1. Challenges

- Co-articulation effects at code-switches
- No dictionary for Singaporian and Malaysian EN
- Limited training data for AM
- Lack of text data for LM training

2. SEAME Corpus

<table>
<thead>
<tr>
<th></th>
<th>Train</th>
<th>Dev</th>
<th>Eval</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Speakers</td>
<td>139</td>
<td>8</td>
<td>19</td>
<td>157</td>
</tr>
<tr>
<td>Duration (hrs)</td>
<td>58.4</td>
<td>2.1</td>
<td>2.3</td>
<td>62.8</td>
</tr>
<tr>
<td>#Utterances</td>
<td>48,080</td>
<td>1,943</td>
<td>2,162</td>
<td>52,245</td>
</tr>
</tbody>
</table>

- #Code-Switches = 2.6/Utt
- Very short monolingual segments
  - > 82% EN, >73% CN segments < 1 sec
  - Average duration of EN segments = 0.67 sec
  - Average duration of CN segments = 0.81 sec

3. Baseline system

- 143 MFCC (adjacent frames) → 42 (LDA)
- HMM – GMM, SAT + bMMIE training
- CMU EN dictionary + MAN dictionary + pronunciation variations (rules-based)
- n-gram LM: Linear interpolation of CN LM and EN LM with LM trained with CS transcriptions
- OOV rate: 1.21%, PPL: 489.4
- 2-pass system (cMLLR + MLLR)
- MER: 37.3%

4. Code-Switching Acoustic Models

- Phone merging:
  - Knowledge based: IPA
  - Data-driven: using Battacharya Distance
- Apply Discriminative Training on top

<table>
<thead>
<tr>
<th>System</th>
<th>Baseline</th>
<th>Knowledge-based</th>
<th>Data-driven</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAT</td>
<td>39.7</td>
<td>39.6</td>
<td>39.6</td>
</tr>
<tr>
<td>+ bMMIE</td>
<td>37.3</td>
<td>37.1</td>
<td>37.2</td>
</tr>
</tbody>
</table>

5. SMT-based Language Modeling

- Analysis of approaches to decide which segments to translate
  - based on the code-switch behavior in the SEAME training transcriptions
  - simple Search&Replace, replacement only if segment occurs at least twice, using information of Trigger Words and Trigger POS tags, adapt frequency of code-switch segments, combinations
- Best approach:
  - Set maximum number of replacements per segment, based on the segment frequency in the SEAME training text
  - Improvement in MER by 0.2% absolute → MER: 36.3%
  - Small improvement due to little data for reliable estimates of code-switch behavior for our approaches (48k utterances in SEAME training text)

6. Language Identification (LID)

- Feature Extraction
- Voice Activity Detection
- Language ID Recognition
- Post Processing
- LID

- Voice activity detection: 5.88% Frame Error Rate
- Two languages LID: 16.64 % FER, Others: 7% FER

7. LID integration into Decoding

- Multistream approach
  - Language Tag (CN, EN) in dictionary
  - LID decision tree, LID weight = 0.1
  - MER: 36.5 % (Oracle experiment: 34.4 % MER)

8. Conclusion

<table>
<thead>
<tr>
<th>Systems</th>
<th>MER(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>37.3</td>
</tr>
<tr>
<td>+ CS AM</td>
<td>37.1</td>
</tr>
<tr>
<td>+ SMT based LM</td>
<td>36.9</td>
</tr>
<tr>
<td>+ LID Integration</td>
<td>36.5</td>
</tr>
<tr>
<td>+ Oracle LID</td>
<td>34.4</td>
</tr>
</tbody>
</table>