1) Introduction

- Frisian Language
  - Regional 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.

- 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.

2) FAME! Project

- Disclose the Omrop Fryslân archives containing recordings from 1950s.
- Develop a user-friendly search interface for spoken documents from Omrop Fryslân archives with more than 2600 hours of radio broadcasts.
- Relevant applications toward building this spoken document retrieval system:
  1. Automatic speech recognition.
  2. Speaker identification.
  3. Flexible search interface.
  4. Project Partners:

3) Basic Frisian ASR System

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

- FAME! Database is created.
- Frysk Akademy text corpus.
- Omrop Fryslân Lan news articles.
- Frisian Wikipedia.
- Transcriptions of training speech.
- Frisky 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 id's, 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 (CGM): ~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, 16 falling diphthongs, 8 rising diphthongs and 6 triphthongs.
  - For bilingual lexicon, Dutch phonemes are mapped to the phonetically closest Frisian phoneme.
  - 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: mono + fall diph.
    - rise: mono + rise diph. -> diph: mono + all diph.
    - trigraph: mono + all trigraph.

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>trigraph</td>
<td>50.23</td>
<td>44.88</td>
</tr>
<tr>
<td>diph</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 diph is explained by the limited amount of training data.
- In the following experiments, mono is adopted.

8) Results (II) – AM, LM and Lexicon

- WERs using monolingual 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.