## General overview

- **Task condition**
  - Training-Test condition: short2-short3 (core condition)
  - System development based on English telephone speech
  - => English telephone speech (Cond. 7) trials will be focused in performance analysis

- **Submission overview**

### Submission

<table>
<thead>
<tr>
<th>Sub-systems</th>
<th>MFCC-GMM</th>
<th>MFCC-SVM</th>
<th>GSV-SVM</th>
<th>LLR-SVM</th>
<th>LP-GMM</th>
<th>LPGSV-GMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIM 1</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIM 2</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIM 3</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

### General techniques

- **Front-end of cepstral systems (MFCC-GMM, MFCC-SVM and GSV-SVM):**
  - Speech activity dectection using an energy threshold along with pitch detection
  - 15 cepstrum + 15 Δ cepstrum + 15 Δ 2 cepstrum + 1 Δ energy + 1 Δ energy (47-D features)
  - Feature mapping and feature warping in SVM sub-systems
  - Feature warping only in the GMM sub-system

- **Intersession variation compensation**

  1) Parallelized training of factor analysis models using Maximum-likelihood estimation (MLE) in MFCC-GMM sub-system
     - Compensation performs in model domain and feature domain
     - Rank 40, SRE04 as training data
  2) Factor analysis with ALIZE toolkit in LP-GMM and LPGSV-SVM sub-systems
     - Compensation performs in model domain and feature domain
     - Rank 10, SRE04 as training data
  3) NAP adopted in other remaining sub-systems
     - Subspace dimension: 50, SRE04 as training data
     - Min-max feature normalization

- **SVM-related sub-systems**
  - Gender-dependent linear-kernel SVM (using SVMTorch from IDIAP)
  - 7092 impostor speaker segments taken from SRE04 and SWB I (English data) in cepstral SVM sub-systems
  - 1923 impostor speaker segments taken from SRE04 (English data) in LPGSV-SVM sub-system

- **New characteristics (compared to LIMSI SRE06 system)**
  - Use of MLLR instead of CMLLR in a sub-system
  - Addition of prosodic systems with Legendre coefficients as prosodic features, and with modelling using GMM and SVM
  - Use of NAP and factor analysis in intersession variation compensation
  - Use of logistic regression in score-level system fusion

## Individual sub-systems

1) **MFCC-GMM system**
   - 1536-component GMMs, with MAP adaptation of gender-dependent UBMs, trained using SRE00 training data, and SRE01 developing and training data (totally around 24 hours)
   - Scoring:
     - Log-likelihood ratio with 20 top Gaussians scoring
     - Perform gender-dependent T-norm using SRE04 data (250 female and 250 male speech segments)

2) **MFCC-SVM system**
   - Polynomial features as features
   - 3rd order polynomial
   - Variance normalization
   - Kernel PCA, 2nd order polynomial kernel
   - SVM classifier

3) **GSV-SVM system**
   - Gaussian mean supervectors as features
   - 512 Gaussians
   - Variance normalization
   - SVM classifier

4) **MLLR-SVM system**
   - 47 cepstral features with feature mapping and feature warping
   - Segmentation by forced alignment using transcripts provided by NIST
   - 3-class MLLR transforms as features
   - consonant/vowel/non-speech classes, non-speech class is dropped
   - SAT models
   - No backoff
   - SVM classifier

5) **LP-GMM system**
   - Pseudo-syllable prosodic contour segmentation (based on valley points on energy contour)
   - 6 Legendre coefficients for energy and pitch contours respectively + segmentation duration (totally 13 features)
   - 256-component GMMs, with gender-dependent UBMs trained using SRE04 data
   - Scoring:
     - Log-likelihood ratio with 20 top Gaussians scoring
     - Perform gender-dependent T-norm using SRE04 data

6) **LPGSV-SVM system**
   - Same prosodic feature extraction (with factor analysis) as the LP-GMM system
   - Gaussian mean supervectors (with factor analysis) as features
   - 256 Gaussians
   - Variance normalization
   - SVM classifier

## Score-level system fusion

- **Linear logistic regression as described in [9] is used**
  - 1conv4w-1conv4w core trials in NIST SRE06 are used to train fusion weights and decision threshold values
- **To deal with invalid scores for certain trials in some sub-systems:**
  - Affine calibration transformation is firstly performed so that transformed sub-system scores can be interpreted as log-likelihood ratios
  - 1conv4w-1conv4w core trials in SRE06 used as the training data for the transformation.
  - After score transformation, trials without valid scores were replaced with scores of zero and the fusion can be trained as usual.
Parallelized factor analysis & Survey of prosodic modelling

- MLE-based factor analysis enables parallelized implementation of model training, which speeds up the process without sacrificing the performance
  - Performance of MFCC-GMM system running on SRE05 data

<table>
<thead>
<tr>
<th>Implementation of factor analysis</th>
<th>Min. DCF (x10)</th>
<th>EER (%)</th>
<th>Running time of model training (hour/iteration)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALD2 toolbox approach</td>
<td>0.231</td>
<td>5.36</td>
<td>10.7*</td>
</tr>
<tr>
<td>MLE approach</td>
<td>0.231</td>
<td>5.36</td>
<td>2.0*</td>
</tr>
</tbody>
</table>

*Note: This test was performed on machines with 3.0-GHz Intel Core 2 and 3.0-GHz Intel Core 2.
ALD2 toolbox approach uses one machine and MLE approach uses 20 machines simultaneously.

- Performance of prosodic systems (LP-GMM and LPGSV-SVM) on development data (lconv4w-lconv4w English trials in SRE06)
- Fusion weights trained using lconv4w-lconv4w English trials in SRE05

<table>
<thead>
<tr>
<th>Systems</th>
<th>EER (%)</th>
<th>Fusion with ML-LIR-SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>LP-GMM</td>
<td>0.6596</td>
<td>17.7</td>
</tr>
<tr>
<td>LPGSV-SVM</td>
<td>0.7097</td>
<td>19.29</td>
</tr>
<tr>
<td>LPGSV + GMM</td>
<td>0.6277</td>
<td>15.86</td>
</tr>
</tbody>
</table>

- Other prosodic systems which we do not use in SRE08 submission also provide performance gain
  - System performance running on SRE05 data

Performance analysis (LIMSI system)

- Best-N fusion test on development data (lconv4w-lconv4w English trials in SRE06)
  - Fusion weights trained using lconv4w-lconv4w English trials in SRE05

<table>
<thead>
<tr>
<th>Systems</th>
<th>EER (%)</th>
<th>Fusion with ML-LIR-SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed-size segment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phone segment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syllable-like segment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean, min, max and delta of pitch</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Delta-pitch and delta-energy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legendre coefficients of pitch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaker model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GMM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-gram</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Systems</th>
<th>EER (%)</th>
<th>Fusion with MFCC-GMM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>21.57 19.01 22.88 22.60 25.16 28.61 19.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.07 7.99 8.27 8.15 8.40 8.52 8.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.2 7.1 3.8 5.2 2.3 0.9 4.8</td>
</tr>
</tbody>
</table>

- Best-N fusion test on SRE08 (English telephone trials; Cond. 7)
  - Fusion weights trained using lconv4w-lconv4w English trials in SRE05

Conclusions

- MFC-GMM, ML-LIR-SVM and LP-GMM sub-systems provide significant contribution in overall performance
- Parallelized implementation of factor analysis using MLE facilitates rapid model training without sacrificing the system performance
- Prosodic features provide significant contribution and it will be one of our ongoing directions

References