3) Basic Frisian ASR system

- Challenges
  - Low resources available
  - Code-switching nature of Frisian
  - Complex vowel system

- Acoustic Model
  - FAME! Database is created
  - Frysk Akademy text corpus
  - Omrop Fryslân news articles
  - Frisian Wikipedia
  - Transcriptions of training speech
  - Fluent Frisian Lexicon
  - Elex Dutch Lexicon

4) FAME! Frisian Radio Broadcast Database

- Preparation
  - Manually annotating the radio broadcasts from Omrop Fryslân
  - Collaboration with Frysk Akademy
  - Annotations include orthographic transcription, speaker ids, spoken language, code-switching details, dialect info
  - A modified annotation protocol has been created

- Some statistics
  - 18.5 hours of radio broadcasts annotated in total
  - Longitudinal data: recordings from 1966 to 2015
  - More than 500 speakers, 309 with known identity
  - 21 speakers appear at least 3 times

- 3939 code-switching cases:
  - 2896 cases: Frisian speaker switches to Dutch
  - 95 cases: Dutch speaker switches to Frisian
  - 848 cases: Speakers use a mixed-word that is neither Frisian nor Dutch

5) Frisian Language Model and Lexicon

- Language Model
  - Frisian text corpus: ~2,375,000
  - Training speech transcription: ~13,750 sentences
  - Dutch text corpus (CGN): ~580,000 sentences
  - Monolingual and bilingual N-gram models are trained

- Lexicon
  - Complete Frisian lexicon: ~340k words
  - Dutch phones are mapped to the phonetically closest Frisian phone.
  - Grapheme-to-Phoneme (G2P) models are learned to handle the out-of-vocabulary (OOV) words in training data

6) Initial Recognition Experiments

- Speech data from Frisian speakers
  - FAME! Database is divided into three parts
    - Training set: 8h 20m
    - Development set: 1h
    - Test set: 1h
  - Acoustic models (AM)
    - KALDI speech recognition toolkit is used
    - GMM-HMM and subspace GMM (SGMM) are trained on LDA-MLLT features
    - Speaker adapted training (SAT): FMLLR-adapted features
  - Language models (LM)
    - 3-gram interpolated modified Kneser-Ney
    - Frisian LM and Bilingual LM are compared
  - Lexicon
    - Frisian lexicon contains ~95k words
    - Bilingual lexicon contains ~150k words
    - Various phonetic alphabets are compared:
      - mono: cons. + monoph.
      - fall: monoph. + all diph.
      - rise: monoph. + all diph.
      - riph: monoph. + all triph.

7) Results (I) – Phonetic Alphabet

<table>
<thead>
<tr>
<th>Lex</th>
<th>GMM</th>
<th>SGMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>mono</td>
<td>50.65</td>
<td>44.56</td>
</tr>
<tr>
<td>fall</td>
<td>50.34</td>
<td>45.01</td>
</tr>
<tr>
<td>rise</td>
<td>50.85</td>
<td>45.31</td>
</tr>
<tr>
<td>riph</td>
<td>50.59</td>
<td>45.10</td>
</tr>
<tr>
<td>riph</td>
<td>50.23</td>
<td>44.88</td>
</tr>
<tr>
<td>dirn</td>
<td>50.82</td>
<td>45.46</td>
</tr>
</tbody>
</table>

- Word error rates (WER) in % on the development set:

8) Results (II) – AM, LM and Lexicon

<table>
<thead>
<tr>
<th>Devel</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lex</td>
<td>GM</td>
</tr>
<tr>
<td>FR</td>
<td>FR</td>
</tr>
<tr>
<td>FR-NL</td>
<td>FR</td>
</tr>
<tr>
<td>FR-NL</td>
<td>NL</td>
</tr>
</tbody>
</table>

- WERs using mono- and bilingual lexicon and LM:

<table>
<thead>
<tr>
<th>Devel</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lex</td>
<td>GM</td>
</tr>
<tr>
<td>FR</td>
<td>FR</td>
</tr>
<tr>
<td>FR-NL</td>
<td>FR</td>
</tr>
<tr>
<td>FR-NL</td>
<td>NL</td>
</tr>
</tbody>
</table>

- WERs with G2P for OOV words (Best results for SI system):

<table>
<thead>
<tr>
<th>Devel</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lex</td>
<td>GM</td>
</tr>
<tr>
<td>FR</td>
<td>FR</td>
</tr>
<tr>
<td>FR-NL</td>
<td>FR</td>
</tr>
<tr>
<td>FR-NL</td>
<td>NL</td>
</tr>
</tbody>
</table>

- WERs with G2P and SAT (Best results for SA system):

9) Conclusion

- Initial recognition results are promising for an accurate spoken document retrieval system
- Future work: Investigating deep architectures and recognition schemes with flexible lexicon for code-switching ASR