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

Justin J. H. Lo

Department of Language and Linguistic Science, The University of York, UK
`jl2355@york.ac.uk`

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

Introduction

The study presented here explores the relationship between language and speaker discrimination. Despite separate investigations in multiple languages, the discriminatory potential of many features used in forensic speaker comparison is rarely considered in bilingual terms in the literature. Disjoint speaker populations and conditions across different studies make it hard to isolate the effect of language from their results. By keeping other factors under tight control and maintaining language as the sole variable, this study looks into the question of how the discriminatory behaviour of the same features carries across languages.

Echoing Alexander et al (2014)'s call for more analysis focusing on the performance of individual speakers, this poster applies a two-level analysis on the utility of the phoneme /s/, whose speaker-specificity has been demonstrated in various languages (Cicres & Galvadà, 2016; Heeren, 2017; Kavanagh, 2012). Apart from assessing system-wide performance with global metrics within the framework of likelihood ratios (LRs), this study turns to the "biometric menagerie" (Doddington et al, 1998; Dunstone & Yager, 2009) and makes use of zooplots to identify how each speaker performs relative to the whole population. In this study, the performance of each speaker is traced across different languages with an aim to understanding variation in the differentiability of a speaker and the source of such variation.

Methods

90 male Canadian French-English bilinguals, who all recorded the same sets of read materials in both languages separately, were chosen from the Voice ID Database (Royal Canadian Mounted Police, 2010–2016). Segmentation was performed using the Montreal Forced Aligner (McAuliffe et al, 2017), followed by manual correction. The first two spectral moments (centre of gravity and standard deviation) were then extracted from each instance of /s/ in both static and dynamic forms: the static measurement taken from a 40-ms window centred at the midpoint; the dynamic trajectory represented by a quadratic polynomial fitted to nine sets of measurements obtained from 10-ms windows evenly spread out across the duration of the phoneme.

Different configurations of spectral moments data (English vs French, static vs dynamic) were tested within the LR framework, following the procedure set out in Enzinger, Morrison and Ochoa (2016). To allow all speakers to be tested and to minimise the effect of random sampling, the procedure was iterated 100 times. The global performance of each configuration was assessed using equal error rates (EER) and log-likelihood-ratio costs (C_{llr}). Zooplots were then constructed by plotting each speaker's mean LR in different-speaker comparisons against their mean LR in same-speaker comparisons.

Results

Results from LR testing suggest that, for the same group of speakers, /s/ does not provide the same overall discriminatory potential in English and in French. In particular, it demonstrates consistently stronger C_{llr}'s in French, regardless of the input form of measurements. Zooplot analysis further shows great variation in the way individual speakers' performance responds to the change in language, both in direction and in magnitude. The poster explores this in greater detail.

References

Alexander, A., O. Forth, J. Nash and N. Yager. (2014). Zooplots for Speaker Recognition with Tall and Fat Animals. In 23rd International Association for Forensic Phonetics and Acoustics Annual Conference. Zürich, Switzerland.

Cicres, J. and N. Gavaldà. (2016). A preliminary analysis of the speaker-specificity of the first four spectral moments of Spanish fricatives. In 25th International Association for Forensic Phonetics and Acoustics Annual Conference. York, UK.

Doddington, G., W. Liggett, A. Martin and M. Pryzbocki and D. Reynolds. (1998). SHEEP, GOATS, LAMBS and WOLVES: A statistical analysis of speaker performance in the NIST 1998 Speaker Recognition. In Proceedings of the 5th International Conference on Spoken Language Processing, 1351–1354. Sydney, Australia.

Dunstone, T. and N. Yager. (2009). Biometric system and data analysis: Design, evaluation, and data mining. Springer.

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.

Heeren, W. (2017). Speaker-dependency of /s/ in spontaneous telephone conversation. In 26th International Association for Forensic Phonetics and Acoustics Annual Conference. Split, Croatia.

Kavanagh, C.M. (2012). New consonantal acoustic parameters for forensic speaker comparison. PhD Thesis, University of York.

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://montrealcorpustools.github.io/Montreal-Forced-Aligner/.

Royal Canadian Mounted Police (2010–2016). Voice ID Database [unpublished audio corpus]. Collected at University of Ottawa.