An exercise in calculating numerical likelihood ratios and the practicalities of their implementation

Erica Gold[1] *and Peter French*[1,2]

[1]*Department of Language and Linguistic Science, University of York, York, United Kingdom*
`erica.gold@york.ac.uk`

[2]*J P French Associates, York, United Kingdom*
`peter.french@jpfrench.com`

In recent years, there have been calls for improvements in the quality of forensic evidence by a number of legal and government bodies. It has been argued that all areas of forensic science need to be more transparent, that forensic examinations should be based on validated methodologies, and that the results should be replicable and expressed in quantitative terms (U.S. National Research Council, 2009; House of Commons' Northern Ireland Affairs Committee, 2009; Law Commission of England & Wales, 2011).

To heed these calls for improvement, a considerable amount of research in forensic speaker comparison has been devoted to the application of the numerical likelihood ratio framework (Morrison 2009). The research presented in this paper serves as an exercise in calculating numerical likelihood ratios for a linguistically-homogeneous population of 100 male, Southern Standard British English speakers (Nolan et al. 2009). This paper considers the discriminant power of four parameters (described as good speaker discriminants by experts in Gold and French 2011) in combination, while evaluating the practicalities of the numerical likelihood ratio framework for forensic speaker comparison casework.

The four parameters analyzed are articulation rate, fundamental frequency, long-term formant distributions, and the incidence of clicks (velaric ingressive plosive). Three of the parameters (clicks are excluded owing to the special difficulties they pose for statistical modeling) are combined into an overall likelihood ratio, where the combined calibrated system achieves an EER of .0554 and a Cllr of 0.2831. These results are equivalent to those achieved using a highly developed ASR on the same data, and could undoubtedly be improved upon further by the incorporation of more parameters into the overall package.

The exercise of calculating numerical likelihood ratios revealed a number of difficulties that surround the framework and its application to forensic speaker comparisons. Five prominent difficulties will be discussed in turn: subjective elements of the methodological process, delimiting the relevant reference population, availability of population statistics, lack of models available to calculate LRs, and appropriate procedures for the combination of parameters. The findings of this research are intended to promote discussion on the practical use of numerical likelihood ratios in forensic speaker comparison casework.

References

Gold, E. & French, P. (2011) International practices in forensic speaker comparison. *International Journal of Speech, Language and the Law* 18(2): 293-307.

House of Commons Northern Ireland Affairs Committee (2009). *Cross-border Co-operation between the Governments of the United Kingdom and the Republic of Ireland: Second Report of Session 2008–2009*. London: The Stationery Office.

Morrison, G.S. (2009). Forensic voice comparison and the paradigm shift. *Science and Justice*, 49: 298-308.

National Research Council (2009). *Strengthening Forensic Science in the United States: A Path Forward*. Washington D.C.: The National Academic Press.

Nolan, F., McDougall, K., de Jong, G. and Hudson, T. (2009). The DyViS database: style-controlled recordings of 100 homogeneous speakers for forensic phonetic research. *International Journal of Speech, Language and the Law* 16(1): 31-57.

The Law Commission (2011). *Expert Evidence in Criminal Proceedings in England and Wales*. London: The Stationery Office.