INCITS 456
Speaker Recognition Format for Raw Data Interchange (SIVR-1)

Judith A. Markowitz, PhD
J. Markowitz, Consultants
Chicago, IL
www.jmarkowitz.com

W3C Workshop on SIV

March 6, 2009

What is INCITS 456?

- Data interchange format
- Storage & exchange of speech/voice data
- CBEFF Biometric Data Block (BDB)
- Draft standard
- First for speech/voice
- First in XML
- Developed jointly by ANSI/INCITS M1 and VoiceXML Forum's Speaker Biometrics Committee

What is INCITS 456?

- Captures, stores, and exchanges RAW data
- Does not capture features or models
- Goal is to provide information that will enable recipient to analyze the data
  - Audio format
  - Input device and channel
  - Speaker (sex, age) but not claim
  - Language/dialect

Uses of INCITS 456

- Data sharing
- Watch list creation
- Internal system audit
- Automatic reenrollment of users
- Multi-biometric fusion
- Product/algorithm testing
- SIV registry/service

Structure of INCITS 456

Two levels

- Session header

  information that should not change during the session (EX: sex of speaker, date of session, device & channel, audio format…)

- Instance header (Interaction Turn)

  information that changes from turn to turn of a dialogue (EX: utterance length, SNR, prompt, content…)

Example: Enrollment

July 14, 2008 Chicago

IVR: Welcome ABC Bank's VoiceSure enrollment system…Please say your account number.

Caller: 357128999 *ASR processes input*

SIV Session begins at 1:14 Central Daylight time

IVR: *Thank you… Please say your password*

Caller : *lollapalooza 2.5 seconds*

IVR: *Please say your password again.*

Caller: *lollapalooza 2.2 seconds*

SIV Session ends at 1:16 Central Daylight time

Example: Enrollment

Session header (partial)

Date & Time start: 2008-07-14T13:14-5:00

Date & Time end:2008-07-14T13:16-5:00

Channel: Digital NonVoIP, 300-3100 Hz

Audio format: ByteOrder=0xFF00, streaming, format=OGG Vorbis, mono, sampling rate=8000, bits per sample=8…

Speaker: female

Input Device: telephone

Example: Enrollment

Instance 1 header (partial)

ASR used: No

Prompt used: prompt1.wav

Utterance: utt1.wav, 2.5 sec. (20000 samples), content unknown, Volume=68.5 dB, SNR 42.1

Quality rating: unknown

Code for Example (Session Header)

```
<Session FormatVersion="SIVR-1">
 <DateAndTime>
    <start>"2008-07-14T13:11-5:00"</start>
    <end>"2008-07-14T13:14-5:00"</end>
 </DateAndTime>
    <Channel>
      <Type>"DigitalNonVoIP"</Type>
      <CutoffTop>3100</CutoffTop>
      <CutoffBottom>300</CutoffBottom>
    </Channel>
<AudioFormatHeader>
    <ByteOrder>0xFF00</ByteOrder>
    <Streaming>0</Streaming>
    <HeaderSize>25</HeaderSize>
    <FileLengthInSamples>13600</FileLengthInSamples>
    <AudioFormat>"OGG Vorbis"</AudioFormat>
    <ChannelCount>1</ChannelCount>
    <SamplingRate>8000</SamplingRate>
    <BitsPerSample>8</BitsPerSample>
    <AudioFullSecondsOf>6</AudioFullSecondsOf>
    <AudioRemainderSamples>5600</AudioRemainderSamples>
</AudioFormatHeader>
```

Audio Format

Code for Example (Session Header cont.)

```
<Speaker>
    <SpeakerMF>”Female”</SpeakerMF>
</Speaker>
<InputDevice>
    <Type>”Telephone”</Type>
</InputDevice>
```

Code for Example (instance #1)

```
<Instance>
    <InstanceNumber>1</InstanceNumber>
    <ASRUsed>”No”</ASRUsed?
    <TypeOfPromptContent>”String”</TypeOfPromptContent>
    <StringPromptContent>”URL EnrollPrompts/Prompt1.wav”</StringPromptContent>
    <Utterance>
          <DataType>”Pointer”</DataType>
          <Data>”20080714-3124554/Utt1.wav”</Data>
          <FileLengthInSamples>20000</FileLengthInSamples>
          <AudioFullSecondsOf>2</AudioFullSecondsOf>
          <AudioRemainderSamples>4000</AudioRemainderSamples>
          <Content>”Unknown”</Content>
          <Volume>68.5</Volume>
           <SNREstimate>42.1</SNREstimate>
           <Quality>
                <Score>254</Score>
                <AlgorithmVendorID>0</AlgorithmVendorID>
                <AlgorithmID>0</AlgorithmID>
            </Quality>
     </Utterance>
</Instance>
```

Implementation

- Application generates format directly
- The generated format can then become part of an EMMA tag

```
<emma:interpretation id="intp1"
      emma:medium="acoustic" emma:mode="voice"
          emma:function="verification">
    <DEFF uri="http://example.com/DEFF-docs/mydoc12345/>
  </emma:interpretation>
```

Implementation

- Or be used as a resource in an EMMA derivation

```
<emma:derivation>
    <emma:interpretation id="better">
     <emma:derived-from
          resource=http://www.INCITS456-1.txt
          composite="false"/>
                    .
                    :
</emma:interpretation>  </emma:derivation>
```