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

- 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 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
- FAME! Database is created
- Frysk Academie 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 Academie
    - Annotations include orthographic transcription, speaker ids, spoken language, code-switching details, dialect info
    - 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 monophones, 16 falling diphthongs, 8 rising diphthongs and 6 triphones
  - 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 (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.
    - dipth: mono + all diph.
    - triph: mono + all triph.
- The choice of the phonetic alphabet has a minor effect on the recognition accuracy
- Inferior performance of dipth is explained by the limited amount of training data
- In the following experiments, mono is adopted

7) Results (I) – Phonetic Alphabet

- Word error rates (WER) in % on the development set:
  - mono: 50.65 GMM, 44.56 SGMM
  - fall: 50.34 GMM, 45.01 SGMM
  - rise: 50.59 GMM, 45.31 SGMM
  - dipth: 50.23 GMM, 44.88 SGMM
  - triph: 50.82 GMM, 45.46 SGMM

8) Results (II) – AM, LM and Lexicon

- Development
  - Lex | LM | GMM | SGMM | GMM | SGMM
  - FR | FR | 50.65 | 44.56 | 49.46 | 43.28
  - FR-NL | FR | 50.71 | 44.78 | 49.26 | 43.08
  - FR-NL-FR-NL | FR | 50.51 | 44.29 | 49.23 | 42.96

- Test
  - Lex | LM | GMM | SGMM | GMM | SGMM
  - FR | FR | 50.44 | 44.97 | 49.26 | 43.08
  - FR-NL | FR | 50.51 | 44.29 | 49.23 | 42.96
  - FR-NL-FR-NL | FR | 50.51 | 44.29 | 49.23 | 42.96

- WERs using mono- and bilingual lexicon and LM:
  - FR | FR | 44.72 | 40.87 | 40.13 | 39.39 | 37.84
  - FR-NL | FR | 45.90 | 41.56 | 40.57 | 44.41 | 40.40 | 38.57
  - FR-NL-FR-NL | FR | 45.64 | 41.42 | 40.30 | 44.17 | 39.99 | 38.38

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