Speaker Recognition for Forensic Applications

Joseph P. Campbell
j.campbell@ieee.org

16 June 2014

LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

This work was sponsored under Air Force contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

Speaker Recognition for Forensic Applications
Introduction

- **In forensic or investigative speaker comparison, speech utterances are compared by humans and/or machines for use in court or investigation**
  - **High-stakes application affecting people's lives demands highest scientific standards**
- **Unfortunately, methods used in practice vary widely – and not always for the better*†**
- **Methods and practices grounded in science are critical for proper application and nonapplication‡ of speaker comparison to a variety of international investigative and forensic applications**
- **Provide a critical analysis of current techniques employed and lessons learned**
- **Crucial to improve communication between automatic speaker recognition researchers, legal scholars, and forensic practitioners internationally**
  - **Legal, policy, and societal questions such as allowing speaker comparisons in court**
  - **Requirements for expert witnesses**
  - **Requirements for specific automatic or human-based methods to be considered scientific**
- **You can help!**

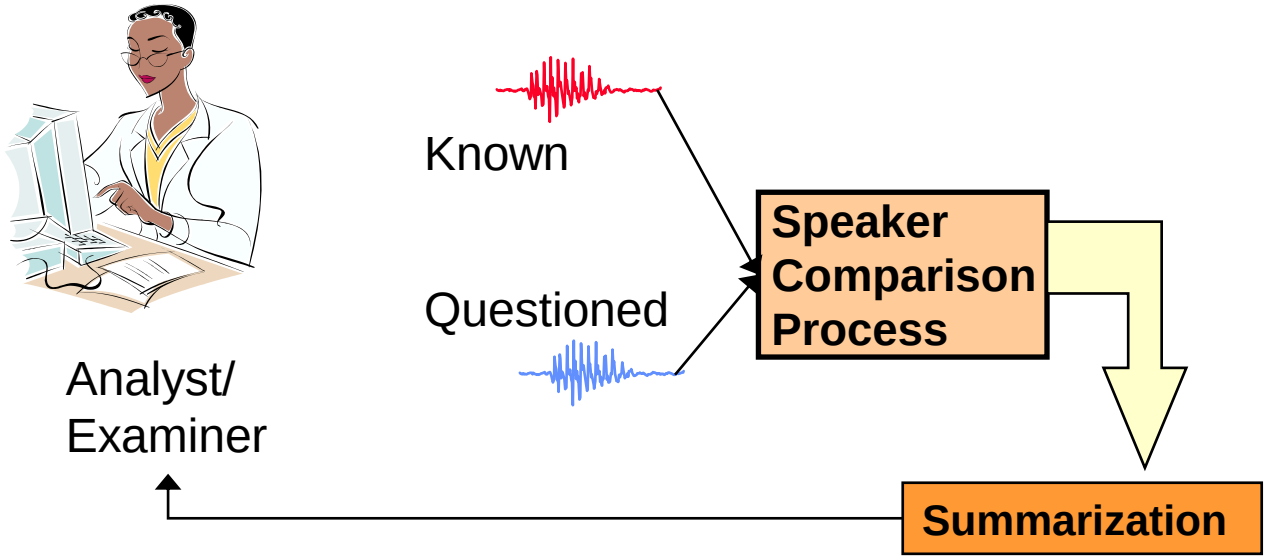
*** P. Rose, *Forensic Speaker Identification,* Taylor & Francis, 2002.**
† E. Gold, P. French, *International practices in forensic speaker comparison,* IJSLL, 2011.
‡ Schwartz, R., et al., *When to Punt on Speaker Comparison?,* 162nd Meeting ASA, 2011.

Outline

- **Background**
- **Approaches**
- **Activities**
- **Request**
- **Future**
- **Conclusion**

Forensics and Investigation

- **Forensics: seeks to establish facts of interest using science and technology in the context of the law or in a court of law**
- **Investigation: systematic inquiry, examination, study, and survey of facts, circumstances, situations, incidents, and scenarios to render a conclusion**



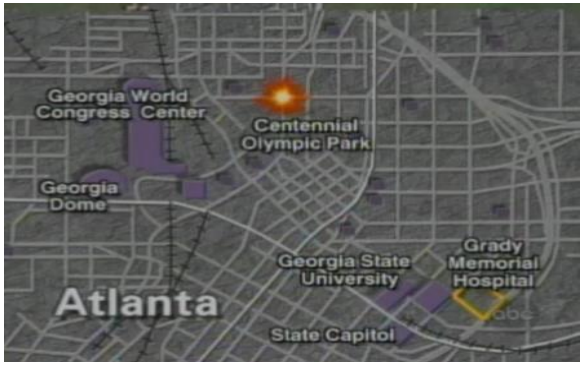
Variations of Speaker Comparison*

	Evidential forensic speaker comparison	Investigatory forensic speaker comparison	Speaker comparison within investigatory voice biometrics (AFIS/ASIS-style)
Presentation in court?	Yes	No	
Number of comparisons	Single comparison (or a relatively small set of comparisons within a complex case)		Large or very large number of comparisons
Methods	Auditory + acoustic; HASR etc. (e.g., see Gold & French†)		Fully automatic;* i.e., the investigator makes the database search without listening to the voices in the database
Reports	In a way accepted by the court, usually in the form of some kind of probability statement (not a categorical yes/no) (see Gold & French†)	Either in a way accepted by the court (although it is not intended for a court) or in a simplified form, which might also include a yes/no statement	In the form of a hit list of one or more speakers from the database according to criteria specified by the user (e.g., specifying the size of the hit list; using certain costs for false identification or false rejection; specifying a threshold for a not-in-the-database decision).

*** Michael Jessen, handout, Forensic Phonetics course, Summer School in Forensic Linguistics, 2013.**
† E. Gold, P. French, *International practices in forensic speaker comparison,* IJSLL, 2011.

Forensic Speaker Recognition Examples

- **Atlanta Centennial Park Bombing (1996)**
  - **"There is a bomb in Centennial Park. You have thirty minutes." – 13-second 911 call**
    - **Are the caller and the suspect in custody the same person?**



- **Trayvon Martin (2012)**
  - **Zimmerman claims justified shooting**
  - ***Orlando Sentinel* hires "voice experts"**
    - **"Who was crying for help?"**

Trayvon Martin shooting: It's not George Zimmerman crying for help on 911 recording, 2 experts say

5:38 p.m. EST, March 31, 2012 |
By Jeff Weiner, Orlando Sentinel

As the Trayvon Martin controversy splinters into a debate about self-defense, a central question remains: Who was heard crying for help on a 911 call in the moments before the teen was shot?

A leading expert in the field of forensic voice identification sought to answer that question by analyzing the recordings for the Orlando Sentinel.

Forensic Speaker Recognition

Is it Really That Difficult?

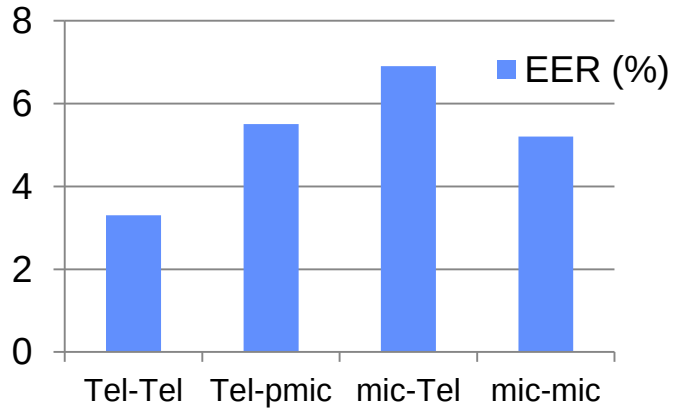
Trial	Truth	Human	Automatic
1	T	FALSE	TRUE
2	F	FALSE	FALSE
3	F	FALSE	FALSE
4	F	FALSE	FALSE
5	T	TRUE	TRUE
6	F	FALSE	FALSE
7	T	FALSE	TRUE
8	F	TRUE	FALSE
9	F	FALSE	FALSE
10	T	TRUE	TRUE
11	F	TRUE	TRUE
12	F	FALSE	FALSE
13	F	FALSE	FALSE
14	T	TRUE	TRUE
15	T	TRUE	TRUE

Incorrect Responses

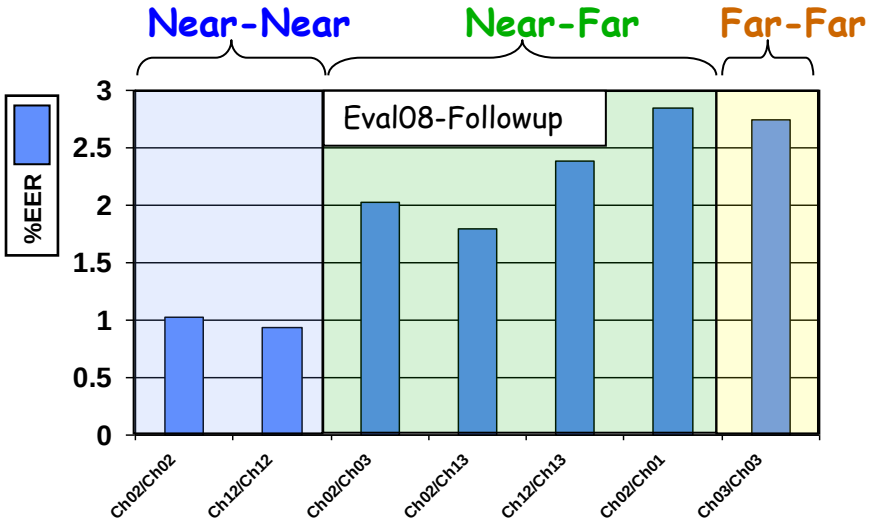
- **NIST Human Assisted Speaker Recognition (HASR)**
- **Conventional NIST SRE uses too many trials (comparisons) for human processing**
- **Select a subset of trials for HASR**
  - **Find most confusable trials using baseline automatic system, then**
  - **Select most confusable trials by professional, not expert, listeners**
- **HASR protocol allows listening**
- **This is difficult, but real forensic data can be more difficult**

Challenges in Speaker Recognition* for Humans and Machines [1]

- **NIST Speaker Recognition Evaluations (SRE & HASR), Netherlands Forensic Institute and Organization for Applied Scientific Research (NFI-TNO), etc. have addressed significant challenges, e.g.,**
- **Channel mismatch**



- **Distance to microphone**



- ***Forensic and investigative speaker recognition has additional challenges...***

*Campbell, et al., "Forensic Speaker Recognition," *IEEE Signal Processing Magazine, Special Issue on Digital Forensics,* v26, i2, Mar 2009, p 95-103.

Challenges in Speaker Recognition for Humans and Machines [2]

To be addressed

- **Talkers**
  - **Unfamiliar to examiner**
  - **Familiar conversants**
  - **Multiple talkers**
- **Stresses**
  - **Emotional, Loading, Physical**
- **Styles**
  - **Conversational, read, orated, loud, yell,… accommodation**
  - **Plotting, deceptive, disguise**
- **States**
  - **Mentally ill, medicated**
- **Situational mismatch**
  - **crime vs interview voice**

- **Language**
  - **Foreign to examiner**
  - **Mismatched samples (dialect, accent too)**
- **Speech samples/segments**
  - **Few, short, noisy, distorted, noncontemporaneous**
  - **Few regions of interest***
- ***Combinations of above factors!***
- **Mismatch galore!**
  - **Between samples, models, background, hyperparams**

*Schwartz, R., et al., "When to Punt on Speaker Comparison?," 162nd Meeting of the Acoustical Society of America, San Diego, CA, 3 Nov 2011.

Challenges in Speaker Recognition for Humans and Machines [3]

To be addressed

- **Presentation**
  - **Scoring**
  - **Decision?**
  - **Opinions**
  - **Court vs Investigation**
  - **Priors?**
- **Calibration**
- **Warnings to users**
- **The Court's questions**
  - **Negotiable?**
- **Assigning voice samples to speaker models dilemma**
  - **Human Automatic systems**
- **Speech detection in noise**
- **Quantify degradations**
  - **Negative factors**
- **Daubert test/factors**
- **Machine & Human**
  - **Fusion**
- **Machine vs. Human**
- **When to punt (not accept a case)***
  - **For Machine?**
  - **For Human?**
  - **For Human & Machine?**

*Schwartz, R., et al., "When to Punt on Speaker Comparison?," 162nd Meeting of the Acoustical Society of America, San Diego, CA, 3 Nov 2011.

Forensic Speaker Recognition

Real Case Data

- **Situational mismatch**


- **Complex situation**

- **Investigative example**


- **…**

- **Running case, triple homicide**
  - **Suspect ran from scene with a victim's cell phone talking via Bluetooth to a friend to get him**
  - **Known recordings are calls in jail from suspect to his friends**

- **Stressed overlapping talkers**
  - **Dangerous situation**
  - **911 call**

- **Threat call**
  - **Prompt action?**

- **It's always something, every case!**

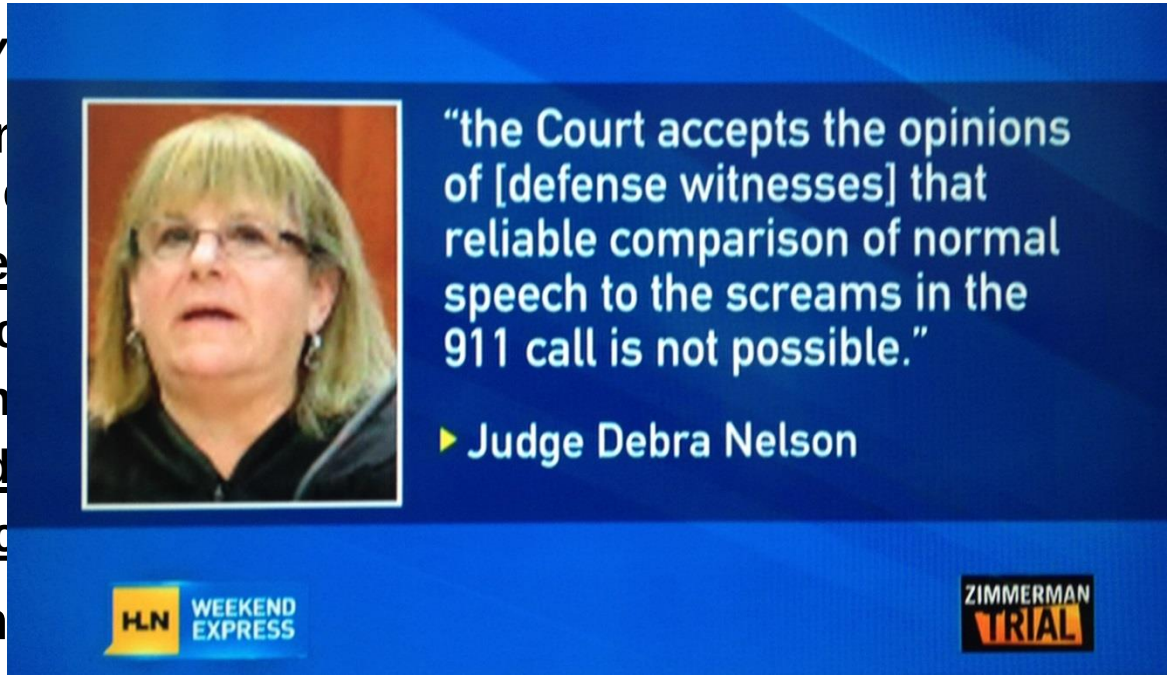
Admissibility and Daubert Standard

- **Judge considers the admissibility of scientific evidence**
  - **Judges are generally not scientists**
- **US Federal Court and ~half US State Courts under FRE 702**
  - **FRE 702 *Testimony by Expert Witnesses* to assist the trier of fact**
- **Daubert* pretrial hearing to assess whether a scientific theory/technique in question**
  1. **Has been or can be tested**
  2. **Has been subjected to peer review and publication**
  3. **Has a known or potential error rate**
  4. **Has existing standards controlling its use that are maintained**
  5. **Has been generally accepted by the scientific community**
- **Other US States use the Frye test (#5) or case-by-case rules**
- **Judge Nelson's Order!**
- **Influencing Canada and UK**

*** Daubert v. Merrell Dow Pharmaceuticals, Inc. (509 U.S. 579), 1993**

Admissibility and Daubert Standard

- **Judge considers the admissibility of scientific evidence**
  - **Judges are generally not scientists**
- **US Federal Court and ~half US State Courts under FRE 702**
  - **FRE 702 *Testimony by***
- **Daubert* pretrial hearin theory/technique in qu**
  1. **Has been or can be te**
  2. **Has been subjected to**
  3. **Has a known or poten**
  4. **Has existing standard**
  5. **Has been generally ac**
- **Other US States use th**
- **Judge Nelson's Order!**
- **Influencing Canada and UK**



Judge Debra S. Nelson, "Order excluding the opinion testimony of Mr. Owen and Dr. Reich," Florida vs. Zimmerman, Circuit Court for 18th Judicial Circuit, 22 June 2013.

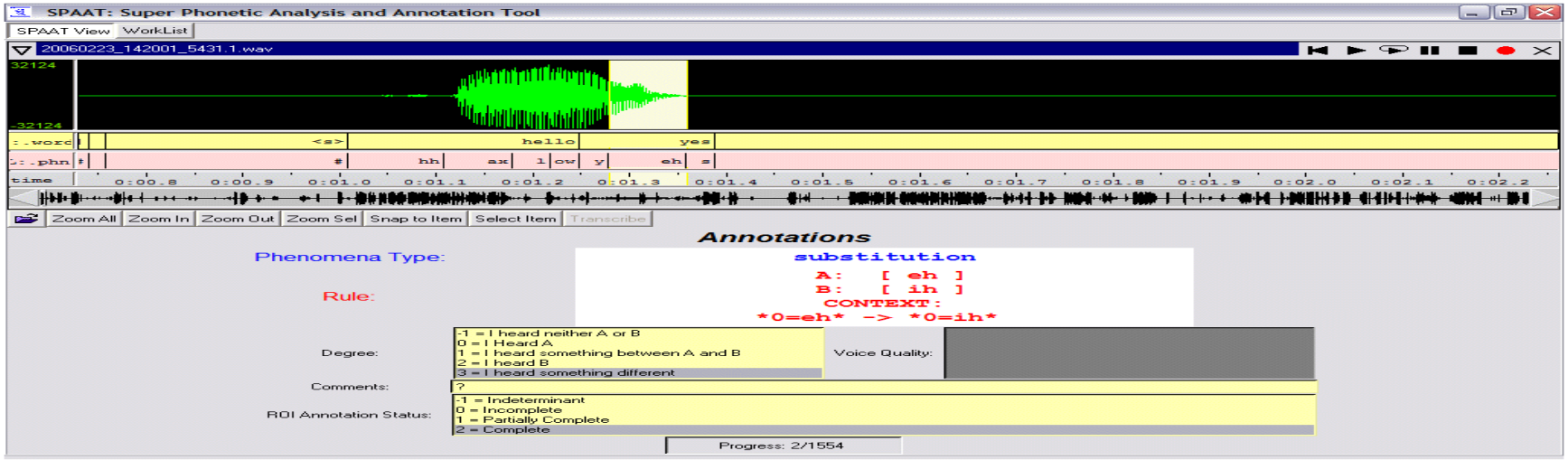
*** Daubert v. Merrell Dow Pharmaceuticals, Inc. (509 U.S. 579), 1993**

Outline

- Background
- Approaches
  - Human example
  - Machine example
  - Human and Machine
- Activities
- Request
- Future
- Conclusion

Forensic Speaker Recognition

Structured Listening



- **Transcribe speech into words**
- **Set of rules for American English transformations (other lx in process)**
- **Apply transformation rules to each Region of Interest (ROI) by listening to the speech to score it and produce likelihood information**
- **Examples: /ih/ → /eh/ substitution, e.g., pin → pen**

  /l/ → (reduced) reduction, e.g., almond → ahmond
- **Final output is a report; e.g., "there is support for the hypothesis that the samples come from the same speaker" (with an explanation)**

Forensic Speaker Recognition
Human Recognition

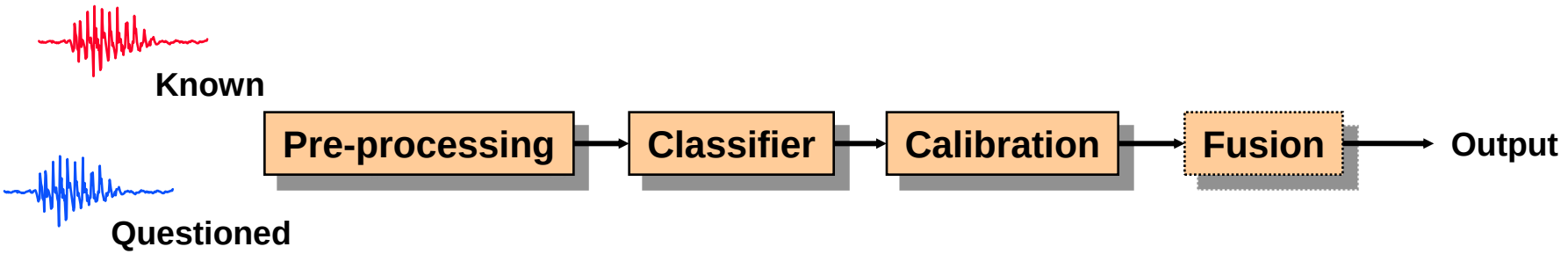
- **Challenges**
  - **Long time to make decision**
    - **Complete detailed analysis can take over a week**
  - **Skilled analyst required**
    - **Native in language of samples**
  - **Might lack required amount of data for each evaluation**
    - **Enough data for Regions of Interest in structured listening**
  - **Reliability?**
  - **Process can be subjective**
  - **Performance of various methods not well quantified**
- **Combine with automatic method?**

Outline

- Background
- Approaches
  - Human example
  - Machine example
  - Human and Machine
- Activities
- Request
- Future
- Conclusion

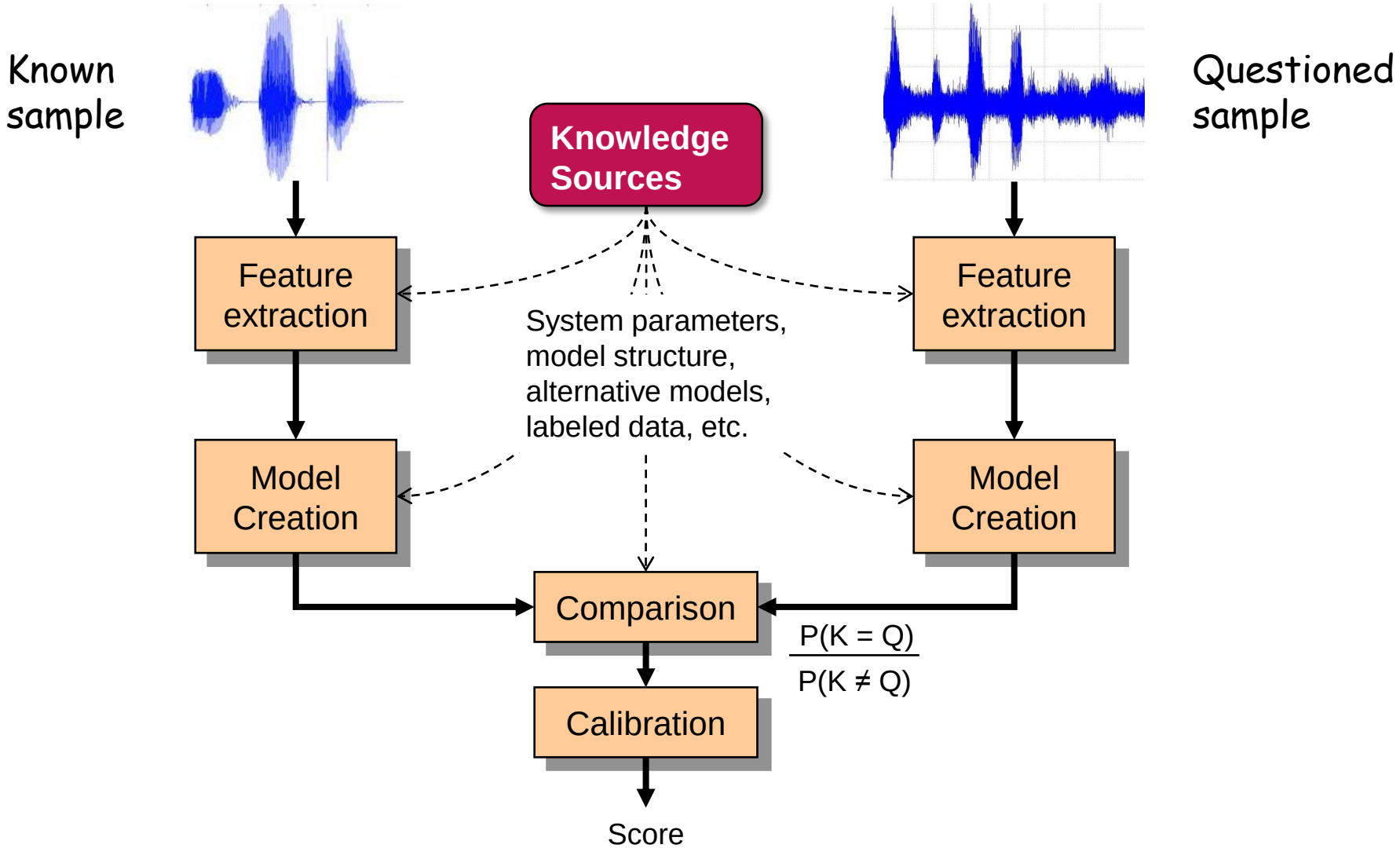
Automatic Speaker Recognition

System Architecture

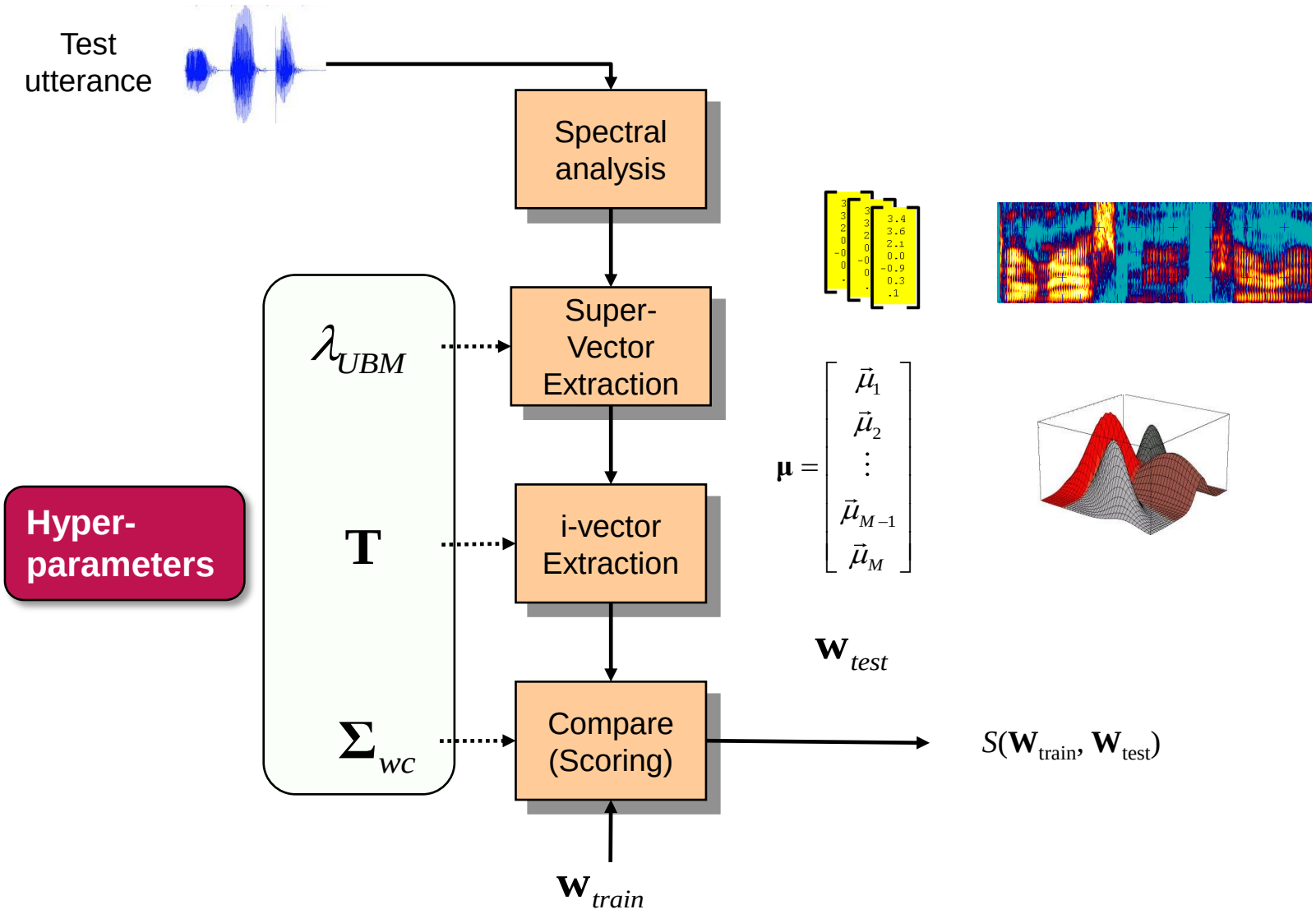


- **Pre-processing**
  - Input: Raw speech signal
  - Output: Salient features about the speaker
- **Classifier**
  - Input: Known and Questioned features
  - Output: Likelihood score
- **Calibration**
  - Input: Likelihood score
  - Output:
    - Likelihood score
    - Match probability
    - Decision
- **Fusion (optional)**
  - Multiple classifier inputs
  - Combined output

Anatomy of a Speaker Comparison System

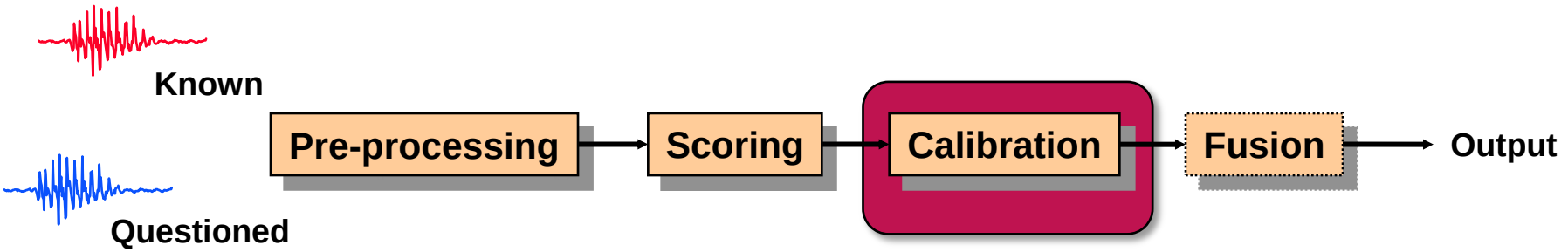


Ex. i-vector Speaker Recognition System

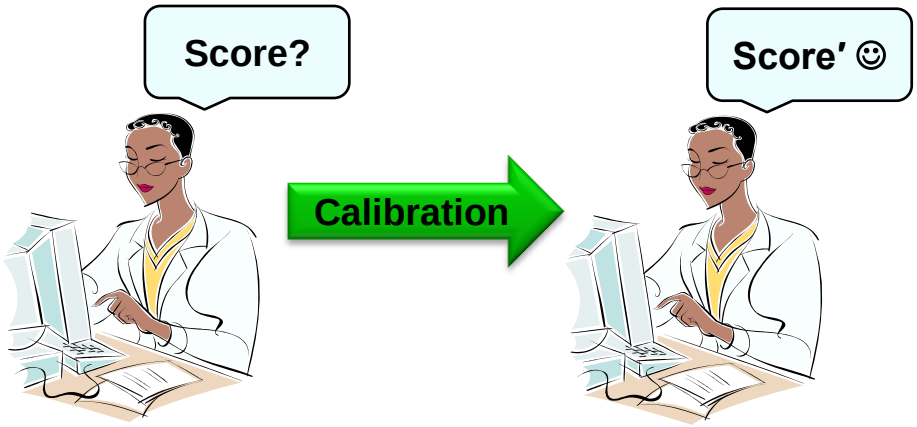


Automatic Speaker Recognition

Calibration



- **Produce consistent results interpretable by humans**
  - **Across conditions**
- **Mitigate mismatches between known, questioned, and multiple training data sets**
- **Calibrator is trained for this purpose**



Outline

- Background
- Approaches
  - Human example
  - Machine example
  - Human and Machine
- Activities
- Request
- Future
- Conclusion

Forensic Speaker Recognition
Combined Human and Automatic Recognition

- How to combine human and automatic speaker recognition?†
  - Separate processes followed by score combination?
    - Weigh each process dynamically?
  - Mitigate observation bias?
  - Deal with variations of subjects, humans, machines, and samples?
  - Consistency and repeatability?
  - Logically consistent results?
- Best Practices are needed to address these questions
- Evaluations of processes are needed
  - NFI-TNO Forensic Speaker Recognition Evaluation, 2003
  - NIST's Human Assisted Speaker Recognition (HASR)
    - Next NIST SRE more like forensic domain samples?
  - Others?
- Progress toward Daubert factors?

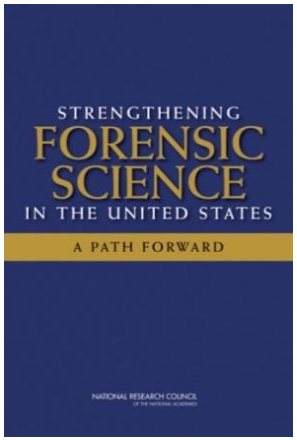
†Schwartz, R., Campbell, J. P., Shen, W., Sturim, D. E., Campbell, W. M., Richardson, F. S., Dunn, R. B., Granville, R., USSS-MITLL 2010 Human Assisted Speaker Recognition, Proc. IEEE ICASSP, Prague, Czech Republic, 26 May 2011.

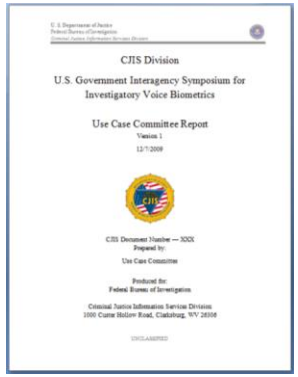
Outline

- Background
- Approaches
- Activities
  - US: SWG-Speaker
- Request
- Future
- Conclusion

US Advances in the Science and Practice of Speaker Recognition

- **NRC/NAS, *Strengthening Forensic Science in the United States: A Path Forward,* 2009**
- **Investigatory Voice Working Group**
  - ***Use Case Committee Report,* 2009**
  - ***Collection Standards Committee Report,* 2009**
- **Schwartz, et al., *When to Punt on Speaker Comparison?,* 162nd ASA, San Diego, 2011**
- **Standard Operating Procedure for Forensic Speaker Recognition, 2013**
- **ANSI/NIST-ITL Type-11 Record Standard, 2013**
- **Scientific Working Group for Forensic and Investigative Speaker Recognition (SWG-Speaker), 2013**
  - **Research, Dev, Test, Eval**
  - **Best Practices**
  - **Science in Law**
  - **Vocabulary**





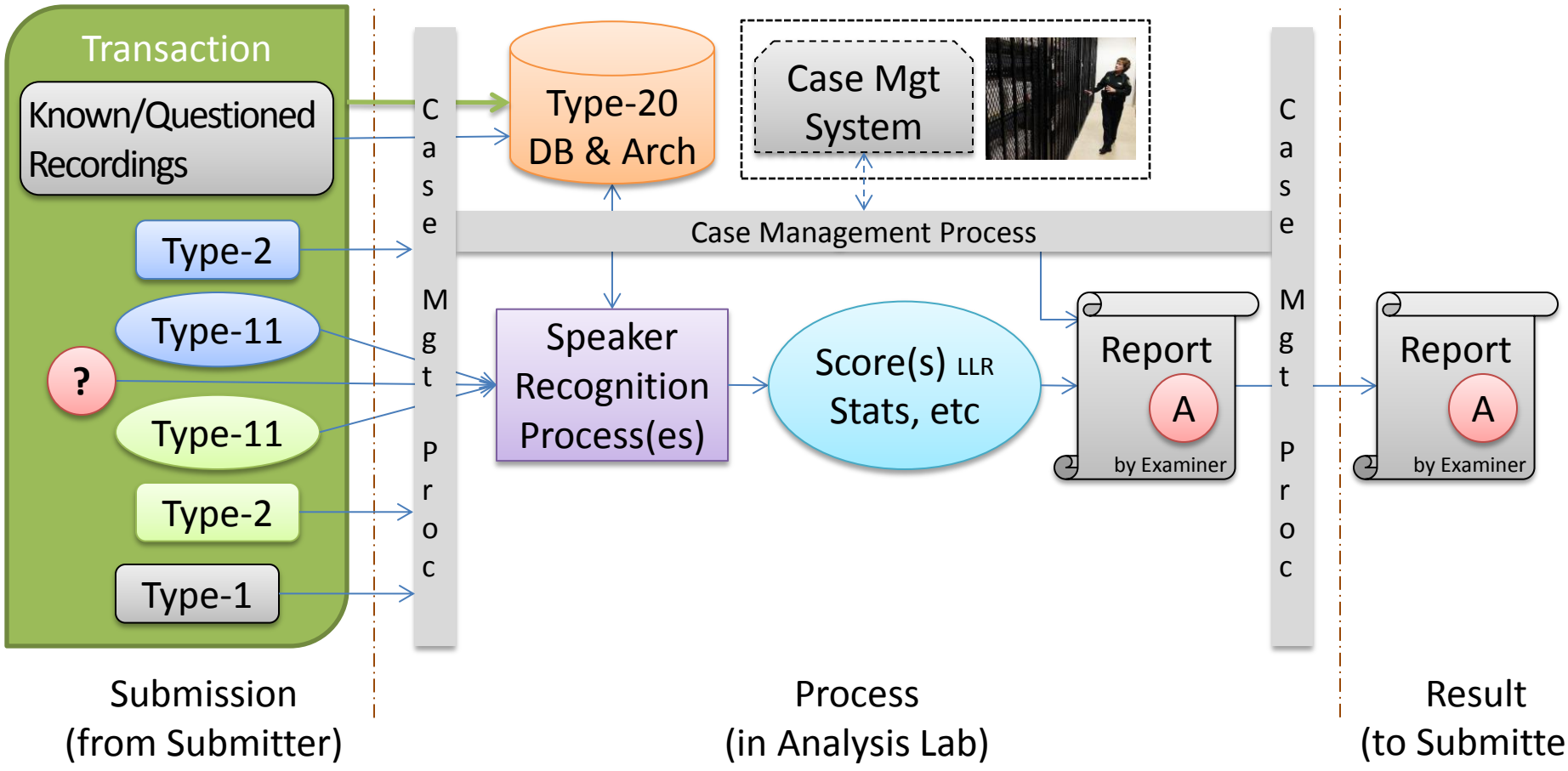
SWG-Speaker

Best Practices Ad-Hoc Committee*

- **The Best Practices Committee seeks to improve forensic science through establishing best practices recommendations**
- **Develop best practices for**
  - **Collection protocols**
    - **Speech materials**
  - **Audio recording and collecting related data used for speaker recognition**
  - **Transmission of this audio and related data**
    - **ANSI/NIST-ITL Type-11 Record and companion Records**
  - **Proper application of speaker recognition technologies**
  - **Examination and reporting guidelines**
    - **Bias concerns**
- **Training, examiner certification, and laboratory accreditation advice**

* *SWG-SPEAKER: Ad-hoc Committees 2013,* 3 March 2013

Ex. "Does the Questioned Voice Recording Share the Same Source as a Known Voice Recording?" ?



Records

Type-1: Mandatory record submitted with each transaction, "transaction header information"

Type-2: Transaction related data; e.g., subject's name and other biographic information, reason for booking, any charges, etc.

Type-11: Voice data and voice metadata for the subject in corresponding Type-2 ("voice data" can be marks for the subject in the original audio stored)

Type-20: Repository of original data if in digital format (original format, nonmanipulated, and unprocessed), which includes raw evidence (without redaction)

Type-xx: Other record types can be transferred that might not be used in the speaker recognition process; e.g., photo of subject, signed papers, etc.

Proposed Best Practices Process

Survey and study current practices
- Nominate promising practices as strawmen
- Develop strawmen practices

Identify gaps in practices
- Develop new practices
- Develop strawmen

Validation
- Evaluation
- Relevant data
- Reproducibility
- Accuracy, Calibration
- Speed, throughput
- Other metrics

Elevate validated, developed practice to a Best Practice
- Consensus process
- Do no harm

Proposed standard?

- **Multiple Best Practices**
- **Repeat**
  - **Ongoing improvement process**
- **Validation methods involving humans?**
  - **Black Box Examiner Study? NIST Forensic SRE? Corpora? Funds?**

Outline

- Background
- Approaches
- Activities
- Request
- Future
- Conclusion

Requests to Research Community

- **Develop candidate Best Practices**
  - **Submit to SWG-Speaker/OSAC for consideration**
- **Pursue Daubert factors**
- **Improve robustness**
  - **Core classifiers**
  - **Calibration**
  - **Limited in-domain training data**
  - **Combined processes**
- **Work with analysts/examiners to improve usability and performance**
  - **Human in the loop**
- **Rise to the challenges of forensic and investigative data**
  - **Handling speaker variability from stress and emotional state**
- **Participate in forensic/investigative-style evaluations**

Engagement Opportunities

Standards, practices, evaluations, and data for forensic and investigative speaker recognition

- NEW! **OSAC Organization of Scientific Area Committees (NIST)**
  - Speaker Recognition Subcommittee (IT/Multimedia Sci Area Committee)
  - SWG-Speaker transition to OSAC
- **ANSI/NIST-ITL 1-2011 Update:2013, Data Format for the Interchange of…**
  - Record Type-11: Forensic and investigatory voice record
- **IAFPA International Association for Forensic Phonetics and Acoustics**
  - Code of Practice, workshops, shared IJSLL journal with IAFL International Association of Forensic Linguists
- **AES Audio Forensics: Techniques, Technologies, and Practice**
- **FAS ASA Forensic Acoustics Subcommittee (Speech Comm TC)**
- **ENFSI European Network of Forensic Science Institutes**
  - FSAAWG Expert Working Group for Forensic Speech and Audio Analysis
  - Monopoly 2011 Methodological guidelines for semi-automatic and automatic speaker recognition for case assessment and interpretation
- **AGSE Arbeitsgemeinschaft Sprechererkennung (WG of German speaking forensic speech and audio specialists from European Labs)**
- **EAB Workshop on Biometrics and Forensics**
- **FSA UK Home Office Forensic Speech and Audio Group**
  - Bring forensic speech and audio under the regulation of ISO 17025

Engagement Opportunities (cont.)

Standards, practices, evaluations, and data for forensic and investigative speaker recognition

- ASSTA Forensic Speech Science Committee Australasian Speech Science and Technology Assoc
- EAFS European Academy of Forensics Science
- IAFS International Association of Forensics Sciences
- ICFIS International Conference on Forensic Inference and Statistics
- AAFS American Academy of Forensic Sciences
- NAS National Academy of Sciences
- Evaluations and data
  - NIST SRE HASR (toward forensic)
  - NFI FRITS (after 2003 NFI-TNO FSRE)
  - Cambridge DyViS corpus
- ISCA
  - Interspeech, ICSLP, Eurospeech special sessions on forensic speech
  - Odyssey The Speaker and Language Recognition Workshop
  - SpLC Speaker and Language Characterization SIG
  - AFCP Association Francophone de la Communication Parlée SIG and JEP
- iberSPEECH
- Programs
  - SIIP Speaker Identification Integrated Project (EU FP7)
  - BBfor2 Bayesian Biometrics for Forensics project
  - ICT COST Action IC1106 Integrating Biometrics and Forensics for the Digital Age

Conclusion

- **Speaker recognition is successfully used today in a variety of applications, but must be applied responsibly with caution***
- **Need to address factors in forensic domain that degrade recognition performance**
  - **Increase robustness and effectiveness**
  - **Improve calibration and efficiency**
- **Real case data can be extremely challenging**
  - **Opportunities for research!**
- **Many challenges to be addressed to satisfy Daubert test**
  - **Additional international challenges and opportunities**
- **Please contact me to share ideas!**

- ***Sauna is a very Finnish way for decision making* – Pasi**

* Campbell, J., Shen, W., Schwartz, R., Bonastre, J.F., Matrouf, D., Forensic Speaker Recognition: A Need for Caution, IEEE Signal Processing Magazine, Vol. 26, Issue 2, p. 95-103, March 2009.

Questions?