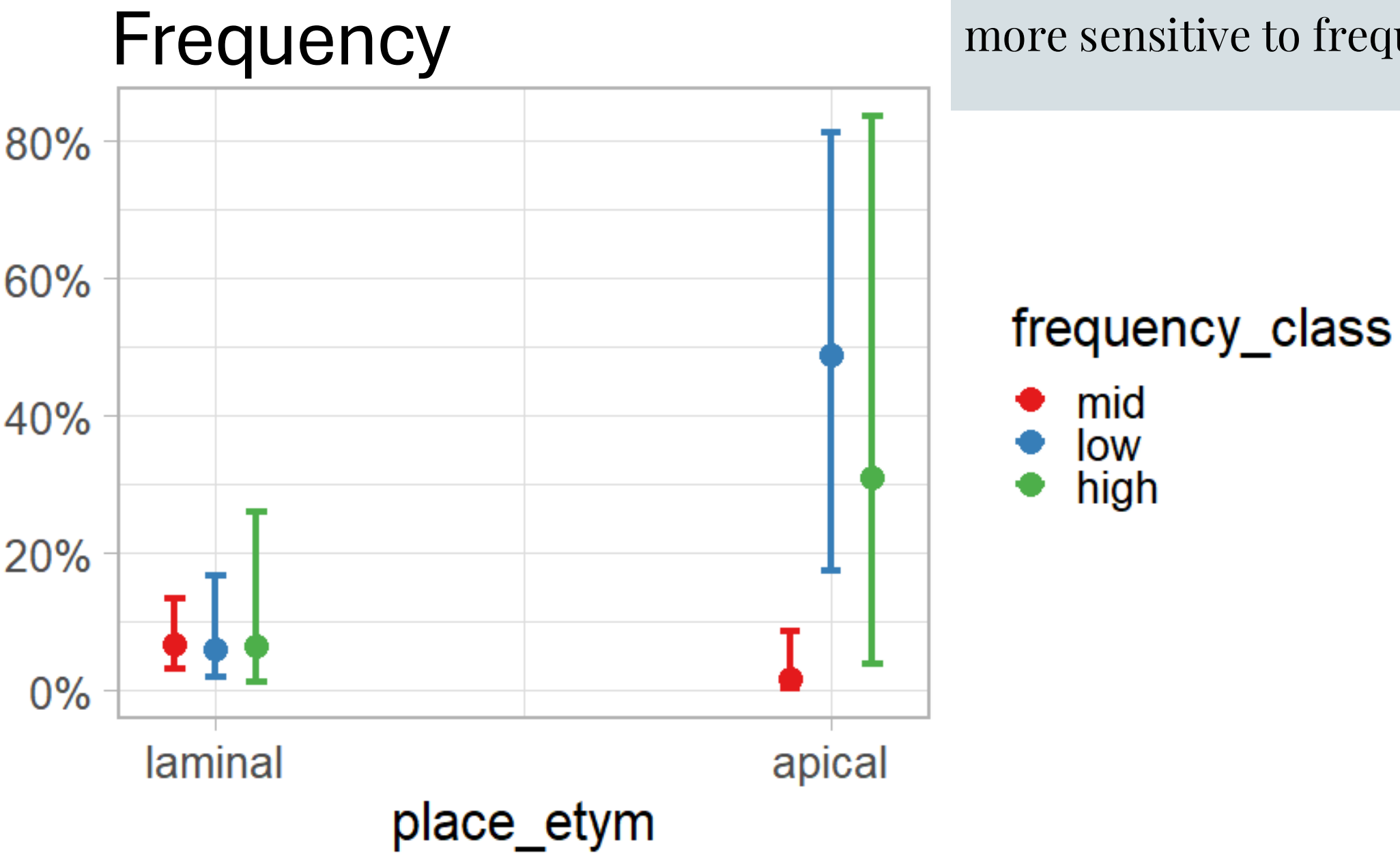
Context



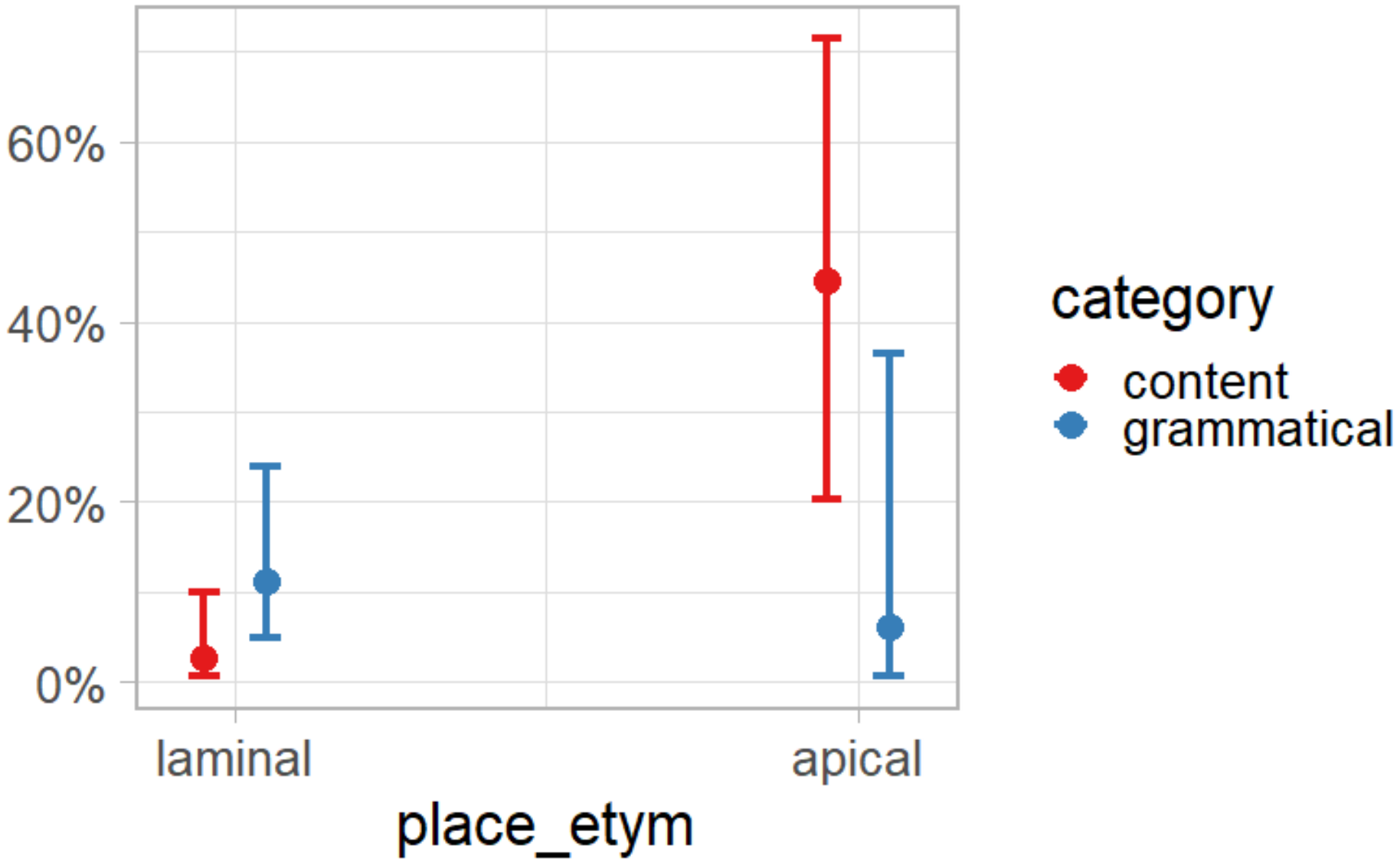
Proportion of innovative vs. conservative spelling suggests apicalization of laminals before /t/



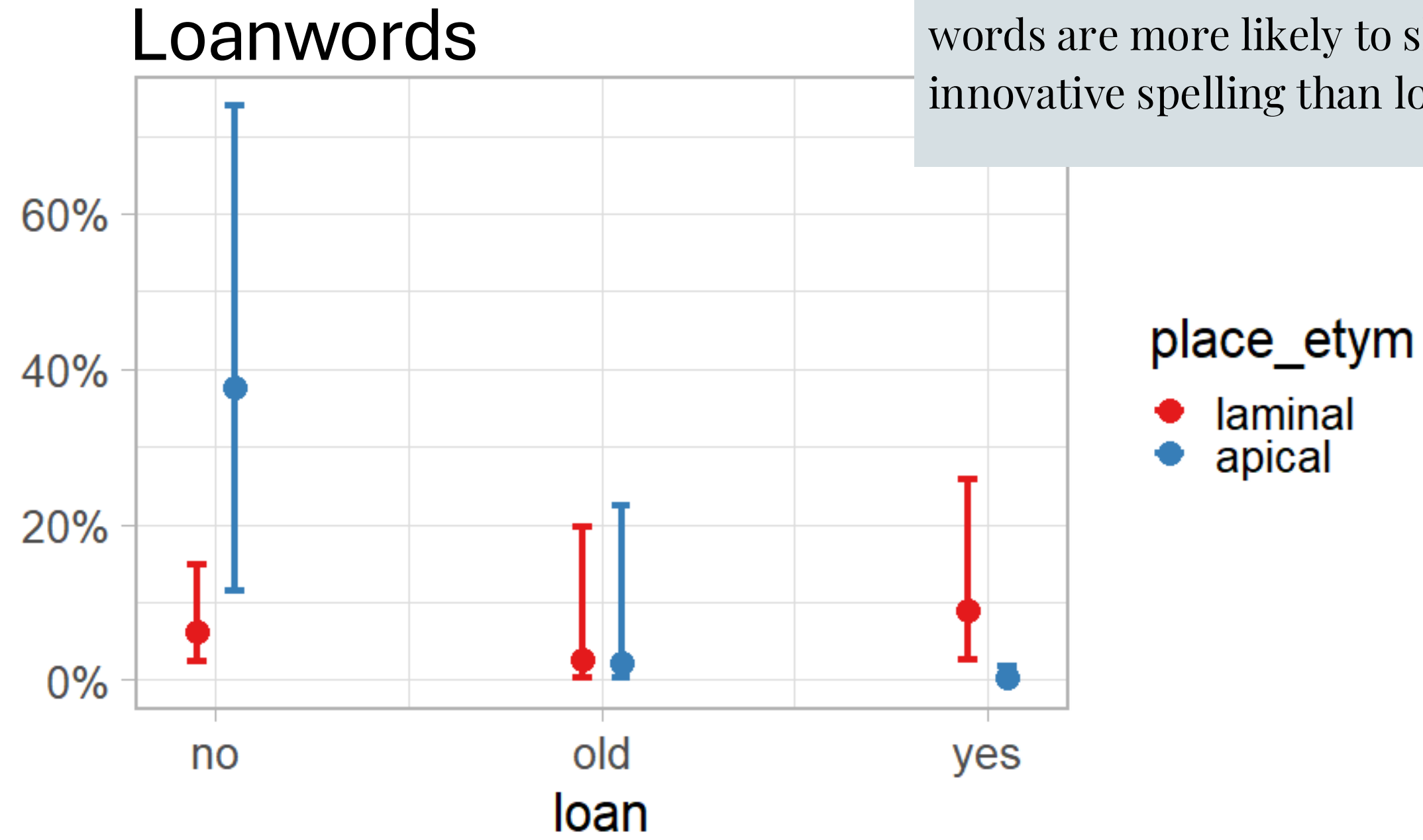
(4) The Central merger could be more sensitive to frequency effects.



Category


60%
40%
20%
0%
laminal
apical
place_etym
category
content
grammatical

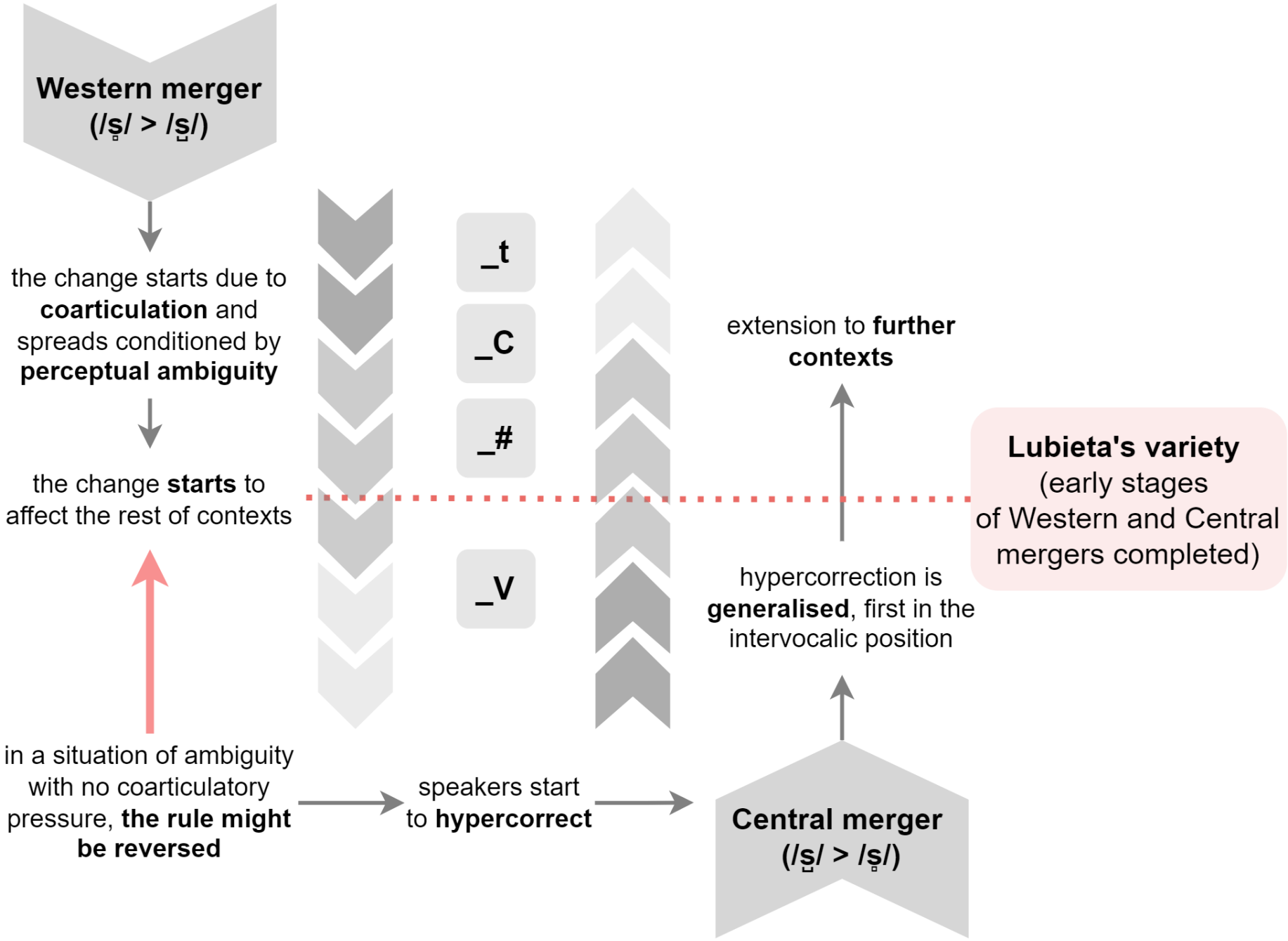
(5) In the Central merger native words are more likely to show innovative spelling than loanwords.



Conclusions

Main results

- Of the variation accounted for by the model, the greatest part is captured by the random effect ***lexeme***.
- The merger is **more advanced in affricates** than in fricatives.
- The variable ***context*** is influential for both the Central and Western mergers.
- The Central merger could be more sensitive to **frequency effects**.
- In the Central merger, **native words** are more likely to show **innovative spelling** than loanwords.



Krajewska, Zuloaga & Egurtzegi 2022: "Sibilant mergers in 18th-century Basque: A quantitative study". *Phonological Data & Analysis* 4.5, 1–31.

Conclusions

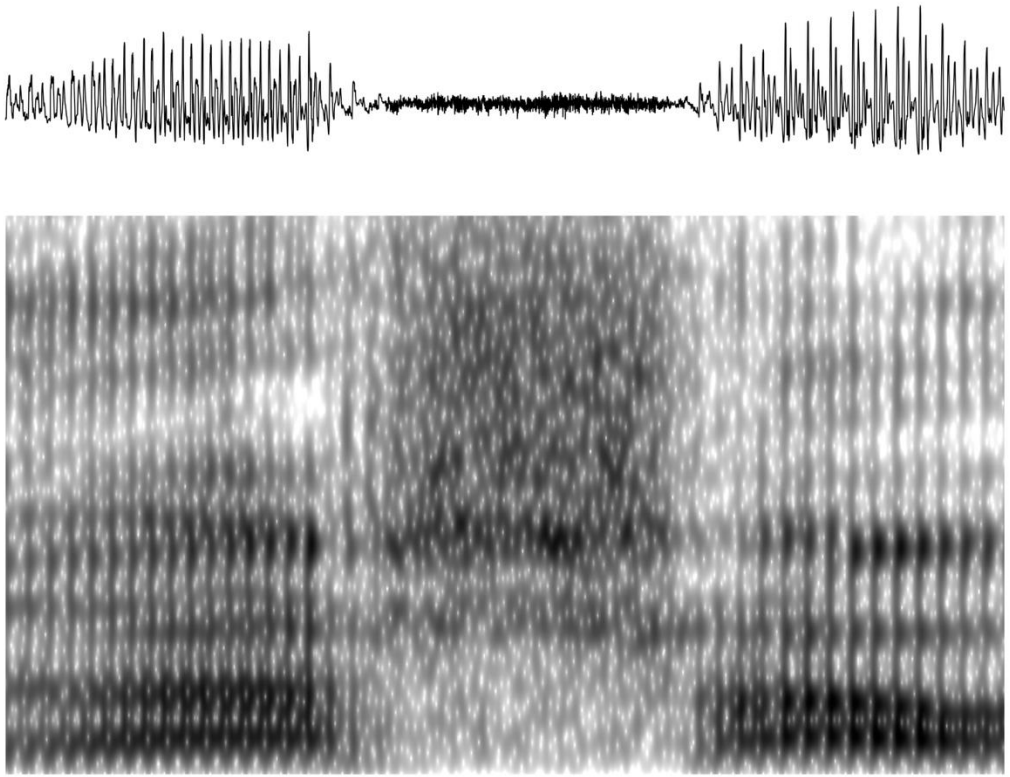
1. The Western merger weakens the contrast between apical and laminal fricatives
   - the change is completed in _C and _#
   - the change is underway in _V, including #_
2. Speakers and writers want to maintain the contrast: They overuse laminal sibilants
   - hypercorrections happen because speakers might not be able to re-establish the original distribution correctly
3. The Central merger pushes the Western merger back as apicals are replaced by laminals in both etymological and non-etymological contexts.

Studying sound change with modern audio data

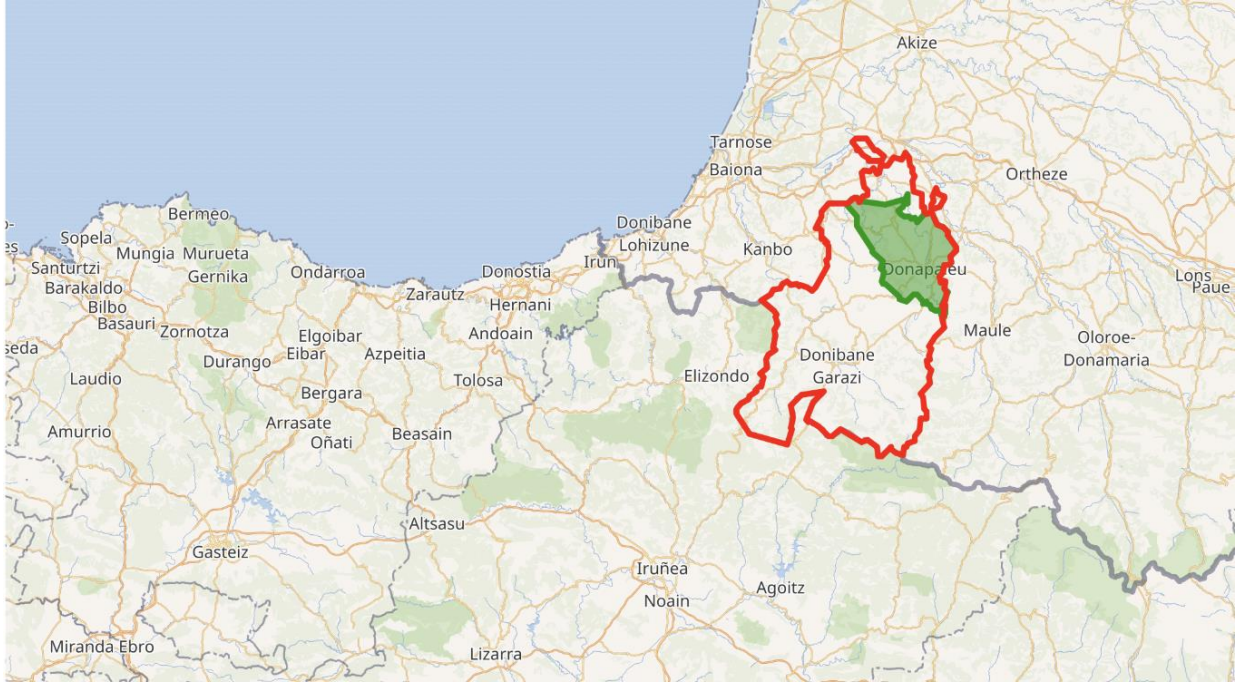
The Eastern sibilant merger in Mixean Basque

Sibilants

- Fricatives with high-frequency spectral energy



Amikuze (Mixe)



- 32 towns; Main town: Donapaleu. (Saint-Palais in French)
- Total population: 7856 in 2015.
- The number of Basque speakers is even smaller.

Amiküzera:

- On its way to disappearing (Camino 2016: 51).
- Children schooled in a Basque-speaking model < 10% (Zabalik 2016).
- The education language = Standard Basque

Mixean Basque phonology

34 contrastive consonants:

- 12 stops: /p, t, c, k, pʰ, tʰ, cʰ, kʰ, b, d, ɟ, g/
- 10 sibilants: /s̪, ʂ̪, ʃ, t͡s̪, t͡ʂ̪, t͡ʃ, z̪, ʐ̪, ʒ, d͡ʒ/
- /f/
- 9 sonorants /m, n, ɲ, l, ʎ, ɾ, r, j, w/
- 2 laryngeals: /h, h̃/

- 6 contrastive vowels: /a, e, o, i, ʉ, u/
- French loanwords: /ʁ, v, ɛ, œ/ and nasalized vowels

Previous acoustic studies on Basque sibilants

- Hualde, J.I. 2010. Neutralización de sibilantes vascas y seseo en castellano. *Oihenart* 25.
- Gandarias, L., Plaza, J & Gaminde, I. 2014. Lekeitioko txistukariez: frikariak eta afrikatuak. *Euskalingua* 24. 6–21.
- Iglesias, A., Gandarias, L. & Unamuno, L. 2016. Euskararen txistukariak aztertzeko indize akustikoez. *Euskalingua* 28. 6–18.
- Muxika-Loitzate, O. 2017. Sibilant merger in the variety of Basque spoken in Amorebieta-Etxano. *Languages* 2(4). 25. https://doi.org/10.3390/languages2040025
- Beristain, Ander. 2018. *Basque dialectal substrate in the realization of /s/ in L2 Spanish*. Champaign: University of Illinois at Urbana-Champaign MA thesis.
- Beristain, Ander. 2021. Spectral properties of anterior sibilant fricatives in Northern Peninsular Spanish and sibilant-merging and non-merging varieties of Basque. *Journal of the International Phonetic Association* 52(3). 1–32. https://doi.org/10.1017/S002510032000274
- Egurtzegi, A. & Carignan, C. 2020. An acoustic description of Mixean Basque. *The Journal of the Acoustical Society of America* 147(4). 2791–2802. https://doi.org/10.1121/10.0000996

Previous acoustic studies on Basque sibilants

Previous acoustic studies on Basque sibilants

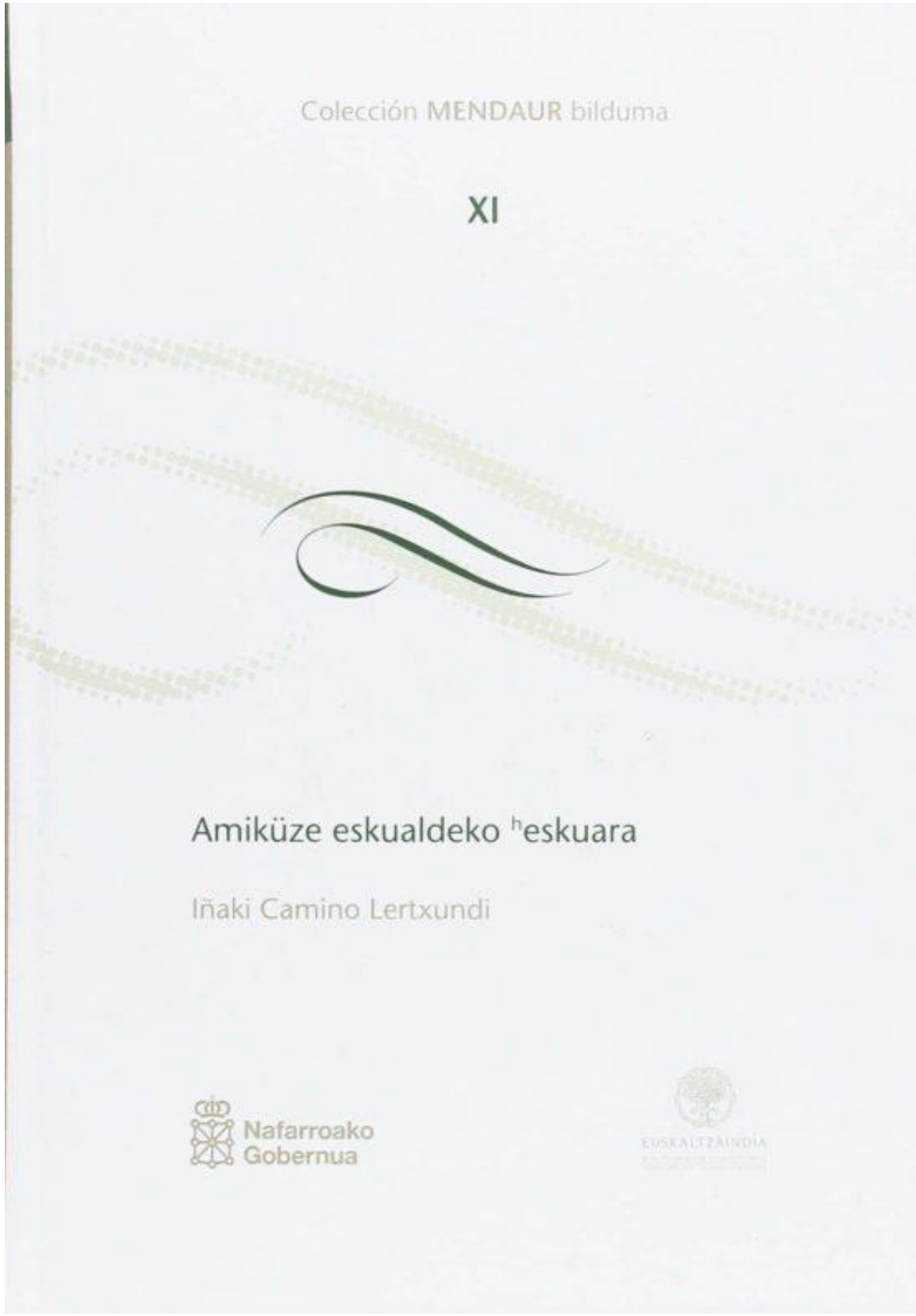
Aims of the study

- Detailed description of the acoustics of Mixean sibilants
- 10 categories based only on acoustical data?
  - Should we consider fewer?
- Provide a model on how to approach the acoustic description of the sibilants in understudied languages/varieties.

Analysis

Data: Source

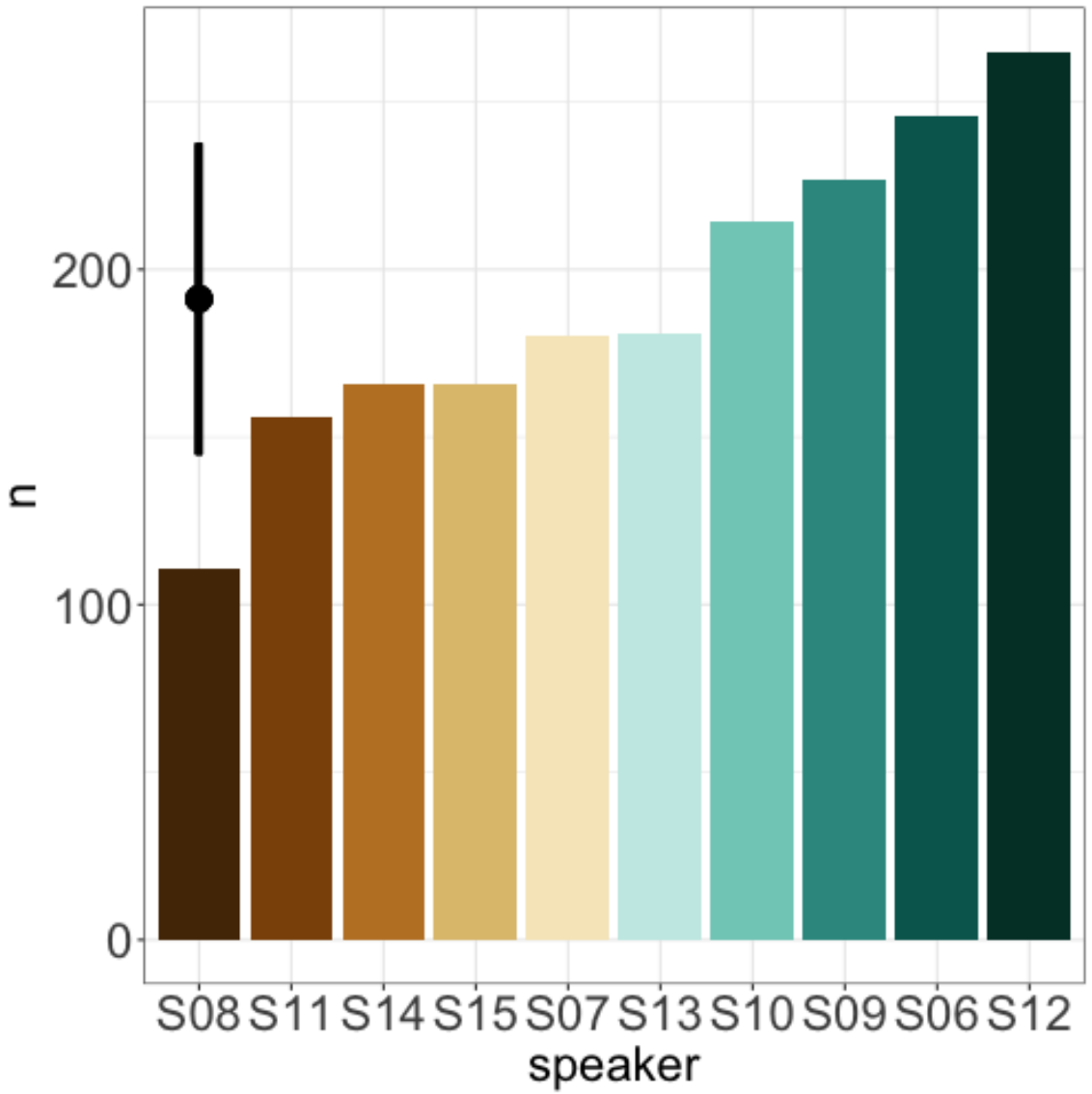
- Time span: 2005-2015.
- Speakers: 10 (3F).
- Range: 80-85 years old.
- Donapaleu, Uhartehiri, Sorhapürü, Arrüeta, Martxüeta, Labetze, Amendüze, Gamue, Zohota & Arberatze.
- Duration: 5.5' (ranging 3.5'-8.5’).



Tokens per participant

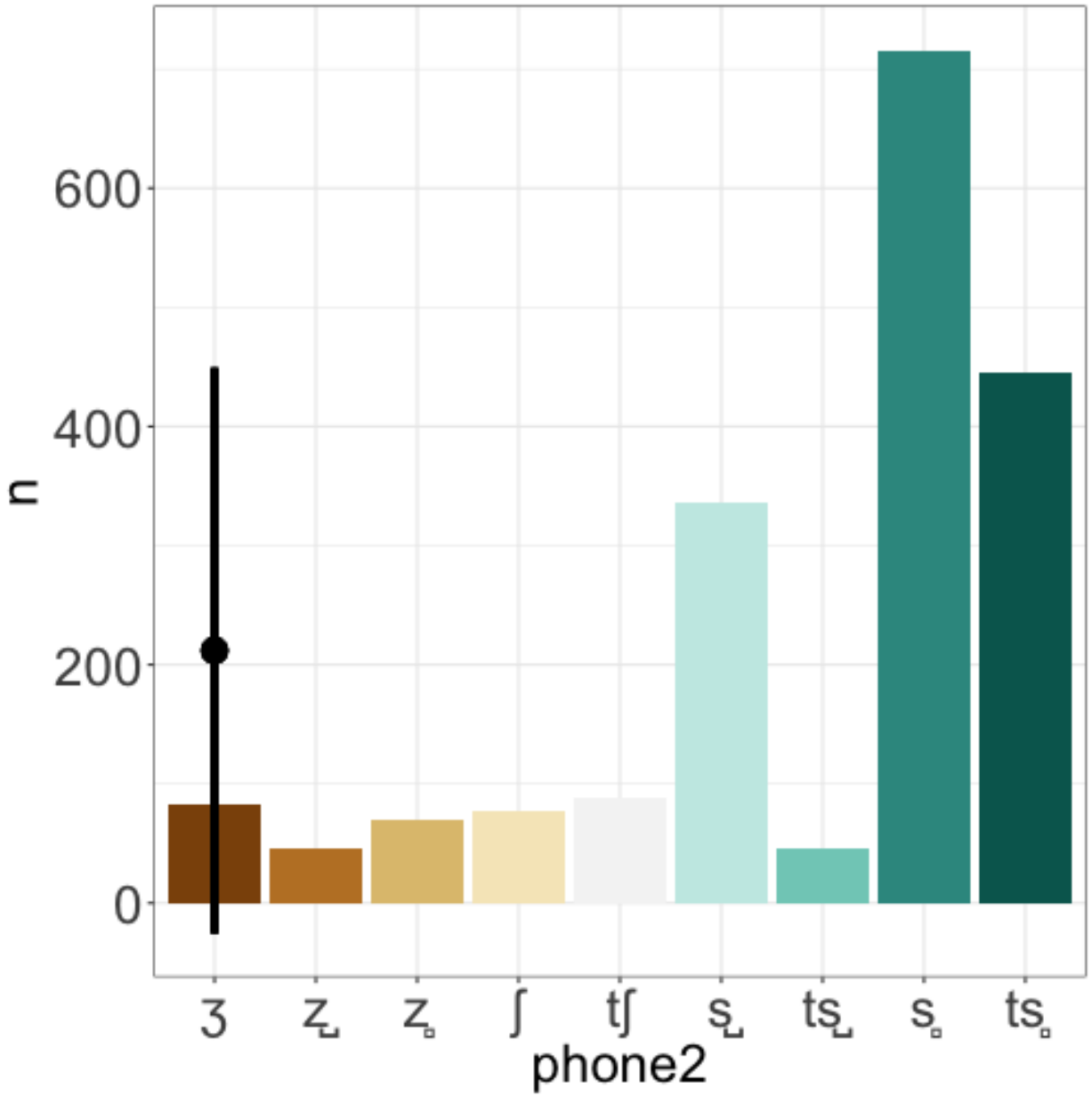
Total n
1912

min	mean	sd	max
111	191	46	265



Tokens per sibilant

min	mean	sd	max
45	212	237	716



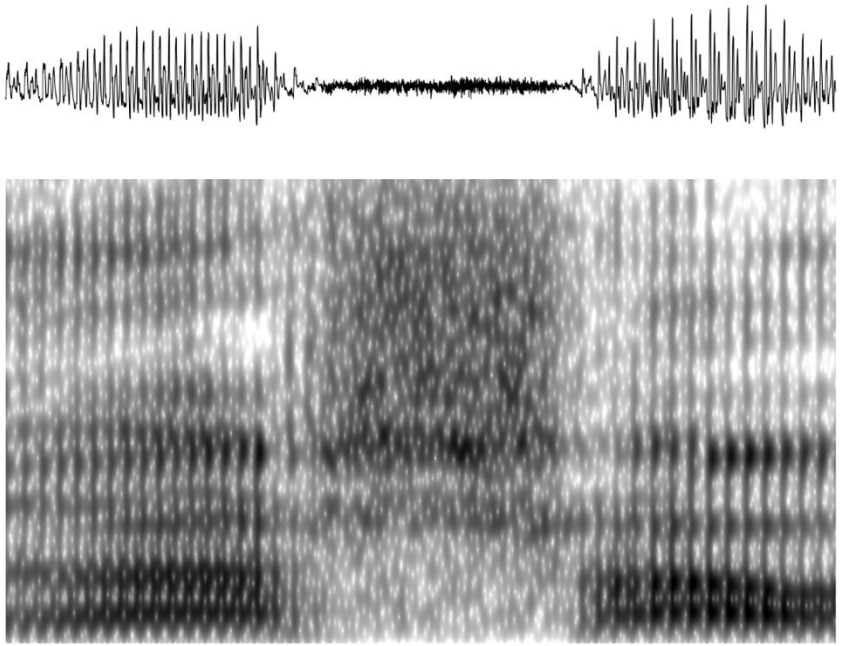
Statistical analyses

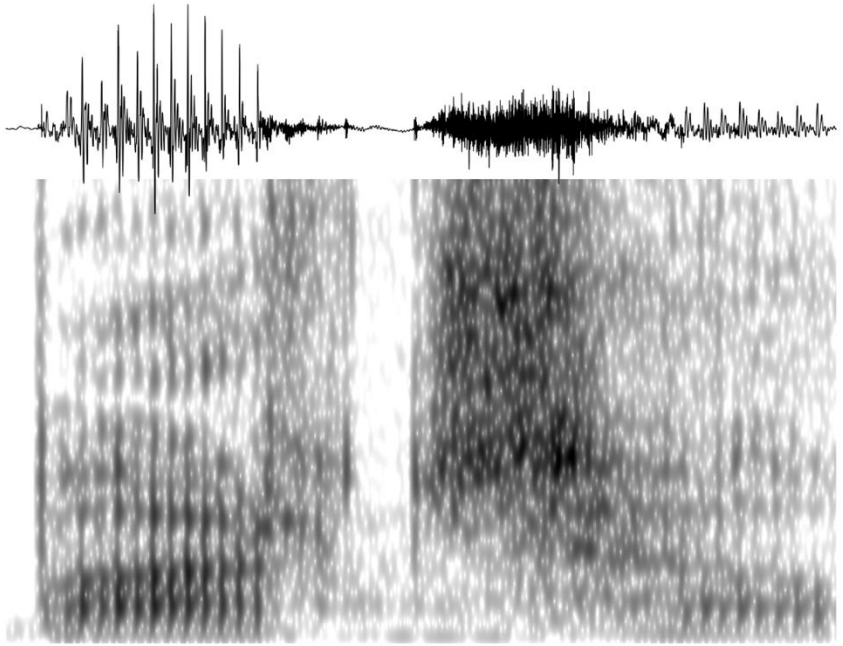
- Bayesian mixed-effects models fitted with brms.
- Priors: Weakly informative.
- Contrast analysis: EMmeans.

Acoustic measurements

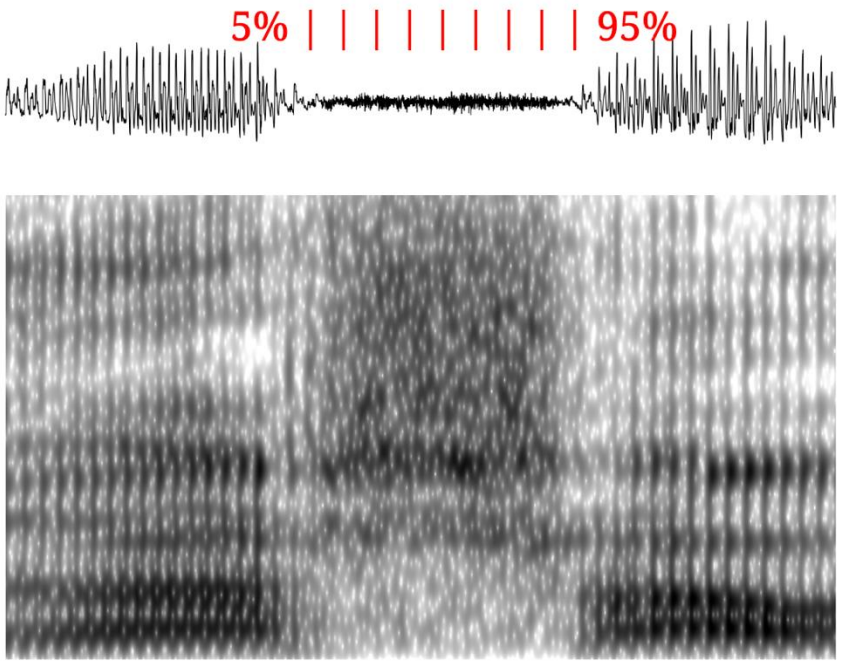
- Static values of spectral moments.
- Temporal dynamics of CoG.
- Voicing.
- Fricative/affricate distinction.

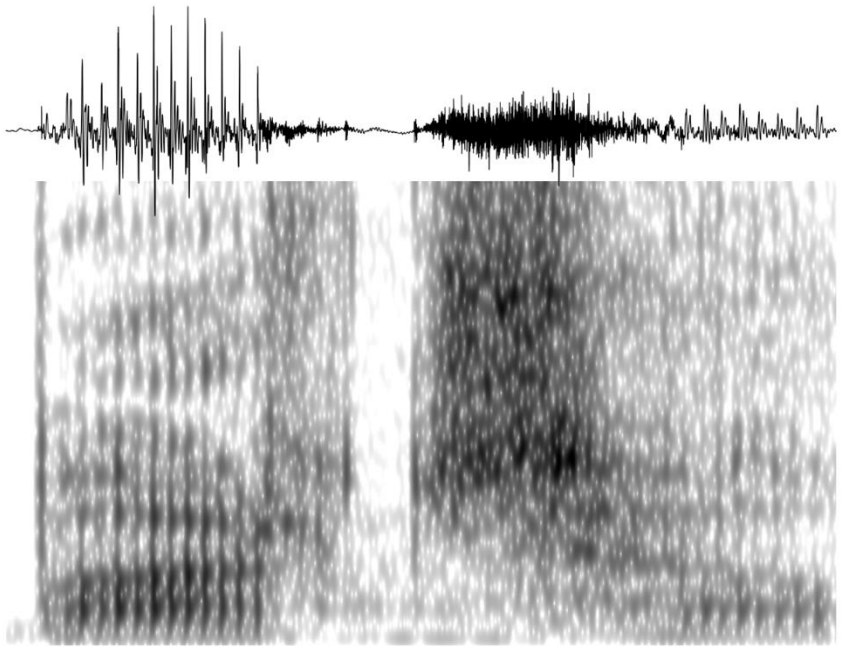
Static values of spectral moments



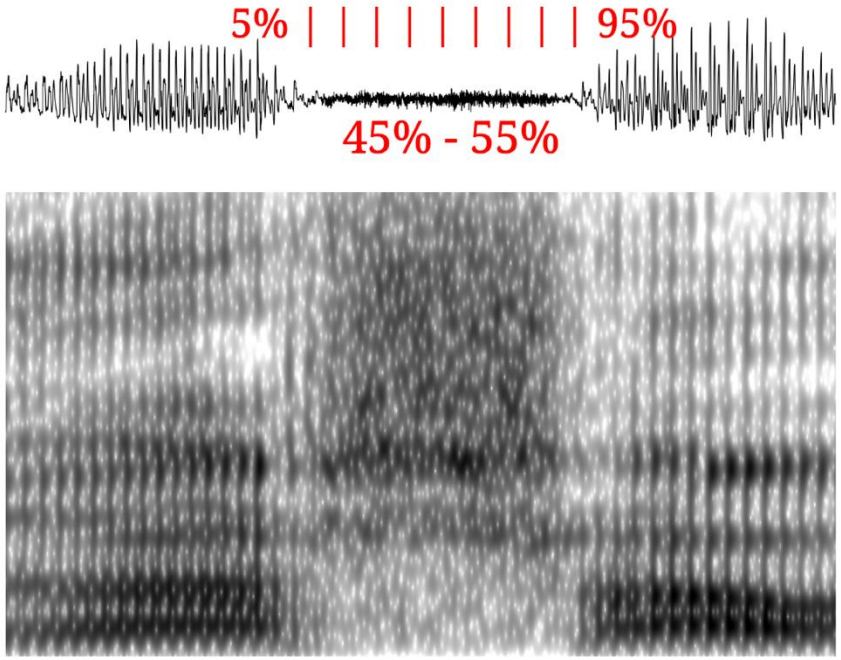


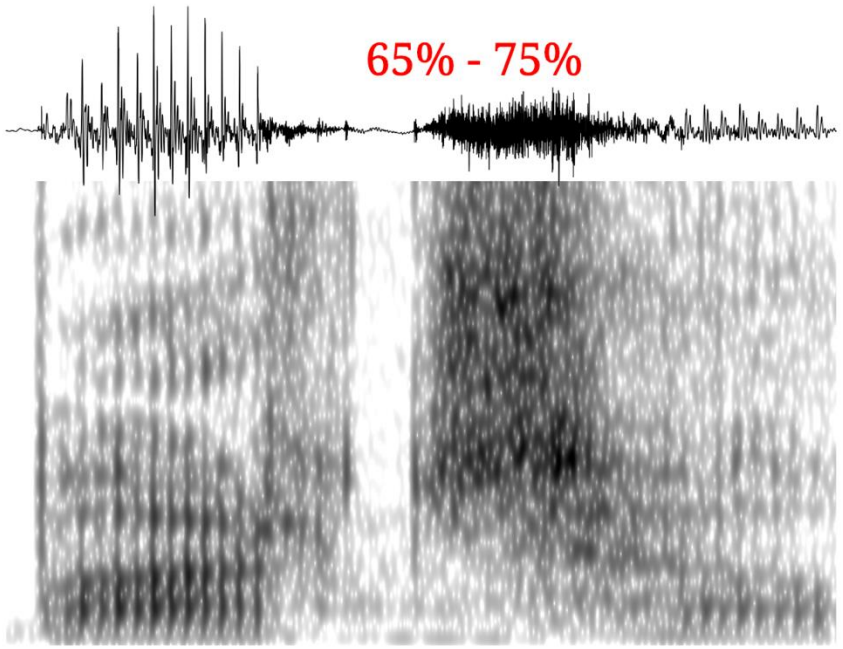
Static values of spectral moments

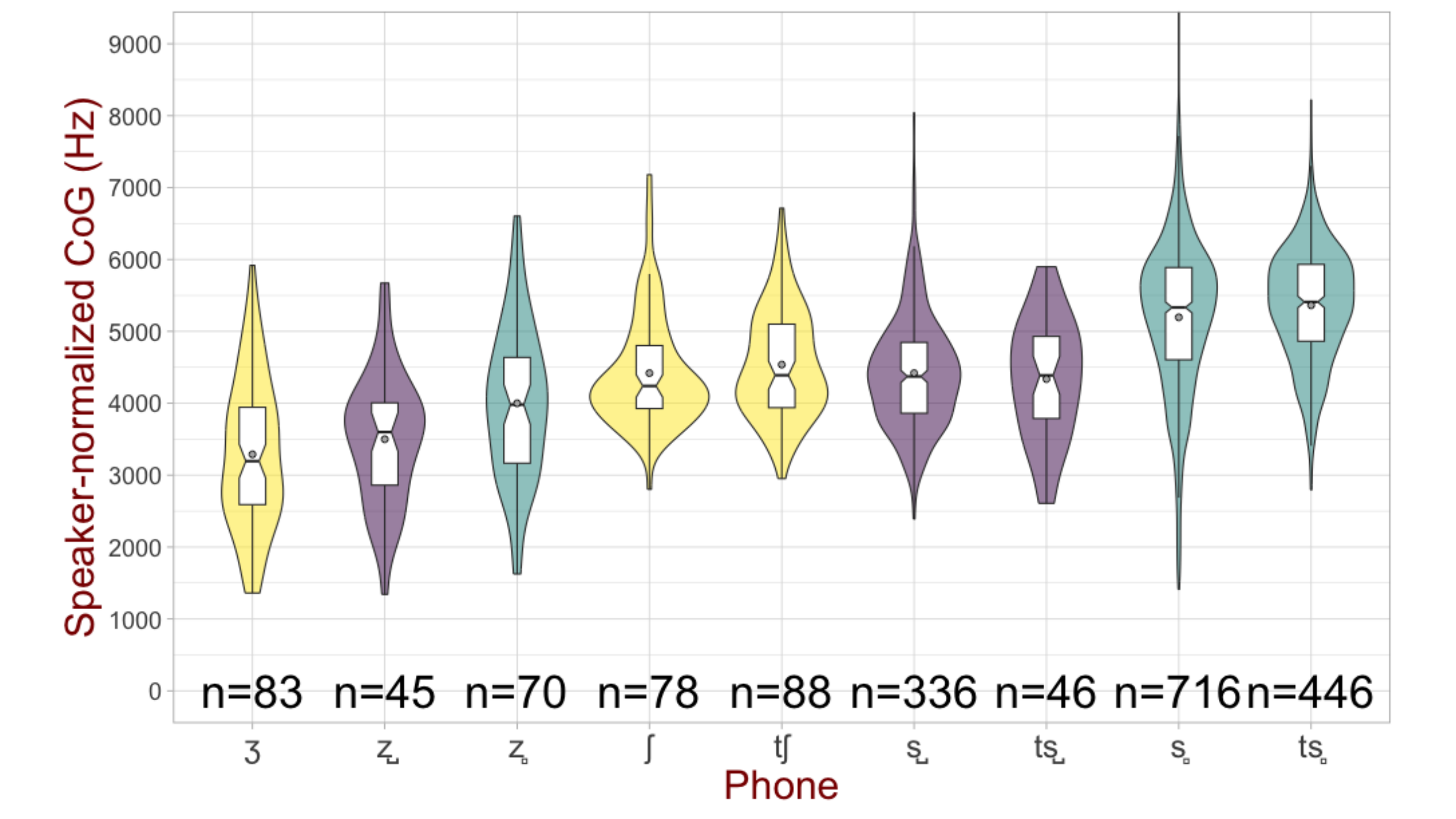


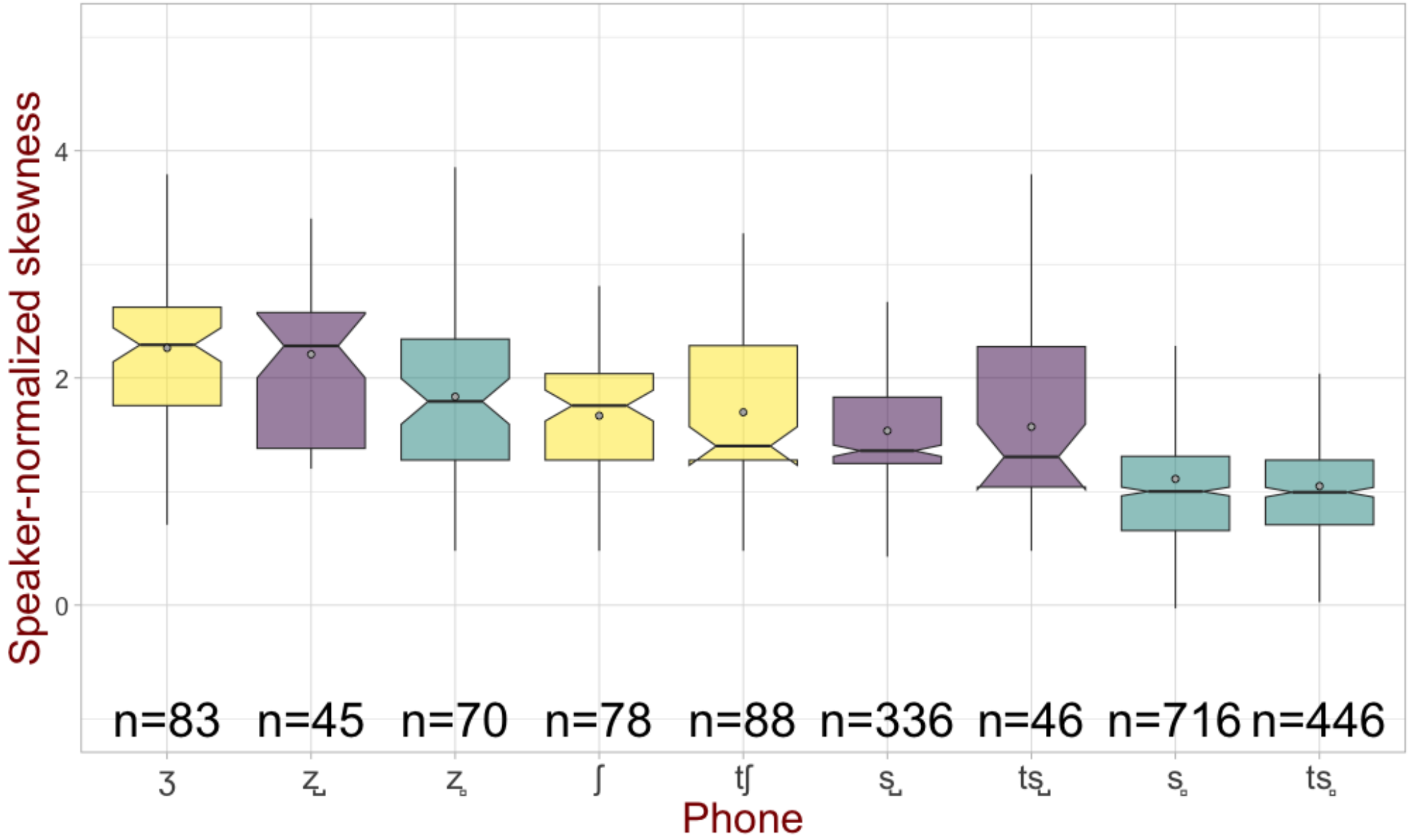


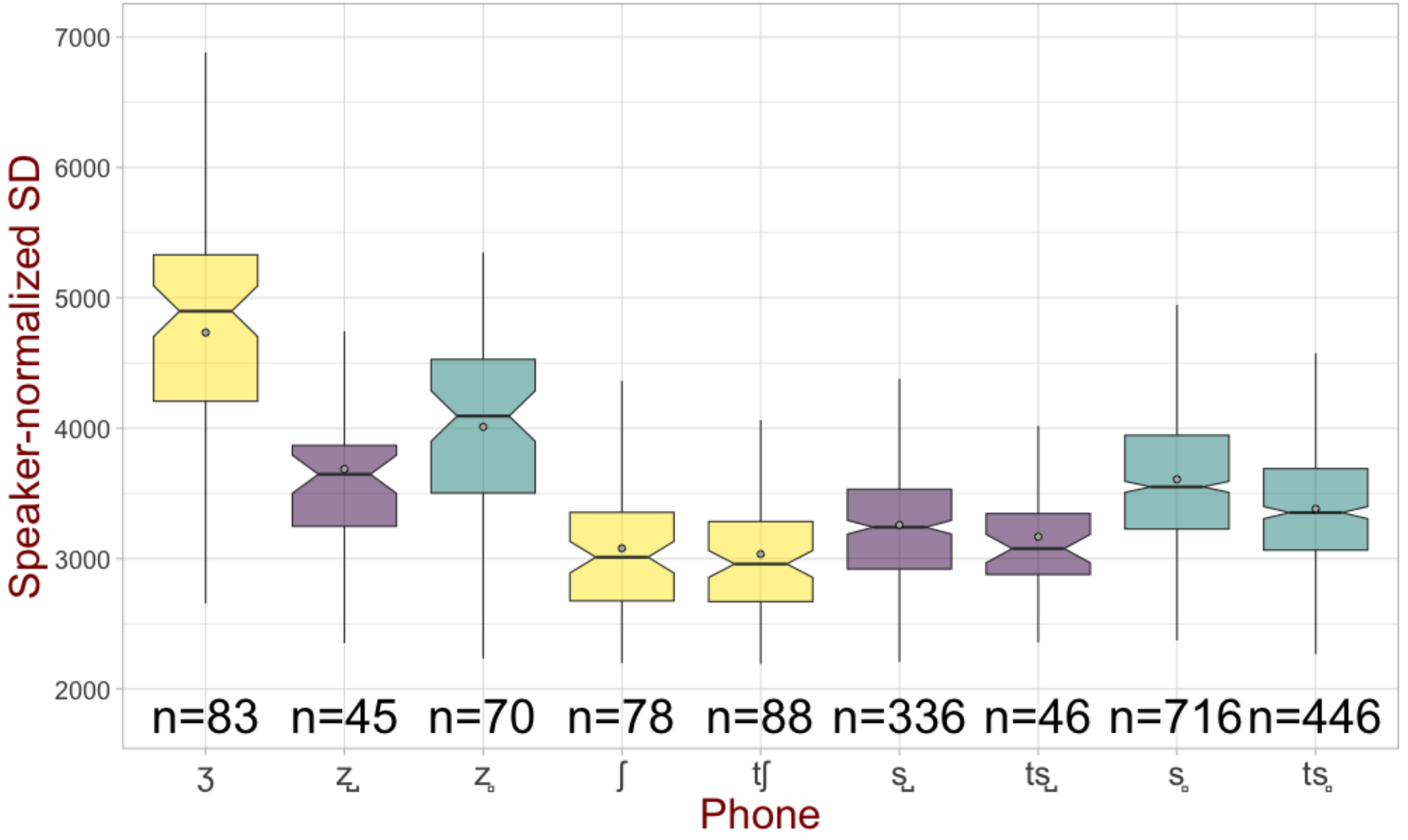
Static values of spectral moments

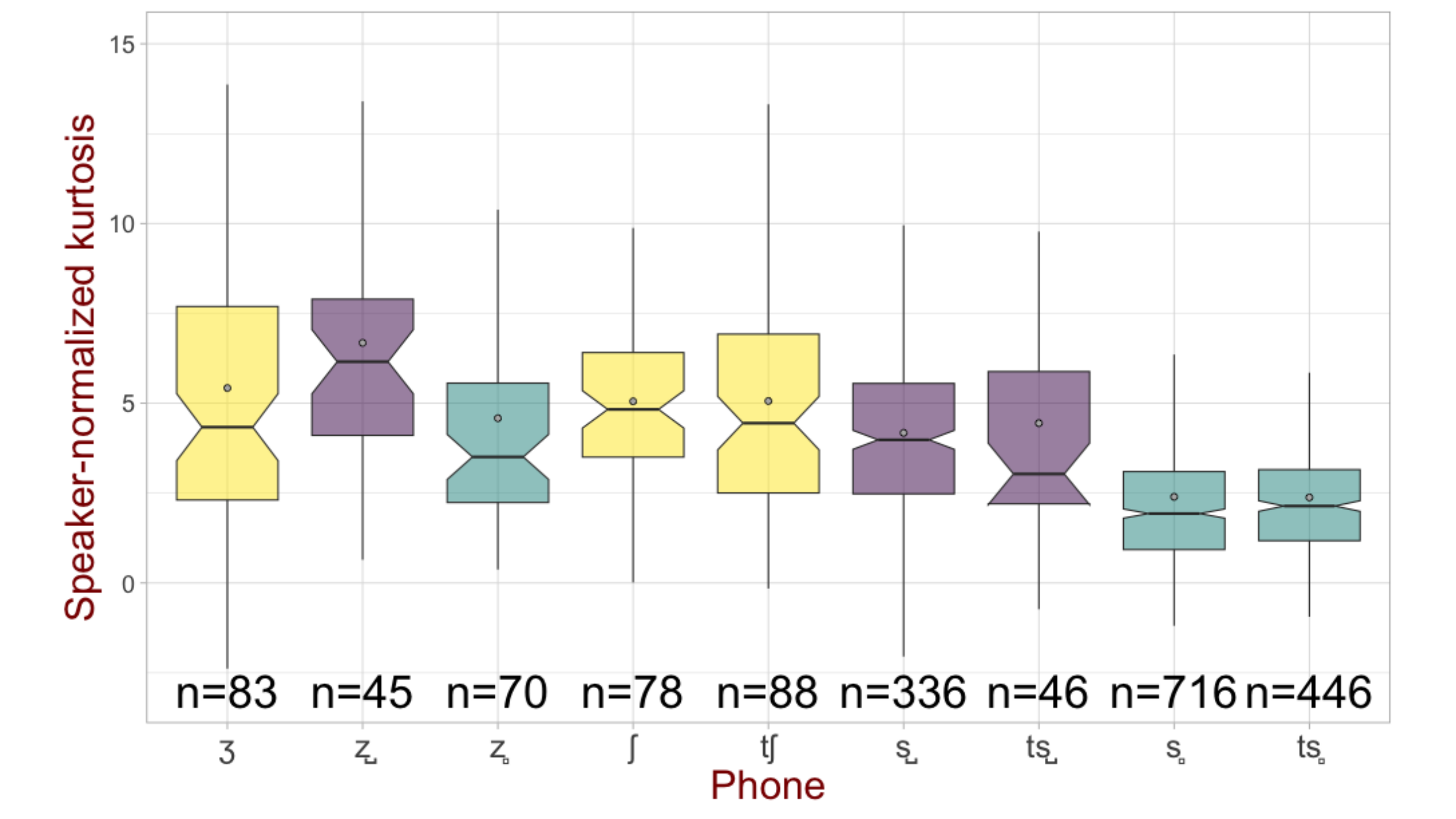





Speaker-normalized CoG (Hz)
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
n=83
n=45
n=70
n=78
n=88
n=336
n=46
n=716
n=446
ʒ
ʐ
ʐ̺
ʃ
tʃ
ʂ
tʂ
ʂ̺
ts̺
Phone


Speaker-normalized skewness
4
2
0
n=83
n=45
n=70
n=78
n=88
n=336
n=46
n=716
n=446
ʒ
ʐ
ʑ
ʃ
tʃ
ʂ
tʂ
ɕ
ʦ
Phone


Speaker-normalized SD
7000
6000
5000
4000
3000
2000
n=83
n=45
n=70
n=78
n=88
n=336
n=46
n=716
n=446
ʒ
ʐ
ʐ̞
ʃ
tʃ
ʂ
tʂ
ʂ̞
tʂ̞
Phone

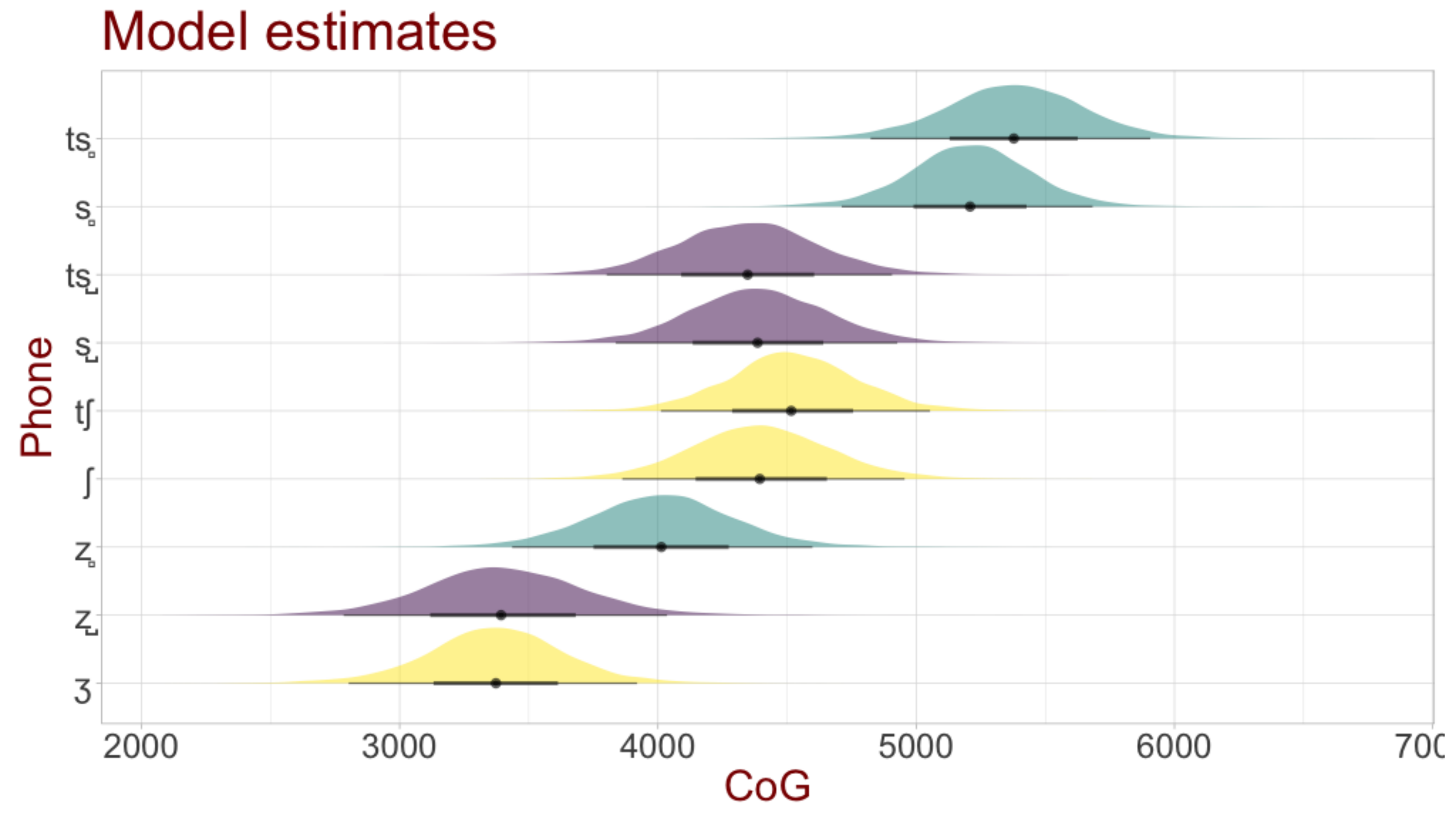

Speaker-normalized kurtosis
15
10
5
0
n=83
n=45
n=70
n=78
n=88
n=336
n=46
n=716
n=446
ʒ
ʐ
ʐ̞
ʃ
tʃ
ʂ
tʂ
ʂ̞
ts̞
Phone

Modeling static CoG

- CoG ~ phone + (phone|speaker) + (1|word).

Priors

- Intercept = normal(4800, 1500).
- b, sd, sigma = normal(0, 1500).


Model estimates
Phone
ts̪
s̪
tsʲ
sʲ
tʃ
ʃ
z̪
zʲ
ʒ
2000
3000
4000
5000
6000
CoG

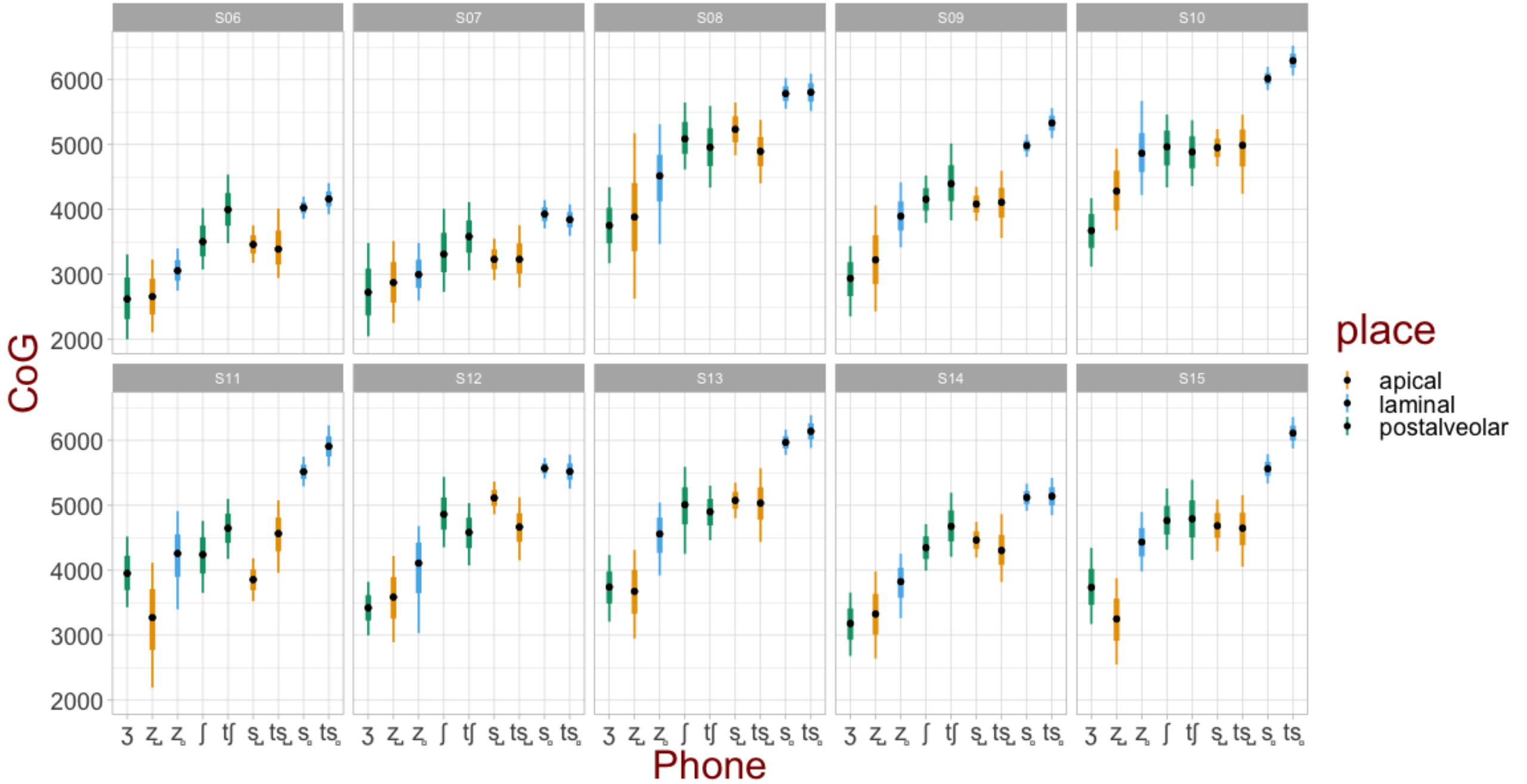
Contrast between phones within manner

contrast	estimate	lower.HPD	upper.HPD	ROPE_Percentage
s_m - S	810.976	476.080	1174.340	0.000
s_m - s_a	822.786	504.243	1145.530	0.000
S - s_a	12.674	-386.022	439.160	0.527
tS - ts_m	-860.289	-1320.531	-433.612	0.000

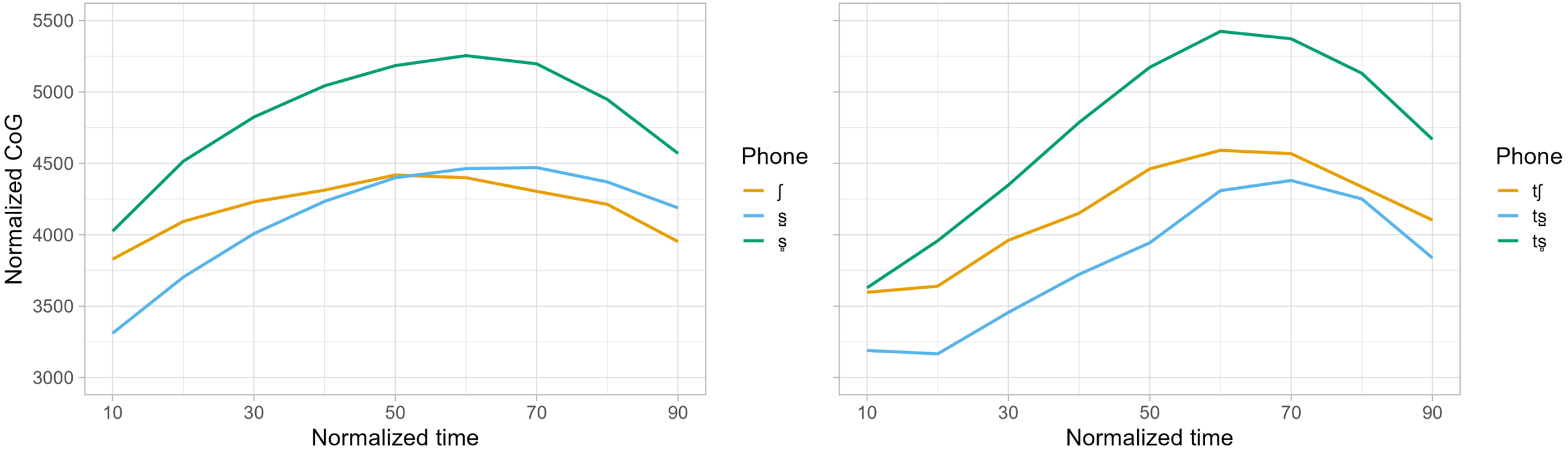
Contrast between phones within manner

contrast	estimate	lower.HPD	upper.HPD	ROPE_Percentage
tS - ts_m	-860.289	-1320.531	-433.612	0.000
ts_a - ts_m	-1028.674	-1449.821	-625.332	0.000
tS - ts_a	170.714	-323.979	676.777	0.341
z_a - z_m	-623.550	-1183.248	18.640	0.038
Z - z_a	-31.055	-656.670	625.280	0.345
Z - z_m	-645.975	-1166.540	-55.740	0.012

Intra-speaker variation



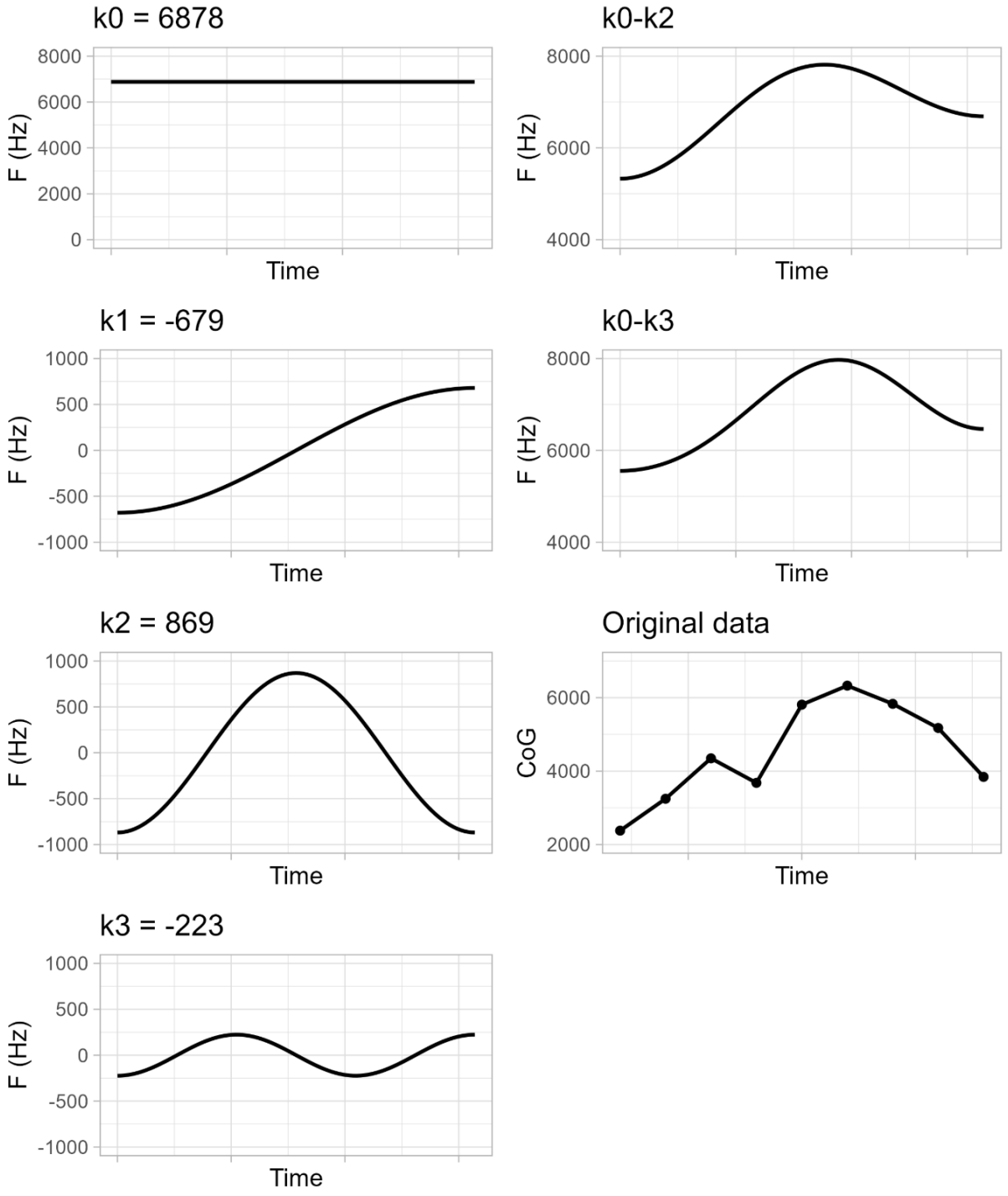
Temporal dynamics of CoG



Discrete Cosine Transformation (DCT)

- Signal into cosinusoidal waves
- Approximate shape of original

- k0 = mean
- k1 = linear slope
- k2 = curvature of the trajectory
- k3 = S vs tS?



DCT modeling

- k0 ~ phone + (phone|speaker) + (1|word)
- k1 ~ phone + (phone|speaker) + (1|word)
- k2 ~ phone + (phone|speaker) + (1|word)
- k3 ~ phone + (phone|speaker) + (1|word)

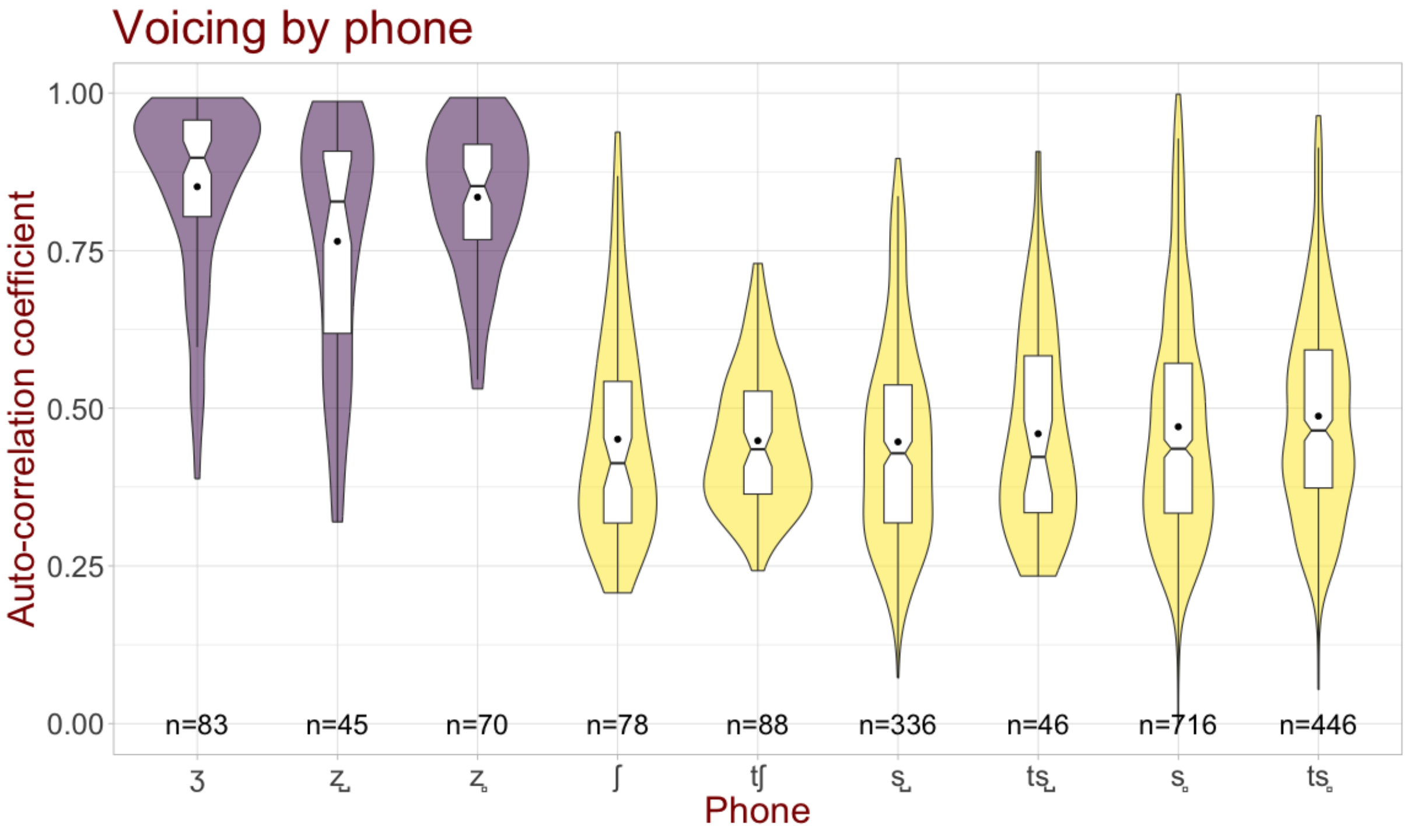
Priors:

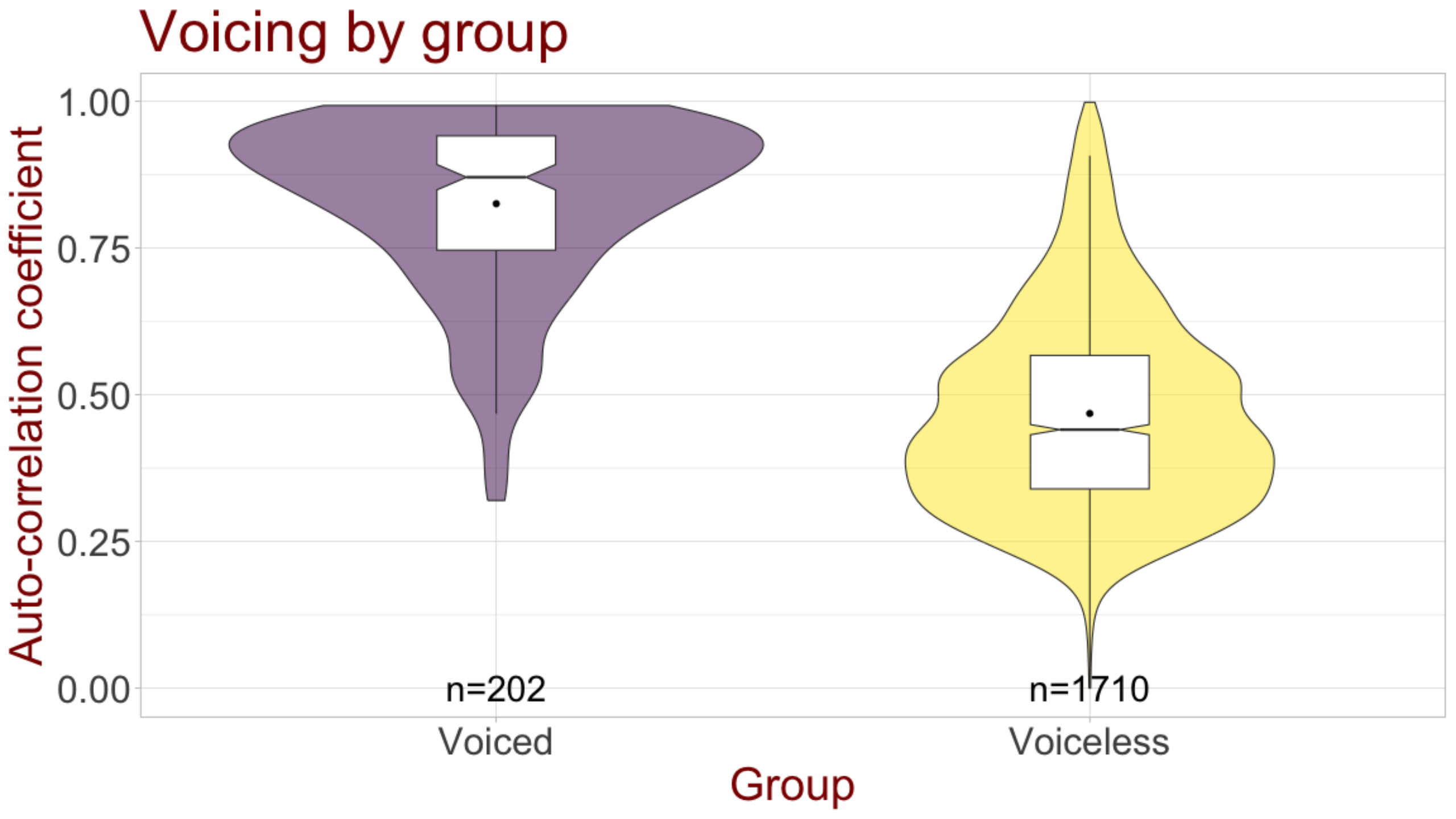
- Weakly informative priors adjusted for each k.

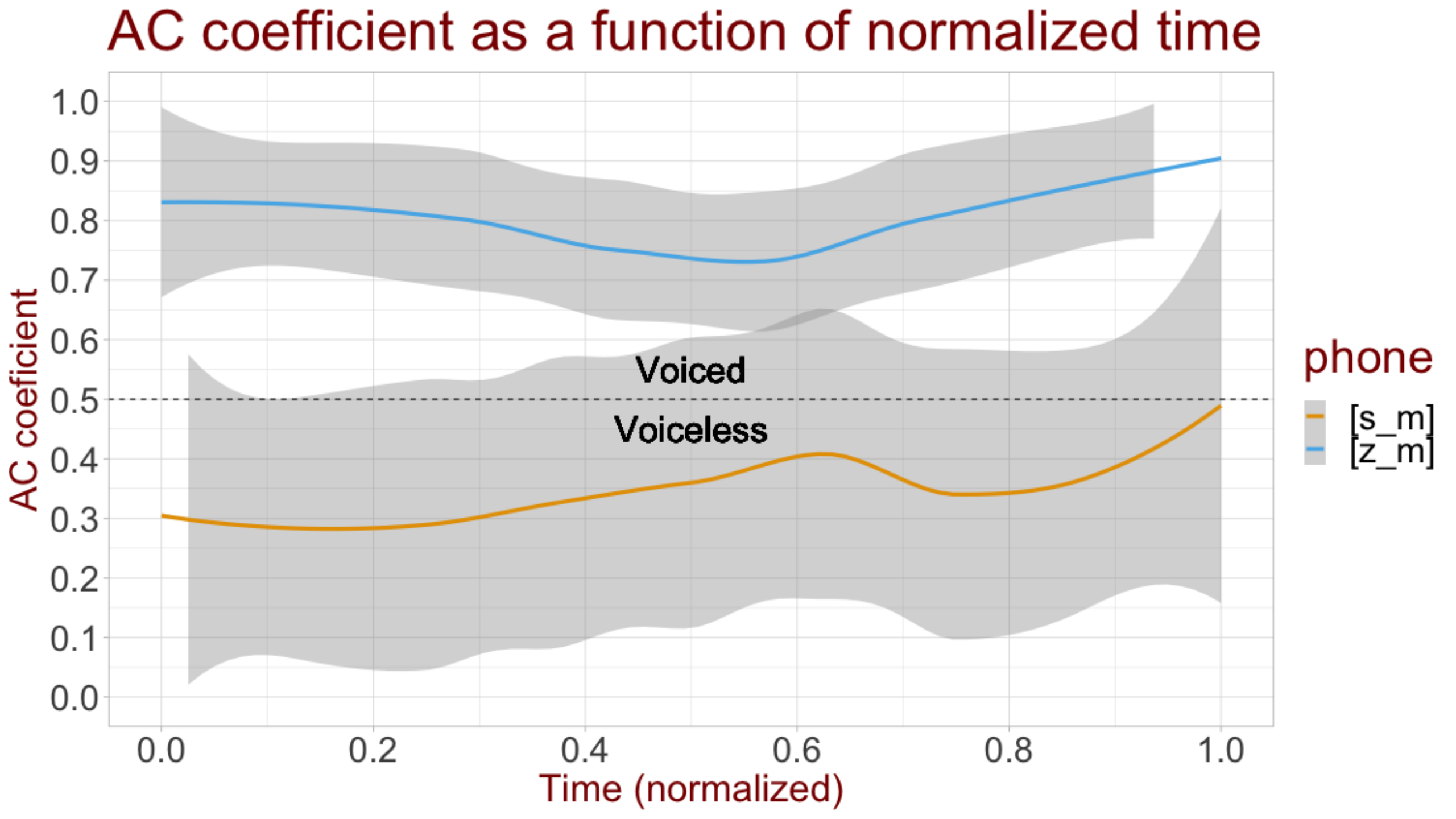
contrast	est_k0	ROPE_k0	est_k1	ROPE_k1	est_k2	ROPE_k2	est_k3	ROPE_k3
s̪ - ʃ	881.1	0	-226.6	0.03	-188.5	0	-2.3	0.58
s̪ - s̺	985.3	0	58.6	0.46	-111.3	0	45.2	0.18
ʃ - s̺	108.3	0.46	283.5	0.04	76.9	0.25	47.7	0.28
t͡ʃ - t͡s̪	-748.3	0	241.7	0.12	181.8	0	-16.3	0.44
t͡s̺ - t͡s̪	-1263.2	0	93.5	0.29	254.2	0	32.5	0.32
t͡ʃ - t͡s̺	517.7	0.10	152.2	0.23	-73	0.27	-49.5	0.24
s̪ - t͡s̪	163.3	0.41	314.3	0.02	79.9	0.10	-110.3	0.00
s̺ - t͡s̺	443.7	0.09	165.4	0.21	-64.3	0.29	-189.0	0.00
ʃ - t͡ʃ	28.5	0.36	294.9	0.04	87.1	0.45	-91.5	0.09
ʒ - ʃ	-1223.9	0	176.3	0.18	147.5	0.23	-4.6	0.42
z̺ - s̺	-1037.7	0	278.2	0.06	102.3	0.17	82.3	0.13
s̪ - z̪	-603.8	0.03	49.8	0.38	46	0.34	58.1	0.23

Voicing

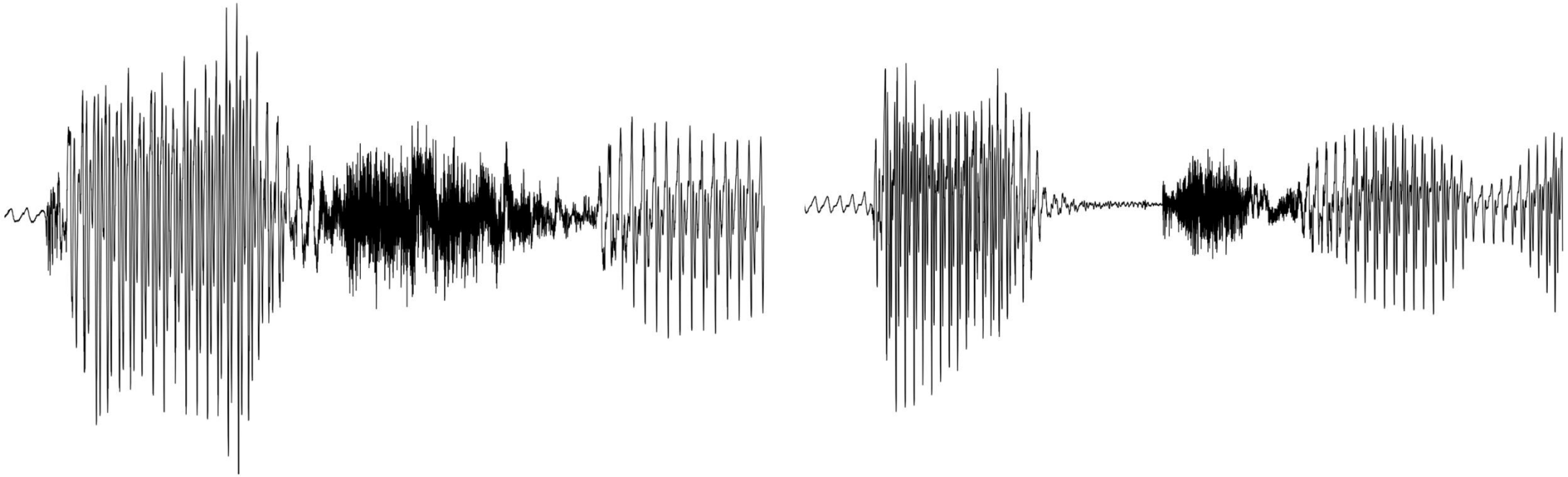
- Auto-correlation (AC)
- In EMU (Harrington 2010)
  - ESPS method
  - Frame spacing: 10 ms
  - Window length: 7.5 ms
  - Pitch ranges: 90–600 Hz♀ & 60–400 Hz♂
- 17957 measurements
- Median of the AC coefficient of each sibilant
  - 0 = no correlation => voicelessness
  - 1 = complete correlation => completely voiced


Voicing by phone
Auto-correlation coefficient
1.00
0.75
0.50
0.25
0.00
n=83
n=45
n=70
n=78
n=88
n=336
n=46
n=716
n=446
ʒ
ʐ
z
ʃ
tʃ
ʂ
ts
s
ts
Phone

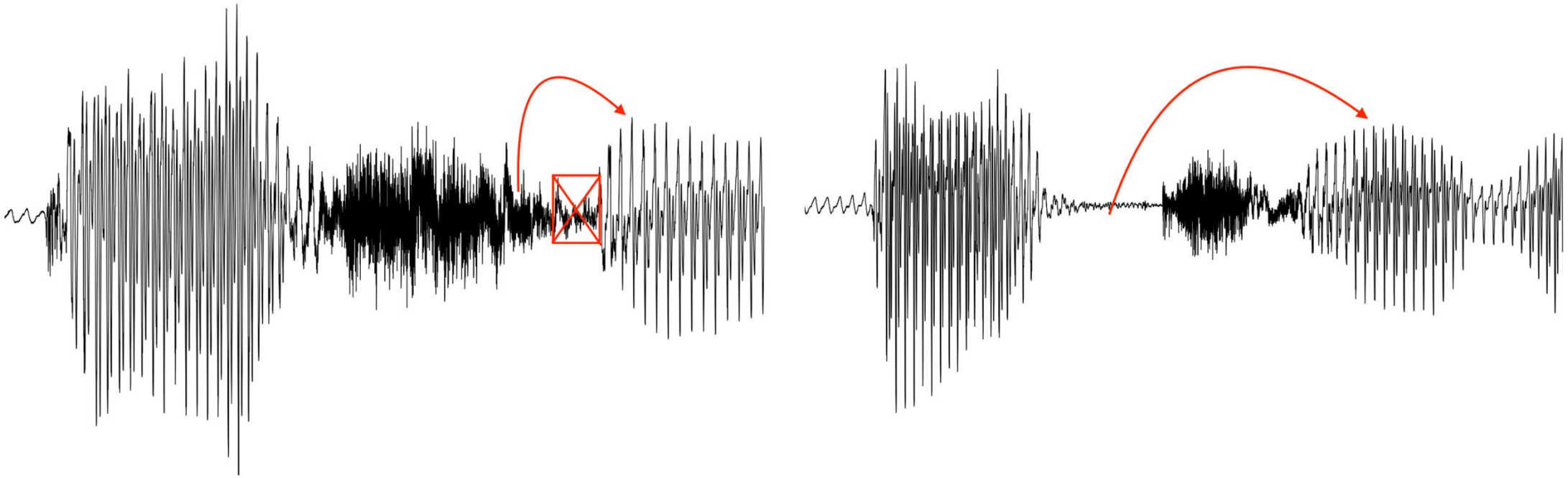

Voicing by group
Auto-correlation coefficient
1.00
0.75
0.50
0.25
0.00
n=202
n=1710
Voiced
Voiceless
Group

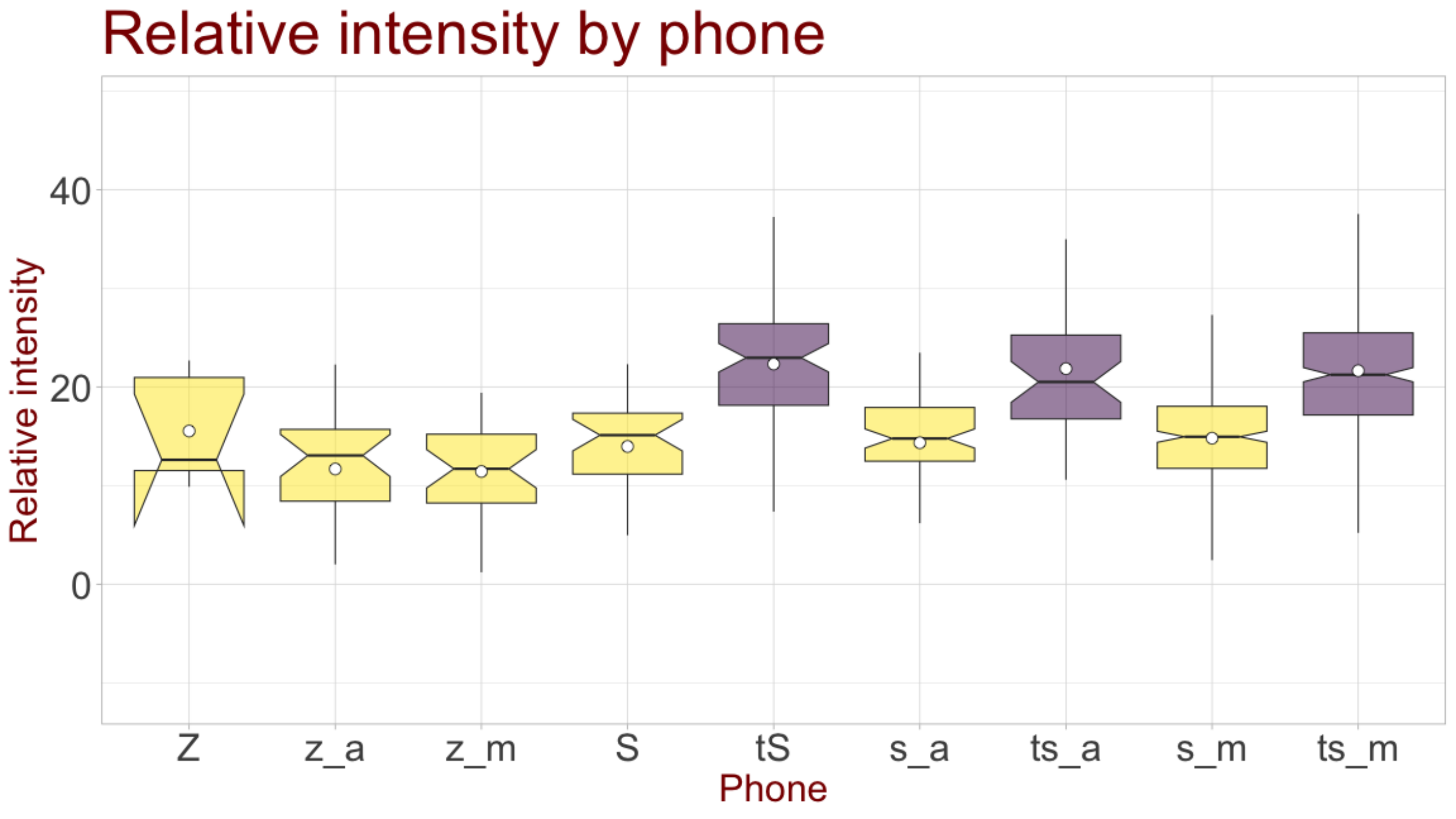

AC coefficient as a function of normalized time
AC coeficient
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Voiced
Voiceless
0.0
0.2
0.4
0.6
0.8
1.0
Time (normalized)
phone
[s_m]
[z_m]

Fricative/affricate distinction



Fricative/affricate distinction




Relative intensity by phone
Relative intensity
40
20
0
Z
z_a
z_m
S
tS
s_a
ts_a
s_m
ts_m
Phone

Relative intensity model

- rel_intensity ~ phone + (phone|speaker) + (1|word)

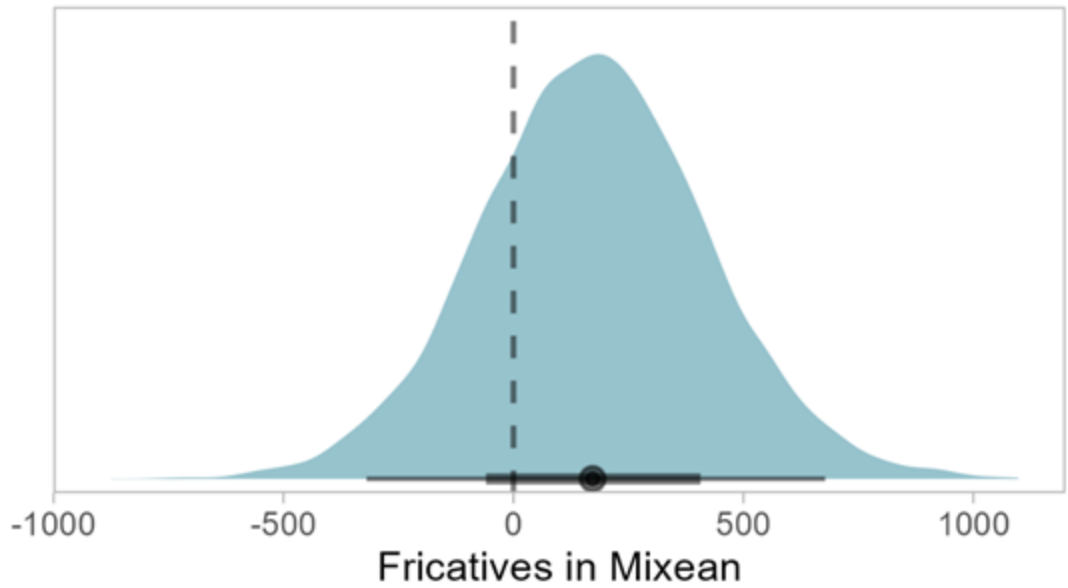
Discussion

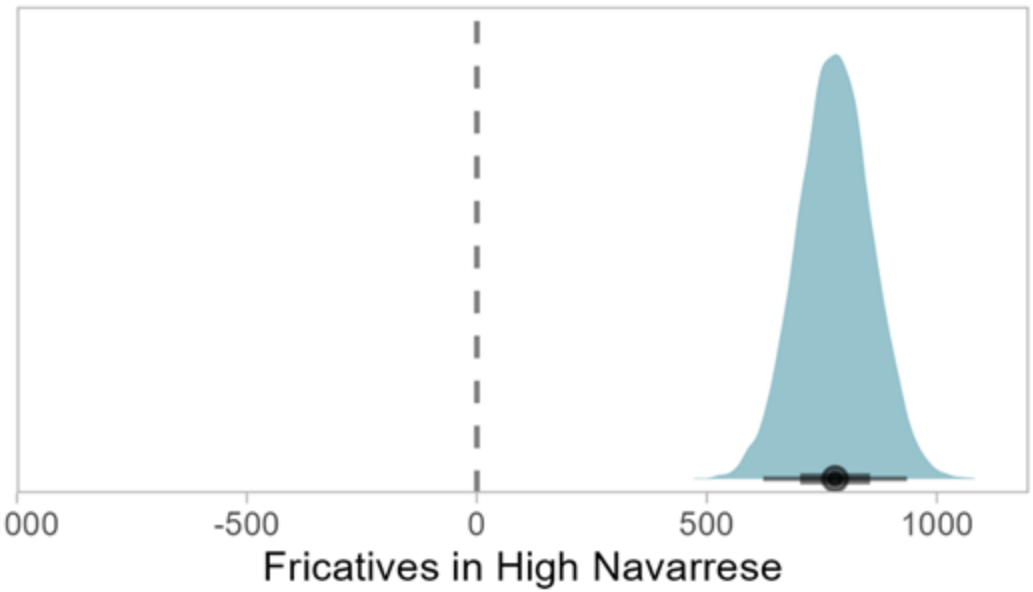
- Static CoG
  - (t)s̪ > (t)s̺, (t)ʃ
  - z̪ > z̺, (d)ʒ

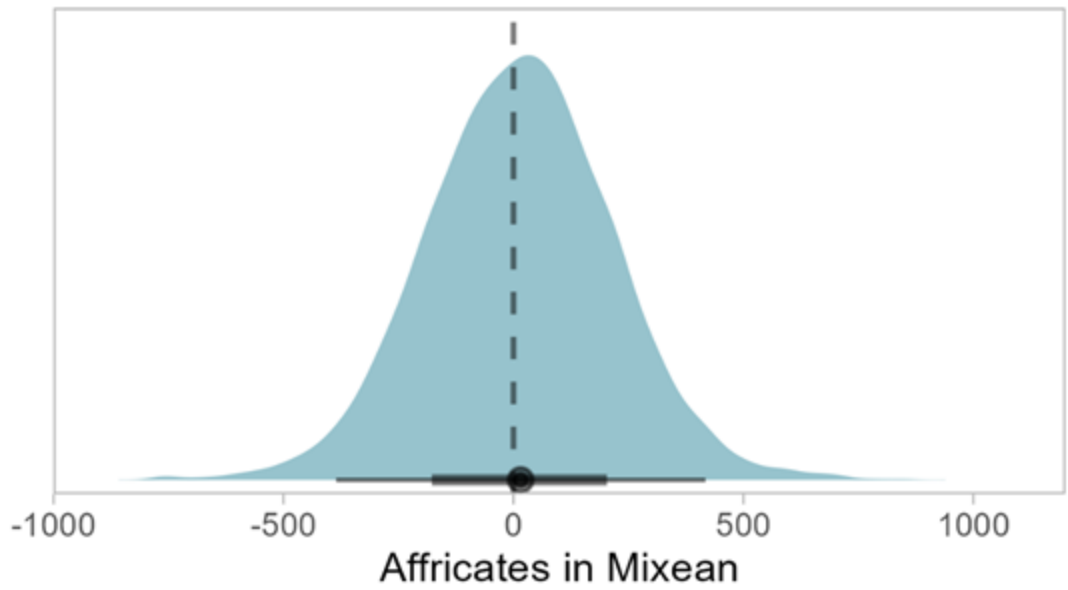
  &

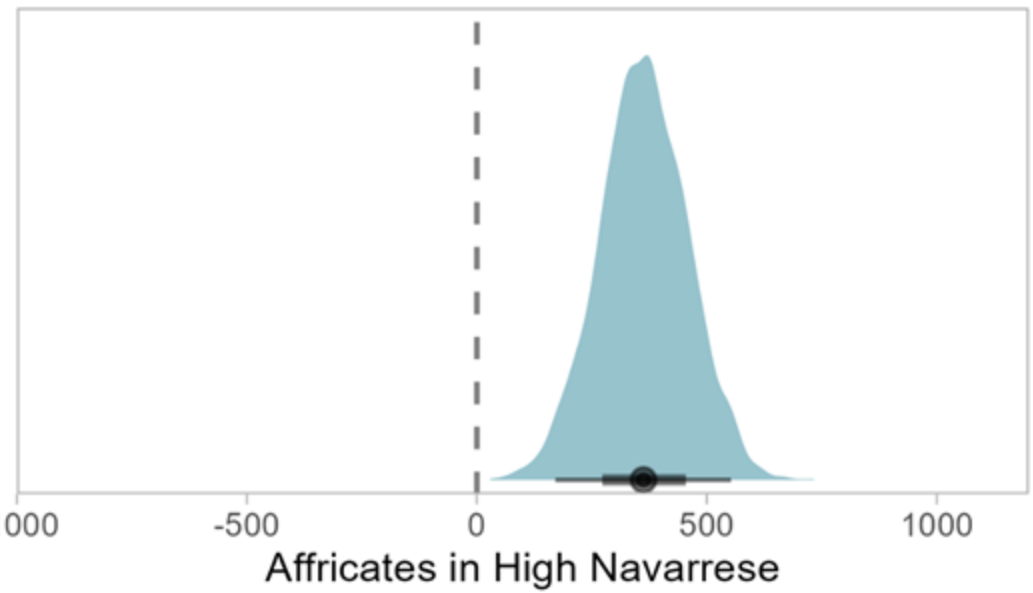
  - (t)s̺ ≈ (t)ʃ
  - z̺ ≈ (d)ʒ
- Dynamic CoG
  - k1: ʃ ≠ (t)s̺
  - k2: laminals > rest
  - k3: fricative ≠ affricate
- Voicing
  - voiced ≠ voiceless
- Relative intensity
  - fricatives ≠ affricates

Contrast distributions of the CoG posteriors of apico-alveolar and postalveolar phones (Hz)









Conclusion

- Acoustically 7 sibilants:

s̪, {s̺/ ʃ}, ts̪, {ts̺/ tʃ}, z̪, {z̺/ ʒ}, dʒ

But…

- Are minimal acoustic differences enough accept/discard a perceptible distinction?

Today's paper:

Krajewska et al. 2022. Sibilant mergers in 18th-century Basque: Aquantitative study. *Phonological Data & Analysis* 4:5, 1–31.

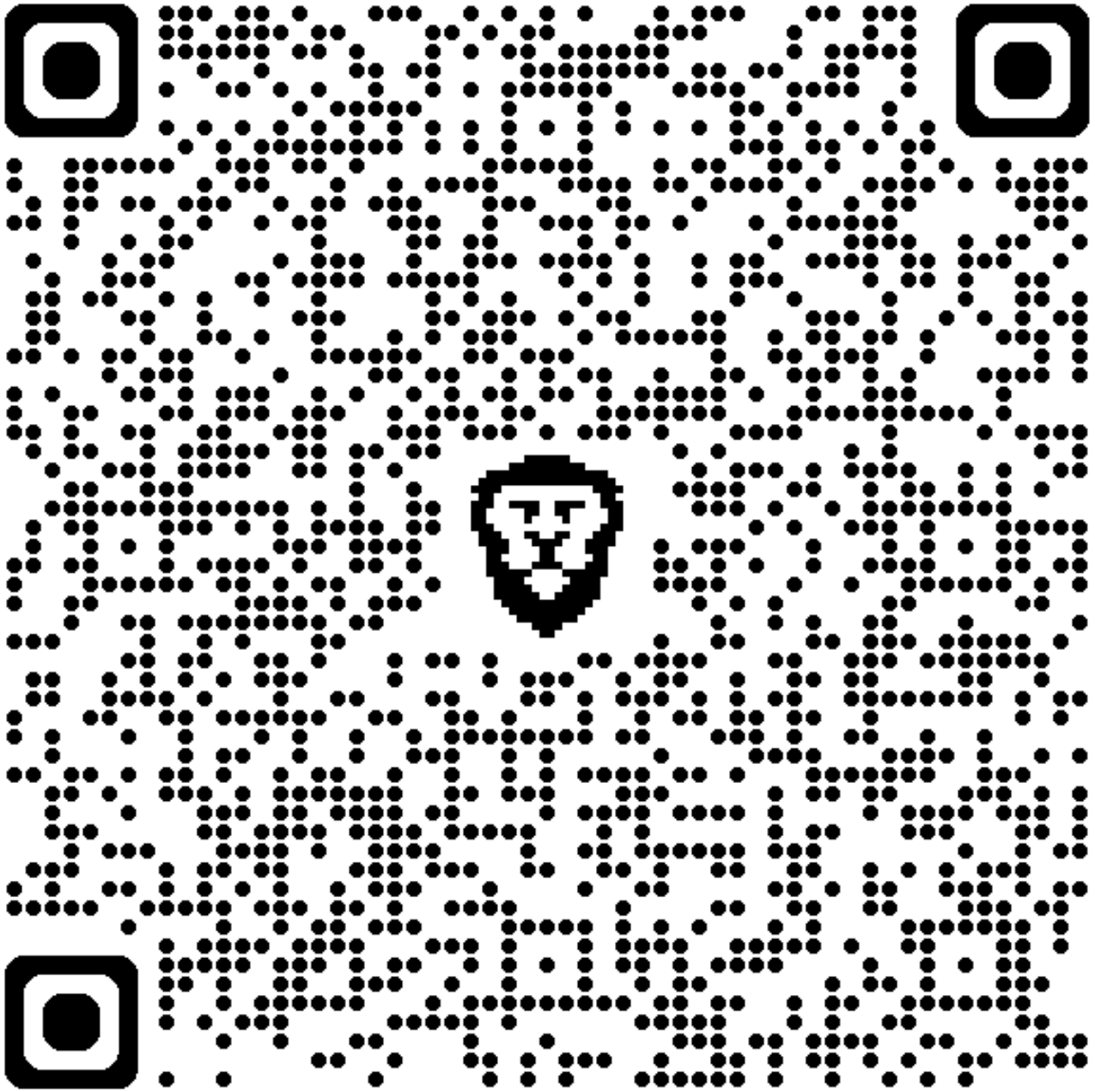
Open access here:
https://phondata.org/index.php/pda/article/view/67/43

All papers are freely accessible here: egurtzegi.github.io/publications

Today's paper:

Egurtzegi et al. 2024. An acoustic exploration of sibilant contrasts and sibilant merger in Mixean Basque. *Journal of the IPA*. First View.

Open access here:
https://www.cambridge.org/core/journals/journal-of-the-international-phonetic-association/article/an-acoustic-exploration-of-sibilant-contrasts-and-sibilant-merger-in-mixean-basque/839C776F241162A02D7505FE4D042A7E

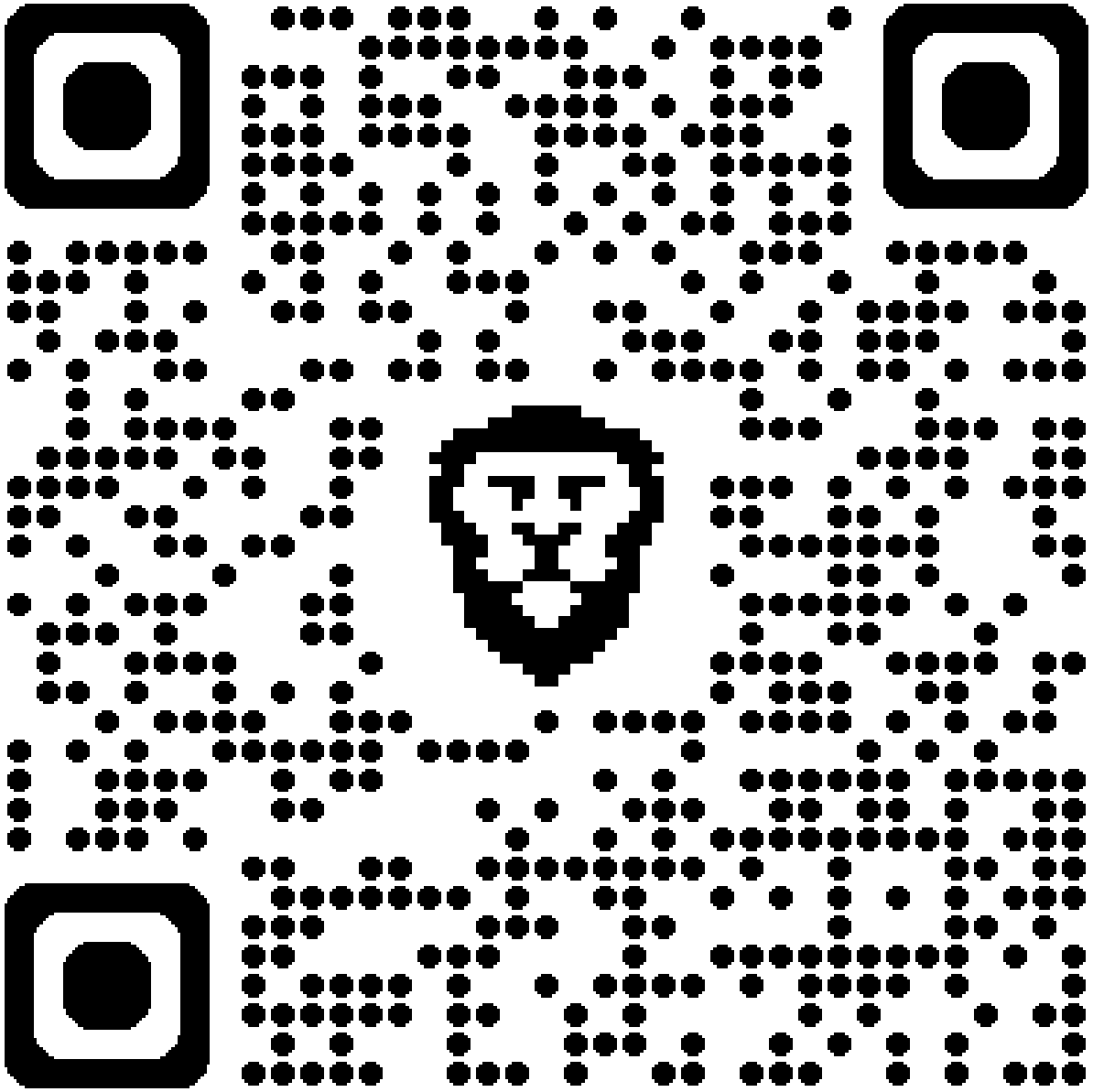


All papers are freely accessible here: egurtzegi.github.io/publications

Next paper:

Blevins & Egurtzegi. 2017.
“Unexpected obstruent loss in initial obstruent–sonorant clusters: An apparent example from Basque”,
Phonology 34, 507-522.

Open access here:
https://egurtzegi.github.io/papers/blevins_egurtzegi2017phonology.pdf



All papers are freely accessible here: egurtzegi.github.io/publications