

Voice Standard Development Plans

*Defining a Voice Record to the ANSI/NIST-ITL Standard
Type 11 Record*

Mark Przybocki
Multimodal Information Group

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

Outline

- Introduce the Multimodal Information Group
- Purpose of Speaking Slot
- Groups Working Towards a Similar Goal
- Current Status
- Way Forward
- Schedule

Multimodal Information Group

Functional Statement:

Researches and develops measurement and evaluation methods to advance and promote the use of technologies that provide more effective access to multimedia and multi-lingual information, and improve human-computer interface modalities.

Technologies include those that recognize and/or transform information in speech, text, video, and other modalities through:

- speech analysis,
- video analysis,
- machine translation, and
- speaker and language recognition capabilities

MIG: Speaker Recognition Activities

- **Speaker Recognition Technology evaluation series**
 - Biennial test of automated text-independent SR capabilities
 - SRE12 being planned
 - SRE11 Technical Symposium in December
- **HASR** (*Human Assisted Speaker Recognition*)
 - Combining human experts with automated SR systems (SRE10)
 - Special session at ICASSP 2011
- **Investigatory Voice Biometrics committee**
 - Standards investigation led us to ANSI/NIST-ITL involvement
- **Biometric Exploitation Science and Technology program**
 - R&D evaluation to support the IARPA BEST Program SR track

Purpose of Talk

- To describe the way forward for adding a Voice Record to the ANSI/NIST-ITL Standard
 - What does this mean
 - What doesn't this mean
- *(Not to begin public debate of specific fields)*
- Type-11 Record of the ANSI/NIST-ITL Standard is reserved for the **Biometric Voice Record**
 - This activity has been discussed for 3-4 years
 - Type 11 working group to be chaired by Dr. Alvin Martin (NIST)
(The establishment of this group is to be voted on during this event)

Groups Working Towards a Similar Goal

(a voice standard)

- ANSI/NIST-ITL Standard
 - Biometric data exchange supporting law enforcement scenarios
- Investigatory Voice Biometrics committee
 - Steering Committee of US Government individuals
 - Investigate standards options (national/international)
 - Investigate issues, make recommendations
- Other U.S. Government agencies
 - Internal Best Practices
- ISO 19794 - 13
 - Wider scope standard including non-Law Enforcement applications

Plenty of Lessons Learned

- Type 11 working group will borrow/build on
- Reduce duplication
- We also have overlapping representation

Type-11 Voice Record

Working Definition

- Used to exchange voice biometric data
 - Along with textual information fields relevant to the sample
- Targets automated voice identification and verification systems
 - Not, evidentiary storage
 - Not, transmission of statements
 - Not, for use in speech recognition systems alone
- A record is to be about a single person
 - A sample (defined by a Type 20 record) containing Voice of multiple individuals may be used by multiple Type 11 records, each corresponding to a separate ANSI/NIST-ITL transaction
- To define a *Transmission Standard*

Current Status

- Discussion is occurring between interested groups:
 - Scope of a voice standard
 - Metadata appropriateness

} *Mandatory and optional fields*
- Some areas of agreement about the Type 11 development:
 - Defining a transmission standard is good
 - Implementation neutral
 - Content based on representation dictionary
 - Not forcing a signal encoding
 - Biometric use case
 - Biometric verification and identification
 - Applications geared to law enforcement scenarios
 - Overlaps with (*but not strictly*) Intelligence and National Security scenarios
 - This will be a complex undertaking

Current Status

- Type 20 Record
 - Source representations employed to create representations used in another record
 - To accommodate audio, video, and DNA
 - In the stage of approval in the ANSI/NIST-ITL Standard
- Effort to account for voice requirements seen in fields:
 - Acquisition source types
 - Source representation format
 - Time index
- Room for future amendments to the Type 20 Record *if* required to support voice

Way Forward

- NIST staff familiar with Speaker Recognition technology are actively participating in the ANSI/NIST-ITL Type-11 record development
- Continue to rely on experts inside/outside of the U.S. Government
 - National and International committees
- Bridge standards requirements from U.S. classified environments into the ANSI/NIST-ITL Standard

Way Forward

- Once the proposed working group for Voice is approved, investigation will begin for Type 11 and Type 20 field requirements such as: (*examples only*)
 1. Recording environment
 - a) Device
 - b) Noise sources
 - c) Recording Date/Time
 - d) Original encoding
 2. Age of subject/recording
 3. Alternative needs of various system algorithms (SR approaches)
 4. Minimum sample size
 5. Language/dialect
 6. Transcript
 7. Segmentation/Diarization
 8. Vocal effort
 9. Quality measurements
 10. ...

Tentative Schedule

- 2011 ANSI/NIST-ITL Standard document
 - To have a placeholder for Type-11 Record
- A future amendment to the 2011 Standard will include the Type 11 record *when it is ready*
 - A proposal, comment, acceptance process will follow the standard voting procedures

Tentative Schedule

- Relevant Meetings (**Update**)
 - February 28th
 - Investigatory Voice Biometrics committee meeting
 - March 1-3
 - ANSI/NIST-ITL Standards workshop
 - March 16th
 - DOD Biometrics Standards Working Group (BIMA)
 - ISO 19794 Standards Group Meeting ...
 - June
 - NIST Voice Symposium II