

Developments in automatic speaker recognition at the BKA

Michael Jessen¹, Franz Bross², and Stefan Gfroerer¹

¹*Department of Speaker Identification and Audio Analysis, Bundeskriminalamt, Germany*

{Michael.Jessen|Stefan.Gfroerer}@bka.bund.de

²*Laboratory for Information and Communication Technology, University of Applied Sciences,
Koblenz, Germany*

bross@fh-koblenz.de

Leaving aside some early approaches initiated the 1970s, developments in forensic automatic speaker recognition at the BKA have restarted in 2002. This contribution presents a chronological overview of the activities in that domain. The chronological stages can be roughly characterised as follows:

1. Implementing and testing a standard automatic speaker recognition system using lab speech data ("Pool 2010", which has also been used in multi-speaker phonetic investigations).
2. Steps towards a specifically forensic approach to automatic speaker recognition (e.g. implementing GMM; working with GSM data and simulating forensically realistic adverse conditions).
3. "Learning from the professionals": getting to know BATVOX (www.agnitio.es); incorporating the principles of ASPIC (e.g. Drygajlo et al., 2003).
4. Implementing the fully forensic automatic speaker recognition system SPES and testing it with BKA case data (Bross et al., 2006) and with the NFI-TNO test (van Leeuwen et al., 2006).
5. Improving the system technologically (e.g. switching from MFCC to PLPCC; optimising parameters such as the number of GMM modules, shape and size of spectral windows etc.).

In the presentation special attention will be given to aspects that are of interest for forensic phoneticians who plan to supplement their inventory of methods by automatic speaker recognition or who made their first experiences in that area. Aspects to be addressed include the following:

- The difference between fully automatic speaker recognition and manually-guided (not fully) automatic speaker recognition
- Influence of the size and shape of the suspect data on automatic speaker recognition
- The difference between Drygajlo-normalisation and the direct method
- Experiences from using automatic speaker recognition in actual casework

A discussion of possible future developments will be provided, including the issue whether and how phonetic features should be taken into account in automatic speaker recognition.

References

- Bross, F., M. Jessen and J.P. Köster (2006). Some experiences from tests of an automatic speaker recognition system under forensic conditions. *Proceedings of the 4th European Academy of Forensic Science Conference*, Helsinki, pp. 45–46.
- Van Leeuwen, D.A., A.F. Martin, M.A. Przybocki and J.S. Bouten (2006). NIST and NFI-TNO evaluations of automatic speaker recognition. *Computer Speech and Language* **20**, 128–158.
- Drygajlo, A., D. Meuwly and A. Alexander (2003). Statistical methods and Bayesian interpretation of evidence in forensic automatic speaker recognition. *Proceedings of Eurospeech 2003*, Geneva, pp. 689–692.