Vowels are “stretchier” than consonants

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

Roger Yu-Hsiang Lo, Melissa Wang, Michelle Kamigaki-Baron, Noah Luntzlara, Márton Sóskuthy

Background

- **Rate** is an important aspect of within- and between-speaker variation
  - Rate := “the number of output units per unit of time” (Tsao, Weisner & Iqbal, 2006)
  - Speaking/speech rate: pause intervals are included
  - Articulation rate: pause intervals are **not** included
- Articulation rate variation is well-documented (e.g., Wood, 1973; Port, 1981; Miller, Grosjean, Lomanto, 1984; Gay, 1978; Crystal & House, 1988; Crystal & House, 1990)
  - Almost all studies focus exclusively on English
  - Articulation rate tends to be measured at the syllable level (e.g., #syllable / time unit)
  - Articulation rate varies substantially at both **global** (i.e., measured over large stretches of speech) and **local** (i.e., measured over a single pause-free utterance) levels (Miller, Grosjean, Lomanto, 1984)
  - Listeners are sensitive to articulation rate at both global and local levels (Port, 1978; Plug & Smith, 2021)

Background

- Relatively little is known about how **segments** respond to changes in articulation rate
  - Early studies have established that both **consonants** and **vowels** are shortened at a higher rate (Crystal & House, 1982, 1988)
  - Conflicting findings when consonants are contrasted with vowels
    - Change in duration that occurs with articulation rate takes place in **vowels** (Kozchevnikova & Chistovich, 1965; Port, 1976)
    - Constant consonant proportion at different articulation rates (Wood, 1973)
  - Involved explicit instructions to speak fast / slowly in a lab setting
  - Limited by the number of participants (e.g., 2 English speaker for K&C [1965]; 1 speaker per language in Wood [1973])

Research Questions

- How does the duration of different types of segments vary in response to **local** changes in articulation rate?
  - By comparing [1] consonants vs. vowels, and [2] among different consonant types (e.g., stops, fricatives, etc.)
  - With corpus data from seven unrelated languages
  - Using recordings of read / semi-spontaneous / spontaneous speech **without** instructions regarding how fast / slowly they should talk
  - At least 20 speakers per language

Methods

Dataset construction

- **American English**
  - Indo-European
  - North American Buckeye corpus
    - 40 speakers (20 f, 20 m)
    - ~40 hours of spontaneous speech in total
- **Kapampangan** / **Seoul Korean** / **Taiwan Mandarin**
  - Austronesian / Koreanic / Sino-Tibetan
  - OoPS-Lab general-purpose speech corpora
    - 20 speakers (10 f, 10 m) per language
    - ~2 hours of read speech and ~2 hours of spontaneous speech per language
- **Swahili** / **Turkish** / **Vietnamese**
  - Niger-Congo / Turkic / Austroasiatic
  - IARPA Babel program
    - 40 speakers (20 f, 20 m) per language
    - ~6.5 hours of spontaneous conversational telephone speech per language

Methods

Dataset construction: OoPS-Lab general-purpose speech corpora

- Languages included so far: Cebuano, French, Kapampangan, Seoul Korean, Taiwan Mandarin
- 20 speakers per language
  - 5 young female + 5 young male speakers (20 - 30 years)
  - 5 old female + 5 old male speakers (50+ years)
- Recordings were made online on participants' own device
- Read speech from reading a short essay + prompted semi-spontaneous monologue

Methods

Data annotation and management

- OoPS-Lab data was transcribed manually by native speakers of respective languages
- With the exception of the English data, all speech was forced-aligned with the Montreal Forced Aligner (McAuliffe et al., 2017a)
- Duration data managed and extracted using PolyglotDB (McAuliffe et al., 2017b)
- Statistical models and visualization carried out with R

Methods

Data annotation and management (cont.)

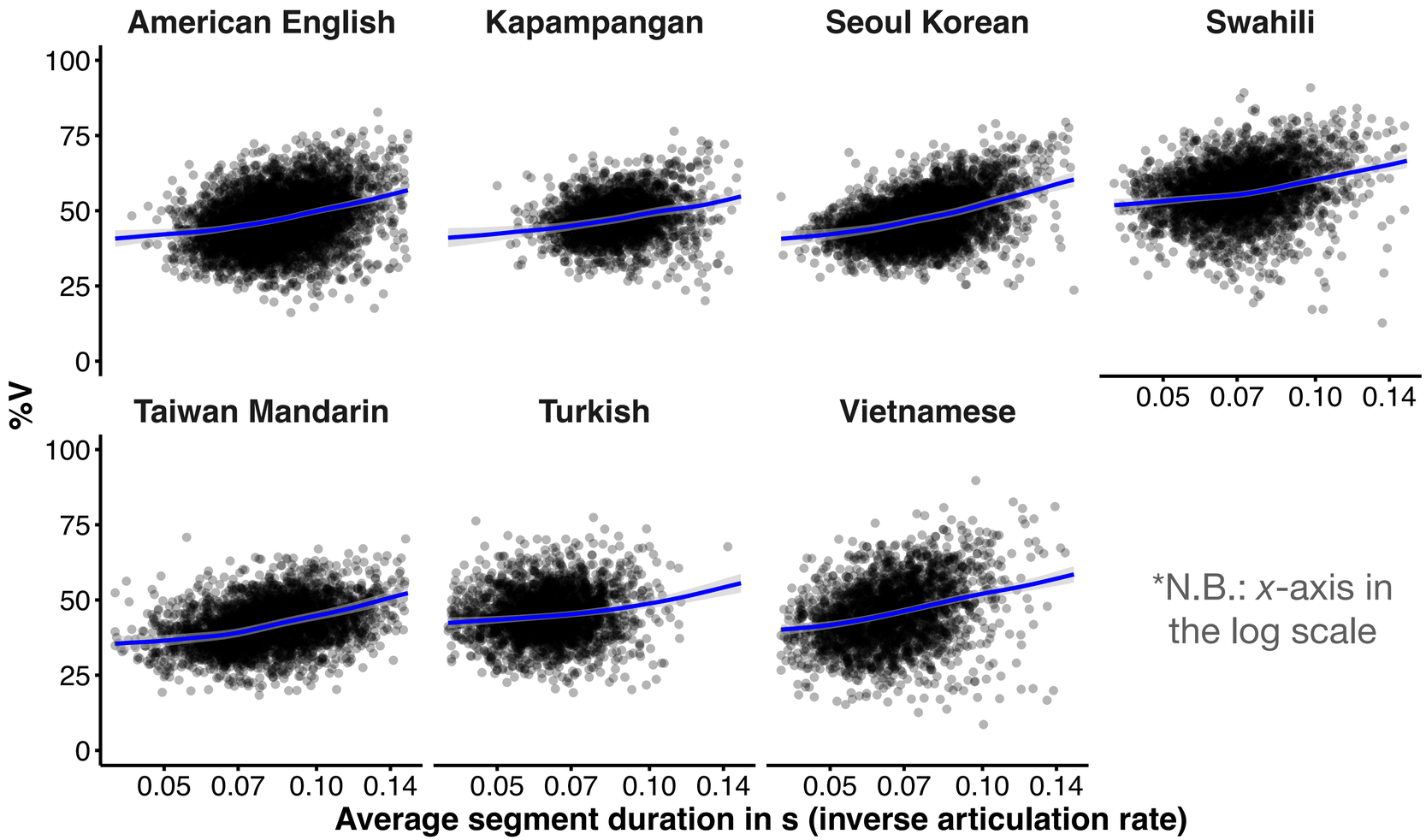
- Recordings were first parsed into utterances
  - Utterances defined as speech segments bounded by non-speech intervals of > 150 ms
- Measurements
  - Local articulation rate := segment rate within each utterance = #segments / utterance duration
  - Average segment duration = 1 / local articulation rate = utterance duration / # segments

  ⇒ higher average segment duration → slower speech
- Utterance inclusion criteria
  - Had more than 5 syllables
  - Had an average segment duration between 40 ms and 250 ms
  - Had a log average segment duration that is within ±3 SD of mean log average segment duration over all utterances across languages
  - In total, A. English: 4,930; Kapampangan: 2,587; S. Korean: 4,259; Swahili: 2,893; T. Mandarin: 3,158; Turkish: 2,453; Vietnamese: 2,329

Methods

Analyses and results

- %V (duration percentage of vowels among all segmental material)
  - Analyzed with a generalized additive mixed model (GAMM; Wood, 2017): %V ~ s(log art. rate)
  - With radon smooths by speakers and languages
- Vowels accounts for more and more portion as speech slows down

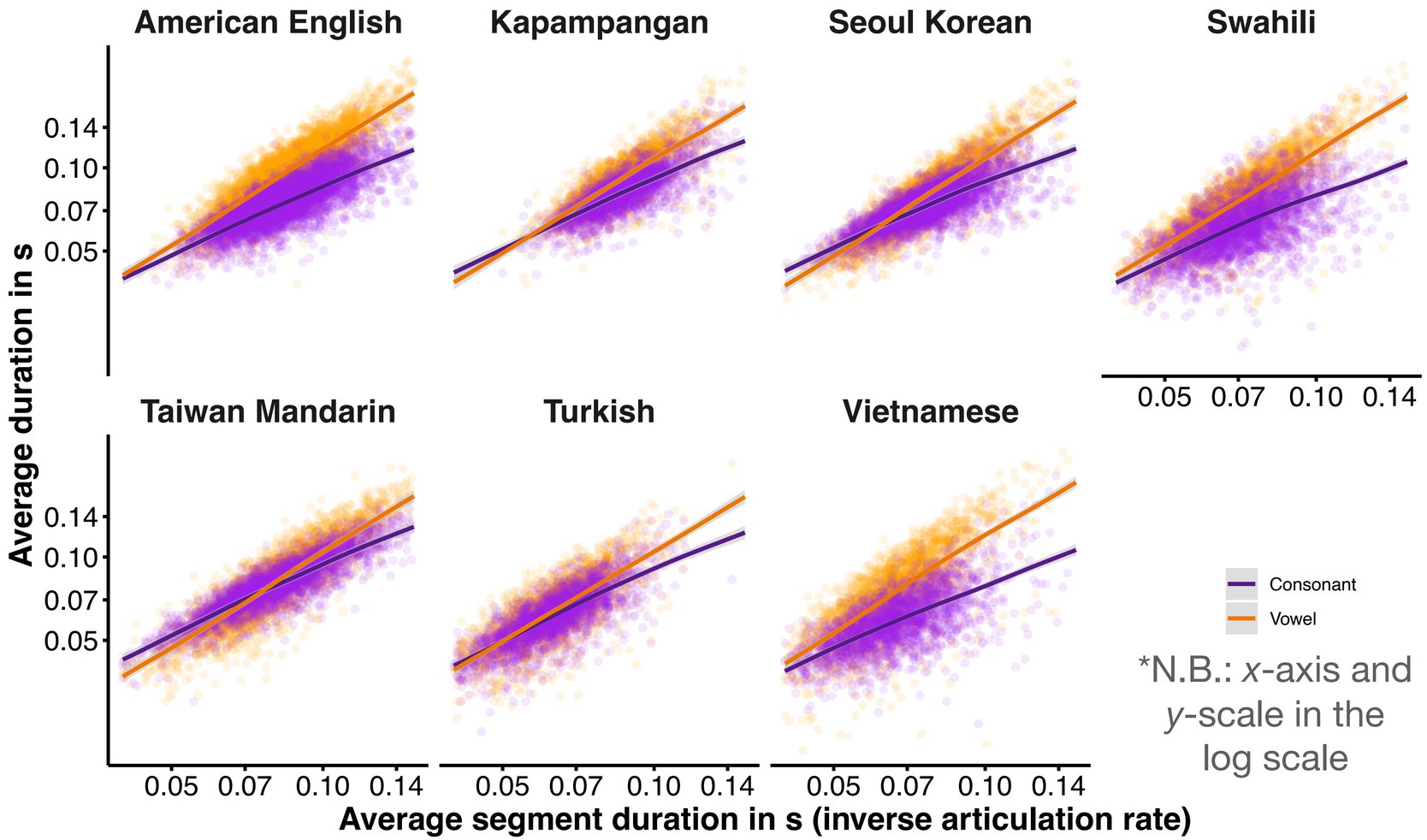


*N.B.: *x*-axis in the log scale

Methods

Analyses and results (cont.)

- Average C and V duration
  - GAMM: log avg. dur. ~ seg. type + s(log art. rate, by = seg. type)
  - With radon smooths by speakers and languages
- Vs undergo greater duration adjustment than Cs
  - Fast: Vs same or shorter than Cs
  - Slow: Vs up to 1.5x longer than Cs

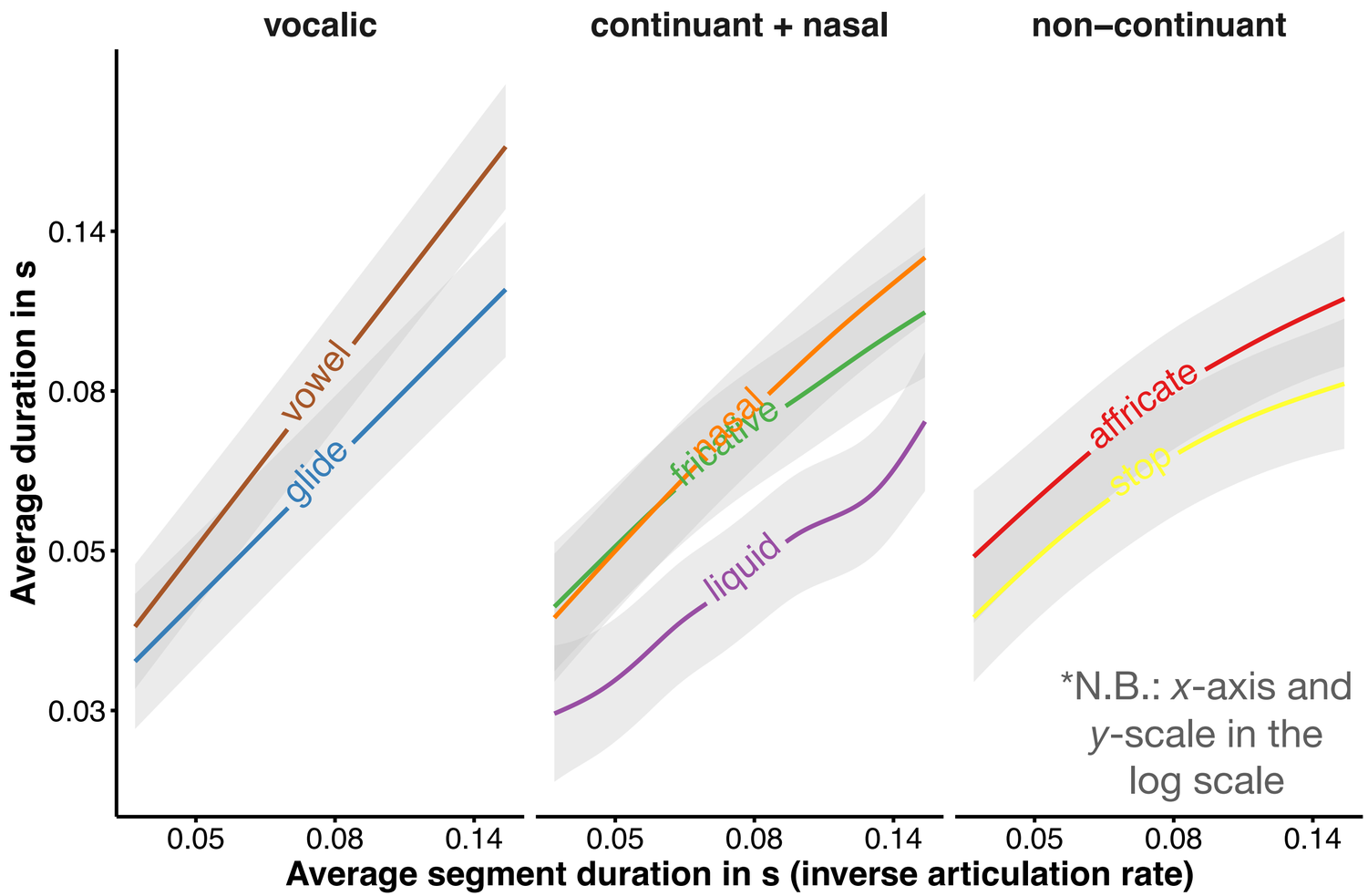


*N.B.: *x*-axis and *y*-scale in the log scale

Methods

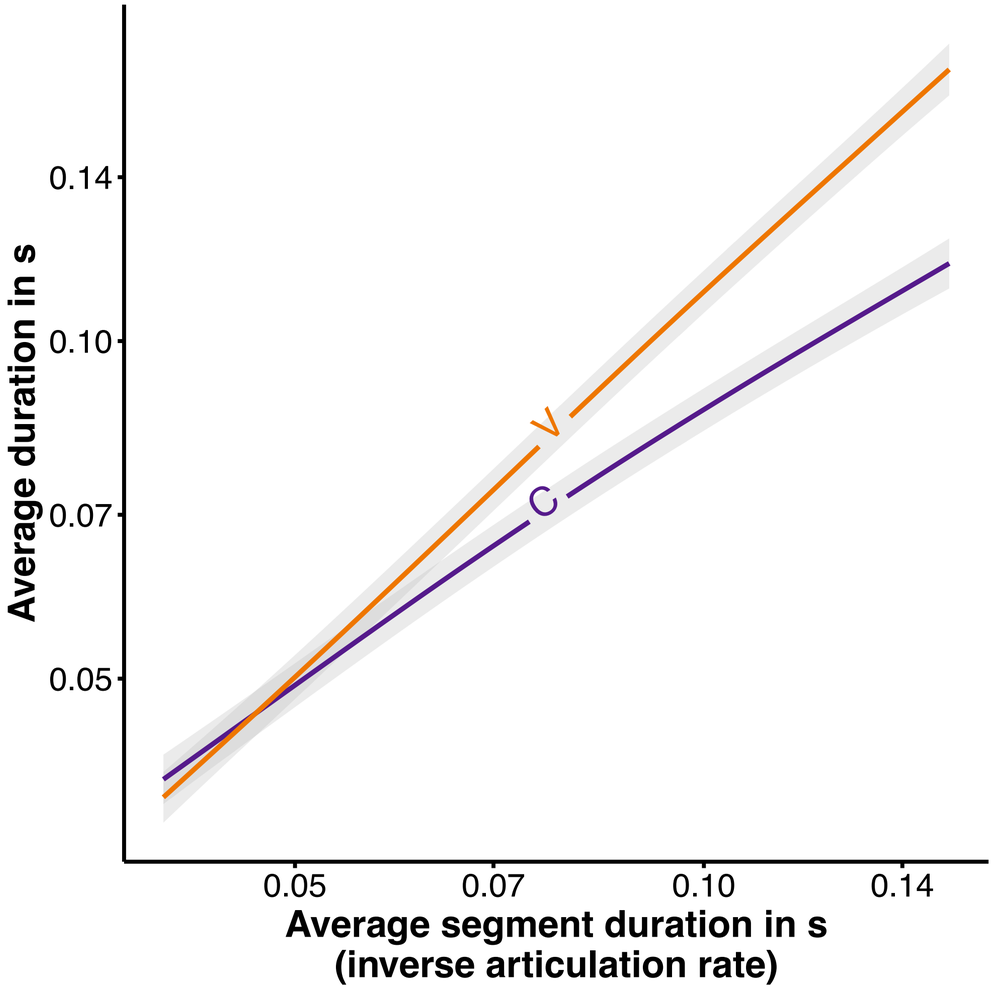
Analyses and results (cont.)

- Average duration of different C types
  - GAMM: log avg. dur. ~ C type + s(log art. rate, by = C type)
  - With radon smooths by speakers and languages
- Duration of non-continuants vary less than continuants
  - Vs still stand apart



Summary and Discussion

- Across all languages, vowels are "stretchier" than consonants
  - Agree with Kozchevnikova and Chistovich (1965) and Port (1976)
- Different consonant types (in particular, continuant vs. non-continuant) display distinct stretchiness
- "Stretchiness" of a segment is primarily determined by its temporal and aerodynamic mechanism
- Future direction: rate perception by varying vowel and consonant durations



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.

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.