

Forensic Speaker Comparison Using Machine and Mind

Jonas Lindh^{1,2}

¹*University of Gothenburg, Sweden*

jonas.lindh@gu.se

²*Voxalys AB, Sweden.*

jonas@voxalys.se

Over a period of 10 years of conducting forensic casework we have developed a framework for working with cases using both AVC (Automatic Voice Comparison) and phonetic, linguistic analysis. Initially the open source system ALIZE was used in some cases (Bonastre et al., 2005) and subsequently, 2 commercial systems have been acquired (Batvox 4.1 and Vocalise).

A correct logically coherent framework uses a likelihood ratio framework for dealing with the evaluation of the forensic analysis at hand (Morrison, 2012). In Sweden, the National Forensic Centre (NFC¹) has implemented a verbal scale where spans of likelihood ratios are applied to each step in a 9-point ordinal scale (Nordgaard et al., 2012), similar to the verbal scale suggested by Champod and Evett (2000).

Morrison (2012) comprehensively explains how there is no conflict in applying subjective judgments together with, or separate from, the use of statistical calculations of likelihood ratios using databases. Most forensic disciplines apply both to some extent. However, in Sweden it has always been important that reports are presented within the same framework, across disciplines, which is why so-called calibration meetings across disciplines have been a frequently occurring standard for several years. We considered and approached our small discipline in the same way and started judging spans of likelihood ratios for different features holistically. When it comes to the automatic systems, we have gathered databases and continuously perform evaluations on data that is as similar as possible to the case data we have experience of. Through testing several systems and techniques in matched and mismatched conditions one becomes very aware of the performance, possibilities and types of scores the systems produce. The more this approach is utilised, the more advantageous information can be retrieved regarding behavioural aspects of both mind and machine, helping the analyses and thereby aiding the course of justice as much as possible.

Morrison (2012) discusses subjectiveness and objectiveness. It seems to us that, occasionally, the reference to statistical models and databases is used to distance an expert from the responsibility of the outcome of the report, which *we* believe to be a problem. Independently of the technique and supposed objectiveness, a transparent report (including limitations) and evaluation (pointed out in Morrison, 2012) can help the different judicial systems to understand and interpret the outcome of an analysis and the ability of a specific laboratory. After proper evaluations, it will be

¹ <http://nfc.polisen.se/en/English/>

a lot easier to start producing proper guidelines. The ratio in certain cases can be calculated with some different techniques such as through an automatic system, formant values and F0. The variation of the specific ratio depends on the editing of files, normalisation techniques utilised, system used etc., and is so great that it is rather a span of ratios that can be calculated and/or judged. Furthermore, the perceiver of the outcome of an analysis has to interpret the number or span, which produces yet another source of variation. An intermediate path is to follow the recommendations in Nordgaard et al. (2012) and help the perceiver and judge of the likelihood ratio by incorporating the outcome into an ordinal scale with a verbal expression value as an aid.

It is not logical to demand that a score-based system should report an objective figure independently of what one feeds it with. Sometimes it seems that analysts accept anything the system presents them with, as if it were a magic black box. We have to learn how to best use and how to interpret the different systems we employ in the same way as we subjectively judge the ability of students or analysts. Whether we calculate or judge a likelihood ratio, this should be transparent in a report and the outcome should be expressed in a logical framework that bridges both our discipline and that of the judicial system. However, this does not mean that each laboratory should stop evaluating themselves using available databases and collect new data. Rather, this suggests a need for more research, evaluations and collection of data alongside the everyday casework.

References

- Bonastre, J-F., Wils, F. and Meignier S. "ALIZE, a free toolkit for speaker recognition." *ICASSP (1)*. 2005.
- Champod, C. and Evett, I. W. Commentary on A. P. A. Broeders (1999) 'Some Observations on the Use of Probability Scales in Forensic Identification', *Forensic Linguistics*, 6(2): 228–41. *The International Journal of Speech, Language and the Law*, 7:238–243, 2000.
- Morrison, G. S. "The Likelihood-ratio Framework and Forensic Evidence in Court: A Response To *R v T*." *The International Journal of Evidence & Proof* 16.1 (2012): 1-29.
- Nordgaard, A., Ansell, R., Drotz, W. and Jaeger, L. "Scale of Conclusions for the Value of Evidence." *Law, Probability and Risk* 11.1 (2012): 1-24.