One Population, Two Languages: How Does Language Choice Affect /s/peaker Di/s/crimination?

Justin J. H. Lo

Department of Language and Linguistic Science, University of York

✉ jl2355@york.ac.uk   🐦 🔗 @justinjhlo

UNIVERSITY of York

Research aims & questions

Bilingualism is highly prevalent, but the discriminatory potential of acoustic-linguistic variables used in forensic speaker comparison is rarely considered in bilingual terms.

1. How does discriminatory power of the same variable vary between different languages?
2. How does individual discriminatory behaviour vary across languages?

Why /s/?

- High between-speaker variability
  - Linked to vocal tract physiology
  - Gestural idiosyncrasy in groove formation[6]
- Low within-speaker variability
  - Precise articulation required[12,15]
  - Less susceptible to coarticulation[3,18]
- Very frequent[11,14] and reliably segmentable[19]
- Forensic implications investigated in English, Spanish and Dutch[4,5,9,10]

Materials

RCMP Voice ID Database[17]

- 90 male Canadian English–French bilinguals
  - Recorded in both languages
  - L1: 38 En, 42 Fr, 10 simultaneous bilinguals
  - No reported knowledge of L3
- Read speech: Phonetically balanced sentences and short passage
- High quality microphone condition

/s/ in English & French[6]

- Similar place of articulation (dental to alveolar) used in En and Fr
- Both use laminal *and* apical /s/ in approximately equal proportions
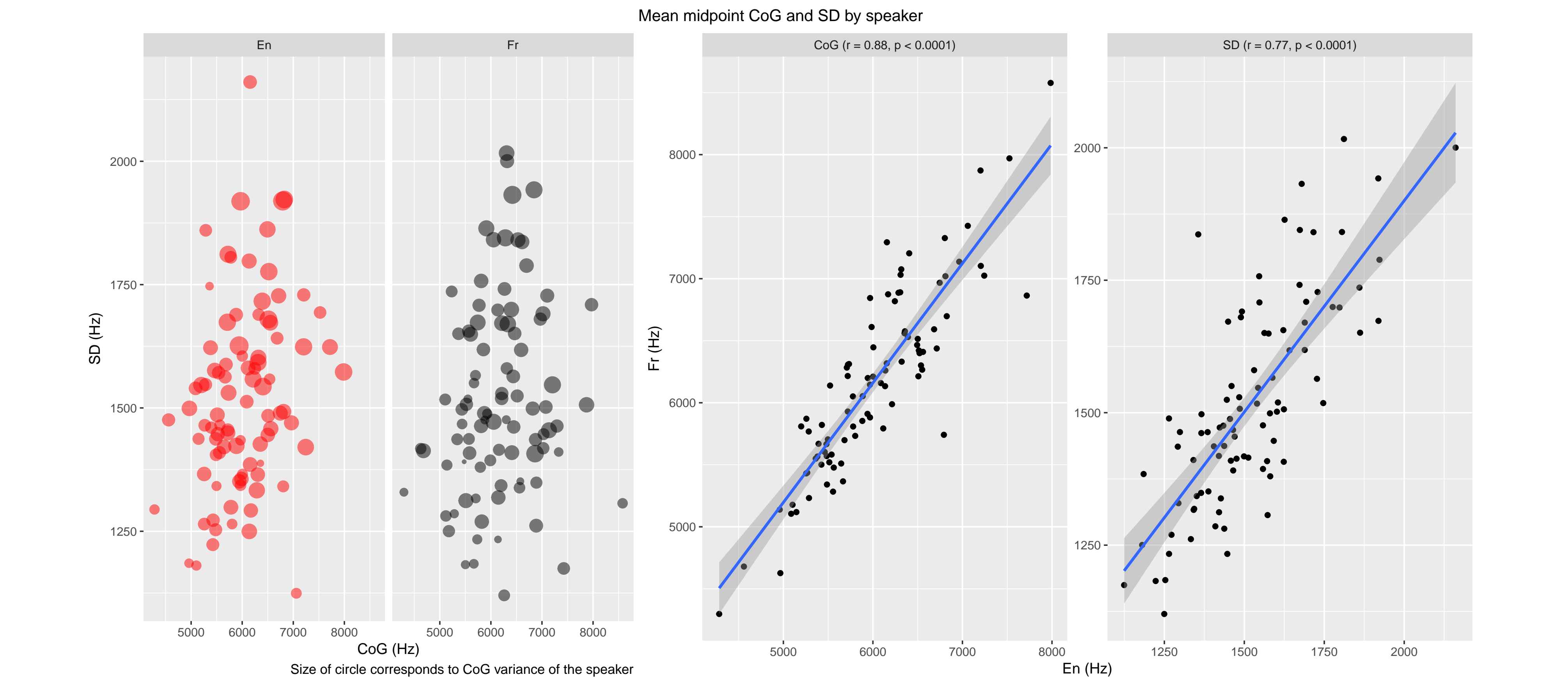
Procedure: Acoustic extraction

1. Phrase-level orthographic transcription
2. Segmentation: Montreal Forced Aligner[13]
3. Extraction of CoG & SD
   - Bandpass filter: 500–11,000 Hz
   - **Static**: 40ms-window centred at midpoint
   - **Dynamic**: 9 10ms-windows spread across duration
     - Quadratic polynomial fitted to each spectral moment and coefficients extracted
4. Excluded: /s/ < 60ms, CoG < 2500 Hz and:

Environment	English	French
Homorganic fricatives	/ss/, /sz/, /zs/	
Assimilation[16]	/sʃ/, /sʒ/	/sʃ/, /ʃs/, /sʒ/, /ʒs/
/s/-retraction[2]	/sr/, /sCr/ (C = /p,t,k/)	–

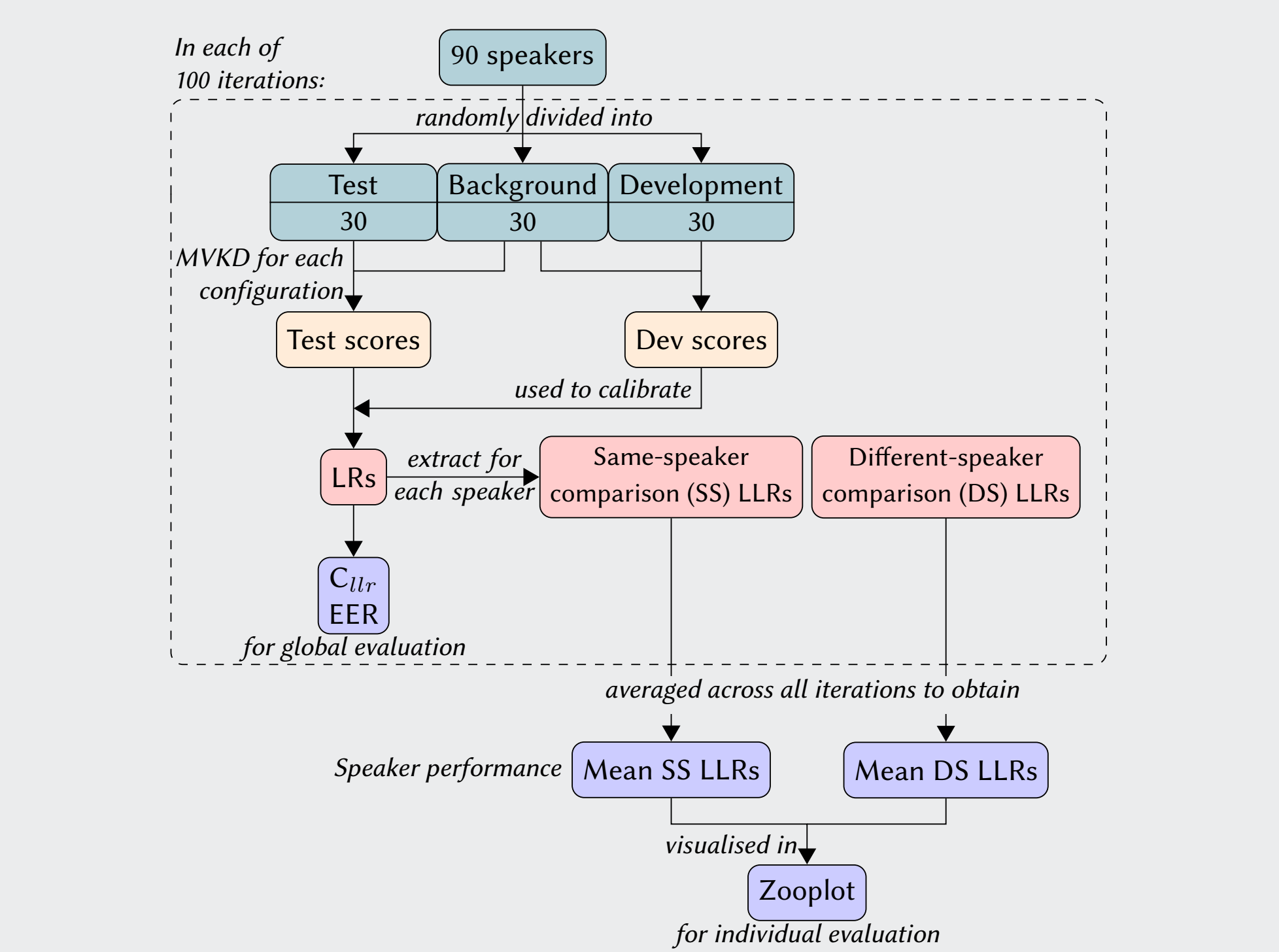- ⇒ 32.5 (En)/43.0 (Fr) tokens per speaker
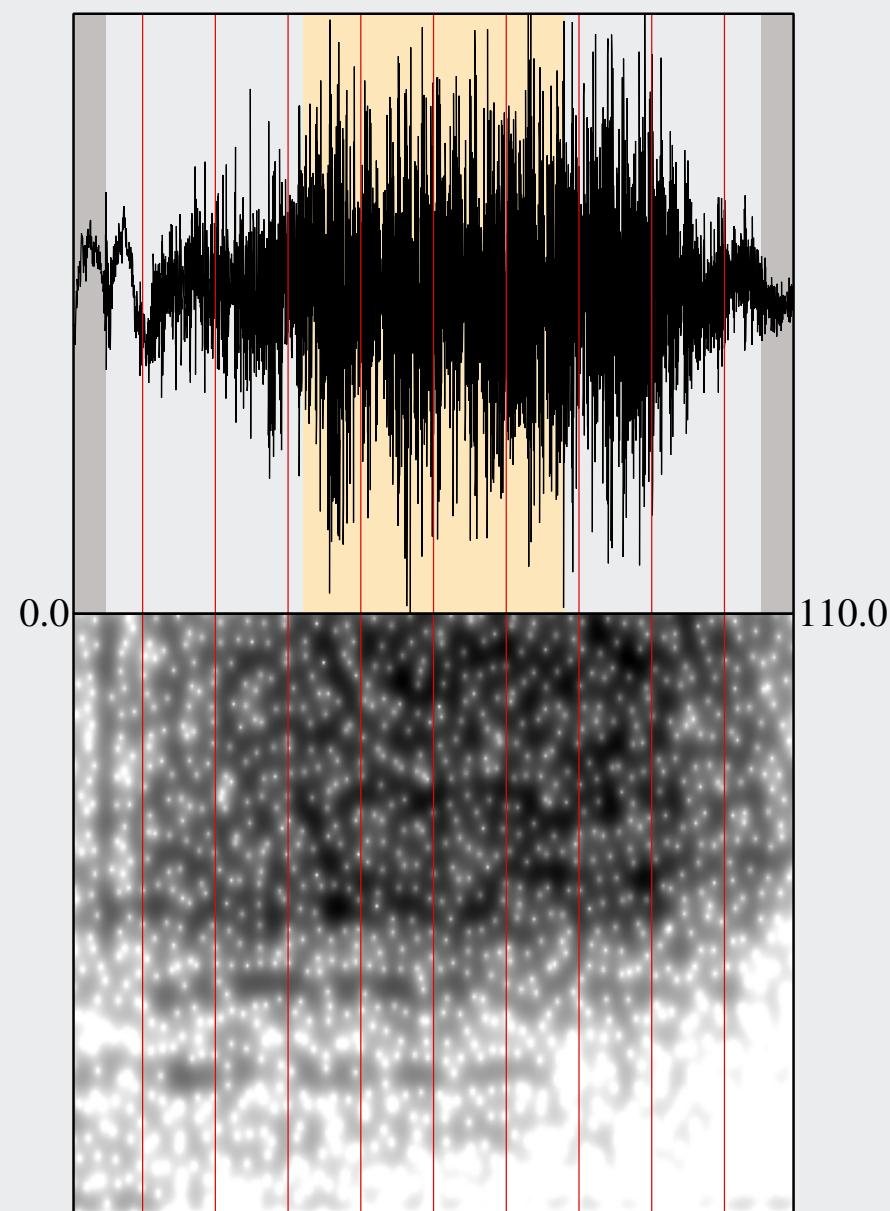
Example of /s/ from "fast": 5ms from each end excluded (grey), static (orange) and dynamic windows (centred at each red line).

Results: Descriptive data

Mean midpoint CoG and SD by speaker. CoG (r = 0.88, p < 0.0001). SD (r = 0.77, p < 0.0001). Size of circle corresponds to CoG variance of the speaker
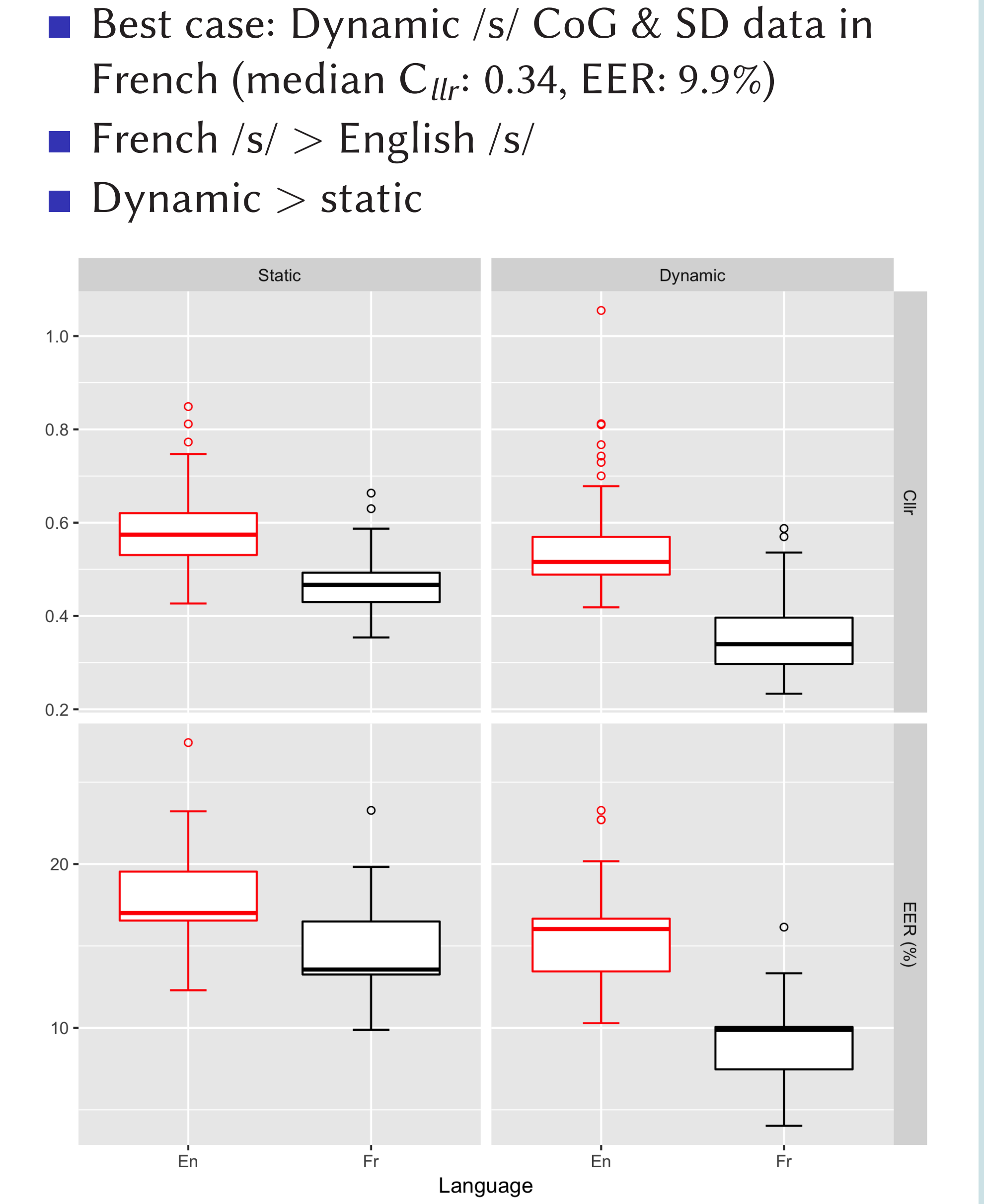
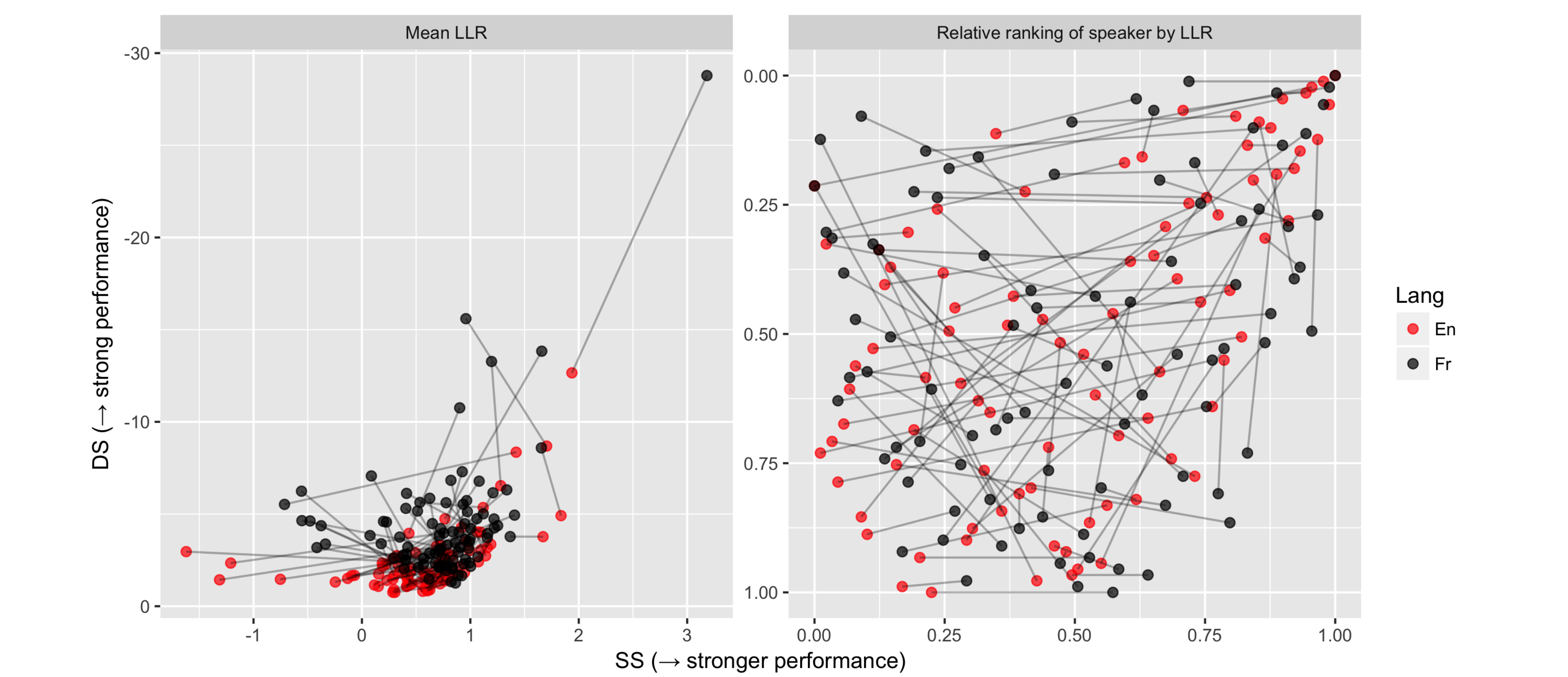## Procedure: Likelihood ratio (LR)-based testing

- Static vs dynamic CoG & SD in <u>En vs Fr</u> as input to MVKD formula[1] for LR calculation
- Tests following procedure in Enzinger, Morrison and Ochoa (2016)[8], with 100 randomised iterations to minimise sampling effects and test all speakers
- C_{llr} and EER to measure global performance
- Zooplots[7] to compare individual performance

Results: Global metrics

- Best case: Dynamic /s/ CoG & SD data in French (median C_{llr}: 0.34, EER: 9.9%)
- French /s/ > English /s/
- Dynamic > static

Results: Individual behaviour (based on static input)

Mean LLR. Relative ranking of speaker by LLR. DS (→ strong performance). SS (→ stronger performance). Lang: En, Fr

Discussion

- Unequal discriminatory power despite broadly similar distribution of spectral data
- All speakers were on average capable of being distinguished from others ($LLR_{DS} < 0$)
- Spectral dynamics of /s/ can convey richer speaker-specific information
- Overall shift to stronger discriminatory behaviour in Fr evident in zooplot
- Within-population distribution does not carry well across languages
- Different speakers contribute to errors in the system in different languages

[1] Aitken, C. G. G. and D. Lucy (2004). Evaluation of trace evidence in the form of multivariate data. *Applied Statistics*, 53 (1), 109–122. [2] Baker, A., D. Archangeli, and J. Mielki (2011). Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change*, 23 (3), 347–374. [3] Bladon, R. A. W. and F. J. Nolan. A video-fluorographic investigation of tip and blade alveolars in English. *Journal of Phonetics*, 5, 185–193. [4] Cicres, J. (2011). Los sonidos fricativos sordos y sus implicaciones forenses. *Estudios Filógicos*, 48, 33–48. [5] Cicres, J. and N. Gavaldà (2016). A preliminary analysis of the speaker-specificity of the first four spectral moments of spectral fricatives. 25th *International Association for Forensic Phonetics and Acoustics Annual Conference*. York, UK. [6] Dart, S. N. (1998). Comparing French and English coronal consonant articulation. *Journal of Phonetics*, 26 (1), 71–94. [7] Dunstone, T. and N. Yager (2009). *Biometric System and Data Analysis: Design, Evaluation, and Data Mining*. Springer. [8] Enzinger, E., G. S. Morrison, and F. Ochoa (2016). A demonstration of the application of the new paradigm for the evaluation of forensic evidence under conditions reflecting those of a real forensic-voice-comparison case. *Science and Justice*, 56 (1), 42–57. [9] Heeren, W. (2017). Speaker-dependency of /s/ in spontaneous telephone conversation. 26th *International Association for Forensic Phonetics and Acoustics Annual Conference*. Split, Croatia. [10] Kavanagh, C. M. (2012). New consonantal acoustic parameters in forensic speaker comparison. PhD thesis. University of York. [11] Malécot, A. (1974). Frequency of Occurrence of French Phonemes and Consonant Clusters. *Phonetica*, 29, 158–170. [12] McAuliffe, M. J., E. C. Ward, and B. E. Murdoch (2001). Tongue-to-palate contact patterns and variability of four English consonants in an /i/ vowel environment. *Logopedics Phoniatrics Vocology*, 26 (4), 165–178. [13] McAuliffe, M. M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger (2017). *Montreal Forced Aligner [Computer programme]. Version 0.9.0., retrieved November 2017.* URL: http://montrealcorpustools.github.io/Montreal-Forced-Aligner/. [14] Mines, M. A., B. F. Hanson, and J. E. Shoup (1978). Frequnecy of occurrence of phonemes in conversational English. *Language and Speech*, 21 (3), 221–241. [15] Mooshammer, C., P. Hoole, and A. Geumann (2007). Jaw and Order. *Language and Speech*, 50 (2), 145–176. [16] Niebuhr, O., M. Clayards, C. Meunier, and L. Lancia (2011). On place assimilation in sibilant sequences—Comparing French and English. *Journal of Phonetics*, 39 (3), 429–451. [17] Royal Canadian Mounted Police (2010–2016). *Voice ID Database [unpublished audio corpus]*. Collected at University of Ottawa. [18] Tabain, M. (2001). Variability in Fricative Production and Spectra: Implications for the Hyper- and Hypo- and Quantal Theories of Speech Production. *Language and Speech*, 44 (1), 57–94. [19] Turk, A., S. Nakai, and M. Sugahara (2006). Acoustic segment durations in prosodic research: A practical guide. In S. Sudhoff et al. (Eds). *Methods in Empirical Prosody Research*. Berlin: Walter de Gruyter, pp. 1–28.