A Speech Corpus for Dyslexic Reading Training

Jakob Schou Pedersen, Lars Bo Larsen
Multimedia Information and Signal Processing section
Aalborg University, 9220 Denmark
E-mail: jsp@es.aau.dk, lbl@es.aau.dk

Abstract

Traditional Danish reading training for dyslexic readers typically involves the presence of a professional reading therapist for guidance, advice and evaluation. Allowing dyslexic readers to train their reading skills on their own could not only benefit the dyslexics themselves in terms of increased flexibility but could also allow professional therapists to increase the amount of dyslexic readers to whom they have a professional contact. It is envisioned that an automated reading training tool operating on the basis of ASR could provide dyslexic users with such independence. However, only limited experience in handling dyslexic input (in Danish) by a speech recognizer exists currently. This paper reports on the establishment of a speech corpus of Danish dyslexic speech along with an annotation hereof and the setup of a proof-of-concept training tool allowing dyslexic users to improve their reading skills on their own. Despite relatively limited ASR performance, a usability evaluation by dyslexic users shows an unconditional belief in the fairness of the system and indicates furthermore willingness for using such a training tool.

1 Introduction

This paper deals with the investigation of dyslexic speech in relation to ASR performance. The ultimate goal is to establish a fully automated reading tutor for dyslexic users, allowing them to improve their reading skills on their own without the need for supervised therapeutic intervention. No such system is commercially available today as most research on reading trainers is targeted towards children or second language learning systems. The main focus of the paper is on the recording and analysis of a database of speech read by dyslexics.

Dyslexic spoken input is quite different from non-dyslexic spoken input to a speech recognizer, rendering a need for a specialized ASR configuration taking this particular type of input into consideration, (Pedersen, 2008). In addition, a suitable (i.e. in terms of applied pedagogical and cognitive principles) training framework is needed in order for a fully automated training setup to become feasible for this particular group of users, (Pedersen, 2008) and (Pedersen et al., 2008).
For what concerns the former need, existing research shows that reading behaviour of dyslexics highly deviates from that of non-dyslexics readers in several aspects, (Høien and Lundberg, 2004; Gupta, 2004; Gregor and Newel, 2000):

- **Frequent regressions**.
  Dyslexics often experience irregular eye movements leading to frequent regressions within the text being read.
- **Difficulties reading long/complex words**.
  Impairments of short-term memory can cause dyslexics to experience difficulties in reading long or syntactically complex texts.
- **Mixing up of similar letters/words**.
  Dyslexics having difficulties perceiving and remembering letters may mix these up as well as words of similar form.
- **Punctuation difficulties**.
  Dyslexics often experience difficulties with recognition of punctuation.
- **Slow decoding**.
  Dyslexics often read out textual information at a very low pace.
- **Problems with long words**.
  Dyslexics often experience problems reading long words. Words misread are furthermore often constituted by correct initial syllables.
- **Addition and omission of words**.
  Dyslexics often add or remove words from texts when reading these.

In order to obtain a more detailed impression of characteristics potentially having an impact on ASR performance a corpus of dyslexic speech has been recorded from eight dyslexic readers (3 males and 5 females, mean age 25 years) prompted with written material, selected and approved by professional therapists. All eight readers were also selected by professional therapists ensuring a representation ranging from mildly dyslexic to severely dyslexic. In order to include representations of varying difficulties for all the readers, the prompting texts were of varying difficulty (achieved by adjusting font-size, word-length, word frequency and line-spacing).

All prompting texts were presented to the readers on individual sheets of paper in increasing order of difficulty in order to allow for natural breaks in between them when reading out aloud for the recording of the corpus. On occurrences of words that were completely illegible to a reader an operator assisted with these by pronouncing them for the reader to repeat.

2 Database Design

A total of approximately 100 minutes (including silence and filled pauses) has been recorded (7578 words in total). All audio has been recorded in a silent environment using an uncompressed 16 bit PCM WAV-format at a sample rate of 22050 Hz. The dyslexic readers were all equipped

with a Plantronics DSP-100 close-talking headset connected to a PC running the freeware OS-independent tool “Audacity” for the recording process, (Mazzoni et al., 2005).
Based on an initial listening to the recordings, a pragmatic approach has been taken for annotating these in a suitable manner. The characteristics shown in Table 1 have been annotated as these were assumed not only to be present in the corpus, but also to deviate significantly from normal reading behavior.

Word omission / deletion	Word insertion
Word substitution	Word splitting
Abnormal pausing	Temporal inconsistency

Table 1: Dyslexia-related characteristics annotated in the corpus.

Compared to traditional annotation of audio this task has been somewhat more complex as miscues also needed to be annotated, thereby necessitating more than one layer of annotation in order to establish links between connected audio events (e.g. word substitutions and the words that were supposed to be read). Occasionally the annotator has needed to rely on acoustic similarities for making these links (i.e. in situations of ambiguity).
The orthographic part of the annotation resembles a traditional annotation task as e.g. for the “SpeechDat II” corpus (SpeechDat, 1999) where recordings of non-dyslexic readers were annotated at an orthographic level. Therefore, at an orthographic level, the annotation conventions from the SpeechDat project are reused.

Combined, this has lead to the annotation structure shown in Table 2. In order to support this layered structure, the open-source tool “Praat” (Boersma and Weenink, 2005) has been used for annotating the recordings.

Layer	Content
‘Ortho’	Orthographic annotation (at word level)
‘Error’	Type of error (e.g. insertion, substitution, etc.)
‘Prompt’	Word intended to be read in case of errors
‘Timing Event’	Temporal inconsistencies (i.e. jumps or pauses in the text)

Table 2: Corpus annotation layers.

3 Corpus Analyses

Analyzing the recordings, a number of characteristics clearly deviating from normal reading behaviour have become evident.
At an overall level, there is (as we expected) a high frequency of misreadings: in average 21% of all words were misread by each speaker. The speaking rate ranges from 38.67 to 115 words per minute with an average speaking rate of 76.86 words per minute. Misreadings are furthermore strongly negatively correlated with the speaking rate (r=-0.84), which is consistent with the hypothesis that a decrease in uncertainty leads to an increase in oral reading rate, (Brubaker, 1972).

3.1 Regression and progression

Regressions occur quite frequently in the readings (in average approximately 100 times per speaker): singular regressions (i.e. re-reading the last word read) are the most frequent type of such temporal inconsistency. Not only incorrectly read words are corrected in this manner – correctly read words often also get repeated – and sometimes even replaced by incorrect words.
Singular progressions (i.e. skipping a single word) occur relatively infrequently in the recordings (in average approximately 12 times per speaker) but appear nevertheless to be positively (although weakly) correlated with word length (r=0.47). This clearly deviates from normal reading behaviour where smaller words are much more likely to be skipped than larger words, (Brysbaert and Vitu, 1998).

3.2 Word length impact

The frequency of misread words and the length of these (measured in number of characters) appears to be positively correlated (r=0.52). In fact, if leaving out words longer than 15 letters (each prompted less than 10 times in the prompting texts) a strong positive correlation exists between word length and misreading frequency (r=0.86). This is consistent with previous findings of children with developmental dyslexia having significant difficulties naming pictures illustrating items with long names, (Swan and Goswami, 1997).

3.3 Abnormal pausing

In total, approximately 40% of the recordings consist of pauses (silent as well as filled). The registered pauses ($\bar{x} = 0.73$ seconds, $\sigma = 0.56$ seconds) indicate that this type of pausing during readout is substantially larger than for normal readers:

In (Brubaker, 1972) an average pause duration $\bar{x}$ of 0.36 seconds and a standard deviation σ of 0.10 seconds is reported for non-dyslexic young readers.
In (Horii and Ramig, 1987) $\bar{x} = 0.46$ seconds and $\sigma = 0.10$ seconds are reported for non-dyslexic readers with a mean age of 28.1 years.
In (Hammen and Yorkston, 1996) $\bar{x} = 0.46$ seconds and $\sigma = 0.17$ seconds are reported for elderly non-dyslexic readers.

Pauses of abnormal duration often emerge immediately before the occurrence of reading mistakes and can thus often be seen as an indication of such upcoming difficulties.

3.4 Summing up

The conducted corpus analyses have explicitly validated numerous of the findings reported by (Høien and Lundberg, 2004), (Gupta, 2004) and (Gregor and Newel, 2000):

- **Frequent regressions**
  A relatively high amount of regressions have been registered.
- **Slow reading**
  A large amount of abnormally long pauses (silent as well as filled) have been registered, which has lead to a slow overall reading process.
- **Problems with long words**
  Word length and misreading frequency have been found to be positively correlated.

In addition, the analyses have provided new insights in the reading behaviour of dyslexics:

- **Word length effect on singular progression frequency**
  A weak positive correlation (r=0.47) between word length and word skipping frequency has been registered. This deviates clearly from normal reading behaviour.
- **Pause duration**
  The registered average pause duration ($\bar{x} = 0.73$ seconds, $\sigma = 0.56$ seconds) is substantially longer than for normal reading.

4 Exploitation

Taking into account the typical characteristics of dyslexic speech, as outlined in chapter 3, a proof-of-concept ASR-based recognition setup for detecting misreadings in dyslexic read speech has been established, (Pedersen, 2008). This is based on a dual-decoder Goodness Of Pronunciation (GOP)-scoring framework described in (Witt, 1999) and illustrated in Figure 1.

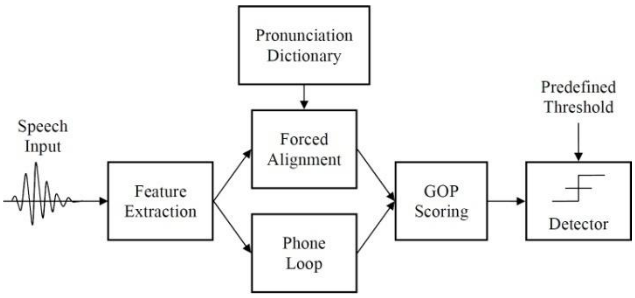

Figure 1: Forced alignment setup, redrawn from (Witt, 1999).

Using multiple sensory stimulation, the user of the system is prompted visually with written text on a computer screen and auditorily via pre-recorded speech through loudspeakers. Following this the user is to read out the same text out aloud into a headset microphone. Any inconsistencies with the prompted text are then to be detected via the recognition setup shown in Figure 1 and pinpointed visually to the user.

In order to accommodate for diverging learning styles, users of the system can select between a set of different assistance forms such as e.g. pre-recorded speech for those who prefer an auditory learning style or visual cues for those who prefer visual learning style.

The ASR configuration has been implemented using the CMU Sphinx IV open source speech recognition engine, (Lamere et al., 2004). In order to cope with the aforementioned characteristics of the spoken input to the system, the finite-state grammar shown in Figure 2 has been used for the forced alignment part in Figure 1.
This grammar is based on the text used to prompt users of the system. Any inconsistencies (e.g. reading mistakes or regressions) with the spoken input from the user are handled by means of filler models ("FM" in Figure 2) and the possibility of skipping words (although associated with a low probability).

Seen from a pedagogical point of view, occasionally accepting wrong input from the user (false acceptance – FA) is more productive than being overly strict and thus falsely rejecting correct input (false rejection – FR). Focus has therefore been on reducing the amount of FRs. This comes however at the expense of decreasing the amount of correct rejections (CR).

In a test of the final system by 15 dyslexic participants (6 females and 9 males, mean age 26.2 years) the results shown in Table 3 were obtained, where the optimal ASR configuration is shown in terms of the ability to actually spot reading mistakes (miscue detection rate – MDR) vs. the tendency to make false rejections (false alarm rate – FAR).

Utterance	ASR	Performance
Incorrect	CR: 4% FA: 6%	MDR = CR/(CR+FA) = 50.07%
Correct	CA: 85.16% FR: 4.8%	FAR = FR/(FR+CA) = 8.77%

Table 3: ASR performance results.

In other words, while erroneously rejecting approximately 9% of correct input, the setup is capable of correctly detecting half of the misreadings by the user – the remaining 50% are (erroneously) accepted by the system.

This will not necessarily constitute a problem, as dyslexic users may actually not explicitly notice this. In fact, in a usability evaluation (n=16) of the proof-of-concept setup 80% of the subjects expressed an unconditional belief in the fairness of the evaluation done by the system, and envisioned furthermore that they would like to use such a system. In total, each participant read out 306 words divided into 23 individual sentences.

5 Conclusions and Further work

We have investigated the characteristics of dyslexic read speech in the context of designing an ASR-based reading training tool. As a step towards this we have recorded a speech corpus and verified how dyslexic reading behaviour differs from normal readers in terms of

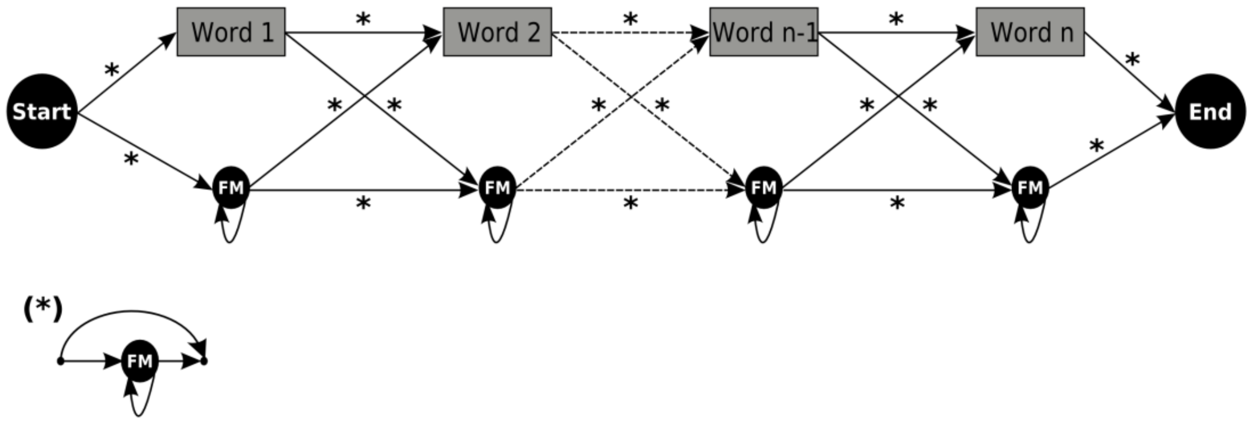

Figure 2: Finite-state grammar used in the forced alignment configuration, (Pedersen, 2008).

pausing. In addition, we have found a high number of regressions and an abnormal correlation between skips and word length. The database has been used in the configuration of an ASR system and to verify the performance. User tests with dyslexic readers have been carried out on the developed reading trainer and have shown that the design and performance are accepted by the target user group. Further improvement and research into dyslexic ASR is currently being carried out and a commercial version of the reading trainer is under consideration.

For questions regarding availability of the annotated corpus please contact the authors.

6 Acknowledgements

The authors would like to thank the professional therapists at the Institute of Language, Speech and Brain Disorders, The County of North Denmark for their kind assistance and guidance. Additionally, the authors would like to thank all dyslexic participants who have participated in the recording of the corpus as well as in subsequent evaluations of the proof-of-concept setup.

7 References

Boersma, P. and Weenink, D. (2005). Praat: doing phonetics by computer. http://www.praat.org.

Brubaker, R. S. (1972). “Rate and pause characteristics of oral reading” Journal of Psycholinguistic Research, 1:141-147.

Brysbaert, M. and Vitu, F. (1998). “Word skipping: Implications for theories of eye movement control in reading. Eye guidance in reading and scene perception”, pages 125-147

Gregor, P. and Newel, A. F. (2000). “An empirical investigation of ways in which some of the problems encountered by some dyslexics may be alleviated using computer techniques”. In Proceedings of the Fourth International ACM Conference On Assistive Technologies, pages 85-91. ACM Press.

Gupta, A. (2004). “Reading difficulties of Hindi-speaking children with developmental dyslexia”. Reading and Writing, 17:79-99.

Hammen, V. L. and Yorkston, K. M. (1996). “Speech and pause characteristics following speech rate reduction in hypokinetic dysarthria”. Journal of Communication Disorders, 29(6):429-445.

Høien, T. and Lundberg, I. (2004). “Dysleksi”. Special-pædagogisk forlag, Herning. ISBN: 87-7399-280-1.

Horii, Y. and Ramig, P. R. (1987). “Pause and utterance durations and fundamental frequency characteristics of repeated oral readings by stutterers and nonstutterers”. Journal of Fluency Disorders, 12:257-270.

Lamere, P., Kwok, P., Gouvêa, E. B., Raj, B., Singh, R., Walker, W., andWolf, P. (2004). ”The CMU sphinx-4 speech recognition System”.http://research.sun.com/sunlabsday/docs.2004/sphinx4.pdf.

Mazzoni, D., Brubeck, M., and Haberman, J. (2005). Audacity: Free audio editor and recorder. http://audacity.sourceforge.net.

Pedersen, J. S. (2008). ”User Centred Design of a Multimodal Reading Training System for Dyslexics”. Ph.d. thesis, Aalborg University. ISBN: 87-92078-57-5.

Pedern, J. S., Larsen L. B. and Lindberg B. (2008). “Usability of ASR-based reading training for dyslexics”, in Iterspeech-2008, 2799-2802.

SpeechDat (1999). “The Danish Speechdat(II) corpus – a spoken language resource”. In Bente Maegaard, C. P. and Wedekind, J., editors, Datalingvistisk Foreningsårsmøde 1999 I København. Proceedings. CST Working Papers, volume Report No. 3, pages 15-32. Center for sprogteknologi (CST).

Swan, D. and Goswami, U. (1997). “Picture naming dificits in developmental dyslexia: The phonological representations hypothesis”. Brain And Language, 56:334-353.

Witt, S. (1999). “Use of speech recognition in computer-assisted language learning”. Ph.d. thesis, Cambridge University Engineering Department, Cambridge,UK. http://mi.eng.cam.ac.uk/reports/svr-ftp/auto-pdf/witt_thesis.pdf.