Voice Biometrics

Progress towards a
Voice Supplement to ANSI/NIST-ITL 1-2011 Standard

Presented by Mark Przybocki
NIST, ITL, Information Access Division
http://www.nist.gov/itl/iad/mig

Contributions from many organizations, including public comment

Much of this work was made possible by support from the FBI Biometric Center of Excellence

ANSI/NIST-ITL Type-11 Record

Motivation

- To address a gap in the ANSI/NIST-ITL standard in order to facilitate the interchange of **voice data** for investigatory and forensic purposes

Voice as a Biometric

- Features of voice have been used to identify individuals for over 50 years
- Voice/Speaker Recognition performance
  - NIST Speaker Recognition evaluations (*not biometric*) provides R&D directions
  - Plenty of challenges identified and many addressed
- Challenges to recognition performance remain
  - Suboptimal recording conditions
  - Intrinsic speaker variations

Voice Supplement to ANSI/NIST-ITL

- ➢ Accelerate the establishment of best practices and standards
- ➢ Begin an iterative process of improvements expanding use-cases

[Review] @BCC 2011 We Covered

- Established a Voice committee
- Identified groups working on similar goals of a Voice standard
- Discussed working concepts for a Voice Type 11 Record
- Identified unique challenges presented by Voice for a biometric
- Discussed a planned way forward to develop the Type 11 Record

Voice Committee

- New Chair: Vincent Stanford, NIST
- Mailing List: voice_std@nist.gov
- First Meeting: March 9th, 2012
  - A thorough review of a initial draft for the Type 11 Record
  - Comments incorporated, gaps filled in
- Second Meeting: August 29th, 2012
  - A comprehensive review of a full draft for the Type 11 Record
  - Document and Comment Form posted on NIST web-space http://www.nist.gov/itl/iad/mig/ivb.cfm
- *Third Meeting: November 20th, 2012*
  - *An open event, all interested parties are welcome to attend*

Draft Voice Supplement to the ANSI/NIST-ITL 1-2011 Standard

- (Draft) 22 August, 2012 Version A-1c
  - Identifies contributors
  - Defines specialized terms
  - Covers the intended scope of the Type-11 Record
  - Reviews Metadata requirements (Administrative, Speaker, Content, and Audio Technology)
  - Provides an overview of the General Organization of the Type-11 Record
  - **Working Definition for the Type-11 Record (~40 pages)**

Potential Scenarios using a Type-11 Record

1. Voice model creation and storage for a known (or unknown) speaker
2. Match detection of speakers in two audio recordings
3. Match detection of a voice in an audio recording to a list of speakers
4. Converting an analog audio recording into digitized voice data file(s)
5. Duplicating or transcoding an audio recording
6. Finding and isolating voice (or just speech) signals in an audio recording
7. Determination of the number of distinct speakers in an audio recording
8. Indexing an audio recording into segments attributable to distinct speakers
9. Creation of a diary, attributing speech segments to a speaker of interest
10. Creation of audio transcriptions
11. Redaction of an audio recording to remove sensitive speech segments
12. Removing non-speech (or speech from “other” speakers) from a recording
13. Enhancing speech prior to a match detection process
14. Audio authentication
15. Transfer voice recordings for permanent archival

Potential Use Cases for Investigatory Voice

1. Identification of voices broadcasted for propaganda purposes
2. Determination of the number of controllers involved in multiple attacks
3. Proof of life of kidnapped victims
4. Identification of individuals involved in illegal activities
   - Weapon sales
   - Financial fraud
   - Serial murder suspects
   - …

- To clear individuals accused of illegal activities

Challenges Facing a Voice Standard

- Content of audio recordings
  - Generally contains speech and non-speech signals, often from multiple individuals
  - Speech (*what was spoken*) has identifying information and therefore carries additional meaning to be preserved (*other protections privacy/security?*)
- Collection of audio recordings
  - Speech is collected in the time dimension and will not have a single "time of collection"
  - In mobile applications a recording may not be linkable to a single geographic location
- Unlike other biometric modalities, voice recordings may reflect the social and behavioral conditions present in the collection environment, including the relationship between the data subject and any other speakers

Concluding Thoughts

- Work on a Voice Supplement began in early 2011, at a time when the Speaker Recognition Research Community continued to make major strides in robustness and accuracy as demonstrated through NIST Speaker Recognition Evaluations
- Draft Voice Supplement to the ANSI/NIST-ITL 1-2011 Standard
  http://www.nist.gov/itl/iad/mig/ivb.cfm
  Comments should be sent to voice_poc@nist.gov
- Anticipate adoption of Voice Supplement in 2013
- (NEXT) contribute to the establishment of Scientific Working Group dedicated to the technical issues surrounding Speaker Recognition

Questions or More Information

- Vincent Stanford, *Information Access Division*
  vincent.stanford@nist.gov

- Mark Przybocki, *Multimodal Information Group*
  mark.przybocki@nist.gov

http://www.nist.gov/itl/iad/mig

National Institute of Standards and Technology
U.S. Department of Commerce