



Vowels are “stretchier” than consonants

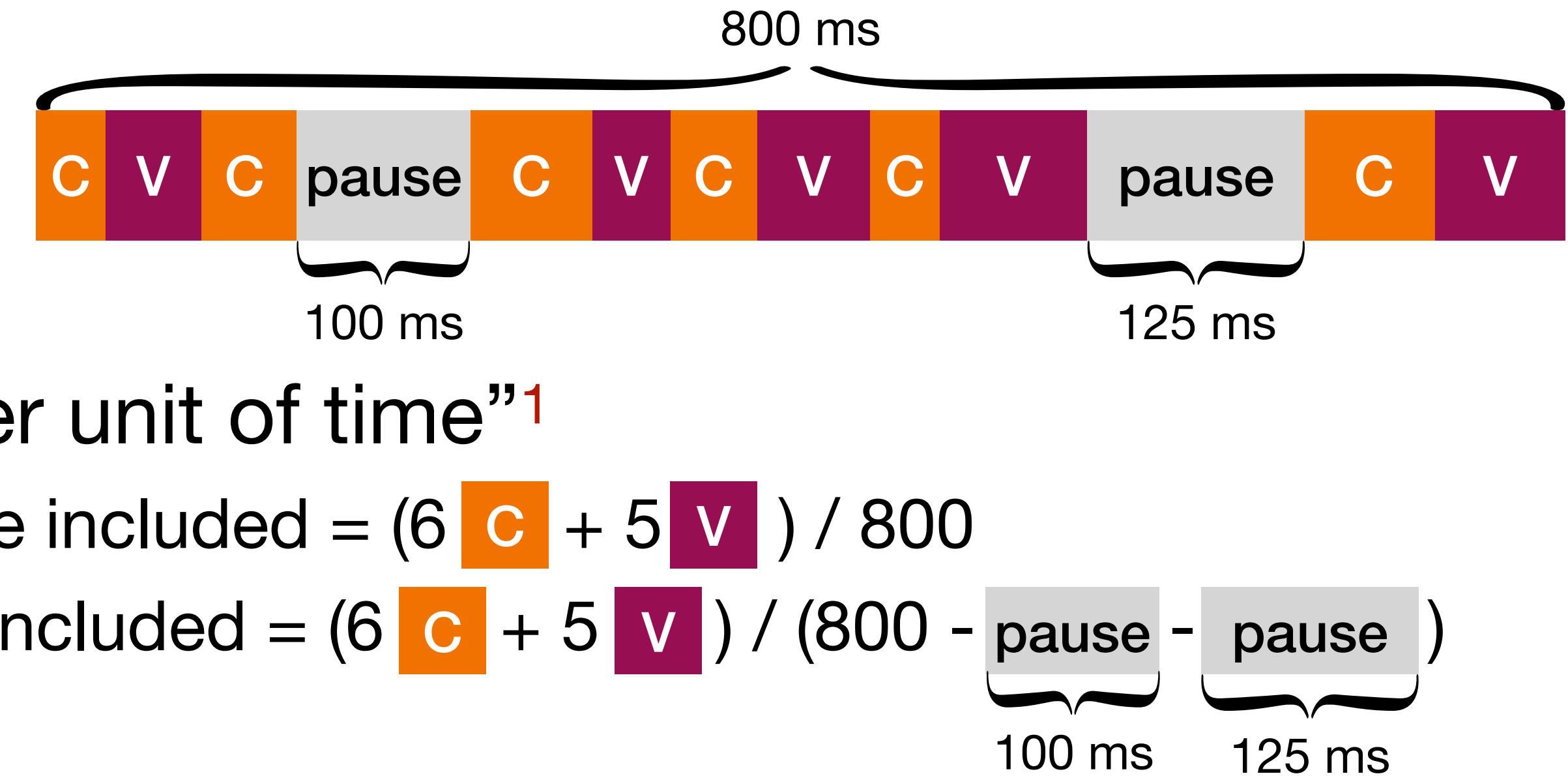
A cross-linguistic corpus study of the segmental implementation of articulation rate

Roadmap

- Background
- Research Questions
- Methods
- Results
- Summary and Discussion



Background



- **Rate:** “the number of output units per unit of time”¹
 - Speaking/speech rate: pause intervals are included = $(6 \text{ C} + 5 \text{ V}) / 800$
 - Articulation rate: pause intervals are **not** included = $(6 \text{ C} + 5 \text{ V}) / (800 - \underbrace{\text{pause}}_{100 \text{ ms}} - \underbrace{\text{pause}}_{125 \text{ ms}})$
- **Articulation rate** variation is well-documented²
 - Tends to be measured at the syllable level (e.g., #syllable / time unit)
 - Varies substantially at both **global** and **local** levels³
 - Correlated with listeners’ speech tempo perception⁴
 - Almost exclusively on English

¹ Tsao, Weisner & Iqbal, 2006

² Wood, 1973; Port, 1981; Miller, Grosjean, Lomanto, 1984; Gay, 1978; Crystal & House, 1988; Crystal & House, 1990

³ Miller, Grosjean, Lomanto, 1984

⁴ Port, 1978; Plug & Smith, 2021

Background

- Relatively little is known about articulation rate at **segmental** level
 - Both **consonants** and **vowels** are shortened at a higher rate¹
 - **Conflicting** findings
 - Change occurs mainly in **vowels**²
 - Constant %C at different rates³
 - **Explicit** instructions to speak fast / slowly
 - Limited by the **number** of participants
 - 2 English speaker for K&C [1965]²
 - 1 speaker per language in Wood [1973]³

¹ Crystal & House, 1982, 1988

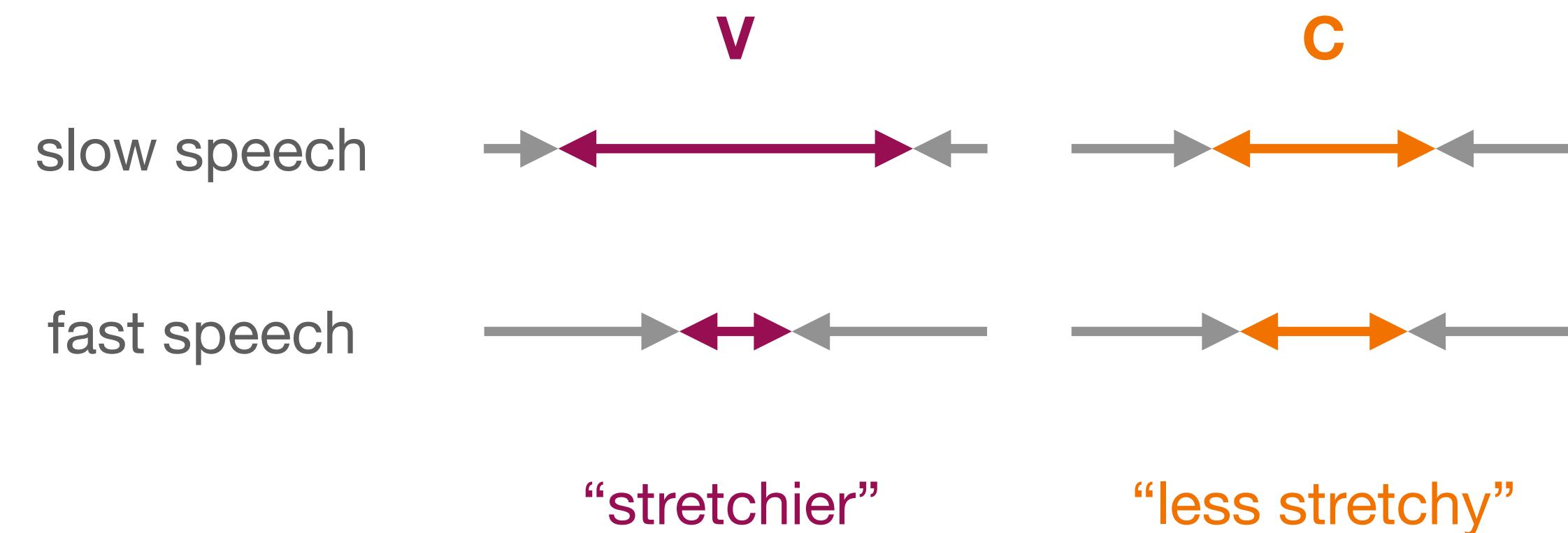
² Kozchevnikova & Chistovich, 1965; Port, 1976

³ Wood, 1973

Research Questions

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

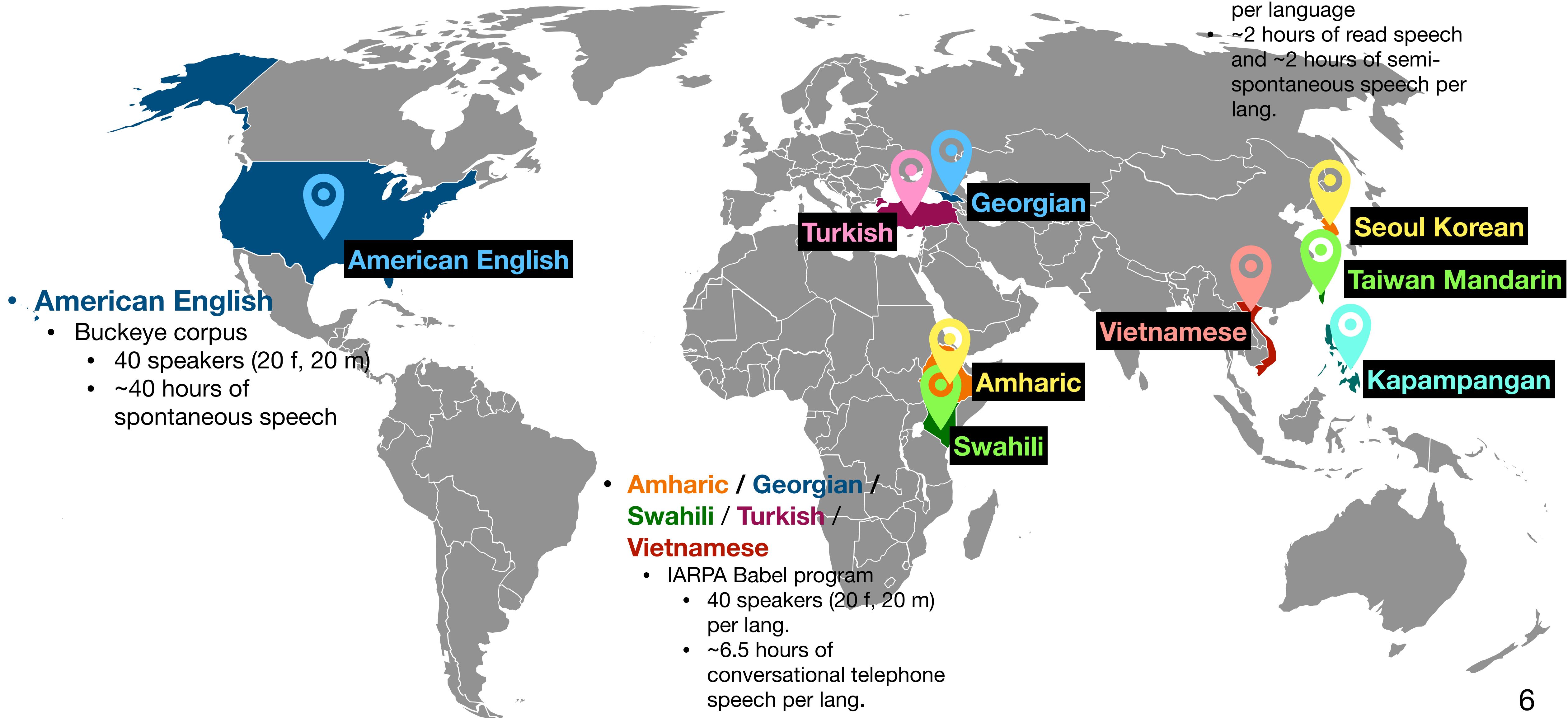
1. Consonants vs. vowels
2. Different consonant / vowel types



- Corpus data from **9 unrelated languages**
- Read / semi-spontaneous / spontaneous speech **without explicit instructions**
- At least **20 speakers** per language

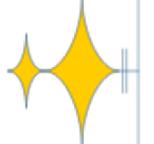
Methods

Dataset construction



Methods

Data annotation and management

- Corpus transcription
 - **OoPS-Lab data:** manually by native speakers of respective languages
 - **Buckeye / IARPA Babel data:** transcriptions come with the corpora
- **Forced-aligned** with the  **Montreal Forced Aligner**¹
 - Except for the Buckeye English data
- **Duration** data managed and extracted using PolyglotDB²
- **Statistical models and visualization** carried out in R



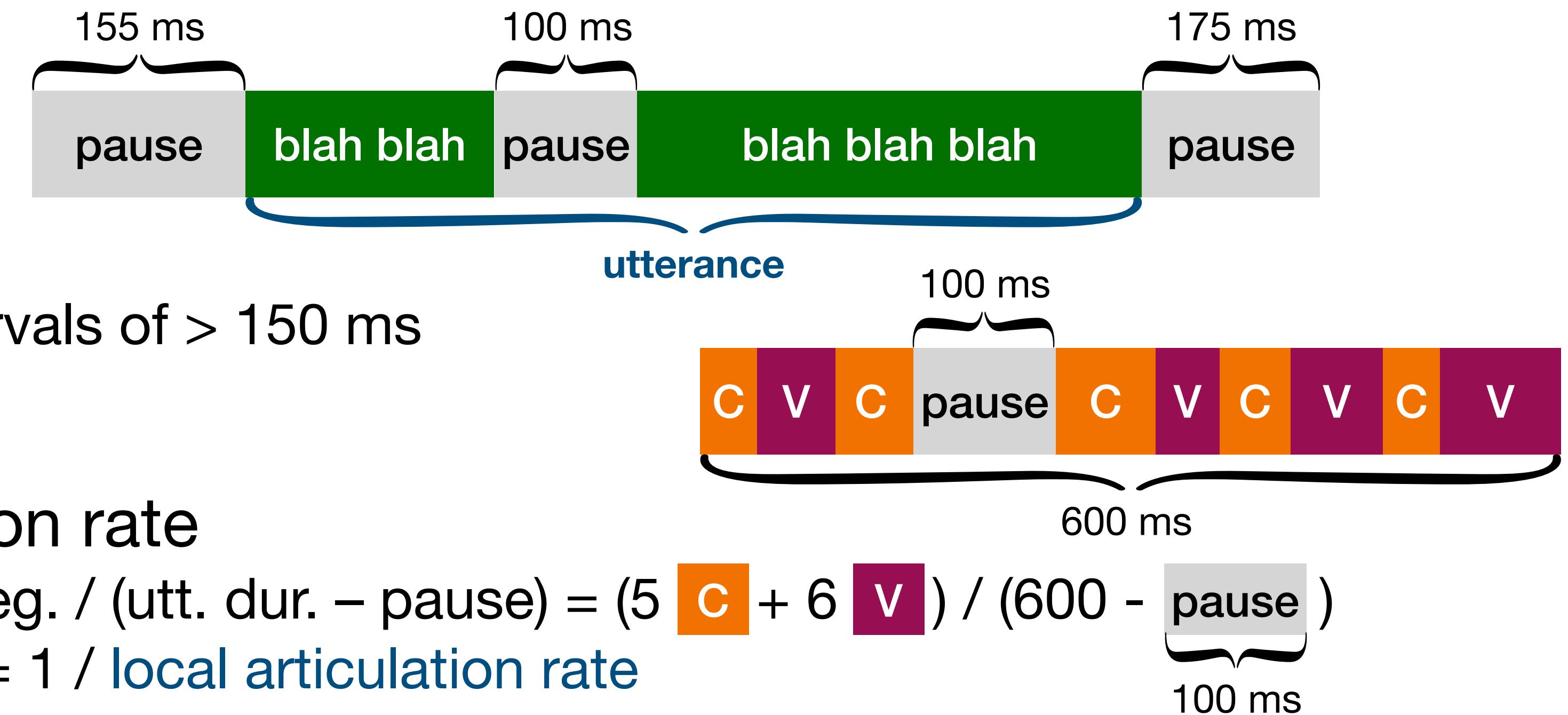
¹ McAuliffe et al., 2017a

² McAuliffe et al., 2017b

Methods

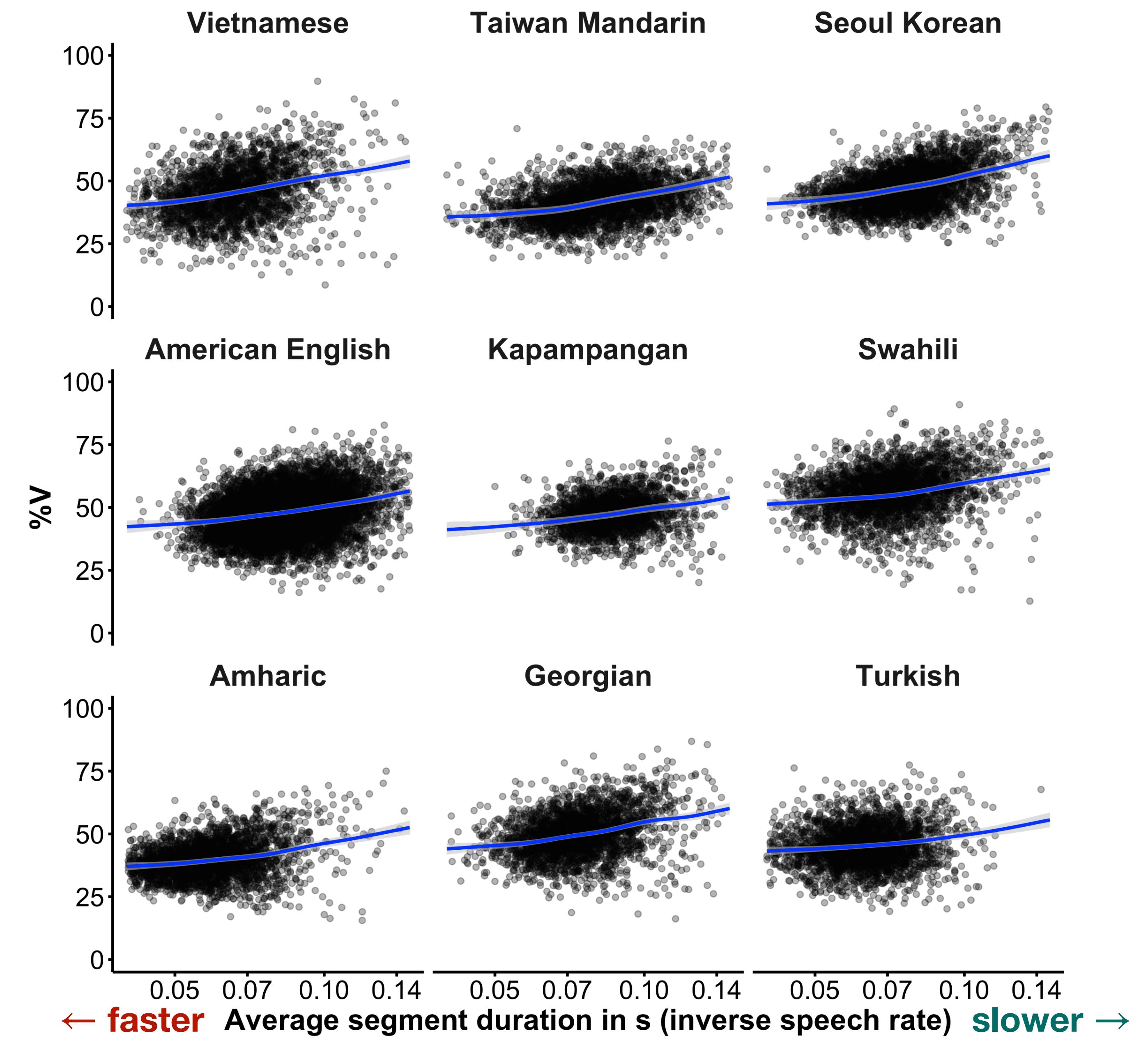
Measurement

- Recordings → utterances
 - Bounded by non-speech intervals of > 150 ms
 - > 5 syllables
- Measurements of articulation rate
 - **Local articulation rate** := #seg. / (utt. dur. – pause) = $(5 \text{ C} + 6 \text{ V}) / (600 - \text{pause})$
 - **Average segment duration** = 1 / local articulation rate
⇒ **higher** avg. segment duration → **slower** speech
- Utterance inclusion criteria
 - Log average segment duration within ± 3 SD of mean log average segment duration
 - **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



Results

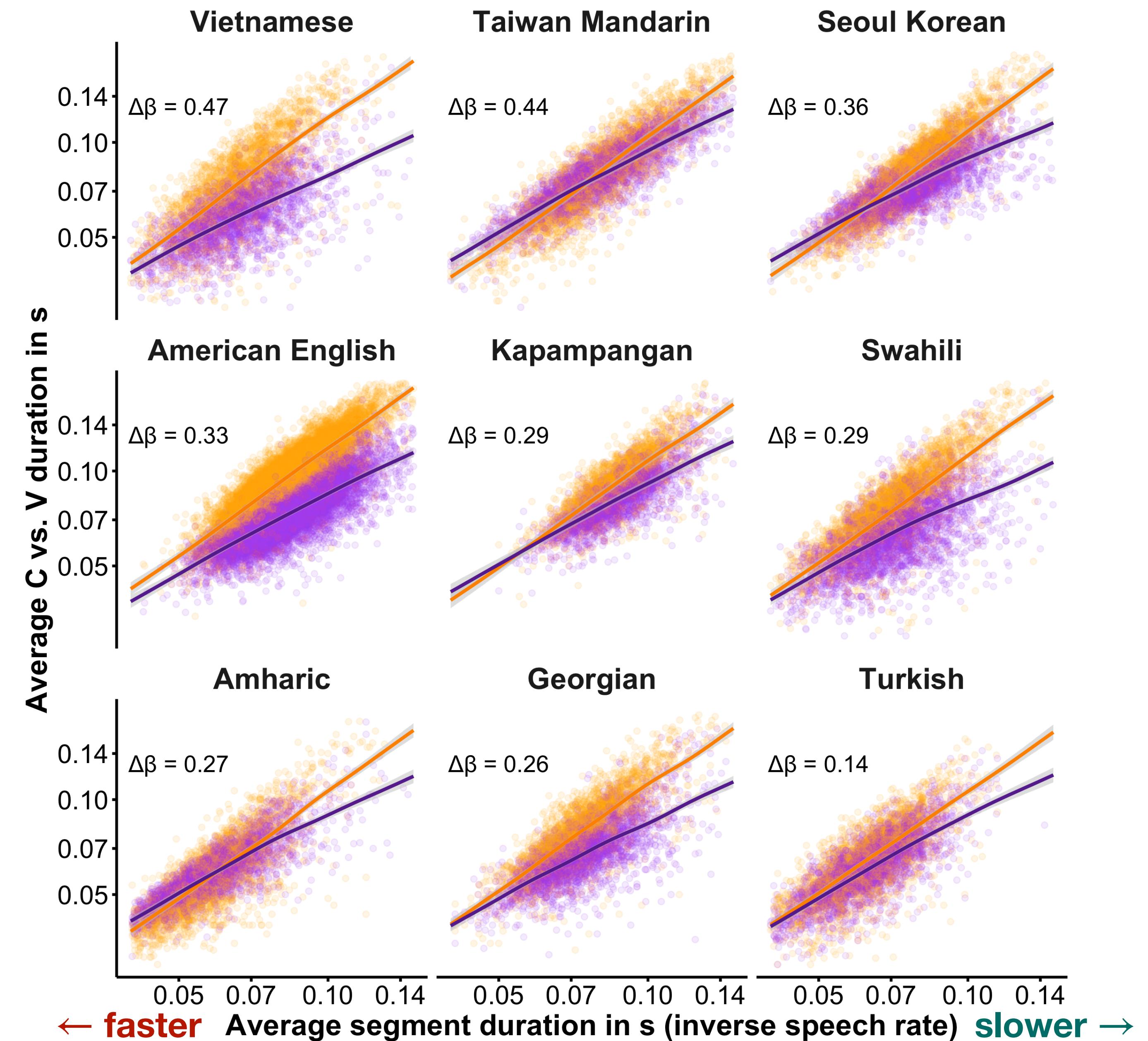
%V (percentage of vowels among duration of all segmental material): goes **up** as articulation rate **slows down**



Results

Vs undergo greater duration adjustment than Cs

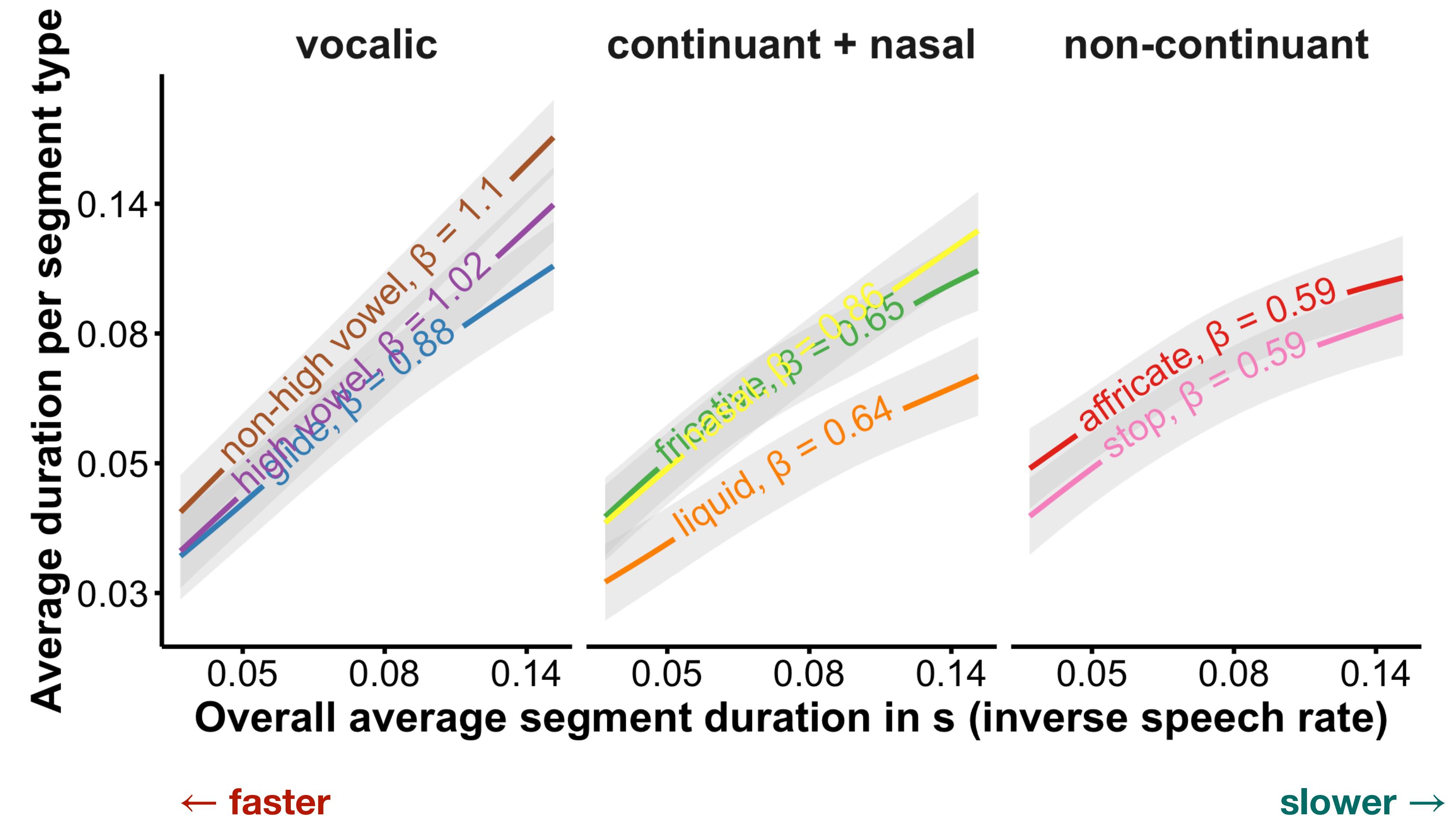
- **Fast:** Vs same or shorter than Cs
- **Slow:** Vs up to 1.5x longer than Cs



Results

“stretchiness” hierarchy:

vocalic
↓
continuant
↓
non-continuant



Summary and Discussion

- Across all languages, vowels are “stretchier” than consonant (cf. 1 & 2)
- Different vowel and consonant types display different stretchiness (partly aligned with sonority hierarchy)
 - determined by temporal and aerodynamic characteristics of segments
- Implications:
 - Interactions with segment reduction / deletion (deletion = “catastrophic failure”; stretchy segments more resistant?)
 - Rate perception by varying V and C durations
 - Implications for sound change (e.g., V length contrasts more stable than C ones)

¹ Kozhevnikova & Chistovich, 1965

² Port, 1976

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 spontaneous 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 medial 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.
- Tsao, Yinh-Chiao, Gary Weismer, and Kamran Iqbal. 2006. Interspeaker variation in habitual speaking rate: Additional 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.