1) Introduction

- Regional official language of the Netherlands.
- Mostly spoken in the province of Fryslân.
- Approximately 20% of the population.
- Mostly 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.

2) FAME! Project

- KALDI speech recognition toolkit is used.
- Omrop Fryslân.
- Regional public broadcaster with a radio station.
- TV channel both broadcasting in Fryslân.

3) Basic Frisian ASR System

- Challenges:
  - Low resources available.
  - Code-switching nature of Frisian.
  - Complex vowel system.
- IAM: Database is created.
- Frysk Akademy text corpus.
- Omrop Fryslân news articles.
- Flexury Fryslân 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.
- Some statistics:
  - 18.5 hours of radio broadcasts annotated in total.
- 2986 code-switching cases:
  - 95 cases: Dutch speaker switches to Frisian.
  - 484 cases: Users 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.
  - Frisian phonetic alphabet contains 20 consonants, 20 monophthongs, 16 falling diphthongs, 8 rising diphthongs, and 6 trinphongs.
  - 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.
  - GMM-HMM and BMM (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 dipth., + rise: mono + rise dipth., + dipth: mono + all dipth., + triph: mono + all triph.
  - Word error rates (WER) on in % on the development set.

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>dipth</td>
<td>50.09</td>
<td>45.10</td>
</tr>
<tr>
<td>trph</td>
<td>50.23</td>
<td>44.88</td>
</tr>
<tr>
<td>dtrph</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 dtrph is explained by the limited amount of training data.
- In the following experiments, mono is adopted.

8) Results (II) – AM, LM and Lexicon

- Development test:
  - WERs using mono- and bilingual lexicon and LM.
- Test:
  - WERs with G2P for OV word (Best results for SI system).

<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>50.71</td>
</tr>
<tr>
<td>FR-NL</td>
<td>FR-NL</td>
<td>50.51</td>
</tr>
</tbody>
</table>

- WERs with G2P for OV word (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.