The following full text is a preprint version which may differ from the publisher's version.

For additional information about this publication click this link. http://hdl.handle.net/2066/157399

Please be advised that this information was generated on 2017-08-25 and may be subject to change.
1) Introduction

- The Frisian language is a regional language of the Netherlands.
- It is spoken in the province of Fryslân and is more prevalent in the northern part of the province.
- The language is spoken by approximately half a million speakers.
- Many speakers are bilingual due to the influence of Dutch during conversations.
- The Frisian language is closely related to English and Dutch.

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 towards 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

4) FAME! Frisian Radio Broadcast Database

- Preparation:
  - Manually annotating the radio broadcasts from Omrop Fryslân.
  - Collaboration with Frysk Academy.
- Some statistics:
  - 18.5 hours of radio broadcasts annotated.
  - 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: Users speak a mixed-word that is neither Frisian nor Dutch.

5) Frisian Language Model and Lexicon

- 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.
- Bilingual lexicon.
- Dutch phones are mapped to the orthographic code.

6) Initial Recognition Experiments

- Speech data from Frisian speakers:
  - FAME! Database.
- Data Preparation:
  - Training: 8h 20m.
  - Development: 1h.
  - Test: 1h.
- Acoustic models (AM):
  - KALDI speech recognition toolkit.
  - GMH-GMM and subspace GMH (GGMM) are trained on LDA-MLLT features.
  - 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.
  - dipth: mono + all diph.
  - triph: mono + all triph.

7) Results (I) – Phonetic Alphabet

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

<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>riph</td>
<td>50.23</td>
<td>44.88</td>
</tr>
<tr>
<td>dtrh</td>
<td>50.82</td>
<td>45.46</td>
</tr>
</tbody>
</table>

- The choice of the phonetic alphabet has a minor effect on the recognition accuracy.
- Inferior performance of dtrh 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></td>
<td>Lex</td>
<td>LM</td>
</tr>
<tr>
<td>FR</td>
<td>FR</td>
<td>FR</td>
</tr>
<tr>
<td>FR-NL</td>
<td>FR</td>
<td>FR</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.

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.