

# Optionality and variability of speech pauses in read speech across languages and rates

1

## Introduction

- prosodic phrasing is variable – between and within speakers [1, 2]
- speech pauses analyzed for
  - optionality, i.e. ab-/presence in given location
  - variability, i.e. variation of duration and breath involvement
- pause modeling should be more than purely punctuation-based [3]
- duration: highly variable for pauses [4] and quite variable for breath noises [5]
- pauses less robust to speech rate increase than other prosodic markers, such as final lengthening [6]

2

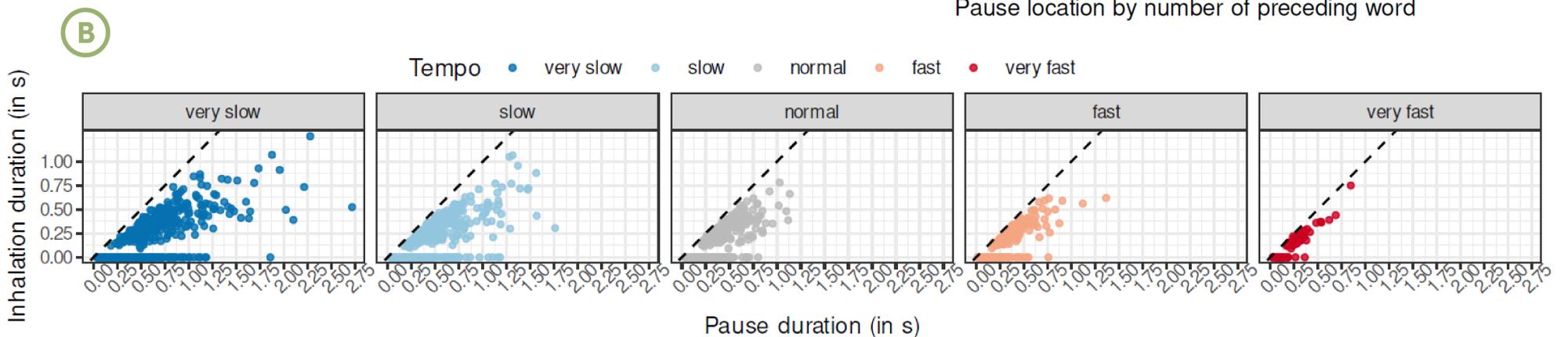
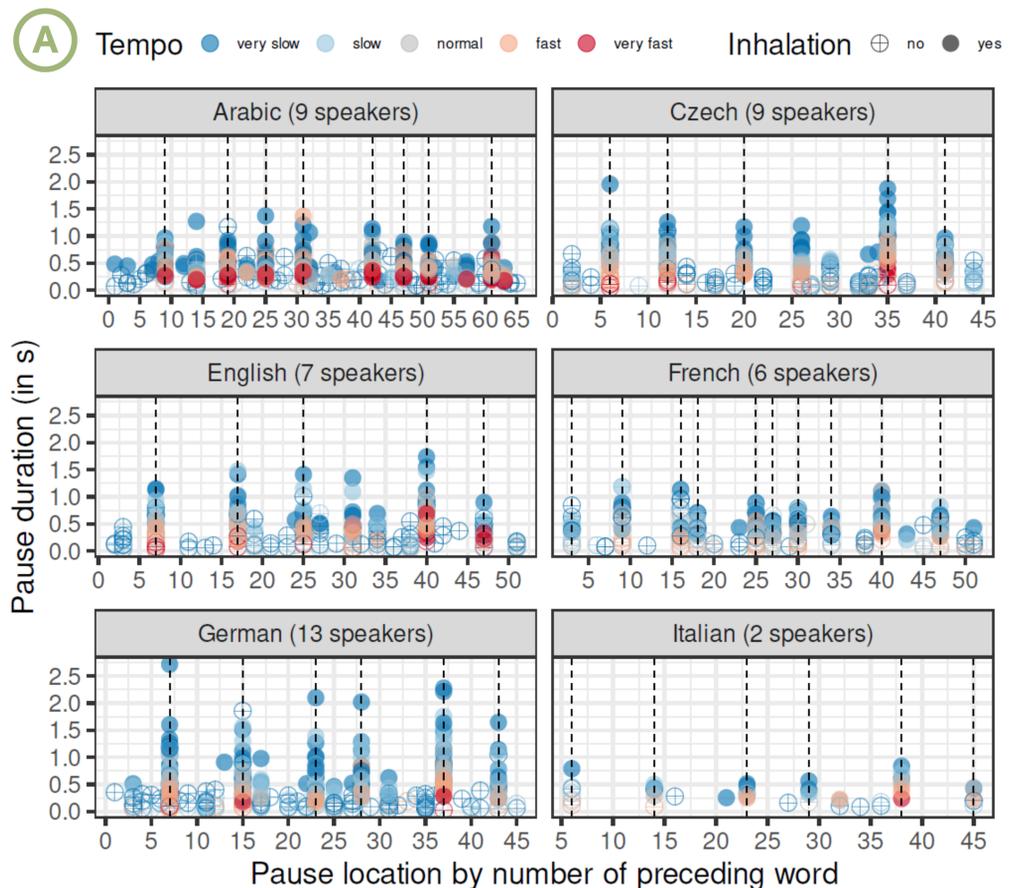
## Methods

- BonnTempo-Corpus: 9 Czech, 7 English, 6 French, 13 German, 2 Italian speakers + 9 Arabic speakers [7, 8]
- 5 intended speech rates: *very slow* to *very fast*
- annotation: pauses (no fixed threshold), pause-internal breath noises and pause-internal silent edges around breath noises [9]

3

## Results

- some pause locations less optional (see Fig. A):
  - many pauses
  - duration highly variable in accordance with speech rate
  - breath pauses mostly there
  - often related to punctuation/conjunctions
- with increasing speech rate (see Fig. A and B):
  - fewer pauses (non-breath in particular), mostly at the less optional pause locations
  - pauses, inhalations and edges around the breath get shorter
  - breath noises fill bigger portions of breath pauses



4

## Discussion and Conclusion

- optionality: some pause locations *less optional*
- variability: high in durations of pauses, breath noises, and edges around them
- speech rate affects optionality and variability → with increase: fewer and shorter pauses, breath noises, and edges
- implications for pause modeling in natural and synthetic speech

5

## References

- [1] F. Goldman-Eisler, "The Distribution of Pause Durations in Speech," *Language & Speech*, 4.4, 232–237, 1961. [2] J. Trouvain & M. Grice, "The Effect of Tempo on Prosodic Structure," *ICPhS*, 1999, 1067–1070. [3] J. P. Gee & F. Grosjean, "Performance structures: A psycholinguistic and linguistic appraisal," *Cognitive Psychology*, 15.4, 411–458, 1983. [4] E. Campione & J. Véronis, "A large-scale multilingual study of silent pause duration," *Speech Prosody*, 199–202, 2002. [5] J. Trouvain, R. Werner, & B. Möbius, "An Acoustic Analysis of Inbreath Noises in Read and Spontaneous Speech," *Speech Prosody*, 2020, 789–793. [6] M. Żygis, J. Tomlinson, C. Petrone, & D. Pfützte, "Acoustic cues of prosodic boundaries in German at different speech rate," *ICPhS*, 2019, 999–1003. [7] V. Dellwo, I. Steiner, B. Aschenberger, J. Dankovi, & P. Wagner, "BonnTempo-Corpus and BonnTempo-Tools: A database for the study of speech rhythm and rate," *Interspeech*, 2004, 777–780. [8] O. Ibrahim, H. Asadi, E. Kassem, & V. Dellwo, "Arabic speech rhythm corpus: Read and spontaneous speaking styles," 2020, 5337–5342. [9] T. Fukuda, O. Ichikawa, & M. Nishimura, "Detecting breathing sounds in realistic Japanese telephone conversations and its application to automatic speech recognition," *Speech Communication*, 98, 95–103, 2018.