The Reading Assistant: Eye Gaze Triggered Auditory Prompting for Reading Remediation

John L. Sibert and Mehmet Gokturk
Department of Computer Science
The George Washington University
Washington D.C. 20052, USA
Tel: 1-202-994-4953
E-mail: *sibert/gokturk@seas.gwu.edu*

Robert A. Lavine
Department of Physiology and
Experimental Medicine
The George Washington University
Washington D.C. 20052, USA
Tel: 1-202-994-3550
E-mail: phyral@gwumc.edu

ABSTRACT
We have developed a system for remedial reading instruction that uses visually controlled auditory prompting to help the user with recognition and pronunciation of words. Our underlying hypothesis is that the relatively unobtrusive assistance rendered by such a system will be more effective than previous computer aided approaches. We present a description of the design and implementation of our system and discuss a controlled study that we undertook to evaluate the usability of the Reading Assistant.

KEYWORDS: eye tracking, eye gaze, reading disability, interaction techniques

INTRODUCTION

In recent years the computer has become a key classroom tool for the remediation of reading disabilities. To assist the reader, multimedia educational software is available that will allow text, displayed on a computer screen, to be sequentially highlighted and spoken by the computer. Existing software either "reads aloud" with the student reading along, or requires the student to request assistance; usually by mouse selection. We feel that an ideal computer-based remediation tool would allow the student to concentrate on the reading task assisted by automated computer response as necessary. This is borne out by anecdotal evidence. Remedial reading teachers have indicated to us that an effective technique for helping children with reading problems is for the teacher to read in a very low voice, or whisper, with a slight delay as the student reads aloud.

In this paper, we present the Reading Assistant; a tool to evaluate the effectiveness of visually activated prompting in the context of a remedial reading program. Potential beneficiaries include the estimated 10 million children with dyslexia in the 50,000 school districts in the U.S., as well as large numbers of adults with learning disabilities.

BACKGROUND

Research using eye movement tracking technologies can be characterized as interaction technique development [1,3,14] or observational research (e.g. in physiology or psychology) [2,8,10,12,14]. In our research we are doing both. We developed the GWGazer system as a test bed for experiments relating eye movements during display scanning to predicting user alertness and attentiveness [4,5]. The Reading Assistant is an extension of GWGazer that tracks the reader's eye movements and, using principles derived from reading research, aids the reader by pronouncing words that appear to be hard to recognize.

UIST '00. San Diego, CA USA

Eye movements in reading research

Eye movement during reading has been investigated for many years, and ranges of values for various eye movement measures have been established for normal readers [12]. Increased text difficulty imposes a greater burden on cognitive processing capacity that is reflected in increased length of fixation and other features of eye movement. During adult reading, eyes move in a sequence of pauses or fixations separated by rapid, intermittent movements or saccades (see Figure 1). The saccades are generally rightward or progressive, but sometimes may be leftward or regressive. Variables that have been measured include number of fixations, duration of fixations, location of fixations in the line of text, number of progressive saccades, number of regressive saccades, and size of saccades [10,11,12,15].

The average progressive saccade encompasses about 7-9 character spaces, equivalent to about 2 degrees of visual angle [10,12], which overlaps the estimated extent of the foveal region of the retina. The typical regressive saccade size is smaller, about 3-4 letters. The number of regressive saccades is about 10-20% of all saccades in skilled readers. These regressions are believed to take place when there is difficulty understanding the text, misinterpretation of the text, or overshooting of the target area. The average fixation duration for skilled readers is reported to be about 250 msec [10,12], but there are great individual differences. For a specific reader, fixation duration may range from 100 to over 500 msec.

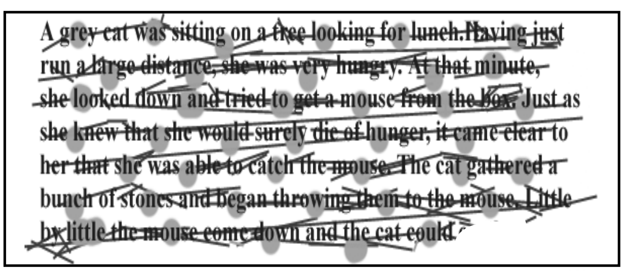


Figure 1: Typical reading pattern. The gray circles represent fixations while the lines represent saccades.

Experiments have explored the effect of fixation position within a word on fixation duration (before the following forward saccade) [10]. The duration of a fixation is correlated with reading difficulty, so that words that are more difficult typically require longer fixations for their identifications. The length of saccades may be influenced by "crude visual clues" [10]. For example, the saccades entering or leaving longer words tend to be longer.

Marked developmental trends in eye movement during reading by children also have been described [12,15]. With increasing grade level in elementary school, there is a decrease in mean fixation duration and an increase in mean saccade length. According to McConkie and Zola [8] "...research involving eye movement monitoring can help in understanding the nature of the mental processes involved in reading, how these develop as one learns to read, and what processing strategies or characteristics are more common in those children who fail to show normal progress in learning to read. Eye movement data are useful in analyzing simultaneously collected data, such as ...oral reading protocols... (and) ...can be used for controlling experimental manipulations during ongoing reading."

Remediation with mouse-activated prompting

Previous research [6,9] has shown substantial improvement in word-identification skills in children with reading disabilities using mouse-activated prompting. Reading from desktop computers with synthetic speech prompting of difficult words has improved timed word recognition in an experimental reading remediation program [18]. Subjects were children in grades 3 to 6, reading at or below the 10th percentile. During daily 25 minute periods for a total of about 10 hours, each student read from the monitor and used the mouse to highlight and then pronounce difficult words or sub word components. Results showed a significant improvement by the experimental group for timed word and non-word recognition. This remedial program is an example of "Perhaps the most promising of all computer-assisted aids for reading acquisition..." [16].

In a more recent publication [19], the basic deficit in dyslexia, the most common reading disorder, is described as word identification associated with impairment in phoneme awareness (segmenting and manipulating sounds within a syllable) and phonological decoding (translating print to sound). After children highlighted a word with the mouse, the word was pronounced as a whole, in syllables, or in segments within each syllable. Improvements were similar "whether the computer speech assistance for 'targeted' words was in whole or segmented form".

Although such mouse-activated acoustic prompting appears to be effective, the need to indicate difficult words by clicking on them with a mouse requires precise eye-hand coordination and adds a significant delay. Furthermore, the user may not always be aware immediately of when he or she is having trouble identifying a word. In addition, the use of the mouse requires the user to interrupt the cognitive process of oral reading and understanding by interjecting a manual task. If this interruption were not necessary, the natural learning process could proceed more naturally and may be enhanced. A technique in which prompting is continuously adapted to the reader's eye movement, eliminates these interrupting factors allowing the natural learning process to proceed more smoothly

THE GWGazer READING ASSISTANT

The Reading Assistant is a visually activated, interactive reading program for use with multimedia PC systems. The system uses eye tracking to trigger synthetic speech feedback as children read text from the monitor. The system takes advantage of (1) the ability of unobtrusive eye tracking systems to follow eye motion and (2) the ability of text-to-speech software to help children learn to read. As students read text displayed on a computer screen, a video camera, mounted below the screen, monitors the students' eye motions. The eye tracking system analyzes the infrared video image of the eye and computes the coordinates of the gaze-point (the point at which the eye is looking) on the screen and sends them to the GWGazer application that we have developed for our research. This application keeps track of the user’s scan of the displayed text in real time. Visual and auditory cues are produced that depend on where the student is looking and whether changes in scan pattern indicate difficulties in identifying a word.

The Reading Assistant is a straightforward extension of the GWGazer system (Figure 2) which we have been developing as part of our eye tracking research program[4,5,13]. GWGazer includes an EYEGAZE ENGINE which communicates with the eye tracker (in this case an Eyegaze Development System from LC Technologies that is running on a second PC) and provides most of the functionality for gaze point tracking and data capture and analysis.

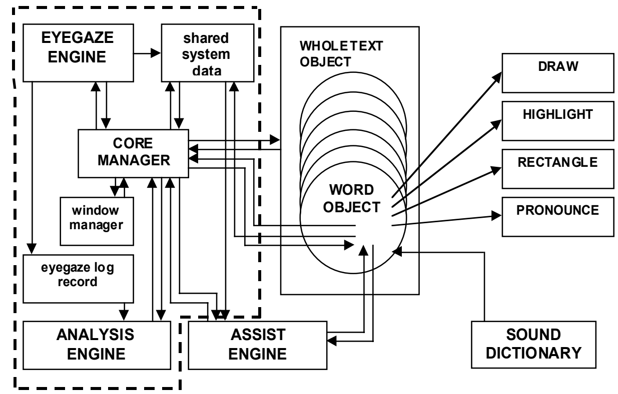


Figure 2: GWGazer Reading Assistant. The dashed line incorporates GWGazer.

The Reading Assistant extension includes an ASSIST ENGINE that determines when to highlight and pronounce a word using the algorithm described below in the section on fixations and dwell time. Words are represented as word objects with methods for drawing, highlighting, indicating their rectangular extent, and pronouncing themselves. A text object is an array of lines where each line is an array of words.

A preprocessing step allows any standard text file to be displayed on the screen in font sizes ranging from 12 pt to 96 pt. The text is displayed in dark blue on a pure white background. The researcher can interactively change the color of the text as well as line and word separation.

Digitized speech corresponding to each word in the text sample is stored in the form of sound files in a sound dictionary. The dictionary currently consists of a directory containing sound files in wave format (*.wav), sampled at 4khz and 8 bit resolution. For example, the word “hello” has a corresponding “hello.wav” file. Each word object contains a reference to its associated sound file. We chose this approach over voice synthesis because of its simplicity and ability to provide tight control over each word, and to provide more natural speech. In English, multiple meanings and pronunciations can be associated with the same spelling of a word. For example, "It is not appropriate to appropriate someone else's paper." Our approach stores two different word objects with the same name but different pronunciations in the array representing this sentence.

Gaze point data is used by the "ASSIST ENGINE" to trigger highlighting and pronunciation of words. When the reader gazes at a particular word longer than a predetermined duration, the word is first highlighted and then spoken by the system.

The system collects eye gaze data during reading at 30Hz or 60Hz sampling rates, depending on which version of the Eyegaze system we use. The data is saved in a special format for post processing and analysis.

In our preliminary studies, all words had the same duration threshold. However, since threshold levels are embedded within the word object, each word is able to have a different threshold, related to the difficulty and frequency of the word in the language.

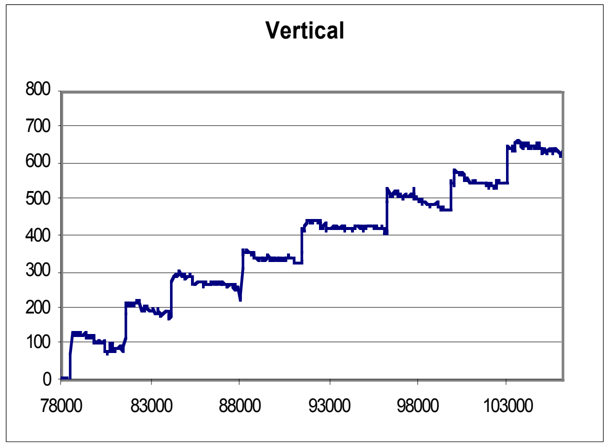
WORD SELECTION ALGORITHM

Determining gaze duration on a single word presents some difficulties. We can only approximate the exact gaze-point. The accuracy of the system is about 0.25 inch on the screen when viewed from a normal reading distance (20 inches). If subjects move around, the gaze-point can drift. Also there is a difference between the physiological gaze-point and the location of letters that are perceived and processed cognitively. For example, a "perceptual span" of approximately 6-7 letters has been found between the gaze-point and the location of cognitive interest [10]. This tendency for readers to "gaze ahead" suggests the need for rapid selection of the correct word so that it can be pronounced before the reader becomes aware of the time lag.

Because of these difficulties, the system must use additional information in order to determine the exact word that the reader is focusing on at any given moment. We try to predict the word of interest by taking advantage of the sequential nature of the reading task.

Previous research has shown that a reader tends to fixate on a word or group of words in sequence while reading a line of text. These visual fixations are separated by small horizontal saccadic eye movements. This pattern is illustrated in figure 3. At the end of each line of text, the reader's gaze "flies back" in a horizontal saccade to the beginning of the next line.

Word sequence determines the next word to be highlighted and pronounced. Words within a line of text are only highlighted and pronounced in left to right order. The current line of text is determined by differentiating between lines using horizontal fly backs that are observable after reading one complete line. The system looks for horizontal fly backs before attempting to change the current line due to vertical variations in eye gaze. If the reader glances briefly away from the text we are less likely to incorrectly deduce a line change.



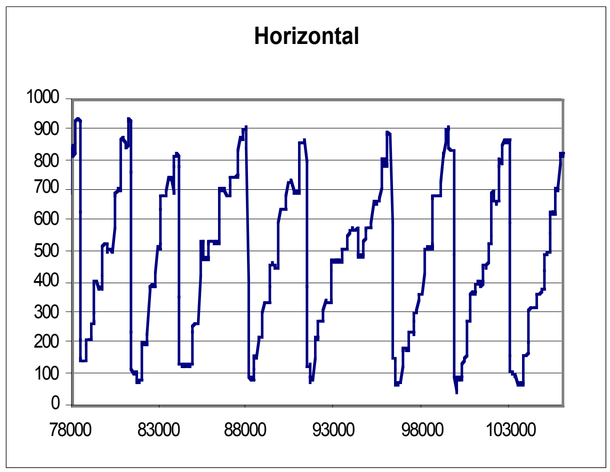


Figure 3: Typical eye gaze movement during reading a text consisting of eight lines.

We have also observed a tendency for users to skip the correct line and briefly fixate on a lower line immediately following a fly back to the left margin of the text. The following word fixations usually return to the correct line. Because the reading task requires sequential reading of the text, we constrain our gaze point track such that the next line is selected even when the Y position drifts down to a lower line. In practice this has turned out to work quite well.

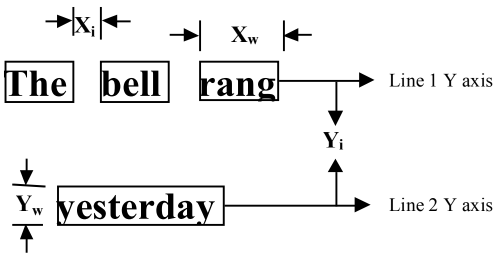


Figure 4: Descriptive parameters for text

Various size and spacing parameters are embedded in word objects and distances can be set interactively by the experimenter (Figure 4):

X_i : Distance between two word objects. This distance is equal for all words and modifiable through the options dialog. The researcher can set the desired word distance according to the experiment.

Y_i : Distance between lines of text. This distance is also modifiable. However, it should be greater than the maximum word height so that lines do not overlap.

X_w: The width of a word object. This value is embedded within each word object.

Y_w: The height of the word object. Similar to the width value, this value is also contained within the word object.

(X_r,Y_r) : The coordinate of the top left corner of the word object's bounding box in screen space. This value is contained in the word object as well. Therefore each word knows where to draw itself on the screen.

Fixations and Dwell

We define a fixation as a sequence of gaze-points (samples) located within a given distance of each other over at least 100 milliseconds duration. Because fixation duration is frequently insufficient for the perceptual and cognitive processing necessary for word recognition, more than one fixation may occur while reading a word. Therefore, we define "Dwell" as a series of one or more fixations separated by short saccadic motions. Now we can define thresholds of Dwell within a given word's bounding rectangle:

Dwell Treshold1: A lower threshold signifying that a particular word is the focus of attention. We use Dwell Treshold1 as a trigger to visually highlight the word.

Dwell Treshold2: A higher threshold suggesting that the subject is having difficulty recognizing the word. If Dwell Treshold2 is exceeded the word is pronounced by the system.

READING AND EYE MOVEMENT OBSERVATIONS

In our initial testing we felt that the system was promising although it would sometimes miss words or even pronounce the wrong word. We were interested to see how well it would perform under more rigorous testing so we designed a pilot study to help us determine the usability of the system. A detailed description of the study and its results may be found in [13].

Subjects were 8 children, aged 10-14, whose parents provided informed consent. Four of them were fifth grade students with a history of reading problems and receiving Learning Disabilities (LD) services. Reading passages used were obtained from the Gray Oral Reading Test, Third Edition [17], each was presented twice in immediate succession to observe effects of familiarization and practice. Trials were video and audio taped to help us localize reading errors.

The study was run using a PC system with MS-Windows 95 and 16" monitor (resolution 1024 X 768). Viewing distance was approximately 20 inches. After repeated trials in order to optimize system performance, parameters were set as follows:

Font Type: Times New Roman
Font Size: 42 point
Line Separation: 70 pixels
Word Distance: 14 pixels
Horizontal Margin: 30 pixels
Vertical Margin: 60 pixels
Dwell Treshold1: 240 msec
Dwell threshold2: 360 msec

Results included recorded samples of simultaneous visual scan and oral reading of text passages that can be replayed and analyzed. Data was obtained on reading speed and accuracy, fixations per second and per word, fixation duration, reading errors, and acoustic prompts provided by the system. Some examples of the data follow.

Table 1 compares performance between the first and second trials on both tasks for subjects 1-4 who were reading at or above grade level.

	5th grade text		10th grade text	
	I	II	I	II
time(s)	37.0	30.0	84.8	68.6
words/s	2.9	3.6	1.3	2.2
errors	0.4	0.0	4.8	3.0
fixations	122.6	112.4	254.8	240.4
fixns/sec	3.3	3.7	3.0	3.5
fixns/word	1.2	1.1	1.7	1.6
fix dur (ms)	231.4	228.4	232.0	241.0
prompts	1.2	0.4	9.0	3.6

Table 1: Trials I and II for subjects 1-4.

As we expected, there was a learning effect. From trial 1 to trial 2, all subjects showed increased reading speed, more fixations/second, and shorter fixation duration. When we compare the results of the first and second presentation of the reading task, practice on this system was associated with improved reading performance. Subjects increased their reading speed on the fifth-grade paragraph from 2.91 to 3.60 words/sec, and from 3.31 to 3.75 fixations/sec. They increased on the tenth-grade paragraph from 1.28 to 2.24 words/sec, and from 3.00 to 3.50 fixations/sec. Most children read the fifth-grade text without errors even the first time. On the tenth-grade text their errors decreased from 5 to 3 on the second presentation.

We can also compare the number of oral reading errors made with the number of prompts provided by the system. Notice that, for the 10th grade text, the ratio of prompts to reading errors was about two to one for the first trial but dropped almost to one to one for the second. This may indicate that, as students become more acquainted with either the system or the text or both, the prompting becomes more efficient.

We used a questionnaire to determine the subjective reactions of the students. In general, they liked the system and found it easy to use and unobtrusive. Interestingly, the most obtrusive part of the system was the video camera.

CONCLUSIONS AND FUTURE DIRECTIONS

Motivated by a desire to improve reading instruction for children, and adults, with reading disability we have implemented a system which uses a reader's visual scanning pattern of the text to identify, and pronounce, words that the reader is having difficulty recognizing. Our pilot study encourages us to believe that our approach is feasible. Although there were some glitches, we were able to run eight young subjects successfully.

We plan a series of parametric studies to investigate the effects of our settable parameters on system usability. For example, in our pilot study, we used the same dwell time threshold for all of our subjects. However we know that the optimum threshold varies with individuals, as well as with the level of difficulty of the reading material. We hope to develop a method for establishing these parameters for individuals at differing levels of difficulty

Of course our pilot study does not establish the effectiveness of the system. In order to do that, we need to carry out a long-term controlled trial with students using the system regularly in their classroom. We have in fact proposed such a study and hope to be able to carry it out within the next few years.

ACKNOWLEDGEMENTS

We gratefully acknowledge the financial support of the Jet Propulsion Laboratory, the technical support of LC Technologies, and the all around support, suggestions, and comments of the HCI research group at GWU.

REFERENCES

1. Bolt, R. A. (1982). "Eyes at the Interface," Proc. ACM Human Factors in Computer Systems Conference, pp. 360-362.

2. Chapman, J.E., Lavine, R.A., Cleveland, D., Joubert, M., and Fontana, J. (1997). Effects of non-optical color absorptive lenses on eye movement patterns during reading and on low contrast vision. 4th World Congress on Dyslexia, Halkidiki, Greece, September 23-26.

3. Jacob, R. J. K. (1993). "Eye-Movement Based Human-Computer Interaction Techniques: Toward Non-Command Interfaces," in H. R. Hartson and D. Hix (eds) Advances in Human-Computer Interaction, Vol 4. Ablex Publishing Co., Norwood, NJ.

4. Lavine, R.A., Dickens, B., McNamara, J., Sibert, J., Gokturk, M., and Heckman, K.(1999). Eye-tracking in a visual vigilance task. In Jensen, R.S., Callister, J.D., Cox, B., and Lavis, R. (eds.), Proceedings of the Tenth International Symposium on Aviation Psychology. Columbus, Ohio, Ohio State University, v.2., 903-909.

5. Lavine, R.A., Sibert, J., and Gokturk, M. (1999). Visual scanning accuracy in a vigilance task: Application to driver alertness monitoring. In Carroll, R., ed., Ocular Measures of Driver Alertness: Technical Conference Proceedings. Herndon, Virginia: Federal Highway Administration and National Highway Traffic Safety Administration, US DOT, 108-117.

6. Lundberg, I. (1995). The computer as a tool of remediation in the education of students with reading disabilities--a theory-based approach. Learning Disability Quarterly 18, 89-100.

7. Lyon, G.R. (1995). Toward a definition of dyslexia. Annals of Dyslexia, 45: 3-27.

8. McConkie, G. W. and Zola, D. (1986). Eye Movement Techniques in Studying Differences Among Developing Readers. Technical Report No. 377. Bolt, Beranek and Newman, Inc. Cambridge, Mass; Illinois University Center for the Study of Reading.

9. Olson, R., Forsberg H., Wise, B., and Rack, J. (1994). Measurement of word recognition, orthographic, and phonological skills. Lyon, G.R., ed., op.cit., 243-268.

10. O'Regan, J.K. (1990). Eye movements and reading. In Kowler, E., ed., Eye movements and their role in visual and cognitive processes. Elsevier.

11. Patchberg, J. P., and Yomas, A. (1978). The Effects of the Reader's Skill and the Difficulty of the Text on the Perceptual Span in Reading. Journal of Experimental Psychology: Human Perception and Performance 4: 545-552.

12. Rayner, K. (1978). Eye movements in reading and information processing. Psychological Bulletin 85: 618-660.

13. Sibert, J., Gokturk, M. and Lavine, R Reading Remediation with Eye Gaze Initiated Auditory Feedback Submitted to ACM CHI2001.

14. Sibert, L. and Jacob, R.J.K.(2000) Evaluation of eye gaze interaction, ACM Press, Proceedings of CHI 2000, 281-88.

15. Taylor, S. E. (1965). Eye Movements in Reading: Facts and Fallacies. American Educational Research Journal 2: 187-202.

16. Torgeson, J.K. and Barker, T.A. (1995). Computers as aids in the prevention and remediation of reading disabilities. Learning Disability Quarterly 18, 76-88.

17. Wiederholt, J.L. and Bryant, B.R. (1992). Gray Oral Reading Tests: Third Edition. Austin, Texas: Pro-Ed.

18. Wise, B., Olson, R., Anstett, M., Andrews, L., Terjak, M., Schneider, V., Kostuch, J., and Kriho,L. (1989). Implementing a long-term computerized remedial reading program with synthetic speech feedback: Hardware, software, and real-world issues. Behavior Research Methods, Instruments, & Computers 21: 173-180.

19. Wise, B.W. and Olson, R.K. (in press), Computer-based phonological awareness and reading instruction. Annals of Dyslexia, 50, p. 95.