Mixer Corpus of Multilingual, Multichannel Speaker Recognition Data

Christopher Cieri1, Joseph P. Campbell2, Hirotaka Nakasone3, David Miller1, Kevin Walker1
1University of Pennsylvania, Linguistic Data Consortium, Philadelphia, PA, USA
2MIT Lincoln Laboratory, Lexington, MA, USA
3Federal Bureau of Investigation, Quantico, VA, USA
ccieri@ldc.upenn.edu, j.campbell@ieee.org, hnakasone@fbiacademy.edu, damiller@ldc.upenn.edu walkerk@ldc.upenn.edu

Forensic Automatic Speaker Recognition
- Minimum Requirements
  - text-independent
  - channel-independent
- New requirements
  - capable of handling multiple languages including bilingual speakers

Plan
- multilingual, multi-channel collection
- dissemination to research sites
- system performance improvement
- system performance evaluation

Component Design

<table>
<thead>
<tr>
<th>Component</th>
<th>Arabic</th>
<th>Mandarin</th>
<th>Russian</th>
<th>Spanish</th>
<th>XC &amp; TR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speakers</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Calls</td>
<td>4</td>
<td>10</td>
<td>20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Generalities
- calls are 6 minutes in duration
- subjects speak to each other
- assigned topics change daily
- robot operator logs ANI
- unique handsets encouraged
- subjects indicate phone and handset type

Components
- Core: 600 subjects*10 calls
- Extended: 100 subjects continue to 20 calls
- Multilingual
  - 100 subjects * 4+ of their calls in Arabic
  - 100 subjects * 4+ of their calls in Mandarin
  - 100 subjects * 4+ of their calls in Russian
  - 100 subjects * 4+ of their calls in Spanish
- Multi-channel Data: 100 subjects * 4 of their calls via multi-channel device
- Transcript Reading: 100 subjects read extracts of transcripts of each others’ and their conversations via multi-channel device

Outcomes To Date
- 4651 subjects recruited
- 12,169 calls (~1200 hours) collected
- >250 new calls each week
- >100 cross channel subjects completed 4 calls
- collection continues

Research
- Supports speaker recognition research with emphasis on forensic-style problems:
  - telephone conversations
  - channel independence
  - language independence & bilingualism
  - transcript reading
- Mixer is the first large scale, publicly available corpus to address all these dimensions.
- FBI vision is to create a corpus that supports speaker recognition system to support research on forensic-style problems.
- Research at MIT-LL aims to produce robust automatic speaker recognition system to support forensic analysis experts.
- Mixer used in 2004 NIST Speaker Recognition Evaluation