Usability of ASR-based Reading Training for Dyslexics

Jakob Schou Pedersen, Lars Bo Larsen and Børge Lindberg

Department of Electronic Systems, Aalborg University, Aalborg, Denmark

{jsp,lbl,bli}@es.aau.dk

Abstract

As an initial step towards self-training of dyslexics, this paper presents the overall design- and usability evaluation results of a simulated reading training system for dyslexics that uses ASR and multi modal presentation techniques as core components. Based on an analysis of dyslexic reading behaviour as well as extensive usability evaluations involving actual dyslexic test subjects it is indicated that current ASR performance in fact suffices for dealing with dyslectic input to an automated training system when the special phenomena present in dyslexic speech are taken into account. Finally, explicit design guidelines for such a system are derived experimentally aiming at assuring a high level of perceived usability when used by dyslectic users.

Index Terms: Dyslexia, multi modality, ASR, WOZ.

1. Introduction

Dyslexia is one of a number of known language disorders that deteriorates reading skills. Dyslexia literally affects millions of citizens all over the world. In Denmark it is estimated that 2-5% [1] of the population has significant difficulties in reading and writing primarily due to dyslexia. As functioning in modern society heavily relies on the capabilities of text processing, dyslexia constitutes a problem of increasing magnitude.

While assistive technologies such as screen readers can aid dyslexics in their everyday life these also burden the users with the demand for carrying around special devices and applying them when having to decipher textual information. Training, which increases the dyslexics’ readings skills to read on their own, is however a resource demanding task that requires skilled personnel and time. It is therefore desirable if this training, or part of it, can be transferred into an automated supplementary tool that can be used by dyslexics on their own.

As many traditional training techniques basically rely on spoken interaction between a dyslexic user and a therapist, this solution calls for the deployment of ASR technology. Despite significant improvements within the area of spoken language technology, most technological developments for dyslexics have so far been constituted by relatively simple combinations of off-the-shelf language technology products and existing tools (e.g. combining regular word processing software with commercial speech-recognition or -synthesis systems, [2]). While such a solution may be applicable for dictation purposes this is not necessarily the case for training purposes as it does not take into account the highly irregular way dyslexics typically read (i.e. heavy usage of filler words, abnormal pausing, restarts, etc.) and the full potential of speech technology is not reached. A dedicated recognition scheme targeted explicitly towards usage by dyslexics seems needed in order to obtain optimal performance when trying to establish an automated variant of reading training for dyslexics. Research in dedication of speech recognition towards dyslectic users has previously shown that such explicit targeting of an ASR system is possible, [3]. This paper presents the Wizard of Oz (WOZ) prototyping results of combining Danish speech recognition dedicated towards usage by dyslexics with traditional pedagogical approaches, into an automated reading training tool, [4]. The dyslexic target group is constituted by Danish adults suffering from developmental dyslexia.

2. Training Scheme

Through interviews with speech therapists working with dyslexics on a daily basis, it was concluded that the widely used pedagogic technique “Book and Tape” [5] illustrated in Figure 1 is applicable as a pedagogical approach in the prototype system.

Figure 1 Outline of the ’Book and Tape’ training technique.

Using this technique dyslexics listen to text-segments being read out aloud (i.e. using a tape recorder) while simultaneously reading the same text from a book. After this, the pupil reads out the same text-segment aloud and progresses to a new segment of text. Parallel to this, a speech therapist listens and corrects any errors and evaluates the performance. Thus, the overall purpose of this technique is to train pronunciation and textual decoding capabilities by the use of multiple sensory channel stimulation.

The idea, presented in this paper, is to transfer this technique into an automated solution in which the performance evaluation is handled via ASR and assistance is provided via multimodal presentation techniques (i.e. adapting to the principle of multiple sensory channel stimulation).

The idea, presented in this paper, is to transfer this technique into an automated solution in which the performance evaluation is handled via ASR and assistance is provided via multimodal presentation techniques (i.e. adapting to the principle of multiple sensory channel stimulation).

3. User Centred Design

The “Book and Tape” scheme is adapted into an automated ASR-based system as outlined in Figure 2, which then does not require the presence of a speech therapist.
As shown in Fig. 2, the tasks handled by a speech therapist are now handled automatically via ASR and (multimodal) assistance. The explicit way of doing this is identified by applying a user centered design (UCD) process followed by an elaborate usability testing of different scenarios through which the general preferences of the target group can be estimated. Thus, the following schemes of feedback, assistance and evaluation (encompassed by steps b, d and e in Figure 2 respectively) are evaluated as these are potentially of vital importance to the perceived usability of such a system.

3.1. Feedback

The purpose of the feedback functionality is to act as a progress indicator to the user, allowing him/her to see what has been read so far, and whether or not this was correctly read. Two overall feedback issues are evaluated: format and timing i.e. how and when to provide feedback.

3.1.1. Format

Words read out correctly by the user need to be marked as such. This marking can be done in a number of ways. As a pragmatic approach, six different formats are evaluated for this purpose: striking out text, font reduction, text removal, text graying, text re-coloring and text underlining.

3.1.2. Timing

The feedback can be provided using one of two timing schemes. It can be presented at a ‘word level’ meaning that while the user reads out words this is reflected, word by word. Feedback can however also be presented at a ‘sentence level’ meaning that all feedback is provided successive to the readout of sentences or other blocks of text.

3.2. Assistance

When the user is incapable of reading out a particular word being prompted on the computer screen assistance functionality may be required.

3.2.1. Initiation (i.e. timing)

Three different types of assistance initiation are evaluated. The first is ‘on-demand’ assistance, where the user explicitly has to request assistance when needed. The second is ‘automatic’ assistance, where the system automatically assesses where and when assistance is needed and provides this accordingly. The third option is ‘multiple-initiation’ where both the system and the user can trigger the provision of assistance.

3.2.2. Format

Whenever there is a need for assistance with a particular word, this needs to be presented in a pedagogical sound way to the user in order to be effective. As a pragmatic approach, five different forms of assistance are evaluated: pre-recorded speech, visual cueing (e.g. a picture of the word), contextual switching (i.e. the word in another context), morphological re-phrasing (i.e. the morphological root of the word) and subdivision of words (e.g. into syllables).

3.3. Evaluation

Naturally, the purpose of using an automated training system is for the user to become better at reading. An evaluation of the user performance is therefore necessary not only for evaluating the individual ‘tasks’ of the training process but also for motivating the user.

3.3.1. Timing

The evaluation of user performance can be provided using one of two timing schemes: continuous or successive. The former presents measures of evaluation (i.e. statistical data) simultaneously to the user reading out words whereas the latter waits for the user to finish a text segment before presenting any evaluation of this.

3.3.2. Format

The performance evaluation needs to be communicated to the user in a comprehensible manner, taking into account the obvious reading difficulties of the user. As a pragmatic approach three different evaluation forms are evaluated: binary word mark-up (words are either green or red), gradual word mark-up (words can be a shade of red depending on the rate of misreading, or green) and numerical outline (numerical data as well as graphical icons).

4. Analysis of Dyslexic Reading Behaviour

In order to investigate the potential ASR performance when dealing with dyslexic input, the reading behaviour of the target group was initially investigated by analyses of recorded reading sessions. These were carried out in a controlled environment by eight subjects - ranging from ‘mildly’ to ‘severely’ dyslexic (as classified by a professional speech therapist). Ten text segments were selected and prepared in close collaboration with speech therapists. In order to comply with the individual reading difficulties of the subjects, the textual complexity of these segments increases from ‘easy’ (text 1) to ‘difficult’ (text 10) by adjustment of the following parameters:

- Word- and sentence length
- Word frequency (i.e. inclusion of rare words)
- Line-spacing and font-size

In total this has resulted in approximately 100 minutes of recorded speech (including intra-word “silence”). Introducing an extended version of the SpeechDat transcription standard [6] as a multi level transcription approach in which traditional orthographic annotations are combined with annotations of
errors, intended words and timing events (i.e. regressions and progressions) these recordings have been manually annotated and analysed.

4.1. Reading Characteristics

Based on analyses of the recorded speech corpus, the following summarized characteristics appear to be descriptive for the target group:

1. Frequent regression/progression.
2. Frequent abnormal intra/inter word pausing.
3. Frequent filled pauses (e.g. ‘eh.’ and ‘ehm.’).
4. Positive correlation between word length and misreading frequency.
5. Frequent correct beginnings of misread words.
6. Variable voice level – the level is often lowered noticeably before encountering difficult words.

Obviously, this deviates substantially from the input normally expected for a speech recognizer and can thus be presumed to have a negative impact on the performance for e.g. a standard commercial ASR dictation engine. A series of offline experiments, using the Sphinx IV open source speech recognition engine [7], have therefore been carried out using the recordings of the target group as training input. [8]. In these experiments a goodness of pronunciation (GOP) approach has been adopted from [9] comparing the output score of a restricted forced alignment decoder to the output of a free phoneme loop decoder. A special grammar targeted towards the typical reading characteristics of dyslexics, has however been developed. Seen from a pedagogical point of view, occasionally accepting wrong input as being correct is more productive than being overly strict. Therefore, focusing primarily on reducing the amount of false rejections as main objective, the parameters shown in Table 1 have been achieved as optimal performance.

<table>
<thead>
<tr>
<th>Word</th>
<th>ASR</th>
<th>Error rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>CA: 85.16% FR: 4.8%</td>
<td>FRRate = FR/(FR+CA) = 5.34%</td>
</tr>
<tr>
<td>Incorrect</td>
<td>CR: 4% FA: 6%</td>
<td>FARate = FA/(FA+CR) = 60%</td>
</tr>
</tbody>
</table>

Table 1 Current optimal ASR performance.

As can be seen, there is still room for improvement in terms of ASR performance: 90% of the words are correctly read and of these 85% are recognized as such (Correct Acceptance – CA). Almost 5% of the correctly read words are falsely rejected (False Rejection – FR). For what concerns incorrectly read words (10%), only 4% of these are identified as being wrong (Correct Rejection – CR) whereas 6% of these are falsely accepted (False Acceptance – FA).

At the time of writing, a real-time implementation of this ASR scheme has not been integrated into the training tool. However, utilizing a WOZ simulation strategy, usability experiments emulating the obtained ASR performance have been conducted in order to investigate if they in fact suffice in terms of perceived usability.

5. Materials and Methods

A prototype tool illustrated in Figure 3 has been developed and used for the experiments.

![Figure 3 Screenshot from the prototype tool.](image)

Similar to the “Book and Tape” approach, the user reads the text after being prompted with a text segment visually and audibly and the system responds with an indication of correct/incorrect words and offers assistance if needed. An integrated administrative interface furthermore allows the test operator to easily switch between the applied assistance schemes, feedback forms, etc. described in chapter 3, to be evaluated during the usability experiments.

The ASR part of the prototype is (transparently to the test subjects) simulated by a test operator (i.e. ‘Wizard’). Thus, all input from the test subjects is evaluated in real-time by the test operator and passed to a simulation proxy with the purpose of artificially introducing ASR errors to emulate the ASR performance shown in Table 1 (i.e. intentionally inducing FRs and FAs). The options listed in chapter 3 have been tested in a semi-balanced manner (within-subjects approach, n=16) in a quiet environment including only a test operator (i.e. ‘Wizard’) and a single test subject at a time. Again, subjects have been selected to range from ‘mildly’ (1) to ‘severely’ (5) dyslexic (estimated by a professional speech therapist) as shown in Figure 4.

![Figure 4 Subject demographics.](image)

6. Test Results

In general, all 16 subjects were quite pleased with the prototype tool and the potential of facilitating self-training by this. Only a single subject felt strange about ‘talking to a computer’ and requesting help from this while the remaining subjects all felt comfortable doing this (none of the subjects...
were aware of the WOZ setup). As shown in Figure 5 relatively heavy usage of the assistance functionalities occurred during the test.

![Figure 5 Misreadings and assistance requests during the test.](image)

From a functional point of view, the following preferences were registered via post-test oral interviews (boldfaced options represent results of statistical significance):

### 6.1. Feedback

Word-level feedback was preferred as timing strategy for receiving feedback, and most subjects preferred the recolouring of text as feedback format. For what concerns the timing of the feedback, the word-level preference appeared to be more distinct for the more severely dyslexic subjects.

### 6.2. Assistance

The multiple-initiation assistance scheme was preferred by most subjects, and when providing assistance this should preferably be done using pre-recorded speech as assistance format. For what concerns the timing of the assistance, the automatic schemes (‘automatic’ and ‘multiple-initiation’) appeared to be preferred by the more severely dyslexic subjects.

### 6.3. Evaluation

Most subjects preferred the successive evaluation scheme, and preferred furthermore to be presented with a numerical outline of their performance. As one subject put it “...this evaluation form allows me to easily compare with previous scores.”

### 7. Discussion

None of the subjects complained about difficulties using the system. They appeared not to notice the applied WOZ strategy and believed they were interacting with a fully automated system, although this was never implied to them. In fact, unintentionally, the system was implemented as being much too strict, as a factor 10 too few FAs were generated.

## 8. Conclusions and future work

Despite the application of too strict testing conditions it is concluded that current state-of-the-art ASR technology can be used as a core component within an automated reading training system for dyslexics, if the special phenomena present in dyslexic speech are taken into account.

Interactive behaviour of both the ASR component and the user is likely to be affected by practical conditions such as precision and sensitivity of the speech detection, recognition delay and sensitivity to environmental noise. More usability evaluations are therefore needed in order to determine if an actual integration of ASR into the automated training system will have an impact on the perceived usability.

Finally, while the evaluation results indicate that the target group (i.e. dyslexics) are obviously capable and prepared to accept using an automated reading training system this does not necessarily mean that they will actually benefit from this in terms of reading proficiency. Thus, the long term effects of such a tool remains undisclosed and needs to be evaluated.

### 9. Acknowledgements

For great help and assistance throughout the work partially documented in this paper, the authors would like to thank the Institute of Language, Speech and Brain Disorders, County of North Denmark (especially Anders Sigh, Thomas Bo Christensen and Birgit Trolle Hansen) as well as the dyslexic subjects that have participated in the design and evaluation of the prototype system. Finally, the authors would like to thank Morten Højfeldt Rasmussen for his involvement in the ASR experiments.

### 10. References