

# Progress Towards a Voice Standard

*Defining the Type 11 Record of the ANSI/NIST-ITL Standard*

Mark Przybocki

NIST, ITL, Information Access Division

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

*Contributions from others at*

*NIST, FBI, MIT-LL and Expert Consultants*

# Setting Context

## ANSI/NIST-ITL Standard

- Used by Federal, State, local and tribal agencies to exchange fingerprints, palm print, face and iris data
- No current provision for voice data
- Type 11 Record – reserved for Voice

## Challenges of Forensic Speaker Recognition

- Broader set of transactions than access control
- Limited control over field collection conditions
- Standard doesn't have a current model suitable for Voice adaption

# The Nuisances of Speech

- Speech data is a time domain signal containing characteristics that change within the recording
  - Background noise
  - Speaker behavior (pitch, vocal effort, language spoken ...)
  - Speech content
  - Geographical location
  - Number of speakers
- Collection variables
  - Multiple devices
  - Transmission channel
  - Room acoustics
- Code switching, band limiting, encoding, segmentation ...

# ANSI/NIST-ITL Standard 2011 Workshop

## Voice Record Recap

- **March 1-3, 2011**
- **Presented voice standard development plans**
  - [http://biometrics.nist.gov/cs\\_links/standard/ansi\\_2011/WORKSHOP/przybocki-Voice\\_Plans.pdf](http://biometrics.nist.gov/cs_links/standard/ansi_2011/WORKSHOP/przybocki-Voice_Plans.pdf)

**Purpose:** Described way forward in bringing experts together to discuss the nuances of speech and the attributes of speech samples that would require representation in a voice standard.

- **Established a committee dedicated to voice record development**
  - Dr. Alvin Martin (NIST) agreed to chair

**Representation:** Open to all.  
Contact NIST to be added to the mailing list.

# Type-11 Voice Record

## Working Concepts

- To define a *Transmission Standard*
- Used to exchange voice biometric data
  - Along with textual information fields relevant to the sample
- Targets automated voice identification and verification systems
  - Not designed for evidentiary storage – but can transmit such data
  - Not transmission of statements
  - Not for use in speech recognition systems alone
- A record is to be about a single person
  - A recording (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

# Groups Working Towards a Similar Goal

*(a voice standard)*

## Voice Standard Goal

- ANSI/NIST-ITL Voice Committee
  - Biometric data exchange supporting law enforcement scenarios
- Investigatory Voice Biometrics Committee
  - Under sponsorship by the FBI/BCOE
  - Steering Committee of US Government representatives
  - Originally charged to investigate standards options (national/international)
  - Investigate issues, make recommendations
- Other U.S. Government agencies
  - Internal Best Practices
- ISO 19794-13
  - Different scope - including non-Law Enforcement applications

## Other Efforts

- Non-Forensic Standards efforts
- Single Speaker Access Control Emphasis efforts

# Progress to Report

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

} *Mandatory and optional fields*
- Several Gov-to-Gov meetings to understand internal best practices
  - An effective standard will enhance current use cases
  - Buy-in across multiple agencies is preferred
- The Investigatory Voice Biometrics Committee
  - Developing a proposal focusing on the voice signature - *least controversial*
  - Proposal to be shaped into the actual type-11 record along with possible requirements to modify the type 20 specification
  - Plan to distribute the proposal to all interested, soon

# Points Considered for the Proposal

- Primary sources of metadata being proposed
  - Government meetings/discussions
  - Investigatory Voice Biometrics symposium reports (2009)
- Sampling of metadata  
(case-related information in Type 2 fields)
  - How much voice in the sample?
  - How, when and where was the recording made?
  - How to present speech location(s) inside a data sample?
  - Differentiate between permanent vs. temporary speaker attributes
  - Differentiate between signal information and content information
  - Plus many, many, more



# Type 11 Record Development

*(way forward)*

- Transition the Investigatory Voice Biometrics proposal to the larger committee for comment
  - Will receive feedback as to where proposal meets/fails under alternative use-cases
- Utilize NIST experts with ANSI/NIST-ITL to formalize the Type-11 Record
- Expand to accommodate data collected in the wild
- When the Type 11 record is ready it will be put forth as an amendment to the current version of the ANSI/NIST-ITL Standard
  - A proposal, comment, acceptance process will follow the standard ANSI voting procedures

# Questions or More Information

- Mark Przybocki, *Multimodal Information Group*  
[mark.przybocki@nist.gov](mailto:mark.przybocki@nist.gov)
- Alvin Martin, *Chair of the Type 11 Voice Committee*  
[alvin.martin@nist.gov](mailto:alvin.martin@nist.gov)

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