AES Standards Webinar

AES-X253 “Speech Collection Guidelines for Speaker Recognition: Interviewing at a Temporary Location”

Marisa T. Déry, Chair, SC-03-12 Working Group
Marisa T. Dry is an audio/video forensic analyst in the Boston area that specializes in audio enhancement, audio authentication and voice identification. She is originally from Ottawa, Ontario, Canada and has been in the audio industry for 36 years as a Mastering Engineer, Audio Restoration Engineer and Forensic Consultant. She is a member of CERT in Massachusetts, as well as the MRC where she assists law enforcement as a volunteer. Her clients include Public Defenders’ Offices, Prosecution, Law Enforcement, Legal firms and the General Public. She is a member of several organizations, including AES, IAI, IEEE, SAFE and CARAS.
History

This Paper was originally started by OSAC

THE ORGANIZATION OF SCIENTIFIC AREA COMMITTEES FOR FORENSIC SCIENCE
History

The Original title of this paper was

- “Speech Collection Guideline for Speaker Recognition: Audio Collection at a Temporary Location”
ABOUT OSAC

OSAC strengthens the nation’s use of forensic science by facilitating the development and promoting the use of high-quality, technically sound standards. These standards define minimum requirements, best practices, standard protocols and other guidance to help ensure that the results of forensic analysis are reliable and reproducible.

OSAC was created in 2014 to address a lack of discipline-specific forensic science standards. OSAC fills this gap by drafting proposed standards and sending them to standards developing organizations (SDOs), which further develop and publish them.
NEW ABSTRACT

This document specifies recommended practices for recording audio intended for use in forensic speaker recognition analyses, focusing on doing so at a temporary, non-laboratory location by possibly a non-professional in the forensic sciences. It includes recommendations for the physical preparation of the location, selection of appropriate recording hardware and audio formats, and possible methods for interviewers to elicit the desired type and amount of speech from subjects. It does not cover the methods used to analyze the resulting recordings and does not deal with details related to the handling, transmission, storage, or preservation of the collected data but will include a checklist to aid in the process.
Key Points of the paper

- Non-Laboratory Environment
- Recordist are not Forensic Examiners
- Includes the minimum needed for a forensic scientist to do their job
Speech Comparison

- Compare KNOWN voice with UNKNOWN voice
- Need to have a long enough Recording
- Needs to be a good quality recording
- Needs to have a good amount of vowels
- Important to have normal speaking voice
Speech Comparison

• Critical Listening
• Pitch (Cepstrum)
• Vocal Formants
• Auditory Analysis
Critical Listening

- Listening to the conversation and the voice
- Looking at signal analysis of files (noise, reverb)
- Determine if authenticity test is recommended
- Determine if length, clarity and content is enough for a test
Pitch
Vocal Formants
Speech Comparison

Person 1

Person 2
Auditory Comparison

- Listening to the conversation and the voice
- Looking for patterns (umms, uh, cough, stutter)
- Emotional state of the interviewee
- Prosody of voice
- Accents, Regional Accents, Unique way of saying something (Again, gonna, nah)
Scope and Collection

- Non laboratory, possibly military or covert operation
- Be an indoor space as free as possible from background noises
- Away from large, flat, hard sound-reflective surfaces
- Allow the subject to be as comfortable as possible
Example of poor speech sample
Collection Equipment

- Uncompressed PCM data (WAV)
- 16 bit
- Minimum rate of 16 kHz
- Mono or Stereo
- Method available to back-up the collected data (optical media, thumb drives).
Speech Collection

- Test the equipment by doing a sample recording
- Check Levels:
  - Not too low (resulting in a noisy recording due to quantization effects)
  - Not too high (which causes clipping and thereby introduces nonlinear distortion into the audio stream).
Speech Collection

- The subject’s microphone should ideally be a headset mic
  - it fixes the location of the mic with respect to the mouth.
  - it reduces interference from the interviewer’s speech and any background sounds.
  - the speaker will quickly forget about its presence.

- Otherwise, a microphone on a stable stand or tripod
For recordings made using laptop or other computers, it is preferred to use an external USB condenser microphone with an on-board analog-to-digital (A/D) converter. This is because the internal microphone or external microphones plugged into a “mic” port can pick up noise from internal circuitry.
Speech Collection

- Ask fixed questions to provide identity information (such as name, address, and date of birth)
- If the interviewer knows what type of speech and the language the subject recordings will be compared against, the collection should be designed to elicit those.
Speech Collection

- The interviewer should avoid interjections while the subject is speaking.
- Avoid making noise with paper/newspaper.
- The subject’s portion of the recording should contain a minimum of two minutes of speech and preferably up to five minutes.
Speech Collection

- Engage the Speaker by discussing a subject of interest
- Ask open-ended questions or prompts
- Ask the subject to describe or interpret an image
Annex A (Informative) - Possible Known-Text Phrases for the Subject

Multiple examples to stimulate discussion

- “My name is [Subject name]”
- “I was born on [Subject date of birth]”
- “I was born in [Subject place of birth]” – this can include town, city, region, country, etc.
- “I currently live in [Current residence location] at [address]”
- “My current job/occupation is [Occupation]”
- “I am [married/single] …”
- “My [height/weight/eye color] is …”
Checklist

width

Height

Length
Checklist:

- Recording Device:
- Microphone used (if external):
- Microphone distance to the Speaker:
- Physical State/ Health (e.g., Dehydration, Sleep Deprived):
- Speaker Traits (e.g., State of Teeth, Stutter, Chemically Induced State):
- Emotional State:
- General images used to encourage spontaneous speech (if used):
Paper that might set up guidelines or Best Practices aimed at helping forensic scientists
Acknowledgements

- SC-03-Subcommittee on Preservation and Restoration of Audio Recording
- SC-03-12 Working Group on Forensic Audio and all its members
- Bruce Olson: AES Standards Committee Chair
- Richard Cabot: AES Standards Manager
- John Hansen: Original Chair for this paper
- Cesar Lamschtein: Translation software (and part of our group)
Thank You!

Questions?

Marisa T. Déry,
mtd@tamarforensics.com