Forensic Automatic Speaker Recognition: Fiction or Science?

Joaquin Gonzalez-Rodriguez

ATVS - Biometric Recognition Group
Department of Computer Science, Universidad Autonoma de Madrid, Spain
joaquin.gonzalez@uam.es

Abstract

Hollywood films and CSI-like movies show a technology landscape far from real, both in forensic speaker recognition and other identification-of-the-source forensic areas. Lay persons are used to good-looking scientist-and-investigators performing voice identifications ("we got a match!") or smart fancy devices producing voice transformations causing one actor to instantaneously talk with the voice of other. The expectation both in juries and fact-finders about what can be achieved with present technology in court is overestimated, a subject which has been properly investigated and become known as the CSI-effect [18].

Simultaneously, Forensic Identification Science is facing a global challenge [17] impelled firstly by progressively higher requirements for admissibility of expert testimony in Court [8], and secondly by the transparent and testable nature of DNA typing, which is now seen as the new gold-standard model of a scientifically defensible approach to be emulated by all other identification-of-the-source areas [2]. The likelihood ratio (LR) approach to the analysis of forensic evidences [9][1], popularised by DNA typing in the last decades, can be used for any forensic identification disciplines to report the strength of the evidence. This LR avoids conclusions in the form of an absolute decision of authorship, but provides a degree of support to an incriminating hypothesis versus an exonerating alternative hypothesis, avoiding typical interpretation fallacies [1][7]. The reported LR value can then be used by the Court in combination with prior information on the case and other possible evidence. It is the role of the judge or jury to determine if, with all the information at hand, a decision "beyond reasonable doubt" can be taken or not, based not only on the speech evidence but on all the available information in the case [12]. It is not the scientist's role to decide on the ultimate issue of guilt or innocence but just to provide information relative to the speech evidence at hand.

Forensic Speaker Recognition has had a sordid history (see [14] for an excellent review), which has provoked controversial positions about its use in Court [4][5]. Current forensic approaches, used by several official labs and practitioners, are usually based on single or combined auditory, acoustic and semi-automatic approaches [13]. However, some steps forward are possible (and are being undertaken [16][10]) in order to comply with Forensic Science [7] and admissibility current requirements.

Although Automatic Speaker Recognition has experienced dramatic improvements in the last decade [19](see [3] for an overview of state of the art), as shown in yearly NIST Speaker Recognition Evaluations (SRE) since 1996, detection performance is not error free, specially in forensic recordings where low quality (noise, distortions, etc.) and severe mismatch (channel, speaking style, etc.) between control and recovered samples is usual. Fortunately, taking decisions about the identity of the speaker is not the objective in forensic speaker recognition [7][11] but the estimation of reliable information about the strength of the evidence, which can be properly done in the form of LR values. Moreover, systems and procedures eliciting LR values can be calibrated [6], as demonstrated by leading systems in SRE'06 and SRE'08 [3] and recently by some traditional phonetic approaches [11][15], reducing the calibration loss for any prior probability (dependent of the forensic case at hand) at the given discrimination loss. However, caution is needed in order to handle adequate populations in each case, because mismatch between the recording conditions of the population at hand and those present in the questioned and suspect speech can lead to likelihood ratios not accurately representing the weight of the evidence.

Putting all the above together, we will show in this presentation how forensic speaker recognition can comply with the requirements of transparency and testability in forensic science [11]. This will lead to fulfilling the court requirements about role separation between scientists and judges/juries, and bring about integration in a forensically adequate framework in which the scientist provides the appropriate information necessary to the court's decision processes.

1. References

[1] C.G.G. Aitken and F. Taroni, *Statistics and the Evaluation of Evidence for Forensic Scientists*, John Wiley & Sons, Chichester, 2004.

[2] D.J. Balding, *Weight-of-Evidence for Forensic DNA Profiles*, Wiley, 2005.

[3] K. Berkling, J.F. Bonastre and J.P.Campbell (eds.), Special Issue on Speaker and Language Recognition, IEEE Transactions on Audio, Speech and Language Recognition, vol. 15, no. 7, pp. 1949-2115, September 2007.

[4] J. F. Bonastre, F. Bimbot, L.-J. Boe, J. P. Campbell, D. A. Reynolds, and I. Magrin-Chanolleau, "Person authentication by voice: A need for caution", Proc. of Eurospeech 2003, pp. 33-36, 2003.

[5] A. Braun and H.J. Künzell, "Is forensic speaker identification unethical? - or can it be unethical not to do it?", Forensic Linguistics 5/1: 10:21, 1998.

[6] N. Brummer and J. du Preez, "Application independent evaluation of speaker detection", Computer Speech and Language, vol. 20, no. 2-3, pp. 230-275, 2006.

[7] C. Champod and D. Meuwly, "The inference of identity in forensic speaker recognition", Speech Communication, vol. 31, pp. 193-203, 2000.

[8] U.S. Supreme Court, "Daubert v. Merrel Dow Pharmaceuticals 509 U.S. 579", 1993.

[9] I. W. Evett, "Towards a uniform framework for reporting opinions in forensic science casework", Science and Justice, vol. 38, no. 3, pp. 198-202, 1998.

[10] P. French *et al.*, ”Position Statement concerning use of impressionistic likelihood terms in forensic speaker comparison cases”, The International Journal of Speech, Language and the Law, vol. 14.1, pp. 137-144, 2007.

[11] J. Gonzalez-Rodriguez, P. Rose, D. Ramos, D.T. Toledano and J. Ortega-Garcia, ”Emulating DNA: rigorous quantification of evidential weight in transparent and testable forensic speaker recognition”, IEEE Trans. On Audio Speech and Language Processing, vol. 15, no. 7, pp. 2104-2115, 2007.

[12] D. Hodgson, ”Speaker Identification: a Judicial Perspective”, HCSNet Workshop on Forensic Speaker Recognition, Dec. 2007, Sydney, Australia.

[13] H.J. Knzell, ”Current Approaches to Forensic Speaker Recognition”, Proceedings of ESCA Workshop on Automatic Speaker Recognition, pp. 135-141, Martigny (Switzerland), 1994.

[14] D. Meuwly, *Reconaissance de Locuteurs en Sciences Forensiques: L'apport d'une Approache Automatique*, Ph.D. thesis, IPSC-Universite de Lausanne, 2001.

[15] G.S. Morrison and Y. Kinoshita, ”Automatic-Type Calibration of Traditionally Derived Likelihood Ratios: Forensic Analysis of Australian English /o/ Formant Trajectories”, Proceedings of Interspeech 2008, Brisbane, Australia, 2008.

[16] P. Rose, *Forensic Speaker identification*, Taylor & Francis, 2002.

[17] M. J. Saks and J.J. Koehler, ”The coming paradigm shift in forensic identification science”, Science, vol. 309, no. 5736, pp. 892-895, 2005.

[18] N.J. Schweitzer and M.J. Saks, ”The CSI effect: popular fiction about forensic science affects public expectations about real forensic science”, Jurimetrics, Vol. 47, p. 357, Spring 2007.

[19] D. van Leeuwen, A. Martin, M. Przybocki, and J. Bouten, ”The NIST 2004 and TNO/NFI speaker recognition evaluations”, Computer Speech and Language, vol. 20, no. 2-3, pp. 128-158, 2006.