**Introduction**

- **Goal**: Identify spoken language from utterances
- **Challenge**: Short and noisy utterances
- **Framework**: Total Variability i-Vector Modeling + SVM

**Total Variability i-Vector Modeling**

- **Complete data distribution**: Gaussian Mixture Model, denoted as UBM
- **Utterance-specific data distribution**: GMM, with UBM component means shifted slightly
- **Model Assumption**: Shift in UBM mean supervector is low-dimensional

**Prior Modification**

- **Motivation and hypothesis**:
  - **Observation**: i-Vectors estimated from ample data form clusters
  - **Standard normal prior**: Penalizes probability of large magnitude i-Vector estimates
  - **Hypothesis**: Better to penalize deviation from cluster centers
  - **Gaussian Mixture** prior is better suited to this purpose

- **Proposed Prior**: GMM, with one component per class
  \[ P(x) = \sum_{i=1}^{M} P_C(i) N(\mu_i, C_i) \]

- **i-Vector estimate with GMM Prior**:
  \[ E[x|F] = \sum_{i=1}^{M} P_C(i) I_t^{-1} b_i \]
  \[ b_i = T^*\Sigma^{-1}nF + C_i^{-1}\mu_i, \quad I_t = C_i^{-1} + T^*\Sigma^{-1}nT \]

- **Prior re-weighting**:
  Provide a parameter \( \lambda \) to tune weight of prior relative to data:
  \[ b_i = \lambda T^*\Sigma^{-1}nF + C_i^{-1}\mu_i, \quad I_t = C_i^{-1} + \lambda T^*\Sigma^{-1}nT \]

**Effect of reducing duration**

- **Variance** of the i-Vector estimate increases
- **Decisions become error-prone**
- **Estimate is driven by the nature of the prior**

**Database and System Description**

- **Database**: DARPA RATS
  - Noisy audio recordings from six classes:
    - Five target languages
    - A class corresponding to 10 non-target languages
  - Audio utterances of length: 120s, 30s, 10s, 3s

- **System Description**:
  - **UBM Size**: 2048 Components
  - **i-Vector dimension**: 400
  - **Inter-session variability compensation**: WCCN
  - **SVM**: Fifth order polynomial kernel
  - **Utterance duration**:
    - i-Vector Model \((T, \Sigma)\) Estimation: 30 s
    - SVM training and Test: 3 s

**Results**

<table>
<thead>
<tr>
<th>System</th>
<th>EER</th>
<th>DCF</th>
<th>P10\text{#miss}</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>15.40</td>
<td>15.21</td>
<td>22.19</td>
<td>69.74</td>
</tr>
<tr>
<td>GMM i-Vector</td>
<td>15.27</td>
<td>14.98</td>
<td>20.76</td>
<td>69.44</td>
</tr>
<tr>
<td>GMM Re-estimation</td>
<td>16.32</td>
<td>15.82</td>
<td>22.32</td>
<td>70.11</td>
</tr>
<tr>
<td>Score Re-estimation</td>
<td>15.14</td>
<td>15.07</td>
<td>21.28</td>
<td>69.96</td>
</tr>
</tbody>
</table>

**Acknowledgment**

This research was supported by the Defense Advanced Research Projects Agency (DARPA) and the National Science Foundation (NSF).