

Forensic experts should focus on uncertainty rather than discriminability

Please indicate here whether you prefer:
poster

Saks & Koehler (2005) described a *paradigm shift* within what they called forensic identification science (now better referred to as forensic comparison science, of which forensic voice comparison is one). This involved a move away from unscientific methods, founded on the principle of *discernible uniqueness* – the notion that patterns can be compared to determine a match or mismatch. Since then, the paradigm shift has been extended to include (i) expression of expert conclusions using likelihood ratios, (ii) data-driven estimation of typicality, and (iii) validation of methods in line with international standards. The issue of validation, in particular, has received considerable attention and experts are now under pressure from policy makers and regulators to validate their methods and systems in order to demonstrate that they work. However, forensic validation (and the same is true of forensic research more generally) tends to focus on the overall performance of methods under casework conditions as evaluated by metrics such as EER or C_{lr} . This implicitly focuses the expert's attention on discriminability with different methods chosen and decisions made in a case based on low values (or low assumed values) for the validity metric used. For example, an expert might choose to analyse F3 in a forensic voice comparison case and attach additional weight to the difference between known and unknown samples on the assumption that F3 is generally a good speaker discriminant. The view that discriminability should be the expert's primary focus has been proposed in Smith & Neal (2021).

We disagree with this view. Rather, we believe that the expert's primary concern should be to reduce uncertainty, rather than maximising potential discriminability (i.e. the possibility that a method could produce a low validity value). This is because reducing uncertainty is directly related to reducing the probability of a miscarriage of justice, which is the ultimate aim of the judicial process. Uncertainty here is defined broadly as variability in the specific conclusion (the LR) or validity value that a method produces (also referred to as reliability). Forensic voice comparison, of any kind, involves a series of decisions that potentially introduce uncertainty – in other fields this is referred to as *researcher degrees of freedom* (Roettger 2019).

In this paper, we discuss the issue of uncertainty in forensic voice comparison and demonstrate how it may be reduced via techniques such as Bayesian calibration (Brümmer & Swart 2014). We also present a series of recommendations for forensic experts. Specifically, experts should:

1. Recognise that forensic comparison is a process involving numerous decisions which introduce uncertainty via both systematic and random factors
2. Be explicit about the decisions made at each stage of the process and the implications of such decisions for uncertainty in terms of the results LRs **and** overall method validity
3. Take steps to measure and minimise uncertainty

The focus on uncertainty also directly relates to issues of reproducibility and replicability. In this paper, we consider the specific challenges these concepts pose for forensic voice comparison.

References

- Brümmer, N. and Swart, A. (2014) Bayesian calibration for forensic evidence reporting. *Proc. Interspeech*, 388-392.
- Roettger, T. B. (2019) Researcher degrees of freedom in phonetic research. *Journal of the Association for Laboratory Phonology*, 10, 1.

- Saks, M. J., & Koehler, J. J. (2005) The coming paradigm shift in forensic identification science. *Science*, 309, 892–895.
- Smith, A. M. & Neal, T. M. S. (2021) The distinction between discriminability and reliability in forensic science. *Science & Justice*, 61, 319–331.