Open Source Speech and Language Resources for Frisian

Emre Yılmaz[1], *Henk van den Heuvel*[1], *Jelske Dijkstra*[2], *Hans Van de Velde*[2], *Frederik Kampstra*[3] *Jouke Algra*[3] *and David Van Leeuwen*[1]

[1]CLS/CLST, Radboud University, Nijmegen, Netherlands
[2]Fryske Akademy, Leeuwarden, Netherlands
[3]Omrop Fryslân, Leeuwarden, Netherlands

{e.yilmaz, h.vandenheuvel, d.vanleeuwen}@let.ru.nl,
{jdijkstra, hvandevelde}@fryske-akademy.nl,
{frederik.kampstra, jouke.algra}@omropfryslan.nl

Abstract

In this paper, we present several open source speech and language resources for the under-resourced Frisian language. Frisian is mostly spoken in the province of Fryslân which is located in the north of the Netherlands. The native speakers of Frisian are Frisian-Dutch bilingual and often code-switch in daily conversations. The resources presented in this paper include a code-switching speech database containing radio broadcasts, a phonetic lexicon with more than 70k words and a language model trained on a text corpus with more than 38M words. With this contribution, we aim to share the Frisian resources we have collected in the scope of the FAME! project, in which a spoken document retrieval system is built for the disclosure of the regional broadcaster's radio archives. These resources enable research on code-switching and longitudinal speech and language change. Moreover, a sample automatic speech recognition (ASR) recipe for the Kaldi toolkit will also be provided online to facilitate the Frisian ASR research.

Index Terms: Open source, Frisian language, speech data, automatic speech recognition

1. Introduction

Contact-induced language change in multilingual countries appears in the form of phonological, morphological, syntactic and lexical changes as a result of various linguistic phenomena such as word borrowing and interference. These language contact phenomena are noticeable in minority languages due to the influence of the majority language or in some majority languages that have been influenced by globally influential languages such as English and French. One prominent mechanism, that is induced in the interacting languages, is code-switching which is defined as the continuous alternation between two languages in a single conversation. This topic has been researched in the field of linguistics for more than 30 years [1–3].

Despite the well-established research line in linguistics, robustness of automatic speech recognition (ASR) systems to code-switching and other kinds of language switches have recently received some interest resulting in some robust acoustic modeling [4–9] and language modeling [10–12] approaches for code-switching speech. One primary reason of this gap is the lack of audio data with high-quality recording and annotation in these under-resourced languages allowing the ASR research. Recently, various flexible and reliable data collection methods have been described including [13–15] and these data collection efforts have resulted in multiple databases for under-resourced languages [16–23].

Investigation of code-switching in the context of automatic speech recognition research has become viable in the last years on account of several code-switching databases [24–27]. These databases contain recordings of Mandarin-English, Hindi-English, Cantonese-English and French-German code-switching speech data. The automatic speech recognition systems applied on these data use bilingual pronunciation dictionary and language models to be able to cope with the language switch. Moreover, several language identification techniques are adopted to label the speech segments with the spoken language and perform accurate acoustic and language modeling based on these labels [28–31].

In this paper, we describe an open source data collection containing a speech database, a language model and a phonetic dictionary for the Frisian language. These resources have been collected in the scope of the FAME! (Frisian Audio Mining Enterprise) Project. This project aims to build a spoken document retrieval system for the disclosure of the archives of Omrop Fryslân[1] (Frisian Broadcast) covering a large time span from 1950s to present and a wide variety of topics. Omrop Fryslân has a radio station and a TV channel both broadcasting in Frisian and is the main data provider of this project with a radio broadcast archive containing more than 2600 hours of recordings. The Frisian speech database described in this paper is a small subset of these radio broadcasts.

The language model and the phonetic dictionary for the Frisian language are also provided as a part of this data collection. The phonetic dictionary is obtained by extracting the phonetic transcription in the Frysk Hânwurdboek (Frisian Dictionary) [32] which has been created by the Fryske Akademy[2] (Frisian Academy) and Afûk[3]. The large text corpus on which the language models are trained is excluded to be in compliance with intellectual property laws. The main aim of these resources is to stimulate Frisian ASR research in the context of an under-resourced language with code-switching phenomenon. The longitudinal and bilingual nature of the material also en-

[1]Omrop Fryslân is the regional public broadcaster of the province of Fryslân. (http://www.omropfryslan.nl)

[2]Fryske Akademy performs fundamental and applied research with both academic and social benefit in the fields of the Frisian language, culture, history and society. (http://www.fryske-akademy.nl)

[3]The Afûk foundation promotes the knowledge and use of the Frisian language and the interest in Fryslân and its culture. (http://www.afuk.frl)

ables to perform research into language variation and change in Frisian over time, formal versus informal speech, dialectology, code-switching trends, speaker tracking and diarization over a large period of time.

This paper is organized as follows. Section 2 introduces the demographics and the linguistic properties of the Frisian language. Section 3 describes the components included in the open source Frisian data. The organization of the database for ASR experiments is described in Section 4 and the baseline recognition results are presented in Section 5. Section 6 concludes the paper.

2. The Frisian Language

Frisian belongs to the North Sea Germanic language group, which is a subdivision of the West Germanic languages. Linguistically, there are three Frisian languages: West Frisian, spoken in the province of Fryslân in the Netherlands, East Frisian, spoken in Saterland in Lower Saxony in Germany, and North Frisian, spoken in the northwest of Germany, near the Danish border. These three varieties of Frisian are mutually barely intelligible [33]. The current paper focuses on the West Frisian language only and, following common practice, we will use the term Frisian for it.

Historically, Frisian shows many parallels with Old English. However, nowadays the Frisian language is under growing influence of Dutch due to long lasting and intense language contact. Frisian has about half a million speakers. A recent study shows that about 55% of all inhabitants of Fryslân speak Frisian as their mother tongue, which comes down to about 330,000 people [34]. All speakers of Frisian are at least bilingual, since Dutch is the main language used in education in Fryslân.

The Frisian alphabet consists of 32 characters including all English letters and six others with diacritics, i.e., â, ê, é, ô, û, ú. The Frisian phonetic alphabet consists of 20 consonants, 20 monophthongs, 24 diphthongs, and 6 triphthongs. Frisian has more vowels compared to Dutch which has 13 monophthongs and 3 diphthongs [35]. Dutch consonants are similar to the Frisian ones. There are three main dialect groups in Frisian, i.e. Klaaifrysk (Clay Frisian), Wâldfrysk (Wood Frisian) and Súdwesthoeksk (Southwestern). Although these dialects differ mostly on phonological and lexical levels, they are mutually intelligible [36].

3. Data Components

The components of the Frisian data collection[4] are speech and language resources gathered for building a large vocabulary ASR system for the Frisian language. Firstly, a new broadcast database is created by collecting recordings from the archives of the regional broadcaster and annotating them with various information such as the language switches and speaker details. The second component of this collection is a language model created on a text corpus with diverse vocabulary. Thirdly, a Frisian phonetic dictionary with the mappings between the Frisian words and phones is built to make the ASR viable for this under-resourced language. Finally, an ASR recipe is provided which uses all previous resources to perform recognition and present the recognition accuracies.

[4]This collection is available via http://www.ru.nl/clst/datasets/

3.1. The Speech Database

3.1.1. General Information

The Frisian speech database consists of 203 audio segments of approximately 5 minutes long extracted from various radio programs covering a time span of almost 50 years (1966–2015), adding a longitudinal dimension to the database. The content of the recordings are diverse including radio programs about culture, history, literature, sports, nature, agriculture, politics, society and languages. The database contains a diverse set of speakers appearing multiple times such as program presenters and celebrities in different recordings.

Two kinds of language switches are observed in broadcast data in the absence of segmentation information. Firstly, a speaker may switch language in a conversation (*within-speaker switches*). Secondly, a speaker may be followed by another speaking in the other language. For instance, the presenter may narrate an interview in Frisian, while several excerpts of a Dutch-speaking interviewee are presented (*between-speaker switches*). Both type of switches pose a challenge to the ASR systems and have to be handled carefully during recognition.

3.1.2. Annotation

The radio broadcast recordings provided by Omrop Fryslân have been manually annotated by two bilingual native Frisian speakers. The annotation protocol designed for this code-switching data includes three kinds of information: the orthographic transcription containing the uttered words, speaker details such as the gender, dialect, name (if known) and spoken language information. The language switches are marked with the label of the switched language. The segments containing background noise/music are also labeled to be able to evaluate their impact on the recognition accuracy. In order to get more precise information about the speaker details, all available meta-information of the annotated radio broadcasts is also provided. Every annotated audio segment is cross-checked by the other annotator to avoid systematic annotation errors and to increase the quality of the annotation.

The annotation has been performed using the PRAAT software [37] and the annotated information is stored in textgrid files. The speaker and spoken language information is stored in the tier names and the orthographic transcription and language switching information are stored in the tiers. The tier names are structured to contain all available information about the speaker and spoken language in `language-dialect/gender/speaker name` format. Focusing on the challenges introduced by the code-switching between the Frisian and Dutch language for the ASR systems, the annotation protocol does not distinguish between different types of language interaction. The switches in the spoken language are marked in the brackets including the acronym of the language. For instance, the Frisian speaker uttering the sentence below switches twice to Dutch and the Dutch words are marked with `[nl ...]`.

```
wy prate [nl namelijk] mei Marijke Nicolai
      en it is folle [nl ernstiger]
```
[5]

When a Dutch speaker switches to Frisian, the Frisian words/sentences are marked using `[fr ...]`. Finally, we use

[5]English translation: "We talk indeed with Marijke Nicolai and it is far more serious."

`[fr-nl ...]` for marking the words that can neither be classified as Dutch nor as Frisian. These kind of words include Dutch words pronounced according to Frisian pronunciation rules, Dutch verbs conjugated according to Frisian grammar, compound words consisting of a Frisian and a Dutch word.

3.1.3. Statistics

The total duration of the manually annotated radio broadcasts sums up to 18.5 hours. The stereo audio data has a sampling frequency of 48 kHz and 16-bit resolution per sample. The available meta-information helped the annotators to identify these speakers and mark them either using their names or the same label (if the name is not known). There are 309 identified speakers in the FAME! speech database, 21 of whom appear at least 3 times in the database. These speakers are mostly program presenters and celebrities appearing multiple times in different recordings over years. There are 233 unidentified speakers due to lack of meta-information.

The total number of word- and sentence-level code-switching cases in the FAME! speech database is 3837. These switches are mostly performed by the Frisian speakers as they often use Dutch words or sentences while speaking in Frisian. These cases comprise about 75.6% of all switches. The opposite case, i.e., a Dutch speaker using Frisian words or sentences, occurs much less accounting for 2.5% of all switches. This is expected as it is not common practice for Dutch speakers to switch between Dutch and Frisian. In the rest of the cases, the speakers use a *mixed-word* which is neither Frisian nor Dutch, for instance adapted loanwords. For further details, we refer the reader to [38].

3.2. Language model

As a part of the data collection effort, we created a large Frisian text corpus with more than 38M words. The Frisian text is mainly extracted from Frisian novels, news articles, wikipedia articles, newpapers, magazines and dictionaries. We have trained a trigram language model with on this text corpus to make the automatic speech recognition research viable using the proposed data package. This language model contains 76,629 unigrams, 2,548,372 bigrams and 4,463,871 trigrams. We did not use the orthographic transcriptions of the speech database in language model training to ensure that the provided language models can still be used for different organization of the speech database.

3.3. Phonetic dictionary

The Frisian phonetic dictionary containing the phonetic transcriptions of the most common Frisian words is extracted from the Frisian dictionary software (Frysk Hânwurdboek). It is stored in a UTF-8 text file containing 77,193 entries belonging to 71,429 words. For international validity, the Frisian phones are represented in the corresponding International Phonetic Alphabet (IPA) symbols. Using this original phonetic dictionary, we also created a lexicon compatible with the Kaldi ASR toolkit [39] in which all diphthongs and triphthongs are represented as a combination of their monophthong constituents for the ASR setup.

As the phonetic transcriptions are extracted from a dictionary, some suffixed versions of the words such as plurals are missing resulting in a large amount of out-of-vocabulary (OOV) words. To remedy this problem, we created a second lexicon which includes the out-of-vocabulary Frisian words in the training set and has been created based on grapheme-to-phoneme (G2P) models trained on the original lexicon. The G2P models are trained using the Phonetisaurus G2P software [40] to obtain up to 3-best phonetic transcriptions of the OOV words. The learned dictionary consists of 12,661 entries for 4715 words. These two lexicons are presented separately in the collection and a combined dictionary is used during the recognition.

4. ASR Experiments

Supporting numerous state-of-the-art ASR techniques, the Kaldi ASR toolkit is nowadays very popular in the ASR community. Therefore, we adapted a generic Kaldi recognition recipe to the aforementioned resources for performing ASR on Frisian speech. The total amount of audio segments containing speech is approximately equal to 14 hours. This data is divided into training, development and test sets to be able to perform ASR experiments. The training data of the database comprises of 8.5 hours and 3 hours of speech from Frisian and Dutch speakers respectively. The development and test sets consist of 1 hour of speech from Frisian speakers and 20 minutes of speech from Dutch speakers each. The training, development and test sets contain 2756, 671 and 410 language switching cases respectively.

Considering the monolingual resources used for the recognition, this recognition recipe performs training and recognition only using the speech data uttered by the Frisian speakers. These utterances still contain the switches of the Frisian speaker to Dutch, but exclude the utterances of the speakers who are labeled as Dutch speakers. For multilingual ASR research, the resources for the Dutch language, e.g. [41], have to be included together with the Frisian resources.

We first train a conventional context dependent Gaussian mixture models-hidden Markov models (GMM-HMM) system with 25k Gaussians using 39 dimensional mel-frequency cepstral coefficients (MFCC) including the deltas and delta-deltas. The number of context-dependent triphone states is 8593. A standard feature extraction scheme is used by applying Hamming windowing with a frame length of 25 ms and frame shift of 10 ms. DNNs with 6 hidden layers and 2048 sigmoid hidden units at each hidden layer are trained on both the 40-dimensional log-mel filterbank features (FBANK) and feature-space maximum likelihood linear regression transformed features (FMLLR) with the deltas and delta-deltas.

The recognition system using the subspace GMM (SGMM) acoustic models [42] is used to obtain the alignments for deep neural network (DNN) training. The DNN training is done by mini-batch Stochastic Gradient Descent with an initial learning rate of 0.008 and a minibatch size of 256. The time context size is 11 frames achieved by concatenating ± 5 frames. We further apply sequence training using a state-level minimum Bayes risk (sMBR) criterion [43]. The trigram language model with interpolated Kneser-Ney smoothing is trained using the SRILM toolkit [44] which provides a perplexity of 256 on the development data spoken by Frisian speakers.

We adopt two performance measures to quantify the recognition performance of the ASR system, namely the Word Error Rate (WER) and Code-Switching WER (CS-WER). The latter performance measure is the ratio of the number of erroneously recognized switched words to the total number of switched words.

Table 1: Word error rates in % obtained on the Frisian part of the development and test sets

	Devel		Test	
Acoustic Models	WER(%)	CS-WER(%)	WER(%)	CS-WER(%)
GMM-HMM+MFCC	54.0	92.5	51.2	90.0
GMM-HMM+LDA-MLLT	50.7	91.6	48.7	89.0
GMM-HMM+SAT	46.8	88.6	43.8	87.8
SGMM	43.8	88.3	39.9	86.8
DNN (6, 2048)+FBANK	42.5	**86.2**	39.4	86.3
DNN (6, 2048)+FMLLR	40.5	86.4	37.8	**85.3**
DNN-SMBR (6, 2048)+FBANK	40.9	86.6	37.8	86.5
DNN-SMBR (6, 2048)+FMLLR	**39.1**	87.1	**36.8**	85.8

5. Baseline Recognition Results

We perform ASR experiments using the Frisian resources described in Section 3 and the recognition results obtained on the development and test sets are presented in Table 1. For each column, the best results are marked in bold. These results can be considered as baseline monolingual results as only Frisian resources are used in the recognition experiments.

The conventional GMM-HMM trained on mel frequency cepstral coefficients (MFCC) provides a WER of 54.0% and a CS-WER of 92.5% on the development set and a WER of 51.2% and a CS-WER of 90.0% on the test set. The very high CS-WERs are expected in this setting as most of the code-switched Dutch words are not included in the resources. The correctly recognized switched words are the small amount of the Dutch words which appear in the provided Frisian lexicon. As the results on the development and test set follow a similar pattern, we will only discuss the results obtained on the test set.

Using discriminatively trained features (Linear discriminant analysis-maximum likelihood linear transform (LDA-MLLT)) and the speaker information by applying speaker adaptive training (SAT) reduces the WER to 43.8% and the CS-WER to 87.8%. The SGMM-based acoustic models which are known to perform well under limited training data scenarios [45] further improves the recognition accuracy by providing a WER of 39.9% and a CS-WER of 86.8%.

The best performing DNN-based system is trained using the FMLLR features by applying sequence training with SMBR criterion. This system provides a WER of 36.8% and a CS-WER of 85.8%. From these results, it can be seen that the CS-WER values do not reduce in parallel with the general recognition accuracy and reaches a lower limit between 85%-86% due to the use of monolingual resources.

These results are provided to serve as the baseline recognition results for follow-up research into ASR of the Frisian language. Relevant research topics include multiple research fields such as acoustic and language modeling in code-switching ASR systems, code-switching detection and development of multilingual ASR systems in the context of under-resourced languages.

6. Conclusion

In this paper, we have described several open source speech and language resources for the Frisian language. This data comprises of a speech database containing more than 10 hours of Frisian speech and 4 hours of Dutch speech, a trigram language model trained on a text corpus containing over 38M words and a phonetic dictionary containing more than 70k words. Moreover, an ASR recipe using these resources is provided and the baseline recognition results have been presented. In the future, we aim to extend the data collection by including also a speaker diarization/recognition experimental setup using the same speech database. Considering the longitudinal character of this database, the extension is going to enable the research of speaker tracking and diarization over a large time period and speaker aging effects on speaker recognition systems.

7. Acknowledgements

This research is funded by the NWO Project 314-99-119 (Frisian Audio Mining Enterprise). We would like to thank Maaike Andringa and Sigrid Kingma for meticulously annotating the radio broadcasts. We also would like to thank Frits van der Kuip, Hindrik Sijens, Derk Drukker and Lysbeth Jongbloed-Faber from Fryske Akademy, Luc de Vries and Teake Oppewal from Tresoar, Wim de Boer from Afûk and Maarten van Gompel from RU Nijmegen for their generous support during the Frisian text collection.

8. References

[1] P. Auer, *Code-switching in Conversation: Language, Interaction and Identity*. London, Routledge, 1998.

[2] P. C. Muysken, *Bilingual Speech: A Typology of Code-mixing*. Cambridge University Press, 2000.

[3] S. Thomason, *Language Contact - An Introduction*. Edinburgh University Press, 2001.

[4] G. Stemmer, E. Nöth, and H. Niemann, "Acoustic modeling of foreign words in a German speech recognition system," in *Proc. EUROSPEECH*, 2001, pp. 2745–2748.

[5] M. Luján-Mares, C.-D. Martínez-Hinarejos, and V. Alabau, *Human Language Technology. Challenges of the Information Society: Third Language and Technology Conference, LTC 2007, Poznan, Poland, October 5-7, 2007, Revised Selected Papers*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, ch. A Study on Bilingual Speech Recognition Involving a Minority Language, pp. 36–49.

[6] D.-C. Lyu, R.-Y. Lyu, Y.-C. Chiang, and C.-N. Hsu, "Speech recognition on code-switching among the Chinese dialects," in *Proc. ICASSP*, vol. 1, May 2006, pp. 1105–1108.

[7] N. T. Vu, D.-C. Lyu, J. Weiner, D. Telaar, T. Schlippe, F. Blaicher, E.-S. Chng, T. Schultz, and H. Li, "A first speech recognition system for Mandarin-English code-switch conversational speech," in *Proc. ICASSP*, March 2012, pp. 4889–4892.

[8] T. I. Modipa, M. H. Davel, and F. De Wet, "Implications of Sepedi/English code switching for ASR systems," in *Pattern Recognition Association of South Africa*, 2015, pp. 112–117.

[9] T. Lyudovyk and V. Pylypenko, "Code-switching speech recognition for closely related languages," in *Proc. SLTU*, May 2014, pp. 188–193.

[10] Y. Li and P. Fung, "Code switching language model with translation constraint for mixed language speech recognition," in *Proc. COLING*, Dec. 2012, pp. 1671–1680.

[11] H. Adel, N. Vu, F. Kraus, T. Schlippe, H. Li, and T. Schultz, "Recurrent neural network language modeling for code switching conversational speech," in *Proc. ICASSP*, May 2013, pp. 8411–8415.

[12] H. Adel, K. Kirchhoff, D. Telaar, N. T. Vu, T. Schlippe, and T. Schultz, "Features for factored language models for code-switching speech," in *Proc. SLTU*, May 2014, pp. 32–38.

[13] T. Hughes, K. Nakajima, L. Ha, A. Vasu, P. Moreno, and M. LeBeau, "Building transcribed speech corpora quickly and cheaply for many languages," in *Proc. INTERSPEECH*, 2010, pp. 1914–1917.

[14] I. Lane, A. Waibel, M. Eck, and K. Rottmann, "Tools for collecting speech corpora via Mechanical-Turk," in *Proc. NAACL HLT*, 2010, pp. 184–187.

[15] N. J. De Vries, M. H. Davel, J. Badenhorst, W. D. Basson, F. De Wet, E. Barnard, and A. De Waal, "A smartphone-based ASR data collection tool for under-resourced languages," *Speech Communication*, vol. 56, pp. 119 – 131, 2014.

[16] "The LIDES database," *International Journal of Bilingualism*, vol. 4, no. 2, pp. 209–246, 2000.

[17] V. B. Le, D. D. Tran, E. Castelli, L. Besacier, and J.-F. Serignat, "Spoken and written language resources for Vietnamese," in *Proc. LREC*, 2004, pp. 599–602.

[18] O. Salor, B. L. Pellom, T. Ciloglu, and M. Demirekler, "Turkish speech corpora and recognition tools developed by porting sonic: Towards multilingual speech recognition," *Comput. Speech & Lang.*, vol. 21, no. 4, pp. 580–593, Oct. 2007.

[19] T. Schultz, N. T. Vu, and T. Schlippe, "Globalphone: A multilingual text & speech database in 20 languages," in *Proc. ICASSP*, 2013, pp. 8126–8130.

[20] E. Barnard, M. H. Davel, C. Van Heerden, F. De Wet, and J. Badenhorst, "The NCHLT speech corpus of the South African languages," in *Proc. SLTU*, May 2014, pp. 194–200.

[21] R. Eiselen and M. J. Puttkammer, "Developing text resources for ten South African languages," in *Proc. LREC*, 2014, pp. 3698–3703.

[22] M. Gales, K. Knill, A. Ragni, and S. Rath, "Speech recognition and keyword spotting for low resource languages: Babel project research at cued," in *Proc. SLTU*, May 2014, pp. 16–23.

[23] B. Sene-Mongaba, "The making of lingala corpus: An under-resourced language and the internet," *Procedia - Social and Behavioral Sciences*, vol. 198, pp. 442–450, 2015.

[24] D.-C. Lyu, T.-P. Tan, E.-S. Chng, and H. Li, "Mandarin-English code-switching speech corpus in South-East Asia: SEAME," *Lang. Resour. Eval.*, vol. 49, no. 3, pp. 581–600, Sep. 2015.

[25] A. Dey and P. Fung, "A Hindi-English code-switching corpus." in *Proc. of the International Conference on Language Resources and Evaluation (LREC)*, 2014, pp. 2410–2413.

[26] J. Y. C. Chan, P. C. Ching, and T. Lee, "Development of a Cantonese-English code-mixing speech corpus," in *Proc. European Conference on Speech Communication and Technology*, 2005, pp. 1533–1536.

[27] D. Imseng, H. Bourlard, H. Caesar, P. N. Garner, G. Lecorv, and A. Nanchen, "Mediaparl: Bilingual mixed language accented speech database," in *IEEE Spoken Language Technology Workshop (SLT)*, Dec 2012, pp. 263–268.

[28] J. Weiner, N. T. Vu, D. Telaar, F. Metze, T. Schultz, D.-C. Lyu, E.-S. Chng, and H. Li, "Integration of language identification into a recognition system for spoken conversations containing code-switches," in *Proc. SLTU*, May 2012.

[29] D.-C. Lyu, E.-S. Chng, and H. Li, "Language diarization for code-switch conversational speech," in *Proc. ICASSP*, May 2013, pp. 7314–7318.

[30] Y.-L. Yeong and T.-P. Tan, "Language identification of code switching sentences and multilingual sentences of under-resourced languages by using multi structural word information," in *Proc. INTERSPEECH*, Sept. 2014, pp. 3052–3055.

[31] K. R. Mabokela, M. J. Manamela, and M. Manaileng, "Modeling code-switching speech on under-resourced languages for language identification," in *Proc. SLTU*, 2014, pp. 225–230.

[32] *Frysk Hânwurdboek*. Ljouwert/Leeuwarden: Fryske Akademy & Afûk, 2008.

[33] A. P. Versloot, "Mechanisms of language change. vowel reduction in 15th century West Frisian," Ph.D. dissertation, University of Groningen, 2008.

[34] Provinsje Fryslân, "De Fryske taalatlas 2015. De Fryske taal yn byld," 2015.

[35] G. Booij, *The phonology of Dutch*. Oxford University Press, 1995.

[36] J. Popkema, *Frisian Grammar: The Basics*. Afûk, Leeuwarden, 2013.

[37] P. Boersma and D. Weenink, *Praat: doing phonetics by computer, Version 5.4.12*, 2015. [Online]. Available: http://www.praat.org/

[38] E. Yılmaz, M. Andringa, S. Kingma, F. Van der Kuip, H. Van de Velde, F. Kampstra, J. Algra, H. Van den Heuvel, and D. Van Leeuwen, "A longitudinal radio broadcast in Frisian designed for code-switching research," in *Proc. LREC*, 2016, pp. 4666–4669.

[39] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely, "The Kaldi speech recognition toolkit," in *Proc. ASRU*, Dec. 2011.

[40] J. R. Novak, N. Minematsu, and K. Hirose, "Phonetisaurus: Exploring grapheme-to-phoneme conversion with joint n-gram models in the WFST framework," *Natural Language Engineering*, pp. 1–32, 9 2015.

[41] N. Oostdijk, "The spoken Dutch corpus: Overview and first evaluation," in *Proc. LREC*, 2000, pp. 886–894.

[42] D. Povey, L. Burget, M. Agarwal, P. Akyazi, K. Feng, A. Ghoshal, O. Glembek, N. K. Goel, M. Karafiát, A. Rastrow, R. C. Rose, P. Schwarz, and S. Thomas, "The subspace Gaussian mixture model - A structured model for speech recognition," *Computer Speech & Language*, vol. 25, no. 2, pp. 404–439, 2011.

[43] K. Vesely, A. Ghoshal, L. Burget, and D. Povey, "Sequence-discriminative training of deep neural networks," in *Proc. INTERSPEECH*, 2013, pp. 2345–2349.

[44] A. Stolcke, "SRILM – An extensible language modeling toolkit," in *Proc. ICSLP*, 2002, pp. 901–904.

[45] D. Imseng, P. Motlicek, H. Bourlard, and P. N. Garner, "Using out-of-language data to improve an under-resourced speech recognizer," *Speech Communication*, vol. 56, pp. 142 – 151, 2014.