STANDARDS AND INFORMATION DOCUMENTS

STANDARDS

AES standard for audio forensics - Speech collection guidelines for speaker recognition: Interviewing at a temporary location

AUDIO ENGINEERING SOCIETY, INC.
132 East 43rd St., Suite 405, New York, NY 10017. US.

The AES Standards Committee is the organization responsible for the standards program of the Audio Engineering Society. It publishes technical standards, information documents and technical reports. Working groups and task groups with a fully international membership are engaged in writing standards covering fields that include topics of specific relevance to professional audio. Membership of any AES standards working group is open to all individuals who are materially and directly affected by the documents that may be issued under the scope of that working group.

Complete information, including working group scopes and project status is available at http://www.aes.org/standards. Enquiries may be addressed to standards@aes.org

The AES Standards Committee is supported in part by those listed below who, as Standards Sustainers, make significant financial contribution to its operation.

This list is current as of 2022/2/18

AES standard for audio forensics - Speech Collection Guidelines for Speaker Recognition: Interviewing at a Temporary Location

Published by
Audio Engineering Society, Inc.
Copyright ©2022 by the Audio Engineering Society

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 does include a checklist to aid in the process.

2022-02-20 printing

Contents

Foreword

This foreword is not part of this document, AES-x253-20200708 AES standard for audio forensics -Speech Collection Guideline for Speaker Recognition: Audio Collection at a Temporary Location.

The document was originally produced by the Organization of Scientific Area Committees for Forensic Science, Digital/Multimedia Scientific Area Committee, Speaker Recognition Subcommittee (OSAC-SPEAKER), as a continuation from SWG-SPEAKER (Scientific Working Group – Speaker Recognition).

Original Authors of the document drafted and reviewed/approved first from SWG-SPEAKER and subsequently from USA OSAC-SPEAKER are: Stephen Gibbs (principal author), Darren Haddad (key contributing author)
Additional OSAC-SPEAKER Contributors: Walt Andrews, Jos Bouten, Joseph Campbell, Christopher Cieri, Kathleen Egan, Kevin Farrell, Patrick Gibbs, John Hansen, Aaron Lawson, David Marks, Johanna Morley, Geoffrey Morrison, Hirotaka Nakasone, Douglas Reynolds, Reva Schwartz, James Wayman. Other unnamed members of the OSAC/QIC committee.

This document was developed under project AES-X253, in the SC-03-12 task group, first led by John Hansen then by Marisa T. Déry.

Contributors in the task group include M.T. Déry, J. Hansen, R. Cabot, D. Boss, G. Reid, C. Lamschtein, K. McElveen, J. Owen, C. Grigoras, R. Maher, J. Woodgate, D. Bergfeld, K. Marr, S. Winner, D. Walsh, E.B. Brixen, S. Beck.

At the time of publication the AES SC-03-12 consisted of: E. Achren, A. Alexander, R. Alexanderson, S. Allen, W. Andrews, S. Bamford, A. Bartle, S. Beck, D. Bell, D. Bergfeld, T. Birkeland, M. Bittle, D. Boss, E. Brixen, N. Broner, J. Broyles, B. Brustad, W. Bulla, R. Cabot, S. Cain, A. Castner, J. Charles, P. Connolly, M. de Cunto, A. de Oliveira, M. Déry, W. Deutsch, T. Edwards, E. Enzinger, T. Erbe, J. Freytag, R. Gaskell, S. Gibbs, G. Gibson, R. Gorsline, J. Griffin, C. Grigoras, B. Gunel, H. Hacihabiboglu, D. Hallimore, J. Hansen, P. Harrison, S. Herla, C. Hordyk, G. Houle, B. Hunt, S. Jacobson, P. Kaeseler, F. Kelly, K. Kim, M. Kleiner, B. Koenig, W. Krafft, C. Lamschtein, T. Lewis, R. Maher, H. Malik, D. Marks, K. Marr, R. Max, K. McElveen, P. Merrill, G. Morrison, J. Narus, W. Newman, D. Notowitz, D. Nyberg, N. O'Brien, P. Olsson, J. Owen, T. Owen, P. Pegas, A. Pellicano, C. Peltier, M. Piper, F. Poza, E. Primeau, J. Ramirez, J. Reames, A. Reich, G. Reid, W. Reynolds, R. Rose, M. Roy, M. Sanfilipo, R. Schlesinger, B. Skye, C. Smith, J. Smith, L. Smrkovski, A. Stanley, J. Stathakis, R. Streicher, G. Stutchman, L. Tjellesen, R. Todd, J. Trš, J. Trautner, D. Tunstall, A. Tuzman, C. Van Winkle, Z. Varga, J. Vermeulen, J. Voetmann, G. Wallach, S. Walsh, J. Wheeler, L. Williams, S. Winner, J. Woodgate, D. Wujcik, J. Zjalic.

Marisa T. Déry
Chair, working group SC-03-12
2022-01-12

Note on normative language

In AES standards documents, sentences containing the word "shall" are requirements for compliance with the document. Sentences containing the verb "should" are strong suggestions (recommendations). Sentences giving permission use the verb "may". Sentences expressing a possibility use the verb "can".

AES standard for audio forensics - Speech Collection Guidelines for Speaker Recognition: Interviewing at a Temporary Location

0 Introduction

0.1 General

In this document, the person overseeing the collection session will be designated the "interviewer" and the individual being recorded the "subject". The intended audience for these guidelines is those interviewers called on to perform speech collection using portable equipment at a temporary location not originally designed or intended for audio recording.

The goal of speech collection is to collect an audio recording containing a combination of subject identifying information (such as their name, date of birth, etc.) and speech which can be used for future comparison against the speech of unknown speakers using unspecified speaker recognition methods. Although the specific method of speaker recognition is left undefined, automated/semi-automated computer-based methods were the primary driver for some of the specific parameters found in these guidelines.

These guidelines should be viewed as providing minimum requirements for usable speech collection in an operational or field environment and are not intended as data collection guidelines for research applications, speech intended for transcription, or other applications.

It is important that before implementation of these guidelines the user coordinate their activities with any elements of their parent organization which will be storing and using the collected data. Issues related to audio channel (microphone, telephone, radio, etc.), desired languages and dialects, data formats, etc. should be worked out beforehand to ensure that the data is maximally useful. Similarly, the proper storage and protection of personal identifying information (if collected) should also be coordinated as required within your organization prior to data collection.

0.2 Documentation conventions

Numerical values are decimal unless otherwise stated.

Where new terminology is first introduced in body text, the term will be set in an *italic typeface*.

1 Scope

These guidelines are intended to be one of a series and covers only one scenario - the collection of speech samples for speaker recognition at a temporary, non-laboratory location. One example of this would be the collection of speech samples from subjects during some type of field activity. The field activity could be associated with a range of purposes, such as law enforcement, intelligence, military or sociological. These guidelines presume portable resources and somewhat limited time to perform the collection. It also assumes that the interviewer is fluent in the subject's language, or that an interpreter is present who is both fluent in the interviewer's and subject's languages and cognizant of the goals of the interview.

These guidelines do not deal with details related to the handling, transmission, storage, or preservation of collected data. Specifically:

- It does not deal with any issues related to collection, storage, or protection of personal information.

- It does not recommend how to protect speaker recognition analysts from seeing or hearing subject identifying information which could result in biased analysis results.

It is the responsibility of the guideline's user to determine what their organization's rules and policies are on these matters and to tailor their implementation of these guidelines to comply with those rules and policies. It is also the guideline user's responsibility to learn how to operate the selected equipment, especially the placement orientation and distance of the microphone and the interpretation of any meters on the recording device.

These guidelines also do not deal with possibly important concerns such as personnel and equipment security, power sources for equipment, etc. The alleviation of such highly situation specific concerns is the responsibility of the guideline's user or their agency.

2 Normative references

There are no normative references.

3 Terms, definitions, and abbreviations

For the purposes of this document, the following terms, definitions, and abbreviations apply.

3.1 A/D

Analog-to-digital

3.2 AGC

Automatic Gain Control. A closed-loop regulating circuit which provides a controlled signal amplitude at its output, despite variation of the amplitude in the input signal.

3.3 Codec

Algorithm designed to encode or decode a stream of digital audio data.

3.4 Hz

Hertz is the international unit of measurement of frequency. One hertz corresponds to one cycle per second.