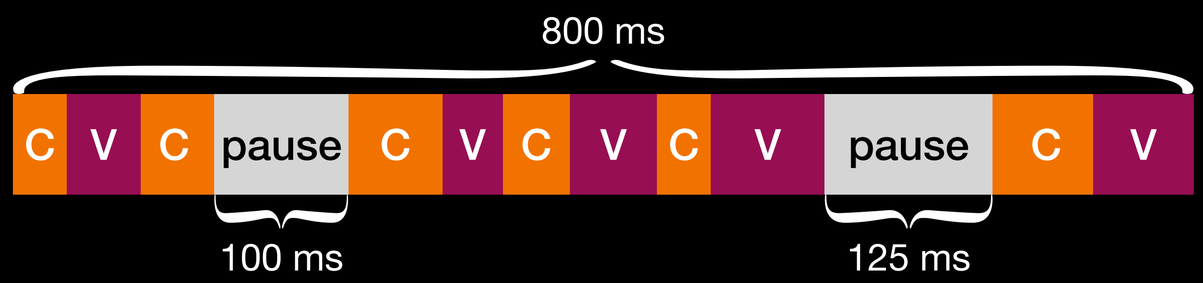
Articulation rate in consonants and vowels

Results and methodological challenges from a cross-linguistic corpus study

Roger Yu-Hsiang **Lo** & Márton **Sóskuthy**
(and thank you to Melissa Wang, Michelle Kamigaki-Baron, Noah Luntzlara)
Origins of Patterns in Speech Lab (OoPS-Lab) | Department of Linguistics | University of British Columbia

Background

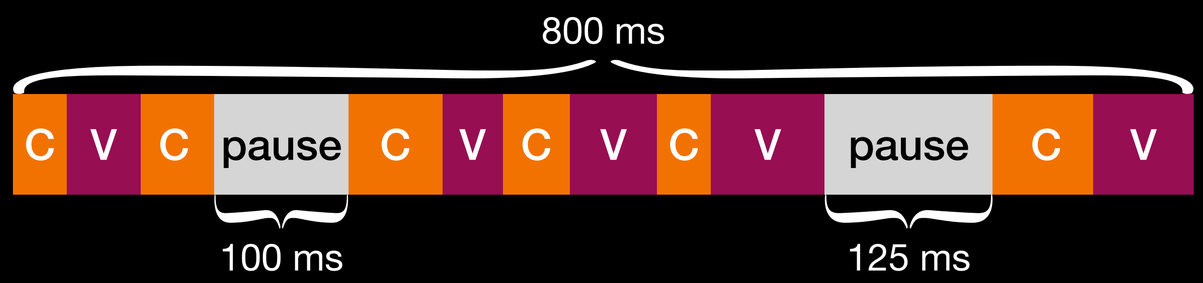
- **Rate**: “the number of output units per unit of time”[1]
  - Speaking/speech rate: pause intervals are included = (6 C + 5 V) / 800
  - Articulation rate: pause intervals are **not** included = (6 C + 5 V) / (800 - pause - pause)



1 Tsao, Weisner & Iqbal, 2006
2 Wood, 1973; Gay, 1978; Port, 1981; Miller, Grosjean, Lomanto, 1984; Crystal & House, 1988; Crystal & House, 1990; Plug & Smith, 2021

Background

- **Rate**: “the number of output units per unit of time”[1]
  - Speaking/speech rate: pause intervals are included = (6 C + 5 V) / 800
  - Articulation rate: pause intervals are **not** included = (6 C + 5 V) / (800 - pause - pause)

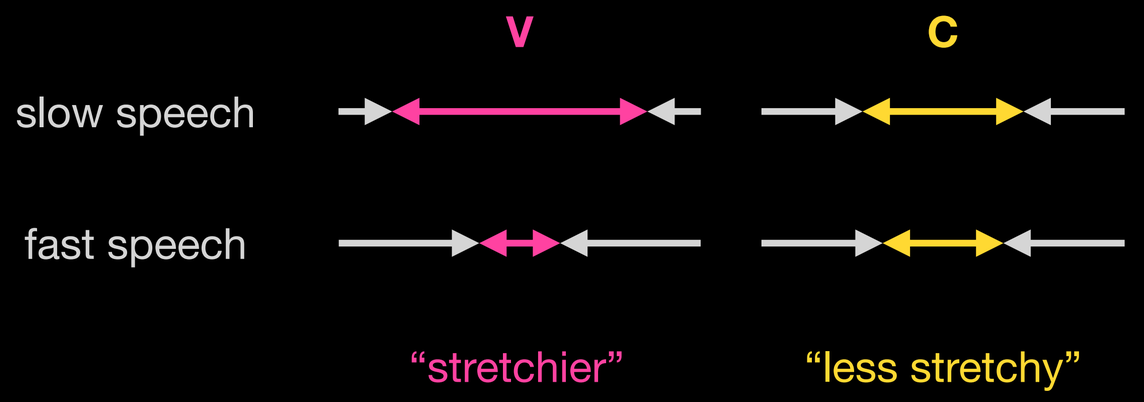


- **Articulation rate** variation is well-documented[2]

1 Tsao, Weisner & Iqbal, 2006
2 Wood, 1973; Gay, 1978; Port, 1981; Miller, Grosjean, Lomanto, 1984; Crystal & House, 1988; Crystal & House, 1990; Plug & Smith, 2021

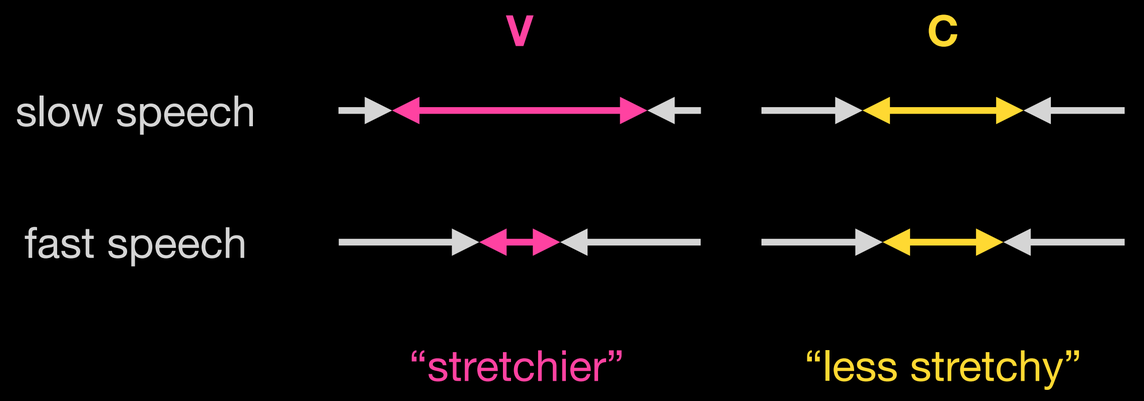
Research Questions

- How does the duration of **different types of segments** vary in response to **local** changes in articulation rate?
  - Consonants vs. vowels



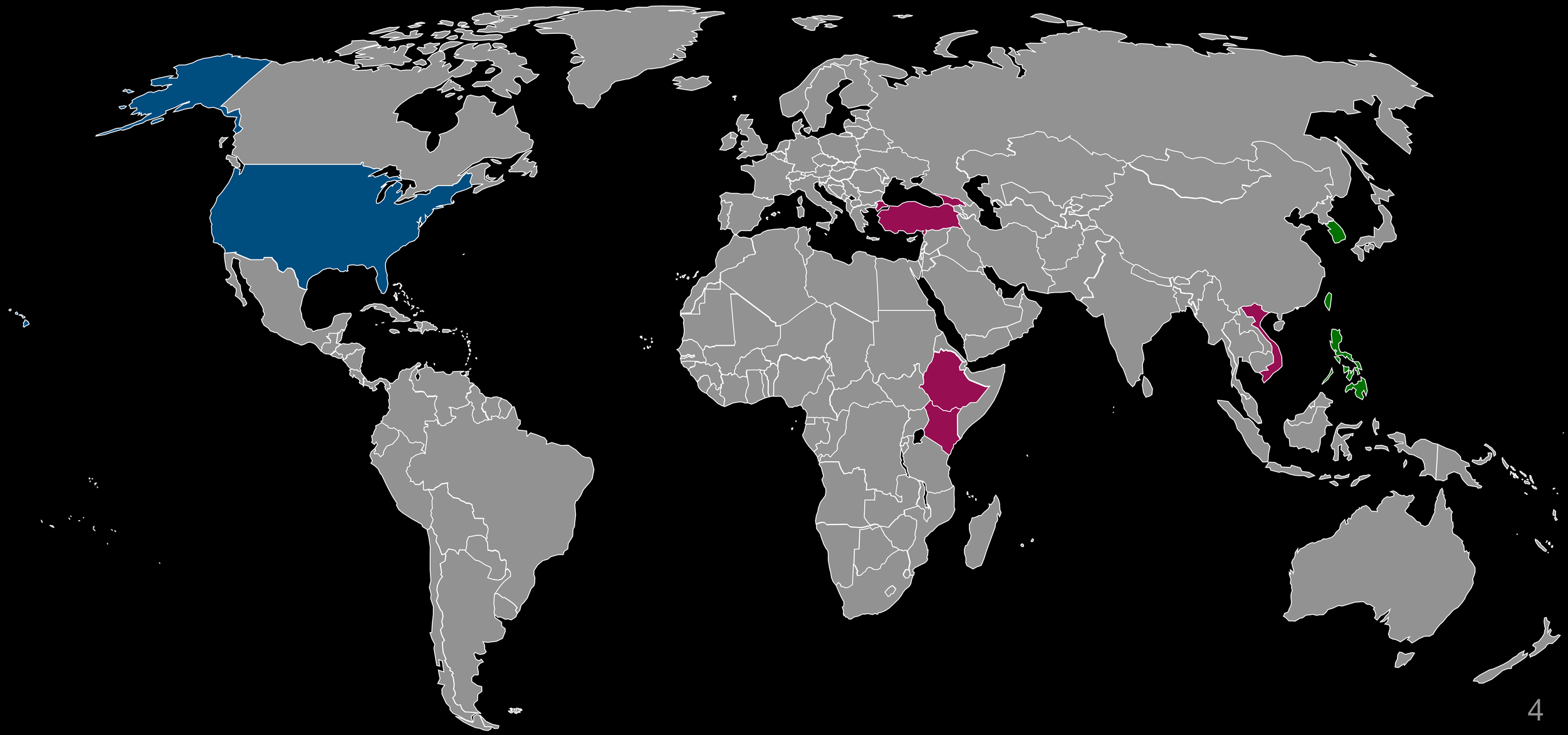
Research Questions

- How does the duration of **different types of segments** vary in response to **local** changes in articulation rate?
  - Consonants vs. vowels



- **Methodological** issues concerning corpus data
  - 9 unrelated languages, and at least 20 speakers per language
  - Control for alignment quality and segment types

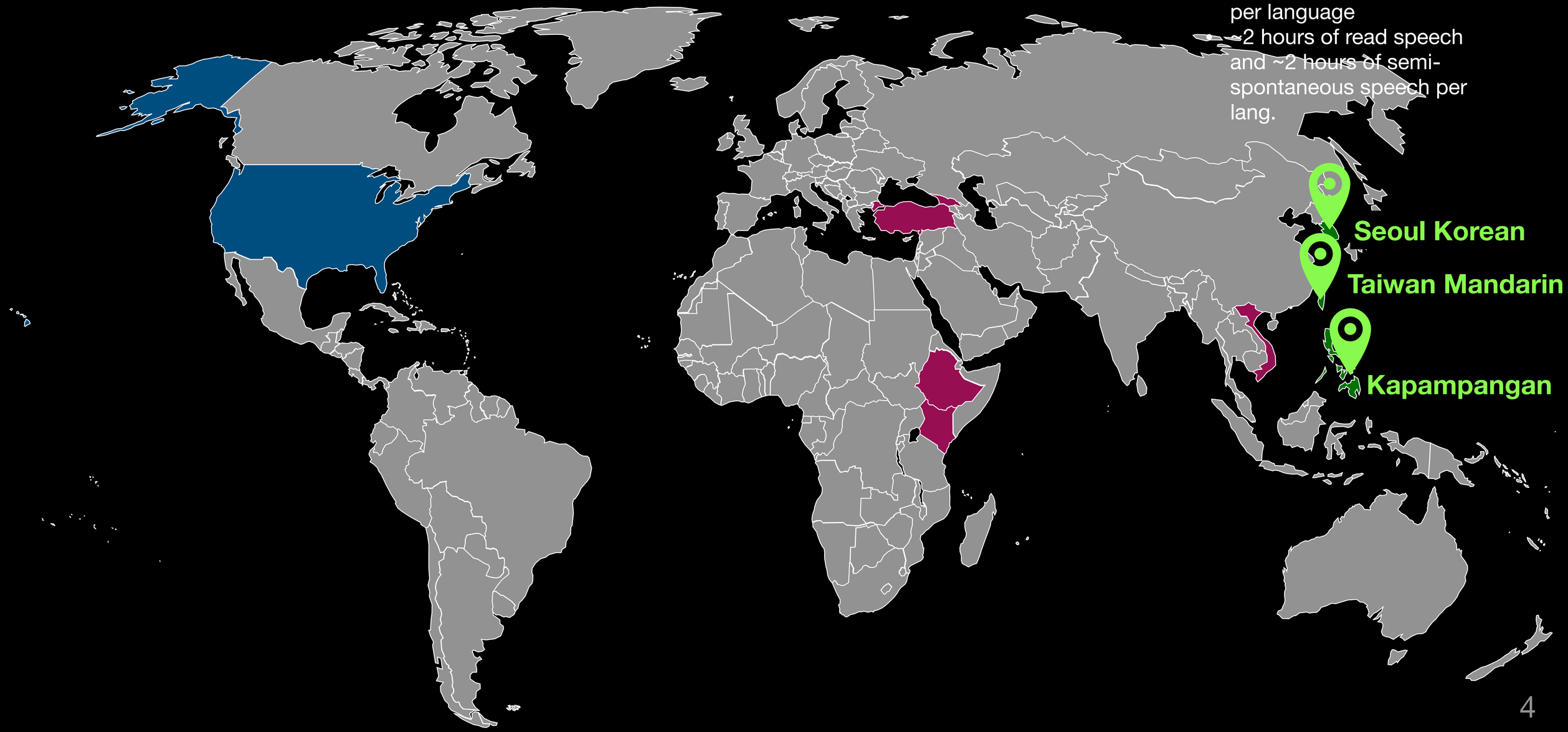
Methods Dataset construction



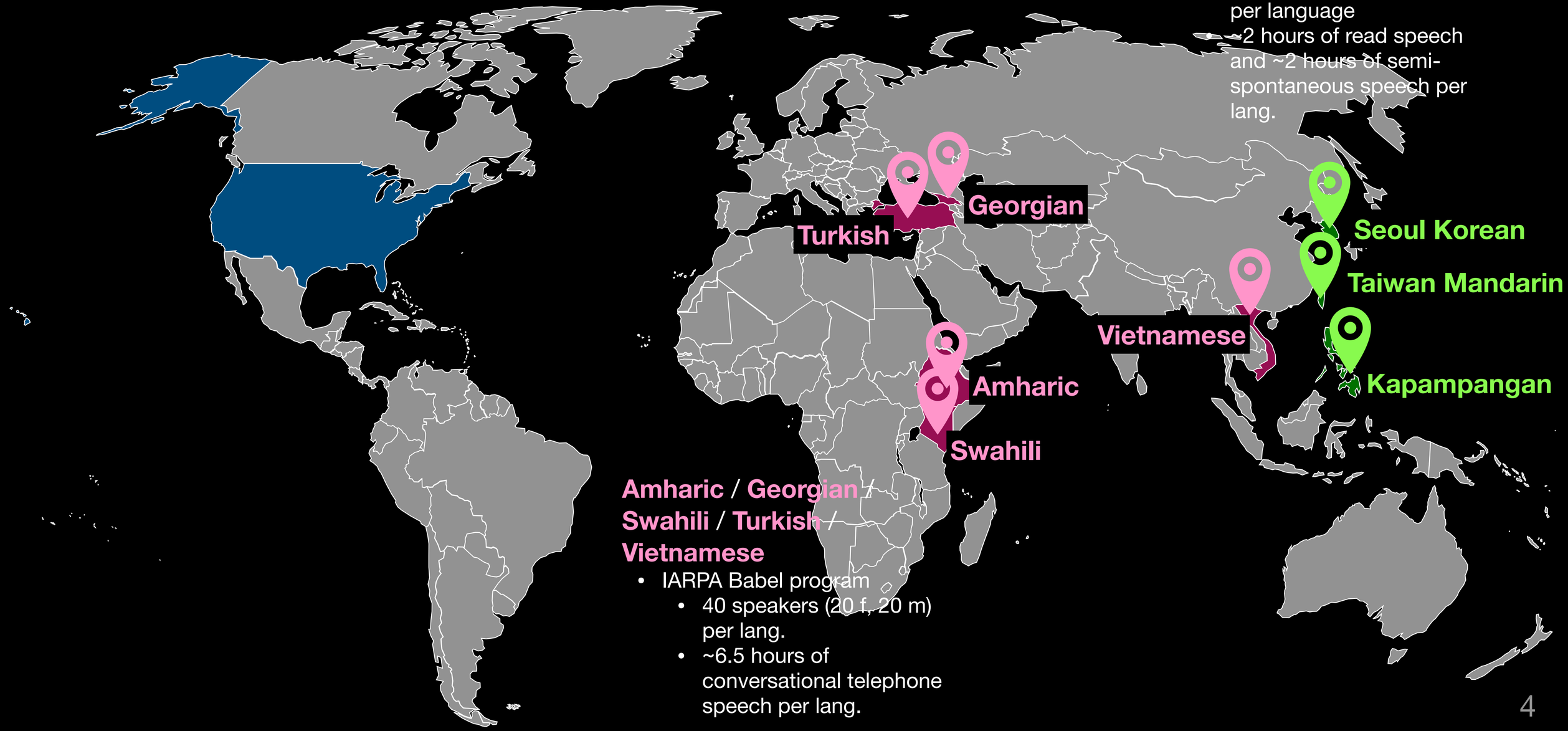
Methods Dataset construction

Kapampangan / **Seoul Korean** / **Taiwan Mandarin**

- OoPS-Lab speech corpora
  - 20 speakers (10 f, 10 m) per language
  - ~2 hours of read speech and ~2 hours of semi-spontaneous speech per lang.



Methods Dataset construction



Kapampangan / Seoul Korean / Taiwan Mandarin

- OoPS-Lab speech corpora
  - 20 speakers (10 f, 10 m) per language
  - ~2 hours of read speech and ~2 hours of semi-spontaneous speech per lang.

Amharic / Georgian / Swahili / Turkish / Vietnamese

- IARPA Babel program
  - 40 speakers (20 f, 20 m) per lang.
  - ~6.5 hours of conversational telephone speech per lang.

Methods Dataset construction

- **American English**
  - Buckeye corpus
    - 40 speakers (20 f, 20 m)
    - ~40 hours of spontaneous speech

Amharic / Georgian / Swahili / Turkish / Vietnamese

- IARPA Babel program
  - 40 speakers (20 f, 20 m) per lang.
  - ~6.5 hours of conversational telephone speech per lang.

Kapampangan / Seoul Korean / Taiwan Mandarin

- OoPS-Lab speech corpora
  - 20 speakers (10 f, 10 m) per language
  - ~2 hours of read speech and ~2 hours of semi-spontaneous speech per lang.

Methods

Measurement

- Forced-aligned with the Montreal Forced Aligner[1]
  - Except for the Buckeye English data

1 McAuliffe et al., 2017a
2 McAuliffe et al., 2017b

Methods

Measurement

- Forced-aligned with the Montreal Forced Aligner[1]
  - Except for the Buckeye English data
- Duration data managed and extracted using PolyglotDB[2]
  - Total
    - Amharic: 2,877; English: 12,515; Georgian: 2,650; Kapampangan: 2,501; Korean: 4,257; Swahili: 2,891; Mandarin: 3,154; Turkish: 2,794; Vietnamese: 2,329

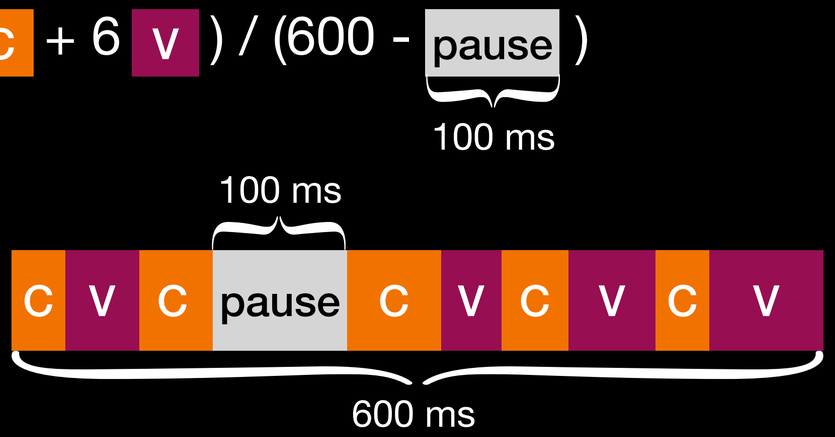
1 McAuliffe et al., 2017a
2 McAuliffe et al., 2017b

Methods

Measurement

- Forced-aligned with the Montreal Forced Aligner[1]
  - Except for the Buckeye English data
- Duration data managed and extracted using PolyglotDB[2]
  - Total
    - Amharic: 2,877; English: 12,515; Georgian: 2,650; Kapampangan: 2,501; Korean: 4,257; Swahili: 2,891; Mandarin: 3,154; Turkish: 2,794; Vietnamese: 2,329
- Measurements of articulation rate
  - Local articulation rate := #seg. / (utt. dur. – pause) = (5 C + 6 V) / (600 - pause)
  - **Average segment duration** = 1 / local articulation rate

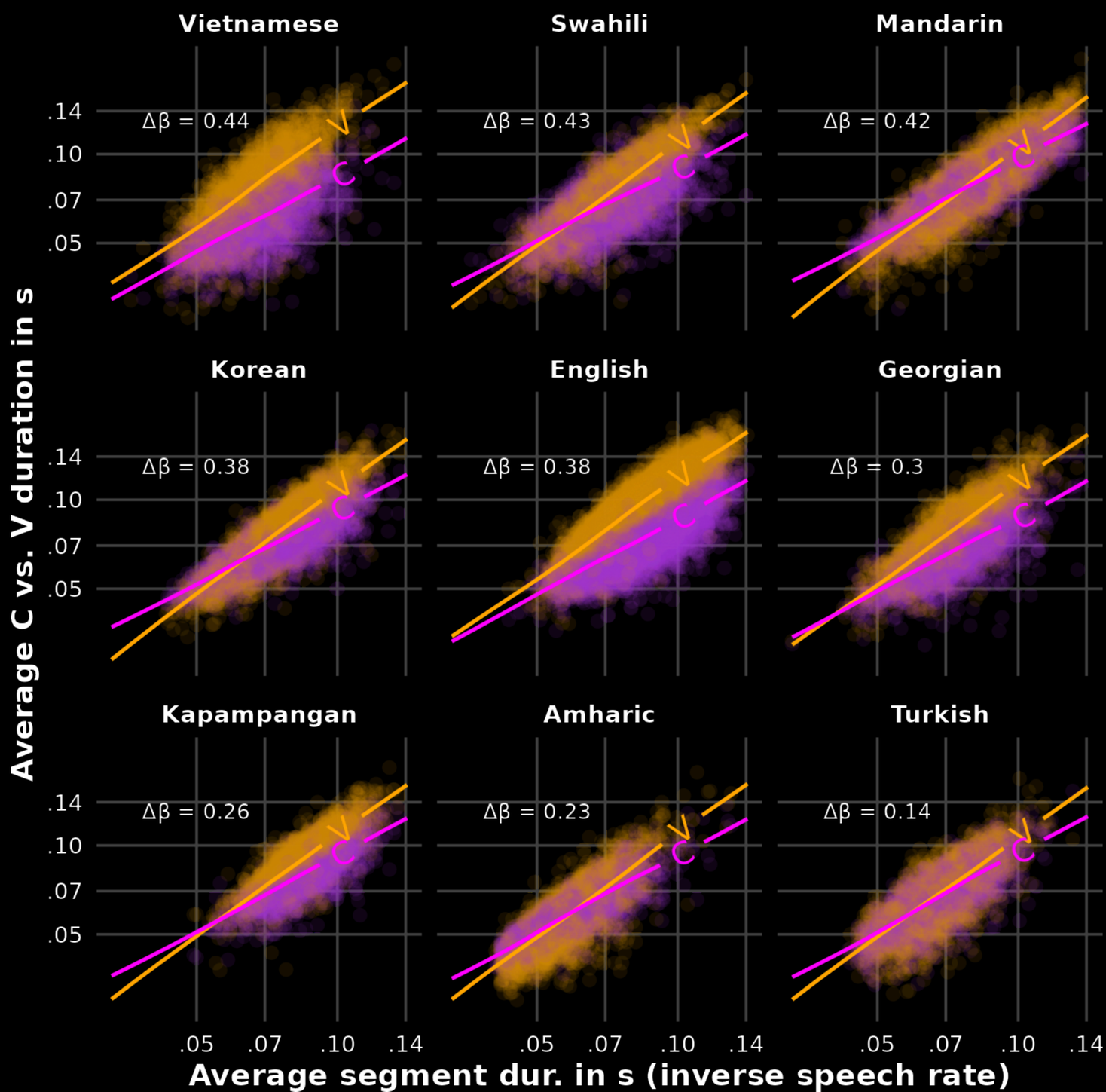
    ⇒ higher avg. segment duration → slower speech



1 McAuliffe et al., 2017a
2 McAuliffe et al., 2017b

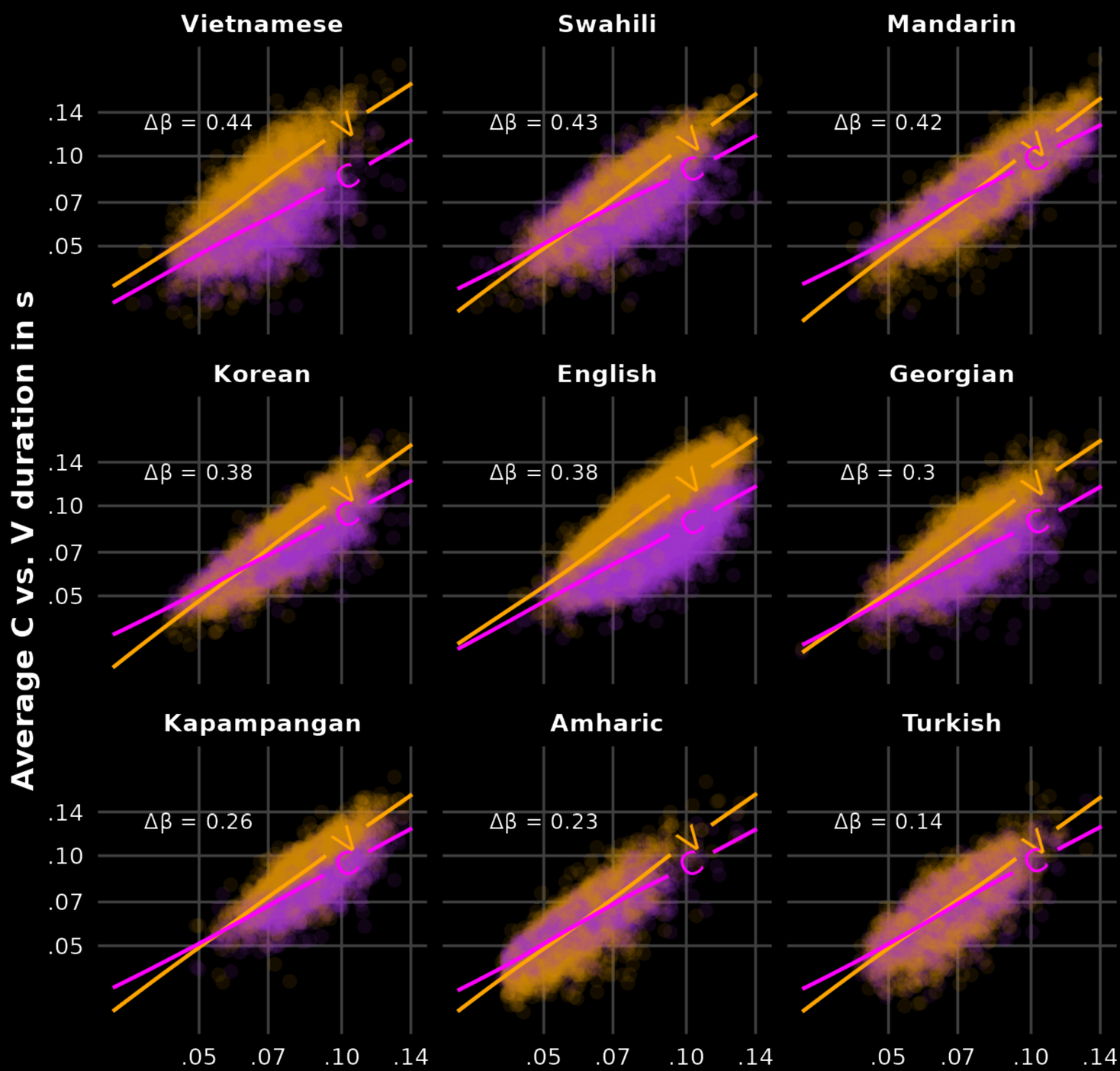
Results

Baseline



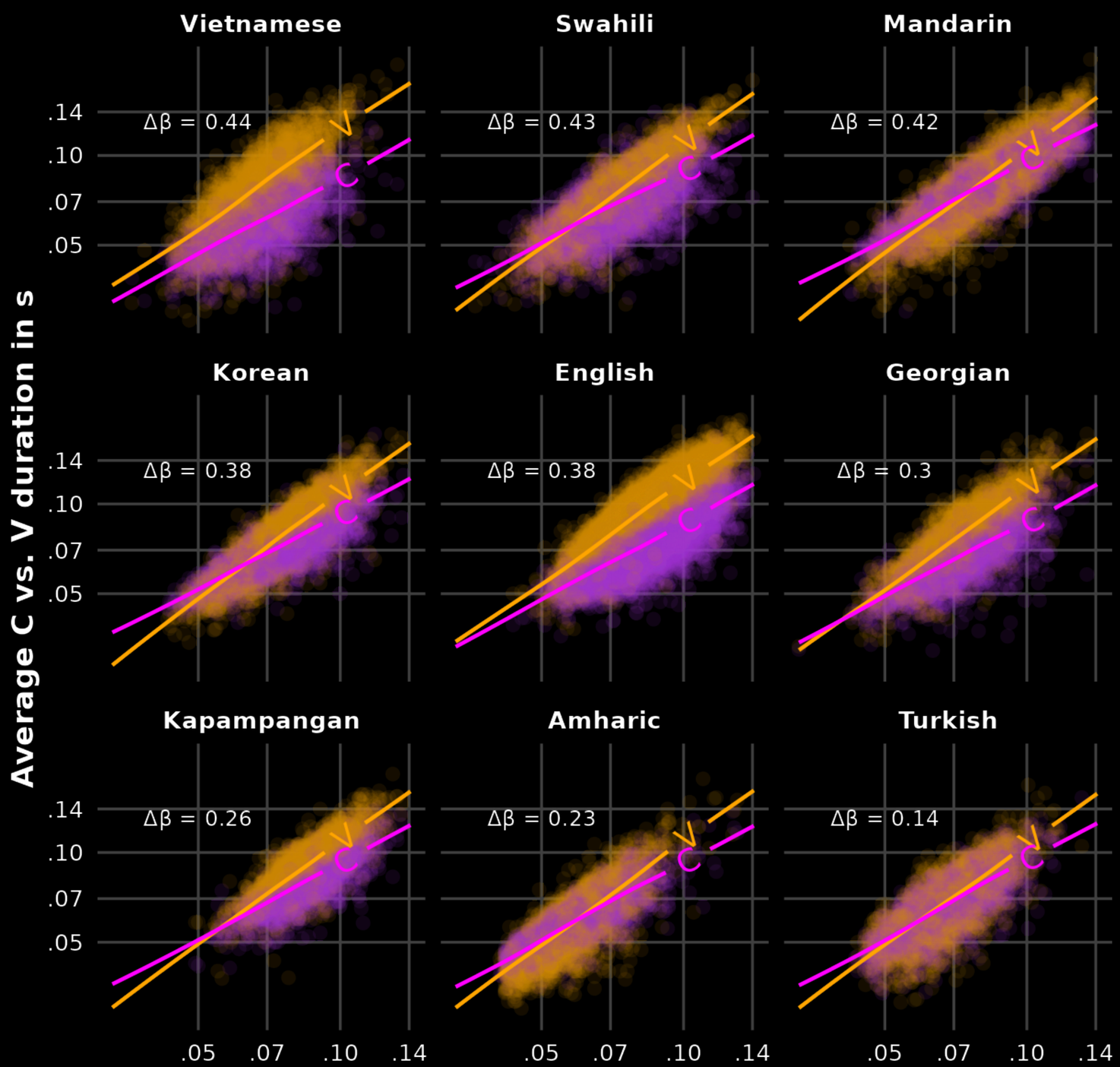
Results

Baseline



Results

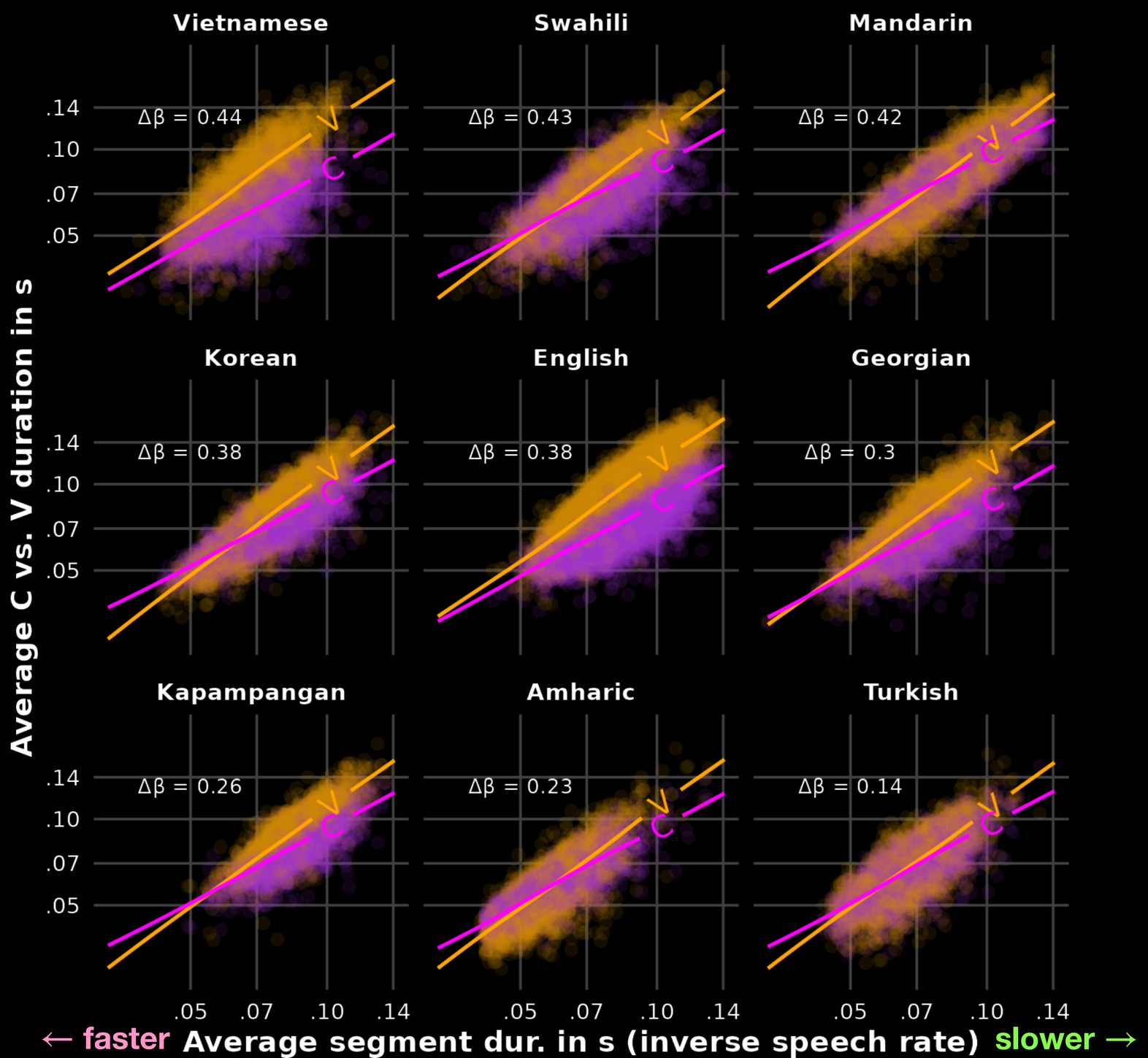
Baseline



Results

Baseline

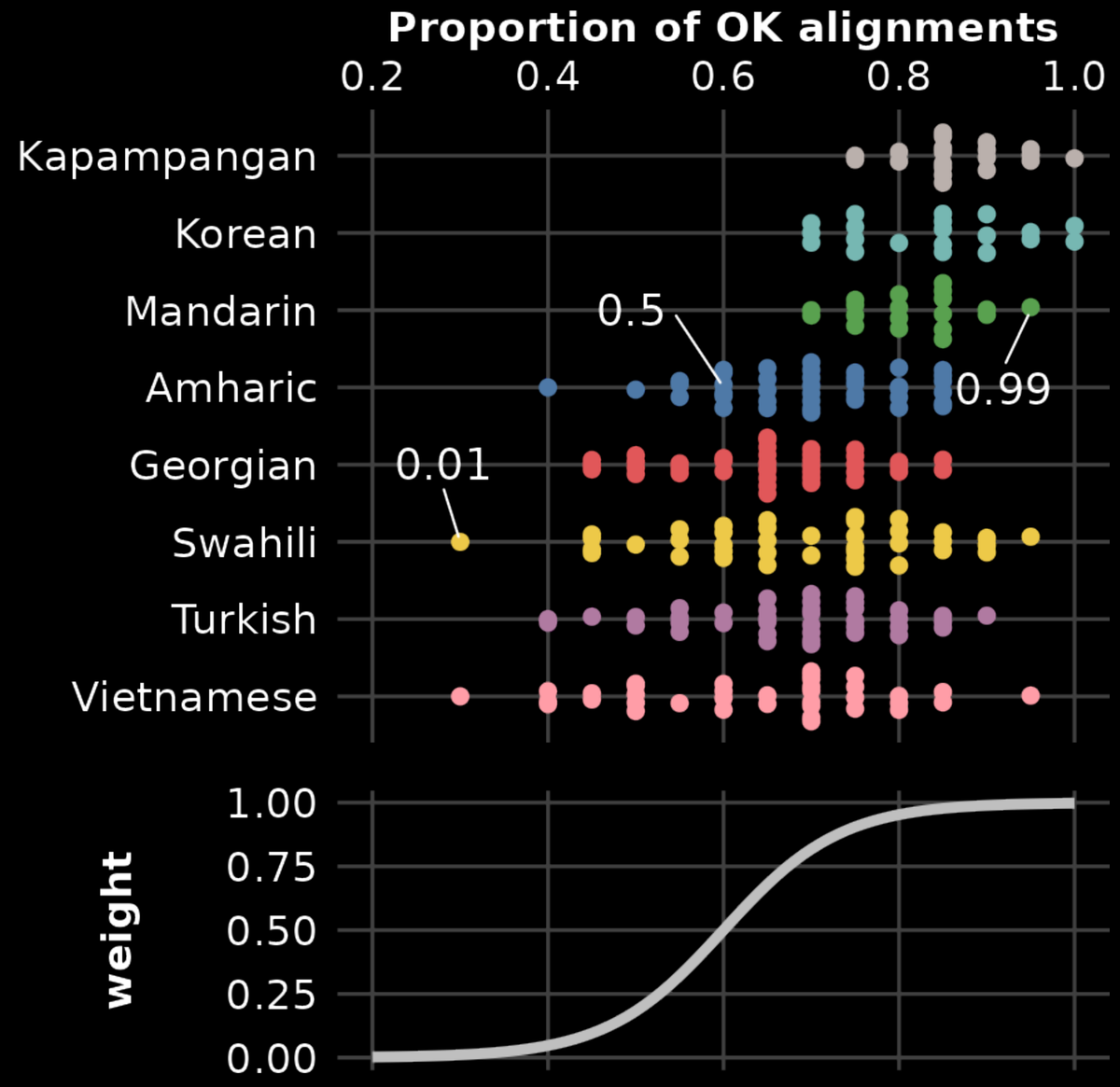
- Vs undergo greater duration adjustment than Cs
  - **Fast**: Vs same or shorter than Cs
  - **Slow**: Vs up to 1.5x longer than Cs



Results

Alignment quality

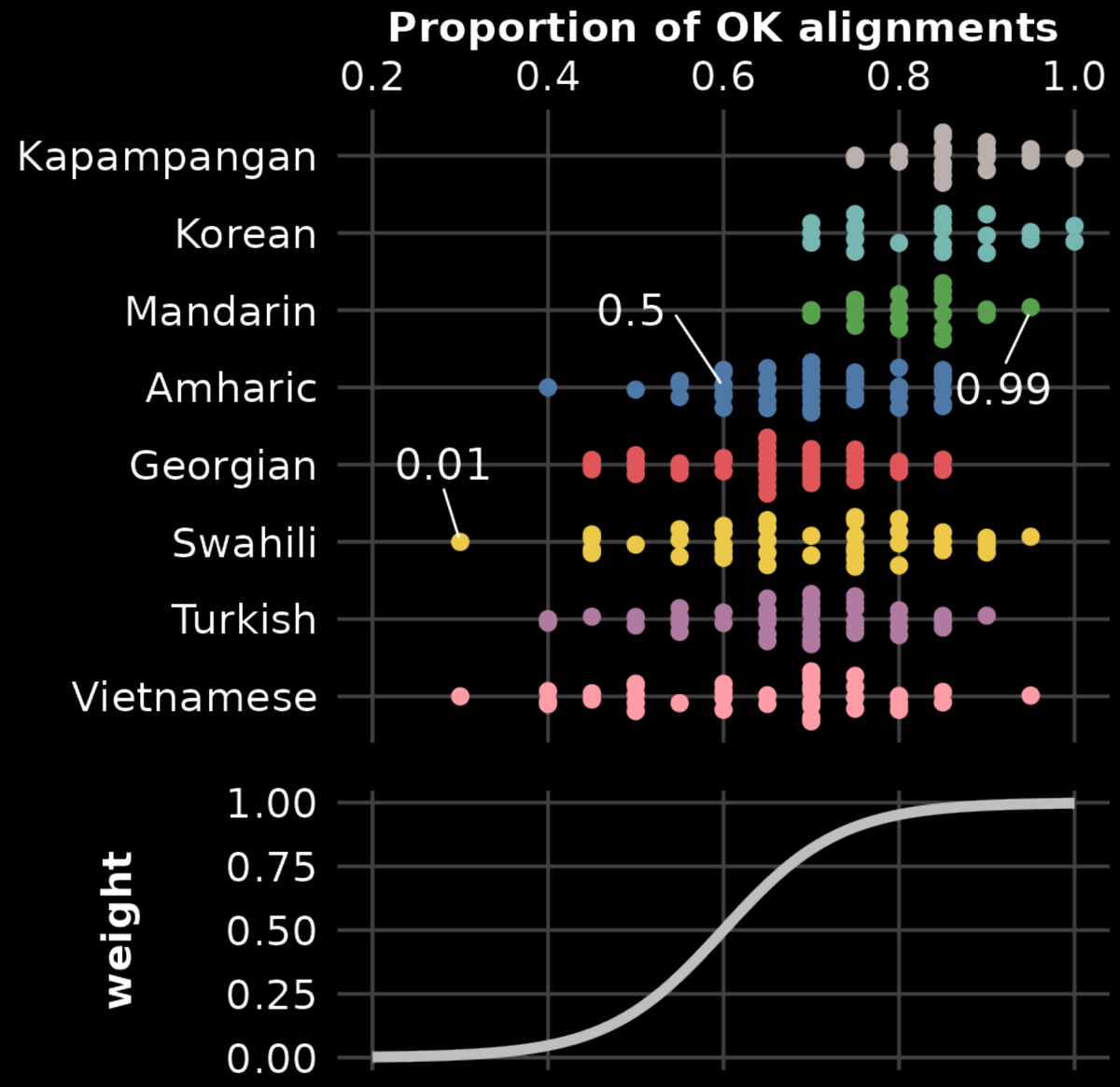
- 20 randomly sampled aligned segments for each speaker
  - **OK**: < 50% aligned interval misplaced
  - **Misaligned**: > 50% aligned interval misplaced



Results

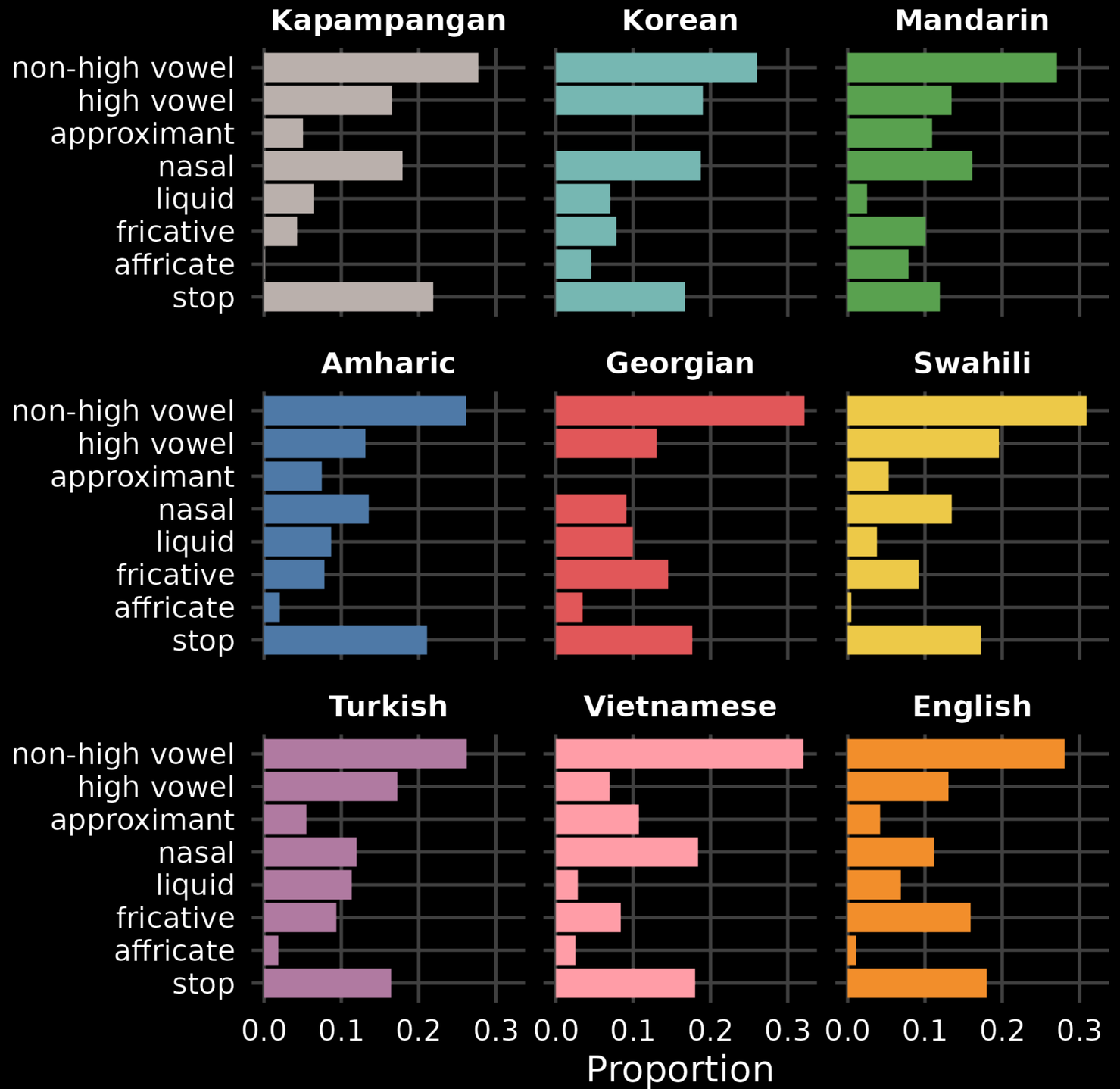
Alignment quality

- 20 randomly sampled aligned segments for each speaker
  - **OK**: < 50% aligned interval misplaced
  - **Misaligned**: > 50% aligned interval misplaced
- Proportion of OK alignment passed through a logistic function to derive a **weight** for each speaker
  - Weighted GAMM



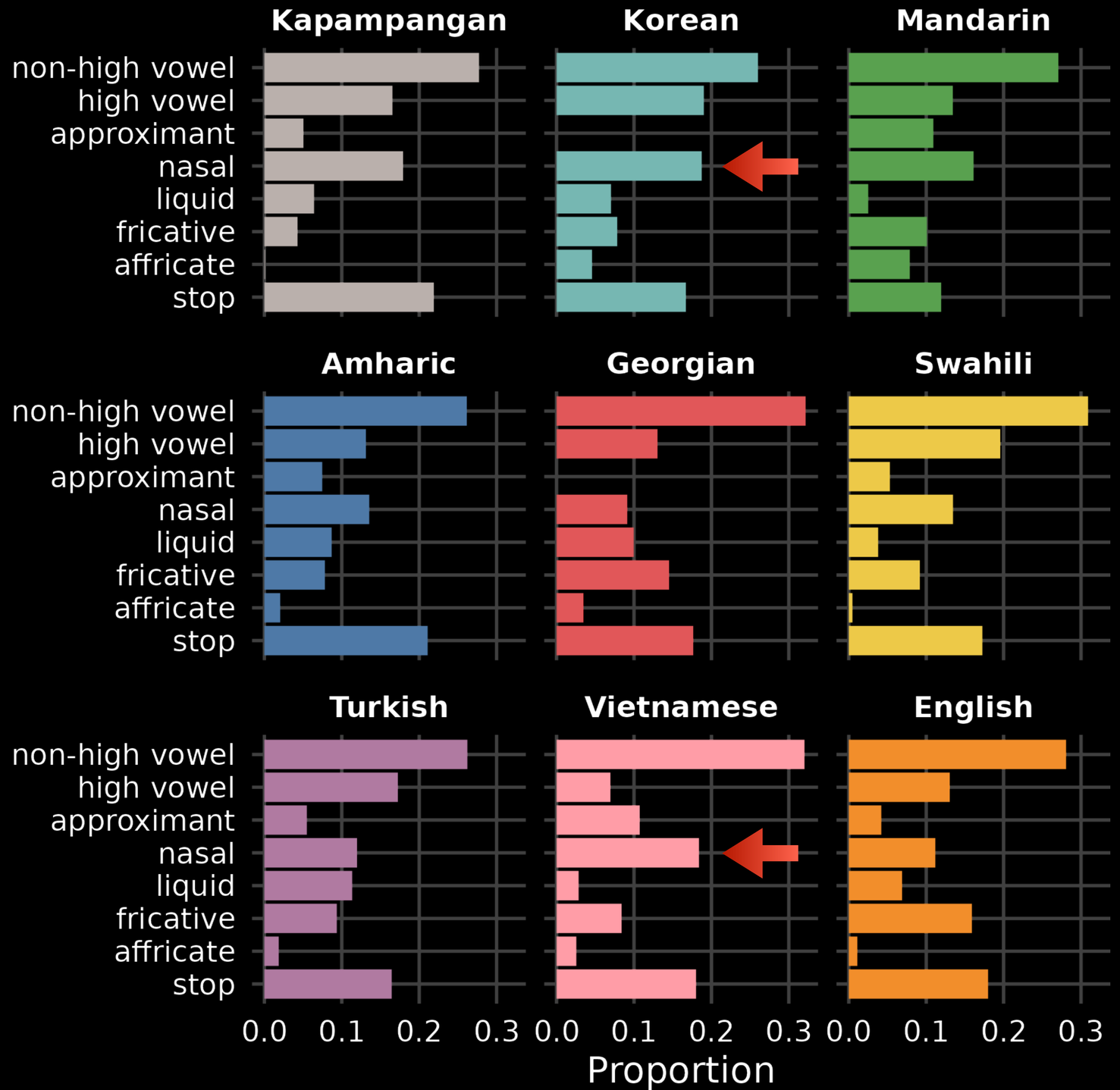
Results

Segment frequency



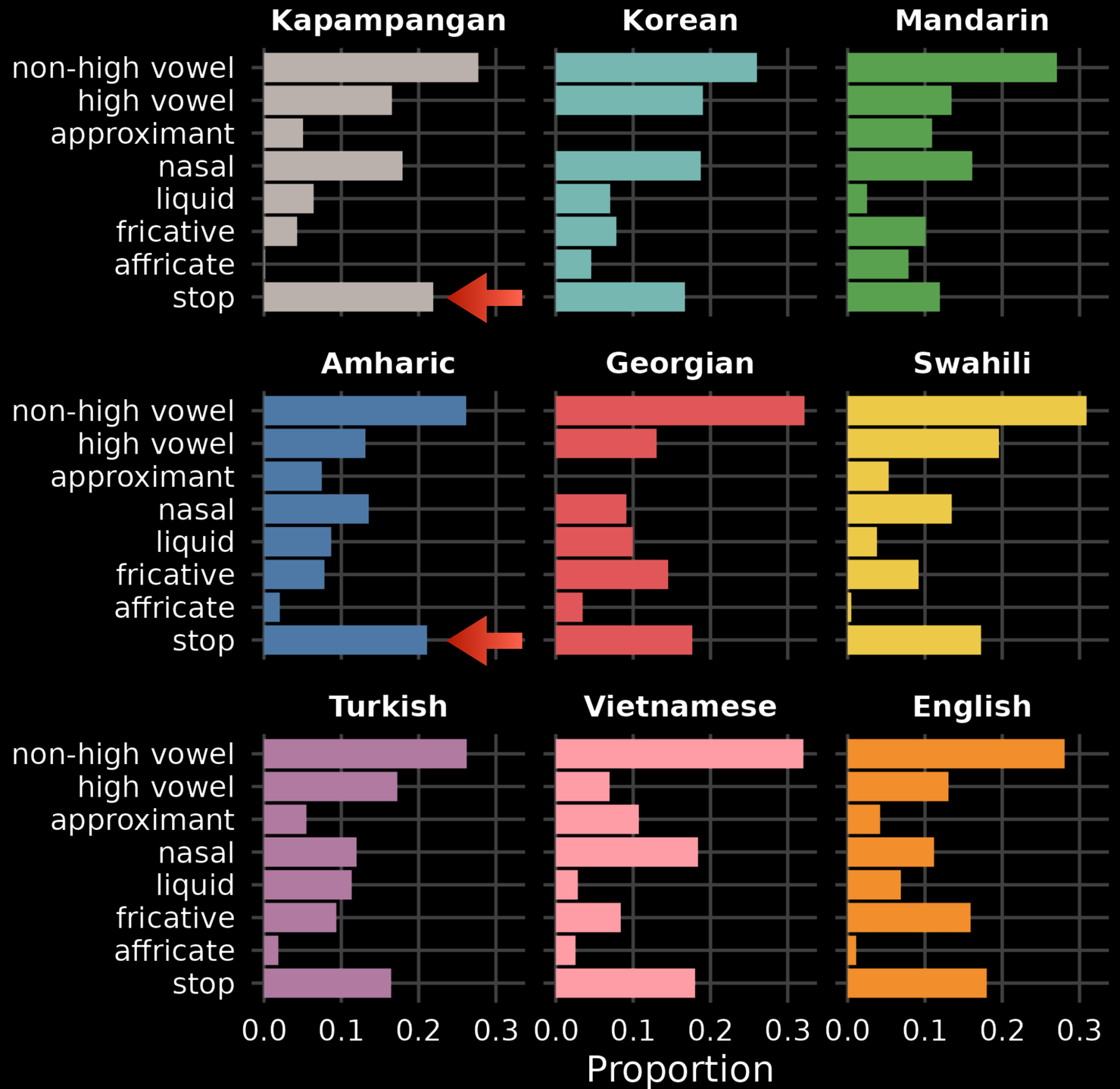
Results

Segment frequency



Results

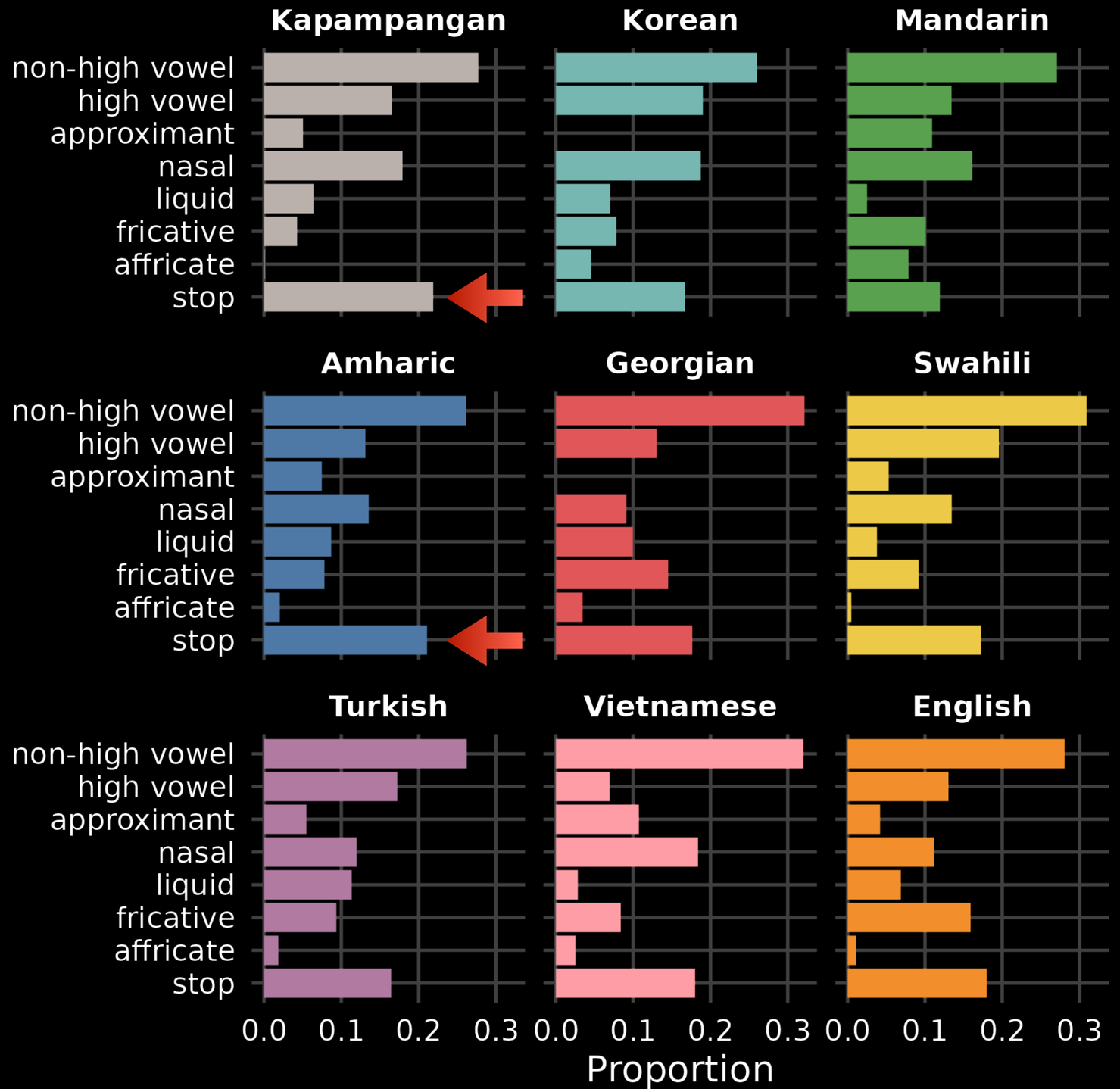
Segment frequency



Results

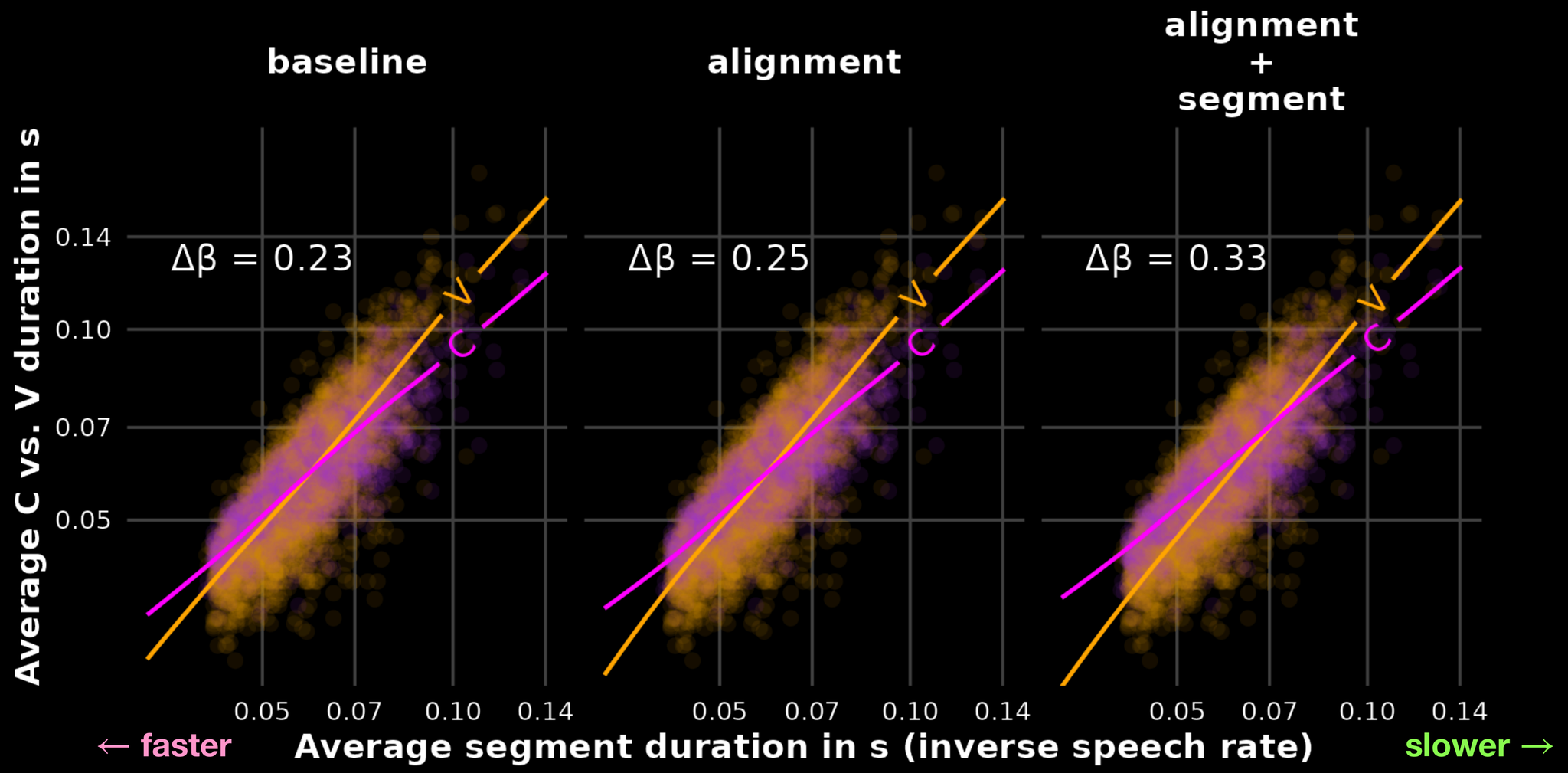
Segment frequency

- Controls for differing segment distributions across languages
  - **Random smooth** by segment type



Results

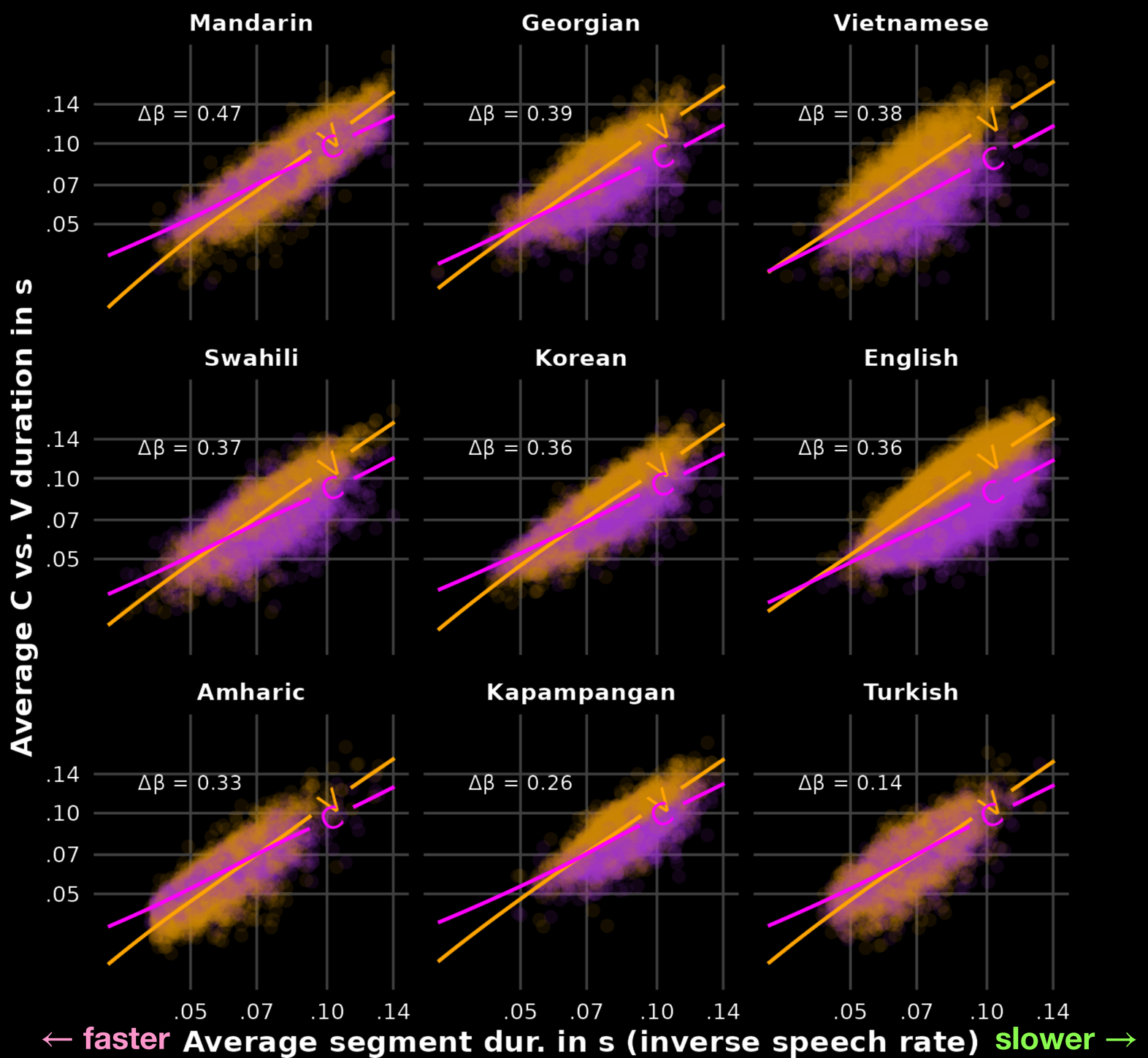
Differences in the effect size for Amharic



Results

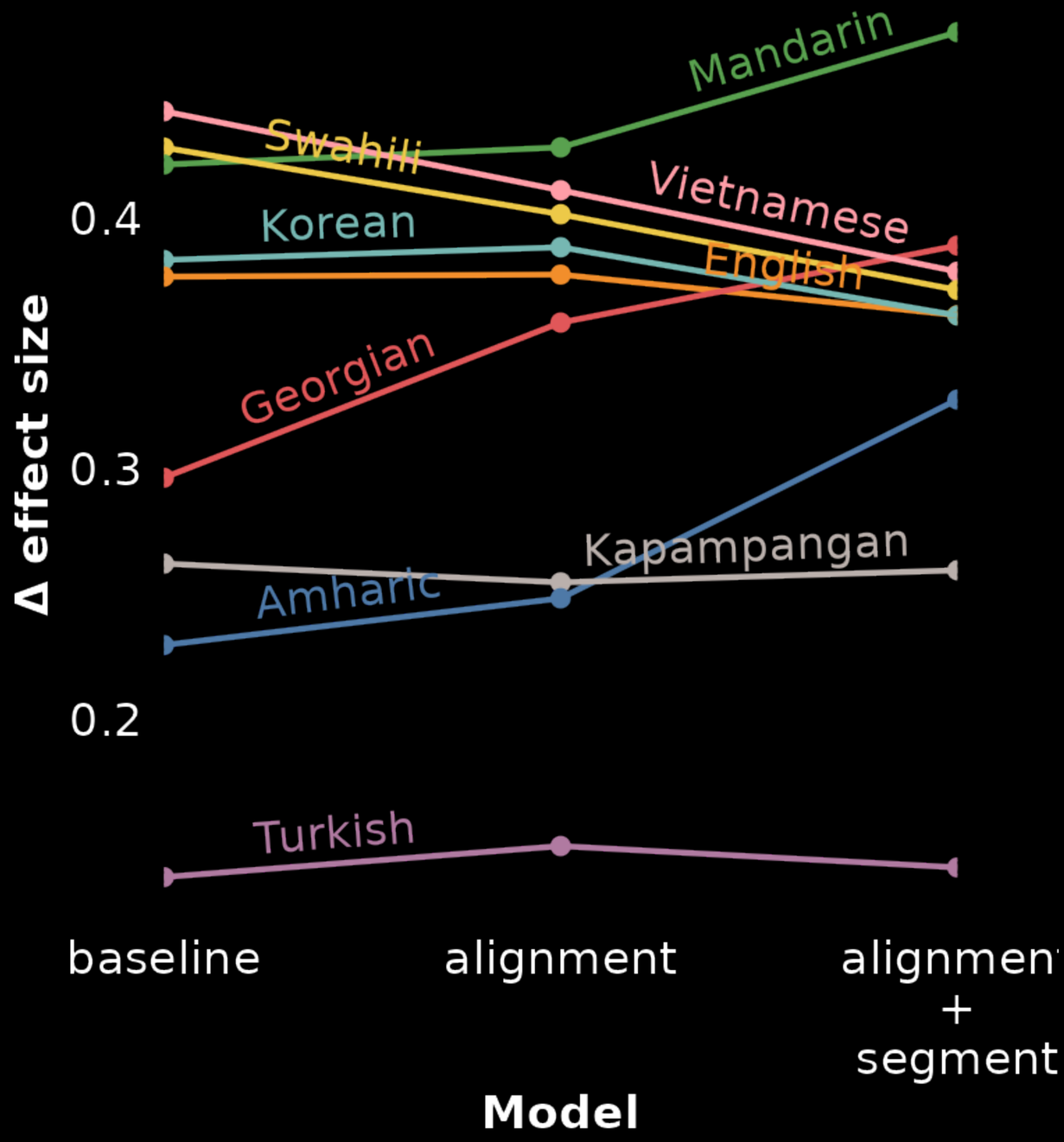
Alignment + segment

- Vs still undergo greater duration adjustment than Cs



Results

Change in effect sizes



Summary and Discussion

- Across all languages, vowels are “stretchier” than consonant (cf. 1)
  - Stable even after controlling for alignment quality and segment types
  - Can be partially explained by reference to articulatory mechanisms

1 Kozchevnikova & Chistovich, 1965; Port, 1976
2 Sundara, 2005; Reidy, 2016

Summary and Discussion

- Across all languages, vowels are “stretchier” than consonant (cf. 1)
  - Stable even after controlling for alignment quality and segment types
  - Can be partially explained by reference to articulatory mechanisms
- Methodological consideration
  - The conclusions do change as **data quality** and **non-phonetic differences** across languages are brought under statistical control

1 Kozchevnikova & Chistovich, 1965; Port, 1976
2 Sundara, 2005; Reidy, 2016

Summary and Discussion

- Across all languages, vowels are “stretchier” than consonant (cf. 1)
  - Stable even after controlling for alignment quality and segment types
  - Can be partially explained by reference to articulatory mechanisms
- Methodological consideration
  - The conclusions do change as **data quality** and **non-phonetic differences** across languages are brought under statistical control
- Implications
  - Subtle cross-linguistic differences in the phonetic implementation of the “same” sound[2]
  - The importance of accounting for systematic differences across data sources

1 Kozchevnikova & Chistovich, 1965; Port, 1976
2 Sundara, 2005; Reidy, 2016

Summary and Discussion

- Across all languages, vowels are “stretchier” than consonant (cf. 1)
  - Stable even after controlling for alignment quality and segment types
  - Can be partially explained by reference to articulatory mechanisms
- Methodological consideration
  - The conclusions do change as **data quality** and **non-phonetic differences** across languages are brought under statistical control
- Implications
  - Subtle cross-linguistic differences in the phonetic implementation of the “same” sound[2]
  - The importance of accounting for systematic differences across data sources

1 Kozchevnikova & Chistovich, 1965; Port, 1976
2 Sundara, 2005; Reidy, 2016

Thank You!

References

Crystal, Thomas H., and Arthur S. House. 1982. Segmental durations in connected speech signals: Preliminary results. The Journal of the Acoustical Society of America 72:705–716.

Crystal, Thomas H., and Arthur S. House. 1988. Segmental durations in connected-speech signals: Current results. *The Journal of the Acoustical Society of America* 83:1553–1573.

Crystal, Thomas H., and Arthur S. House. 1990. Articulation rate and the duration of syllables and stress groups in connected speech. *The Journal of the Acoustical Society of America* 88:101–112.

Gay, Thomas. 1978. Effect of speaking rate on vowel formant movements. *The Journal of the Acoustical Society of America* 63:223–230.

Kozhevnikov, V. A., and L. A. Chistovich. 1965. *Speech: Articulation and perception*. Washington, D.C.: Joint Publications Research Service.

McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner, and Morgan Sonderegger. 2017a. Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. In *Proceedings of INTERSPEECH 2017*, 498–502.

McAuliffe, Michael, Elias Stengel-Eskin, Michaela Socolof, and Morgan Sonderegger. 2017b. Polyglot and Speech Corpus Tools: A system for representing, integrating, and querying speech corpora. In *Proceedings of INTER- SPEECH 2017*, 3887–3891.

Miller, Joanne L., François Grosjean, and Concetta Lomanto. 1984. Articulation rate and its variability in sponta- neous speech: A reanalysis and some implications. *Phonetica* 41:215–225.

Plug, Leendert, and Rachel Smith. 2021. The role of segment rate in speech tempo perception by English listeners. *Journal of Phonetics* 86:1–16.

Port, Robert F. 1976. The influence of speaking tempo on the duration of stressed vowel and medial stop in English trochee words. Doctoral Dissertation, University of Connecticut, Storrs, CT.

Port, Robert F. 1978. Effects of word-internal versus word-external tempo on the voicing boundary for medici stop closure. *The Journal of the Acoustical Society of America* 63:S20.

Port, Robert F. 1981. Linguistic timing factors in combination. *The Journal of the Acoustical Society of America* 69:262– 274.

Reidy, Patrick F. 2016. Spectral dynamics of sibilant fricatives are contrastive and language specific. *The Journal of the Acoustical Society of America* 140:2518–2529.

Sundara, Megha. 2005. Acoustic-phonetics of coronal stops: A cross-language study of Canadian English and Canadian French. *The Journal of the Acoustical Society of America* 118:1026–1037.

Tsao, Yinh-Chiao, Gary Weismer, and Kamran Iqbal. 2006. Interspeaker variation in habitual speaking rate: Ad- ditional evidence. *Journal of Speech, Language, and Hearing Research* 49:1156–1164.

Wood, Sidney. 1973. What happens to vowels and consonants when we speak faster? *Working Papers in Linguistics,* Lund University 9:8–39.

Wood, Simon N. 2017. Generalized additive models: An introduction with R. Boca Raton, FL: CRC Press, 2 edition.