C4 Closing the Performance Gap in Forensic Speaker Recognition

Reva Schwartz, MA, NIST, Forensic Science Research Program, 100 Bureau Drive, Stop 8102, Gaithersburg, MD 20899*

The goal of this presentation is to educate the broader forensic community about current capabilities in forensic speaker recognition and National Institute of Standards and Technology (NIST) activities to support progress in this field.

This presentation will impact the forensic science community by providing an overview of what is possible in the field of forensic speaker recognition, an update on research activities, and a description of future directions.

Speaker recognition is the process of determining whether two (or more) speech samples (live or recorded) are from the same person or different people. While speech recognition seeks to answer what is being said, speaker recognition attempts to answer who is talking. Forensic speaker recognition isn't called upon very often as a forensic technique, but when it is, there are a variety of methods in use.[1]

In forensic casework, analysts are asked to perform speaker comparison tasks from evidentiary recordings that typically exhibit a high degree of variability. The forensic speaker recognition research community has spent a great deal of time and resources in an attempt to study and address these sources of variability. Most of the benefits have played out in algorithm development and, when tested with the type of high-quality speech data that is usually unseen in forensic settings, technology performance has improved greatly.[2] Yet, very often these tools do not perform well when confronted with the type of data encountered in forensic casework.[3] This performance gap is real and significant when compared to performance under optimal conditions. The path forward to address this gap has taken time to come into view.

The Organization of Scientific Aea Committees (OSAC) Speaker Recognition Subcommittee and the European Network of Forensic Science Institutes (ENFSI) Forensic Speech and Audio Analysis Working Group have been developing guidelines and best practice documents, separately, for various aspects of forensic speaker comparison.[4] Their work also seeks to answer some of these lingering questions about how to harmonize practice across the discipline due to the myriad forms of variability present in case data and the differing approaches used in comparison.

Ultimately, any forensic system must be tested under the conditions seen in casework.[5] This means we need to assess performance of these systems using a variety of data types.

While the use of automation in forensic practice has grown, laboratories tend to rely on human practitioners for various aspects of the examination process. How any of these systems — human, automation, hybrid — perform under forensic conditions, remains unclear.[6-8]

To illuminate future paths for forensic speaker comparison, NIST will focus on a series of research activities in the near term, which will be described in detail during this presentation: (1) improvement of underlying technology; (2) assessment of performance across listener types; (3) comparison of performance between listeners and state-of-the-art automated systems; (4) focus on data that bears a close resemblance to that seen in forensic conditions, through the use of: (a) new experimental data collections dedicated to a different set of variables than in the past; (b) operational data for: (i) system testing; (ii) new development of reference datasets; and, (iii) system calibration; and, (5) development of a new Speaker Recognition Evaluation focused on issues inherent to forensic practice.

Reference(s):

1. G.S. Morrison, F.H. Sahito, G. Jardine, D. Djokic, S. Clavet, S. Berghs, C. Goemans Dorny. (2016). INTERPOL survey of the use of speaker identification by law enforcement agencies. *Forensic Science International*. 263, 92–100.
2. C.S. Greenberg, D. Banse, G.R. Doddington, D.G. Romero, J.J. Godfrey, T. Kinnunen, A.F. Martin, A. McCree, M. Przybocki, D.A. Reynolds. The NIST 2014 Speaker Recognition i-Vector Machine Learning Challenge, Odyssey 2014: The Speaker and Language Recognition Workshop, 16-19 June 2014, Joensuu, Finland.
3. A. Alexander, F. Botti, D. Dessimoz, A. Drygajlo. (2005). The effect of mismatched recording conditions on human and automatic speaker recognition in forensic application. *Forensic Science International*. pages S95–S99.
4. A. Drygajlo, M. Jessen, S. Gfroerer, I. Wagner, J. Vermeulen, T. Niemi. (2015). Methodological Guidelines for Best Practice in Forensic Semiautomatic and Automatic Speaker Recognition, for the ENFSI Expert Working Group Forensic Speech and Audio Analysis.
5. G.S. Morrison. (2009a). Forensic voice comparison and the paradigm shift. *Science & Justice.* 49: 298–308.
6. V. Hautamaki, T. Kinnunen, M. Nosratighods, K. Lee, B. Ma, H. Li. (2007). Approaching human listener accuracy with modern speaker verification, *Interspeech*. 2007. pp. 950-954, Antwerp.
7. G. Sell, C. Suied, M. Elhilali, S. Shamma. (2015). Perceptual susceptibility to acoustic manipulations in speaker discrimination. *J. Acoust. Soc. Am*. 137 (2) 911-922.
8. C.S. Greenberg, A.F. Martin, G.R. Doddington, J.J Godfrey. (2011). Including human expertise in speaker recognition systems: Report on pilot evaluation. In Proceedings of ICASSP, pp. 5896–5899.

Speaker Recognition, Reference Data, System Performance