Cross-linguistic Stability of the Discriminatory Power of Long-term Formant Distributions: Testing with Bilingual Speakers

Justin J. H. Lo *Department of Language and Linguistic Science, University of York*
jl2355@york.ac.uk @justinjhlo

LTFDs as speaker discriminants

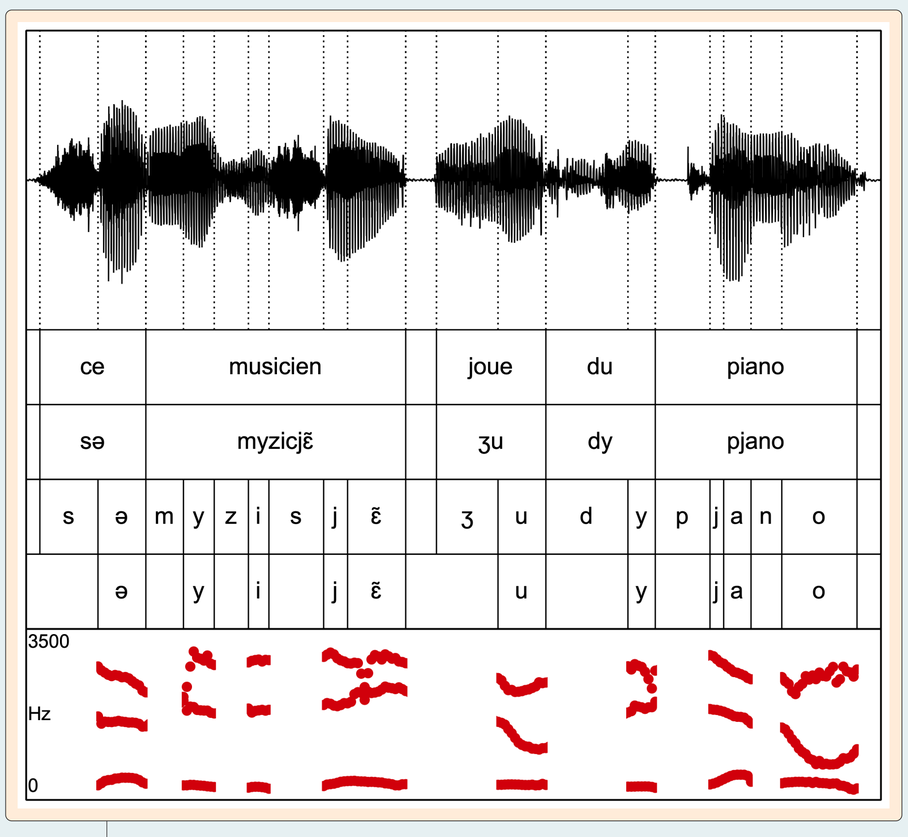
- Overall collection of formant estimates in sample for speaker
- Reflect anatomy of vocal tract and capture articulatory habits[9]
- Both average frequencies and shapes of the distributions provide speaker-specific information[3]
- Strong performance in English and German shown in likelihood ratio (LR)-based testing[1,4,6]

Research aims & questions

- How language-independent are LTFDs?
  - LTFDs for same speaker may be comparable across languages[3,5,7]…
  - …*but* language-specific effects may persist[3]
- How language-independent is the discriminatory power of LTFDs?
  - Assess both languages in LR framework using the same set of conditions
  - Control for speaker effects by using the same bilingual speakers
  - Aim: Isolate language as the only variable

Materials & methods

Speakers
60 adult male speakers from Voice ID Database[11]



Materials (Fr & En)
- Phonetically balanced sentences
- Short passage

Segmentation
- Montreal Forced Aligner[8]
- Manual correction in Praat[2]

Phonemes
- ✔ Vowels (including Fr /$\widetilde{\text{V}}$/)
- ✔ Semivowels (/j, w/ + Fr /ɥ/)
- ✖ Liquids (/l, r/)
- › Total V time: 26.2s (En)/34.6s (Fr)

Feature extraction (*in situ*)
- F1–F3, BW1–BW3
- 25-ms windows every 10 ms
- Max. 6 formants up to 5500 Hz

In each of 100 iterations:

randomly divided into

Test	Background	Development
20	20	20

Modelling GMM–UBM[10]

Speaker comparisons within each language

Test scores

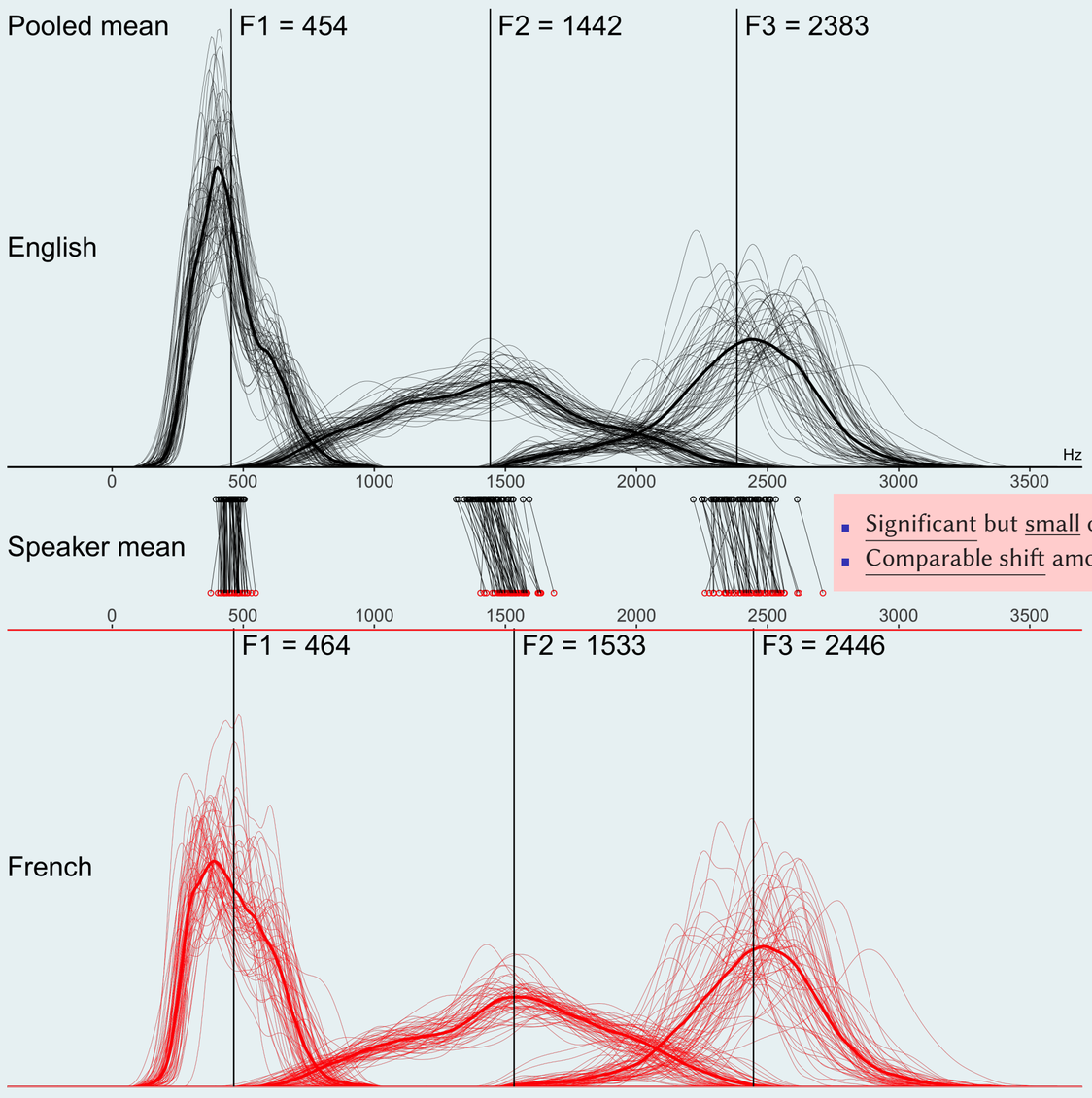
Dev scores

used to calibrate

LRs

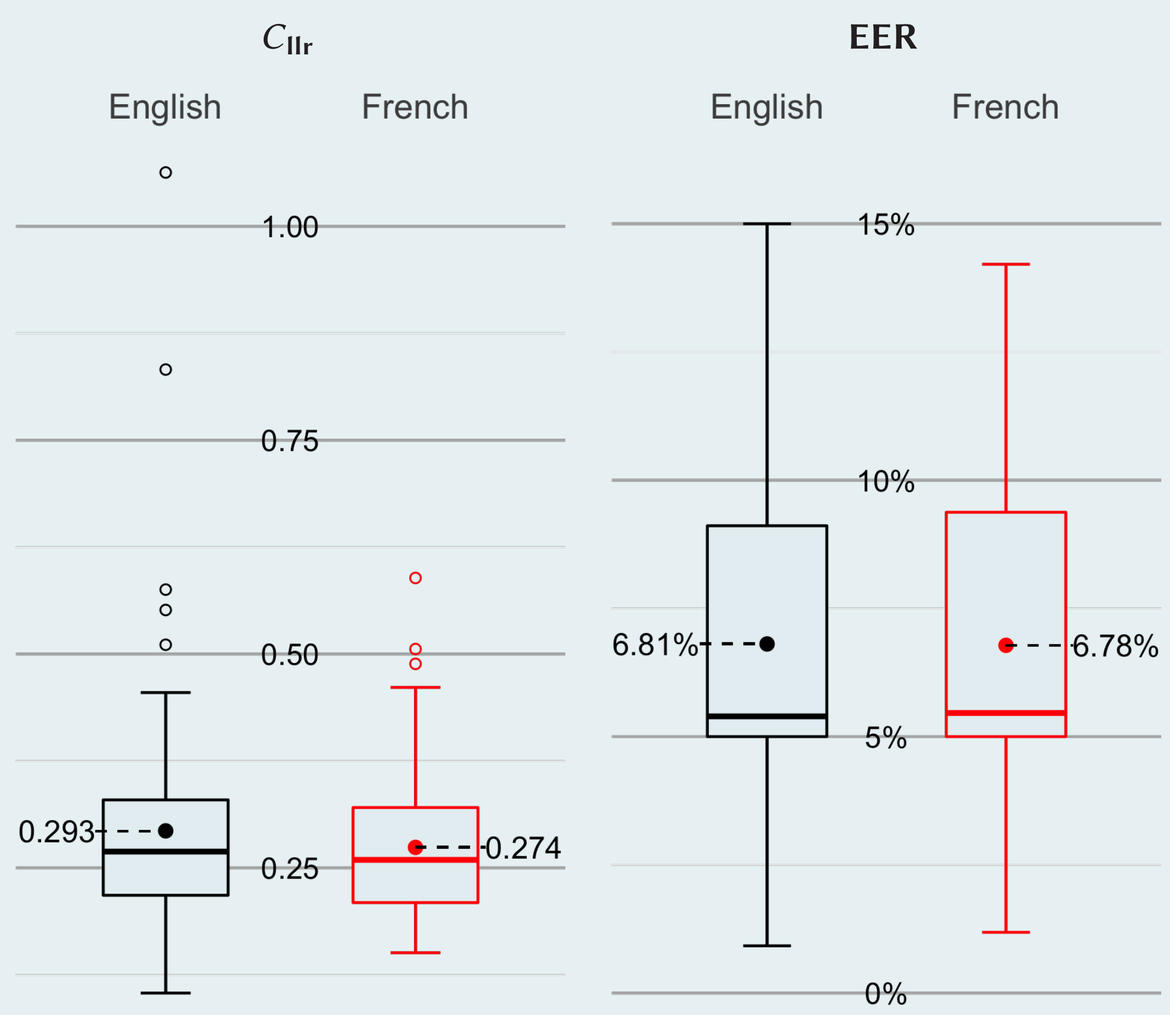
C_{llr} & EER *for performance evaluation*

LTFDs



- Significant but small differences
- Comparable shift among speakers

LR testing (over 100 iterations)



No significant differences in C_{llr} and EER between En and Fr

Discussion

- Shapes show rich speaker-specific information, especially in F3
- Materials may have an effect on LTFDs, e.g. more back vowels in En
- Useful variable in both En and Fr, as shown by low C_{llr} and EER
- Similar discriminatory power in En and Fr despite LTFD differences
- Performance of LTFDs resilient to choice of language

Future research

- Testing in forensically realistic conditions
  - Spontaneous speech data
  - Excluding F1 or bandwidths
- Analysis of individual discriminatory behaviour across languages
- Cross-language speaker comparisons

[1] Becker, T., M. Jessen, and C. Grigoras (2008). Forensic speaker verification using formant features and Gaussian mixture models. *Proceedings of Interspeech 2008.* Brisbane. [2] Boersma, P. and D. Weenink (2016). *Praat: doing phonetics by computer [Computer Programme] Version 6.0.19.* URL: http://www.praat.org. [3] Cho, S. and M. J. Munro (2017). F0, long-term formants and LTAS in Korean-English Bilinguals. *Proceedings of the 31st General Meeting of the Phonetic Society of Japan.* Tokyo, Japan, pp. 188–193. [4] Gold, E., P. French, and P. Harrison (2013). Examining long-term formant distributions as a discriminant in forensic speaker comparisons under a likelihood ratio framework. *Proceedings of Meetings on Acoustics.* Vol. 19. Montreal, p. 060041. [5] Heeren, W., D. van der Vloed, and J. Vermeulen (2014). Exploring long-term formants in bilingual speakers. *Presented at 23rd IAFPA Annual Conference.* Leiden. [6] Hughes, V., P. Harrison, P. Foulkes, P. French, C. Kavanagh, and E. S. Segundo (n.d.). Mapping across feature spaces in forensic voice comparison: The contribution of auditory-based voice quality to (semi-)automatic system testing. *Proceedings of Interspeech 2017.* Stockholm, Sweden, pp. 3892–3896. [7] Jessen, M. and T. Becker (2010). Long-Term Formant Distribution as forensic-phonetic feature. *Presented at 2nd Pan-American/Iberian Meeting on Acoustics.* Cancún, Mexico. [8] 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/. [9] Nolan, F. and C. Grigoras (2005). A case for formant analysis in forensic speaker identification. *International Journal of Speech, Language and the Law,* 12 (2), 143–173. [10] Reynolds, D. A., T. F. Quatieri, and R. B. Dunn (2000). Speaker Verification Using Adapted Gaussian Mixture Models. *Digital Signal Processing,* 10 (1–3), 19–41. [11] Royal Canadian Mounted Police (2010–2016). *Voice ID Database [unpublished audio corpus].* Collected at University of Ottawa.