1) Introduction

- **Frisian Language**
  - regional official language of the Netherlands
  - mostly spoken in the province of Fryslân
  - approximately half a million speakers
  - most speakers are bilingual due to the extensive influence of Dutch
  - code-switching is common practice in daily conversations
  - Member of West Germanic language family
  - closely related with English and Dutch

- **Phonology**
  - Frisian consonants are more or less similar to Dutch consonants
  - Frisian has more vowels (falling and rising diphthongs, triphthongs, rassialization)

- **Linguistically well-researched, however few speech and language technology applications are available**
  - Frisian speech synthesizer, Google Translate...
  - Omrop Fryslân
  - regional public broadcaster with a radio station and a TV channel both broadcasting in Frisian

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
  - Frysk Wikipedia
  - Transcriptions of training speech
  - Fluency 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, 369 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

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: mono + fall diph.
    - rise: mono + rise diph.
    - riph: mono + all diph.
    - iriph: mono + all triph.
    - ditri: diph + all triph.

7) Results (I) – Phonetic Alphabet

<table>
<thead>
<tr>
<th></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>53.34</td>
<td>45.01</td>
</tr>
<tr>
<td>rise</td>
<td>55.85</td>
<td>45.31</td>
</tr>
<tr>
<td>riph</td>
<td>50.09</td>
<td>45.10</td>
</tr>
<tr>
<td>riph</td>
<td>50.23</td>
<td>44.88</td>
</tr>
<tr>
<td>ditri</td>
<td>50.82</td>
<td>45.46</td>
</tr>
</tbody>
</table>

- Word error rates (WER) in % on the development set -
  - The choice of the phonetic alphabet has a minor effect on the recognition accuracy
  - Inferior performance of *ditri* is explained by the limited amount of training data
  - In the following experiments, *mono* is adopted

8) Results (II) – AM, LM and Lexicon

<table>
<thead>
<tr>
<th></th>
<th>Devel</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lex</td>
<td>LM</td>
<td>GMM</td>
</tr>
<tr>
<td>FR</td>
<td>FR</td>
<td>50.65</td>
</tr>
<tr>
<td>FR-NL</td>
<td>FR</td>
<td>50.01</td>
</tr>
<tr>
<td>FR-NL</td>
<td>FR-NL</td>
<td>50.51</td>
</tr>
</tbody>
</table>

- WERs using mono- and bilingual lexicon and LM -

<table>
<thead>
<tr>
<th></th>
<th>Devel</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lex</td>
<td>LM</td>
<td>GMM</td>
</tr>
<tr>
<td>FR</td>
<td>FR</td>
<td>49.26</td>
</tr>
<tr>
<td>FR-NL</td>
<td>FR</td>
<td>49.39</td>
</tr>
<tr>
<td>FR-NL</td>
<td>FR-NL</td>
<td>49.16</td>
</tr>
</tbody>
</table>

- WERs with G2P for OOV words (Best results for SI 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