Forensic voice comparison at the intersection of linguistics and automatic speaker recognition

Vincent Hughes

Department of Language and Linguistic Science, University of York, UK

vincent.hughes@york.ac.uk

Within the field of forensic speech science, there has been growing interest in integrating traditional linguistic methods with automatic speaker recognition (ASR) systems. This work has two aims. The first is to better understand what linguistic information is captured by increasingly 'black boxy' ASR systems. The second is to empirically combine the results of linguistic analysis with ASR output, to reduce overall error rates. Some studies have shown promising results. For example, Gonzalez-Rodriguez et al. (2014) and Hughes et al. (2017) found that ASR misclassifications can be resolved by trained phoneticians primarily using laryngeal voice quality analysis. However, key questions remain:

· As systems continue to produce marked improvements in overall performance with each new paradigm (usually every 3-5 years), will we reach a stage where forensic voice comparison is conducted entirely using ASR?

· If so, what role will linguistic methods have in forensic casework in the future?

I will argue that to answer these questions we must recognise that forensics is a unique application of ASR. As such, we have different concerns and priorities from developers of ASR systems for other commercial applications. Specifically, this means that: (i) Features for analysis should be determined on a case-by-case basis; (ii) The context in which forensic recordings are made is unique, making replication difficult; (iii) Our focus should be on reducing uncertainty rather than maximising potential discriminability. This involves identifying, reporting, and attempting to mitigate for sources of variability in system performance (e.g. sample size); (iv) The state-of-the-art ASR system isn't necessarily the best choice for every forensic case.

In this talk, I will review the current state of knowledge at the intersection of linguistics and ASR, and make proposals for ways forward in the quest to find the best ways of analysing voices in the specific context of forensic comparison.

References

González-Rodriguez, J., Gil, J., Perez, R. and Franco-Pedroso, J. (2014). What are we missing with i-vectors? A perceptual analysis of ivector-based falsely accepted trials. In Proceedings of Odyssey: The Language and Speaker Recognition Workshop, Joensuu, Finland. pp. 33–40.

Hughes, V., Harrison, P., Foulkes, P., French, J. P., Kavanagh, C. and San Segundo, E. (2017) Mapping across feature spaces in forensic voice comparison: the contribution of auditory-based voice quality to (semi-)automatic system testing. In Proceedings of Interspeech, Stockholm, Sweden. pp. 3892–3896.