Forensic and Investigative Speaker
Recognition: Situation BKA

Odyssey 2016
June 21-24, Bilbao, Spain

Michael Jessen
BKA, Forensic Science Institute: Forensic Speech & Audio Department
The outcome of a FASR or FSASR method/system is a Likelihood Ratio.

Validation of a LR method in FASR or FSASR has to be performed with speech samples that are typical representations of the speech material the forensic laboratory is confronted with in everyday work.

http://enfsi.eu/about-enfsi/structure/working-groups/speech-and-audio
Collecting forensic validation data

- Collecting validation data from previous casework at BKA (Solewicz et al., Odyssey 2012) and ongoing

- Collaborating with Netherlands Forensic Institute: NFI-FRITS (van der Vloed et al., Odyssey 2014)

  
  So far, best coverage of matching conditions involving telephone interception data

Still very sparse validation data on **mismatched conditions**.  
**Typical example:** terrorist video (questioned-speaker recording) vs. Telephone-intercepted recording (suspected speaker recording)

Further challenge: cases in several **languages**. Collecting appropriate validation data but also investigating predictable language effects (e.g. LR-shifting effects when Reference Population from another language is used).
Combining quantitative with qualitative evidence

Different types of evidence:

- Quantifiable evidence (LRs) coming from automatic and semiautomatic systems/methods (domain of the Guidelines)

- Qualitative evidence coming from the auditory-phonetic-and-acoustic-phonetic method (Gold & French 2011)

How to combine the two? It has to be a verbal strength of evidence statement, since the combined result is not entirely quantitative.
The courts in Germany expect posterior statements:

**Identity / Non-Identity**

- cannot be assessed
- is probable
- is highly probable
- is very highly probable
- can be assumed with near certainty

Inertia against switching to a Bayesian framework, where

- Speech expert supplies LR or verbal equivalent.
- Other forensic experts do the same.
- The court supplies the prior odds and calculates the posterior odds from the priors odds and all the LRs (or performs the verbal equivalent of these steps).
Interpretation of a case upon the background of a relevant validation
Speaker recognition (voice comparison) at BKA: feature-based classification

Language variety / idiolect
- Regional patterns
- Second language acquisition / ethnolect

Voice characteristics
- F0-level
- Auditory voice quality
- Acoustic vocal tract characteristics
- Formants
- Cepstral coefficients

Habitual speaking patterns
- Tempo
- Dysfluencies
- Breathing patterns
- Intonation patterns
- Speech rhythm

Report: Overall verbal strength-of-evidence statement

Analysed within auditory-phonetic-and-acoustic-phonetic approach
Analysed within automatic speaker recognition
The outcome of a FASR or FSASR method/system is a Likelihood Ratio (LR) – preferably calibrated, otherwise uncalibrated (score).

Each FASR or FSASR method used in a case has to undergo a method validation (test with a set of known same-speaker and different-speaker trials, satisfying a validation criterion) or a case-specific evaluation (same, but without validation criterion). These forms of testing result in performance metrics such as Cllr and performance characteristics such as Tippett plots.

It is crucial that the speech samples used in the method validation or case-specific evaluation are representative of the casework. This can be achieved by building validation databases drawn from previous casework.

The LR outcome of a comparison has to be interpreted upon the background of the relevant method validation or case-specific evaluation (next slide).

If in addition to FASR or FSASR there are findings from the domain of the auditory-phonetic-and-acoustic-phonetic method (AuPA & AcPA; see Gold & French 2011), they can be combined into a final verbal statement.
Methodological guidelines for best practice in forensic semiautomatic and automatic speaker recognition

- The guideline project is an accepted proposal to the 2011 call of the so called Monopoly programme of ENFSI (European Network of Forensic Science Institutes).

- Between 2013 and 2015 several meetings
  - Meetings with input from invited experts in practical FSR and from R&D (Lausanne; Wiesbaden; Warsaw)
  - Internal leading team meetings in order to discuss progress in the writing of the document(s) (Helsinki; den Haag; Wiesbaden)
  - Dissemination conference in Warsaw Sep 2015

- After considering the expert feedback and performing internal corrections the document was completed by the end of 2015.

http://enfsi.eu/about-enfsi/structure/working-groups/speech-and-audio
The guideline project: participants

- EPFL Lausanne (A. Drygajlo, project leader)
- Bundeskriminalamt (BKA) (S. Gfroerer, M. Jessen, I. Wagner)
- Finnish Police (T. Niemi)
- Netherlands Forensic Institute (NFI) (J. Vermeulen)
- Polish Police

Further participation from forensic institutes/police units in Spain and France, which were withdrawn during the project for independent reasons.
1. Aims
2. Scope
3. Methodology of FASR and FSASR
4. Method validation
5. Case Assessment
6. Evaluation and Interpretation
7. Case File and Reporting
8. Quality Assurance

Appendix 1. Annotated Bibliography
Appendix 2. List of Abbreviations

There is also a second, smaller, document on proficiency tests and collaborative exercises for quality management purposes that is not addressed in this talk.