Exposing the anatomy of articulation rate across English dialects and speakers

James Tanner[1], Morgan Sonderegger[2], Jane Stuart-Smith[1], Tyler Kendall[3], Jeff Mielke[4], Erik Thomas[4], Robin Dodsworth[4], and SPADE Data Consortium[1,2,4]

[1]*University of Glasgow (UK)*, [2]*McGill University (Canada)*, [3]*University of Oregon (USA)*, [4]*North Carolina State University (USA)*

Introduction: Speech rate – the most-studied component of speech timing – has been found to reflect both low-level aspects of speech production planning such as utterance length [6, 1] as well as speaker-related attributes, such as dialect, age, and gender [7, 4, 5, 3, 2]. While the effect of utterance length – where longer utterances result in both higher average speech rate and lower variance in rate – have been shown to be the strongest predictor of speech rate [6], variation across speaker-related properties have been of most interest in previous work. In addition, the relationship between utterance length effects and these speaker-specific factors is not clear, particularly in terms of how utterance length effects are structured both *across* and *within* dialects and individual speakers, which may provide insight into the sources of these effects. This study takes advantage of a large multi-dialectal dataset of English spontaneous speech to explore the structure of speech rate variation – particularly focusing on both the size and robustness of utterance length effects and speaker factors (dialect, age, gender) – as a window into the extent to which variability in speech timing is determined by low-level linguistic constraints versus speaker-specific attributes.

Methods: 119,736 utterances (3+ syllables, delimited by 150ms pauses) were extracted from 14 corpora of spontaneous English speech, corresponding to 15 dialects of British and North American English (1115 speakers: 539 female). A Bayesian distributional regression model was fit to the mean and variance of log-transformed articulation rate (syllables/second within an utterance). Log-transformed utterance length was split into two measures, reflecting each speaker's average and the deviation from that average, and were modelled as cubic splines with 5 knots to approximate the non-linear effect of utterance length. Speaker gender & age (as linear effects) were included as fixed effects. Dialects were modelled with random intercepts and slopes, with speakers included as random intercepts and slopes for utterance length deviation.

Results: The length of the utterance has a significant effect on both the mean and variance of articulation rate, though utterance length effects do not appear to meaningfully differ between dialects (Fig. 1). Dialects are estimated to vary in articulation rate by about 40% ($\hat{\sigma}_{dialect}$ = [0.02,0.10]), corresponding to differences of $\sim$2 syllables per second, and North American dialects are generally slower than British varieties (Fig. 2, left). Older speakers are estimated to speak more slowly than younger speakers ($\hat{\beta}_{age}$ = [-0.05,-0.02]) and female speakers estimated to speak more slowly than male speakers ($\hat{\beta}_{gender}$ = [0.03,0.07]), though these effects are relatively small compared with the effect of utterance length (Fig. 2, right). Variation across individual speakers ($\hat{\sigma}_{speaker}$ = [0.10,0.11]) is greater than the variation between dialects ($\Pr(\hat{\sigma}_{speaker} > \hat{\sigma}_{dialect})$ = 0.98), and the sizes of the age and gender effects ($\Pr(\hat{\sigma}_{speaker} > \hat{\beta}_{age})$ = 1; $\Pr(\hat{\sigma}_{speaker} > \hat{\beta}_{gender})$ = 1), indicating that there remains a substantial amount of speaker-specific variation in rate after accounting for the typically-studied speaker-related variables. These findings illustrate that while utterance length and speaker factors both have distinct influences on articulation rate, the structure of these effects are clearly different. The relative invariance of utterance length effects across dialects suggests that the temporal modulation of syllables may reflect low-level articulatory or perceptual constraints [6]. Overall, this study shows that utterance length plays a key role in the structure of English articulation rate: after controlling for utterance length, differences by dialect, age, and gender are apparent, though these are still much smaller than variation across individual speakers.

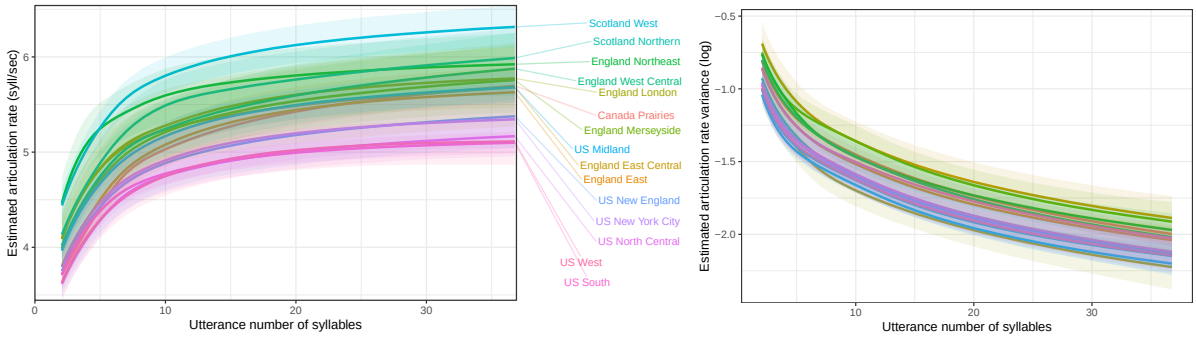

Figure 1: Estimated articulation rate (left) and variance in articulation rate (right) for each dialect as a function of utterance length. Lines indicate posterior medians with shaded areas representing 95% credible intervals.

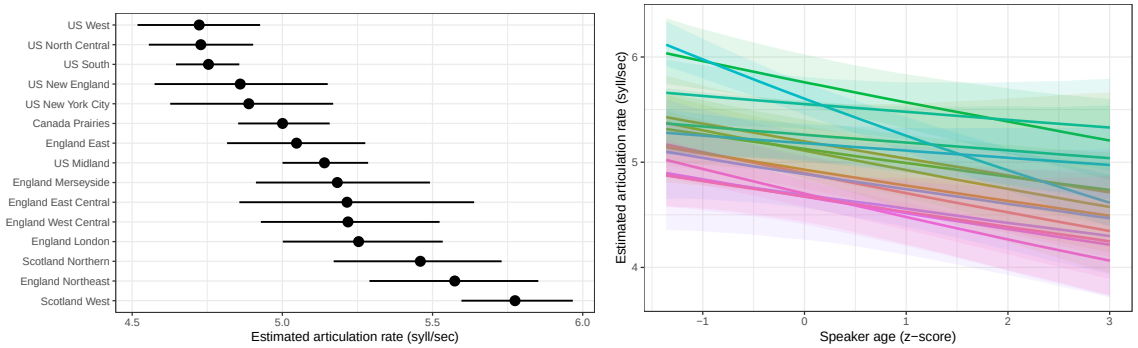

Figure 2: Left: estimated articulation rate for each dialect at average utterance length. Right: estimated articulation rate as a function of speaker age across dialects, held at average utterance length. Values indicate posterior medians with 95% credible intervals.

References

[1] Bishop, J. & Kim, B. (2018). Anticipatory shortening: Articulation rate, phrase length, and lookahead in speech production. In *Proceedings of Speech Prosody 2018* (pp. 235–239).

[2] Cichocki, W., Kaminskaïa, S., & Hagar, L. (2023). Regional variation in articulation rate in French spoken in Canada. *Journal of the International Phonetic Association*, (pp. 1–20).

[3] Coats, S. (2020). Articulation rate in American English in a corpus of YouTube videos. *Language and Speech*, 63, 799–831.

[4] Jacewicz, E., Fox, R. A., & Wei, L. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *Journal of the Acoustical Society of America*, 123, 839–850.

[5] Kendall, T. (2013). *Speech rate, Pause, and Sociolinguistic Variation*. London: Palgrave MacMillan.

[6] Quené, H. (2008). Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *Journal of the Acoustical Society of America*, 123, 1104–1113.

[7] Ray, G. B. & Zahn, C. J. (1990). Regional speech rates in the United States: A preliminary analysis. *Communication Research Reports*, 7, 34–37.