**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

- **FAME! Database is created**
  - Frysk Academy text corpus
  - Omrop Fryslân news articles
  - Frisian 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 Academy
  - 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 sentences
  - 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
  - Complete Dutch lexicon: ~1.1M words
  - Frisian phonetic alphabet contains 20 consonants, 20 monophthongs, 18 falling diphthongs, 8 rising diphthongs and 6 triphthongs
  - For bilingual lexicon, Dutch phones are mapped to the phonetically closest Frisian phone.
  - Grapheme-to-Phoneme (G2P) models are learned to handle the out-of-vocabulary (OVO) 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
  - GMAM-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 lexiconcontains ~95k words
  - Bilingual lexicon contains ~150k words
  - Various phonetic alphabets are compared:
    - mono: cons. + monoph. -> fall: mono + fall diph.
    - rise: mono + rise diph.
    - diph: mono + all diph.
    - triph: mono + 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>50.34</td>
<td>45.01</td>
</tr>
<tr>
<td>rise</td>
<td>50.85</td>
<td>45.31</td>
</tr>
<tr>
<td>diph</td>
<td>50.09</td>
<td>45.10</td>
</tr>
<tr>
<td>triph</td>
<td>50.23</td>
<td>44.88</td>
</tr>
<tr>
<td>dtrn</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 dtrn 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>GMAM</td>
<td>SGMM</td>
</tr>
<tr>
<td>FR</td>
<td>FR</td>
<td>FR</td>
</tr>
<tr>
<td>NL</td>
<td>NL</td>
<td>NL</td>
</tr>
<tr>
<td>GMAM</td>
<td>GMAM</td>
<td>GMAM</td>
</tr>
<tr>
<td>SGMM</td>
<td>SGMM</td>
<td>SGMM</td>
</tr>
<tr>
<td>GM</td>
<td>GM</td>
<td>GM</td>
</tr>
<tr>
<td>SG</td>
<td>SG</td>
<td>SG</td>
</tr>
<tr>
<td>Devel</td>
<td>Devel</td>
<td>Devel</td>
</tr>
<tr>
<td>FR</td>
<td>FR</td>
<td>FR</td>
</tr>
<tr>
<td>NL</td>
<td>NL</td>
<td>NL</td>
</tr>
<tr>
<td>GM</td>
<td>GM</td>
<td>GM</td>
</tr>
<tr>
<td>SG</td>
<td>SG</td>
<td>SG</td>
</tr>
</tbody>
</table>

- WERs using mono- and bilingual lexicon and LM

**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