Institute of Phonetic Sciences,
University of Amsterdam,
Proceedings 26 (2005), 1-5.

FRYSS: A FIRST STEP TOWARDS FRISIAN TTS[1]

Jelske Dijkstra, Louis C.W. Pols, R.J.J.H. van Son

Abstract

In this MA-project a Dutch TTS system based on Festival, NeXTeNS, has been changed step by step into a Frisian system. As for many minority languages, Frisian too has very few digital resources. So, the challenge of this project is to make the system as intelligible as possible with minimal resources. The resulting TTS system is called FRYSS. At the end of the thesis period an evaluation was carried out with native speakers of Frisian. The results show that the TTS system sounds reasonably intelligible. This project demonstrates that a TTS prototype can be built rapidly using existing modules and voices.

1 Introduction

With the rise of computerization and the increasing number of digital applications, it is important for every language, with large or small numbers of speakers, to take part in this process. Not having access to digital tools, for example to speech-related applications such as Text-to-Speech systems, can be a major setback for a language. After all, if a speaker is forced to use such a tool in another language, he might find himself looking down upon his mother tongue and perhaps even considering it unfit for the modern age. One could argue that this phenomenon affects a speaker's language attitude. On a higher level, it means exclusion from an important and ever growing language domain: the so-called digital domain. If these exclusions progress into other domains as well, the language will end up being used exclusively at home. Perhaps even worse, the language could die out — together with its valuable culture (Dijkstra et al., 2004).

In this MA-project, an existing Dutch TTS system called NeXTeNS, or "Nederlandse Extensie voor Tekst Naar Spraak" (Dutch Extension for Text to Speech), was chosen as a framework to create a Frisian TTS system. NeXTeNS (Marsi & Kerkhoff, 2003) is based on the multi-lingual TTS of Festival (Black et al., 1999), which is built up in modules. Each module deals with a certain language aspect, such as for example sentence accent, pause breaks or sentence melody. In this way, it is relatively easy to change the system step by step into another language, such as for example Frisian, viz. the West-Frisian variant, spoken in the province of Fryslân in the north of the Netherlands. As holds true for many other minority languages, Frisian has few or no digital resources. Consequently, the main challenge

[1] Full thesis is available from: http://www.fon.hum.uva.nl/IFA-publications/Others/Other_papers.html

lies in making the system as intelligible as possible. The result is called FRYSS: “Fryske Spraak Synteze” (Frisian Speech Synthesis).

2 Frisian

Fig. 1. (Part of) the Netherlands. Gray area: (West-)Frisian speaking region. (Versloot carthography 1996 in: Visser, 1997)

Of the Dutch province of Fryslân, 94% of the inhabitants are able to understand the language, 74% are able to speak Frisian, 65% are able to read Frisian and 17% are able to write in Frisian (Gorter & Jonkman, 1995). In addition, 55% of the inhabitants of Fryslân have learned Frisian as their mother tongue, which roughly boils down to 350,000 native speakers (Gorter, 2003). Over the last decades, we have seen a small decline in people's ability to speak Frisian.

Despite of the fact that there is a variety of Frisian dialects, all variants are mutually comprehensible.

3 From NeXTeNS to FRYSS

Through the system of NeXTeNS it will be explained briefly how TTS works and what was done to change it into the Frisian system of FRYSS.

3.1 Tokenization

The first step - after loading the text or utterance into the TTS system - was scanning it for tokens such as numbers or abbreviations. These were replaced by the written words. Since NeXTeNS only contained a standard tokenization file, the language specific material had to be replaced by Frisian. In an earlier pilot project a copy of the number-to-word conversion for the Spanish *el*-voice was adapted to Frisian. This adaptation was also inserted in the Frisian tokenization file. Now the system could pronounce positive integer numbers and abbreviations in Frisian.

3.2 Sentence accents and pauses

Since there is no Part-of-Speech tagger for Frisian, accents and breaks had to be assigned in another way. Sentence accents were placed on content words. Every word was run through a self-made list of Frisian function words. This list was based on function words from a Frisian grammar (Tiersma, 1999) and a Dutch function word

list (Quené & Kager, 1990). If the word was not on this list, it was treated as a content word, in order for it to get a sentence accent.

For the placement of pauses the default option of the punctuation cart tree (Black et al., 1999; Black & Lenzo, 2003) was chosen. So wherever commas, dots, colons, etc. occurred, these punctuation marks were replaced by a pause.

3.3 Grapheme-to-phoneme conversion

After that, it was time for the actual transformation into phoneme strings: the grapheme-to-phoneme conversion. Every word was run through a self-made pronunciation lexicon. This lexicon was based on a pre-final digital version of the "Frysk Hânwurdboek"-dictionary from the Fryske Akademy. The pronunciation lexicon was created with a computer script. It contains well over 63,000 words together with their pronunciation and information about lexical stress. If a word from the synthesized text was not available in the pronunciation lexicon, it was built by letter-to-sound rules. These rules have been written for Frisian manually, since their spelling is quite phonetic (Dijkstra et al., 2004).

Unfortunately, an evaluation test of these rules carried out at the end of the thesis period, yielded disappointing results: only 28% of the 1,000 randomly chosen words from the lexicon had been built correctly by the letter-to-sound rules.

3.4 Duration and sentence melody

One of the last steps in the synthesis process was adding duration and sentence melody to the phoneme strings. The duration file of NeXTeNS was used to define the duration of every phoneme. This file also included special rules that can shorten or lengthen the default duration, e.g., shortening in a consonant cluster.

As for sentence melody, only little research has been carried out for Frisian. Most grammars assume the Frisian intonation to be the same as in Dutch (Cohen et al., 1961; Tiersma, 1999). One of the few studies on Frisian intonation has been done by Hoekstra (1991) who concentrates on sentence accents. He claims that lexical and specific functional prepositions are more frequently stressed in Frisian than in Dutch, and less than in English. Because of the so-called similar intonation structure in Dutch and Frisian, the Dutch intonation of ToDI (see also Gussenhoven et al., 2003; Gussenhoven, 2004) was used. This intonation structure was already implemented in NeXTeNS.

3.5 Using Dutch phoneme set to create Frisian phones

Subsequently, the sounds from the phoneme strings were activated in the phoneme set of the Dutch voice, and the text or utterance was pronounced. As there was insufficient time to construct a Frisian phoneme set, it was decided to use the Dutch phoneme set instead. However, Frisian contains several sounds that do not exist in Dutch, so for these sounds a creative solution had to be found. All examples below are stated in Worldbet annotation, unless stated otherwise.

The long Frisian vowels were created by doubling duration: for example doubling the duration of [u:] in "lûd" [lu:t] (sound, noise).

The nasal vowels, also absent in the Dutch set, were replaced by their non-nasal variants plus the sound [n], e.g., [u:_~] in “jûns” [ju:_~s] (in the evening) was replaced by [u:] and [n].

The Frisian diphthongs, e.g., [i&] in “iepen” [i&p&n] (open), were mapped to the components they consist of, namely [i] and [&].

The triphthong [U&_i], e.g., “moai” [mU&_i] (beautiful), was mapped to the Dutch diphthong [oi] (SAMPA-annotation).

4 Evaluation

At the end of the thesis period, an evaluation was carried out with 35 native speakers of Frisian over the internet. It has to be mentioned that most of the respondents were personal contacts. They had to judge 20 sentences (randomly chosen sentences from Frisian internet sites) on six aspects: intelligibility, general quality, naturalness, lexical stress, sentence melody and speaking rate. They judged these aspects on a 5-point scale where “1” stands for bad and “5” for good, or, in case of speaking rate, the lowest score stands for slow and the highest for rapid. The stimuli contained features where synthesis would go wrong, e.g., derived forms of stems (e.g., plurals), wrong placement of (default) lexical stress, diphthongs, nasalized vowels, the feature of breaking, etc. In breaking, vowel change takes place in derived forms of the stem, cf. “doar” [dU&r] (door) versus “doarren” [dwAr&n] (doors) and “doarke” [dwArk&] (small door); “hier” [hi&r] (hair) versus “hierren” [jIr&n] (hairs) and “hierke” [jIrk&] (small hair); “foet” [fu&t] (foot) versus “fuotten” [fwUt&n] (feet) and “fuotsje” [fwUtsj&] (small foot); “beam” [bI&m] (tree) versus “beammen” [bjEm&n] (trees) and “beamke” [bjEmk&] (small tree) (Tiersma, 1999). Breaking is a feature which cannot always be gathered from spelling.

The stimulus sentences varied between 7 and 21 words. The results are showns in Figure 2.

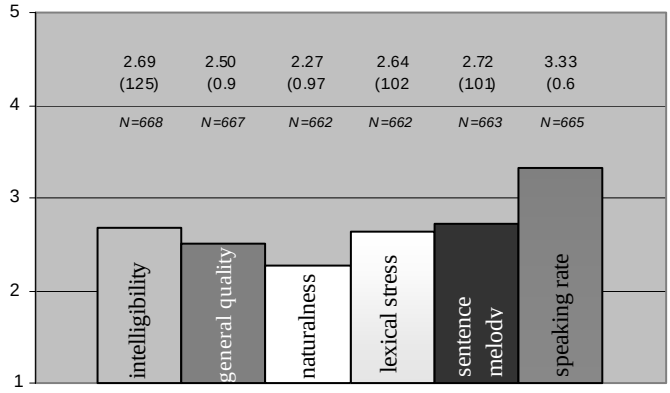

Fig. 2. Mean judgments, standard deviation (between parentheses), and number of responses. Judgments on a 5 point scale; higher is better. For speaking rate higher is more rapid.

The mean scores of intelligibility, general quality, lexical stress and sentence melody are judged below the centre value of 3. Naturalness was judged lowest. Low scores are probably due to features as diphthongs (where components were just knocked together), mistakes in pronunciation, and wrong accent placement, sometimes of

lexical stress, sometimes of sentence accents. The speaking rate over all stimuli was considered normal, or sometimes slightly too fast.

All scores are above minimal (1) and, looking at the scores per utterance of all aspects, except speaking rate, some judgments have even the maximum score of 5. This leaves potential for improvement. The system is intelligible, although far from ideal.

5 Conclusions

In conclusion, one can say that it is possible to construct a reasonably intelligible TTS system for a new language with little digital material. This case study can be considered a challenging example for other minority languages. With the help of this thesis, one can estimate the costs and time needed to develop such a system. In addition, although the system is far from ideal, the initiative can act as a stimulus for further pioneering projects: for instance, research concerning the construction of a phoneme set for Frisian, or research into Frisian sentence melody – to mention just a few options.

Acknowledgments

We would like to thank Erwin Marsi and Joop Kerkhoff for their advice and patience when answering our questions about NeXTeNS. We would also like to thank Hugo Quené for providing us with a Dutch function word list, and Hindrik Sijens, Willem Visser, David Weenink and Paul Boersma for their advice. Further we would like to express our gratitude to the Fryske Akademy for the opportunity to work with a pre-final publication of the Frysk Hânwurdboek.

References

Black, A.W., & Lenzo, K.A. (2003): *Building Synthetic Voices*, FestVox 2.0 Edition.

Black, A.W., Taylor, P. & Caley, R. (1999): *The Festival Speech Synthesis System*, Edition 1.4, for Festival Version 1.4.0.

Cohen, A., Ebeling, C.L., Fokkema, K. & Holk, A.G.F. van (1961): *Fonologie van het Nederlands en het Fries. Inleiding tot de moderne klankleer*. Martinus Nijhoff, 's Gravenhage, 2nd edition (1st edition 1959).

Dijkstra, J., Pols, L.C.W., Son, R.J.J.H. van (2004): Frisian TTS, an example of bootstrapping TTS for minority languages. In: *Proceedings SSW5*, Pittsburgh PA, USA.

Gorter, D. & Jonkman, R.J. (1995): *Taal yn Fryslân op 'e nij besjoen*, Fryske Akademy, Ljouwert.

Gorter, D. (2003): Nederlands en Fries op gespannen voet? In: Stroop, J. (ed.) *Waar gaat het Nederlands naar toe?* Uitgeverij Ben Bakker, Amsterdam.

Gussenhoven, C., Rietveld, T., Kerkhoff, J. & Terken, J. (2003): *Transcription of Dutch Intonation, courseware*, http://todi.let.kun.nl/ToDI/home.htm (1st edition 1999).

Gussenhoven, C. (2004): Transcription of Dutch Intonation. In: Sun-Ah Jun (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, Oxford: Oxford University Press (in press).

Hoekstra, J. (1991): Oer it beklamjen fan ferhâldingswurden yn it Frysk, it Hollânsk en it Ingelsk. In: *Us Wurk*, Volume 40, p. 67-103, Fryske Akademy, Ljouwert.

Marsi, E.C. & Kerkhoff, J. (2003): *NeXTeNS*, http://nextens.uvt.nl.

Quené, H. & Kager, R.W.J. (1990): *PROS*, Research Institute for Language and Speech, Utrecht

Tiersma, P.M. (1999): *Frisian Reference Grammar*, Fryske Akademy, Ljouwert, 2nd edition (1st edition 1985, published by Dordrecht Foris Publications).

Visser, W. (1997): *The syllable in Frisian*. HIL Dissertations 30, HAG, The Hague.