

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, UK

j12355@york.ac.uk

This paper is eligible for the 'Best Student Paper Award'

Introduction

The case for the use of long-term formant distributions (LTFDs) in forensic voice comparison, first advanced in Nolan and Grigoras (2005), has been independently explored in several languages, where researchers have demonstrated that LTFDs hold speaker-specific information in both average frequencies and shapes of formant distributions (Cho & Munro, 2017; Hughes et al, 2017; Moos, 2010). Investigation within the framework of likelihood ratios (LRs) has substantiated the potential of LTFDs as powerful speaker discriminants, finding similarly strong performance in both English and German (Becker et al, 2008; Gold et al, 2013; Hughes et al, 2017). LTFD statistics have further been suggested to be language-independent (Jessen & Becker, 2010), a proposition supported by the finding that bilingual speakers display comparable distributions across languages (Heeren et al, 2014).

While findings thus far raise the possibility that the efficacy of LTFDs as speaker discriminants is independent of the choice of language, these results may not be easily generalised due to different recording conditions, methodologies and speakers (and hence vocal tract characteristics) used in various studies. The current study explores the extent to which the discriminatory power of LTFDs is influenced by language itself, by using a corpus of bilingual speakers and keeping factors other than language constant.

Methods

High-quality microphone recordings of 60 male adult bilingual speakers of Canadian English and French from the Voice ID Database (Royal Canadian Mounted Police, 2010–2016) were analysed. All recordings, consisting of phonetically balanced read speech in both languages, were automatically segmented with the Montreal Forced Aligner (McAuliffe et al, 2017), followed by manual correction. F1–F3 frequencies and bandwidths were automatically extracted from all vowel and semivowel segments, using 25-ms windows at intervals of 10 ms and fixed settings of a maximum of 6 formants up to 5500 Hz.

To test the validity of LTFDs, the speakers were randomly partitioned into equal sets of test, development and reference speakers. All extracted parameters of LTFDs were modelled and compared by GMM–UBM (using the same combination of speaker sets in both languages) to generate output scores, which were calibrated by logistic regression to obtain log-LRs. The procedure was repeated 100 times to minimise the effects of random sampling. The validity of each system was evaluated using equal error rates (EER) and log-LR costs (C_{lr}).

Results

Table 1 shows the performance of LTFDs in English and French. In line with previous findings that LTFDs behave as useful speaker discriminants, the use of F1–F3 frequencies and bandwidths combined yielded low EER and C_{lr} in both languages. Paired t-tests showed no significant differences between the languages in both C_{lr} ($t(99) = 1.41$, $p = .16$) and EER ($t(99) = 0.09$, $p = .93$). These results suggest that the discriminatory potential of LTFDs is not influenced by language choice. The poster will further explore the relationship between these findings and the speakers' LTFDs themselves.

Table 1. Mean (standard deviation) C_{lr} and EER based on 100 iterations.

	C_{lr}	EER
English	0.293 (0.130)	6.81% (2.62%)
French	0.274 (0.087)	6.78% (2.40%)

References

- Becker, T., M. Jessen and C. Grigoras. (2008). Forensic speaker verification using formant features and Gaussian mixture models. *Proceedings of Interspeech 2008*, Brisbane, Australia, 1505–1508.
- 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, 188–193.
- 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, 060041.
- Heeren, W., D. van der Vloed and J. Vermeulen. (2014). Exploring long-term formants in bilingual speakers. *Presented at the 23rd IAFPA Annual Conference*, Leiden, The Netherlands.
- Hughes, V., P. Harrison, P. Foulkes, P. French, C. Kavanagh and E. San Segundo. (2017). 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, 3892–3896.
- Jessen, M. and T. Becker. (2010). Long-term formant distribution as forensic phonetic feature. *Presented at the 2nd Pan-American/Iberian Meeting on Acoustics*, Cancún, Mexico.
- McAuliffe, M., M. Socolof, S. Mihuc, M. Wagner and M. Sonderegger. (2017). *Montreal Forced Aligner* [computer programme]. Version 0.9.0, retrieved 30 November 2017 from <http://montrealcorpus-tools.github.io/Montreal-Forced-Aligner/>.
- Moos, A. (2010). Long-term formant distribution as a measure of speaker characteristics in read and spontaneous speech. *The Phonetician*, **101**, 7–24.
- 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.
- Royal Canadian Mounted Police. (2010–2016). Voice ID Database [unpublished audio corpus]. Collected at University of Ottawa.