

A system for tracking gaze on handheld devices

KRISTIAN LUKANDER

Finnish Institute of Occupational Health, Helsinki, Finland

Many of the current gaze-tracking systems require that a subject's head be stabilized and that the interface be fixed to a table. This article describes a prototype system for tracking gaze on the screen of mobile, handheld devices. The proposed system frees the user and the interface from previous constraints, allowing natural freedom of movement within the operational envelope of the system. The method is software-based, and integrates a commercial eye-tracking device (EyeLink I) with a magnetic positional tracking device (Polhemus FASTRAK). The evaluation of the system shows that it is capable of producing valid data with adequate accuracy.

In natural environments, eye movements are made toward task-relevant targets even when high spatial resolution is not required. This set of eye movements, directing visual attention without conscious intervention, can reveal attentional mechanisms and provide a window into cognition. Thus, monitoring observers' eye movements during a task can provide better understanding of visual perception (Pelz & Canosa, 2001).

The study of eye movements has been utilized widely in multiple areas, including vision and oculomotor research (Carpenter & Robson, 1999), cognitive and psychological research (Hyönä, Radach, & Deubel, 2003; Näsänen, Ojanpää, & Kojo, 2001), reading research in healthy and dyslexic subjects (Eden, Stein, Wood, & Wood, 1994; Rayner & Pollatsek, 1989), providing assistance for the disabled (Barea, Boquete, Mazo, López, & Bergasa, 2000; Majaranta & Riihämä, 2002), and, at least to some extent, usability research (Crosby & Peterson, 1991; Stanford Poynter Project, 2000).

The term "eye movements" is commonly used to describe ocular movements used to fixate targets in the environment. This article uses the term "gaze tracking," defined as tracking the point of view currently fixated on a target surface.

In spite of the inherent complexity of visual perception, the majority of past studies have been carried out with subjects performing relatively simple tasks under constrained laboratory conditions. So far, researchers have been content with the study of eye movements in isolation, excluding head movements. This has been inspired, in part, by a reductionist attitude, but has been dictated, even more, by the equipment available (Pelz & Canosa, 2001).

A number of eye-tracking devices exist on the market. An old, but still accurate review of the techniques used in eye tracking can be found in Young and Sheena (1975).

A comprehensive list of currently available commercial eyetrackers can be obtained from Oyekoya (2004). Many of the current systems impose strict restrictions on the freedom of movement of the subject and the interface, typically resulting in nonrealistic use scenarios and studies of gaze and eye movements without natural head, hand, and body movements.

Eye-tracking devices can be roughly divided into two categories: head-mounted devices and remote devices. Head-mounted devices have the sensory element, typically video camera(s), attached to a helmet or a headband worn by the user, whereas remote systems use remote cameras fixed near the user. Remote systems track the user's gaze and report gaze position calibrated on a planar, stationary stimulus surface, typically a computer monitor. Head-mounted devices come in two varieties: Devices with scene video have a camera pointing forward, delivering roughly the same view of the world as the user sees. These devices report the point of gaze as an overlaid cursor on the scene video. This gives more freedom to the user, because the gaze is not recorded in relation to any fixed coordinate system, but this type of device also requires laborious, subjective evaluation of the test sessions. Other head-mounted devices have a method of compensating for a limited amount of head movement in relation to the target stimulus surface, reporting the point of gaze as coordinates on that surface.

Although some manufacturers—for example, Tobii (www.tobii.se), SmartEye (www.smarteye.se), and SeeingMachines (www.seeingmachines.com.au)—offer equipment that allows for moderate subject movement by using model-based tracking of the subject and the subject's eyes, these systems are unable to track a moving user interface device or stimulus. In addition, with the current configurations and camera technology, the accuracy of these systems varies between 1° and 5° of visual angle.

Applied Science Labs (www.a-s-l.com) offers a somewhat similar approach with their mobile systems. These systems can also be equipped with a magnetic tracker to measure the gaze in relation to predefined areas in 3-D space, such as monitors and the surrounding environment.

This project was funded by the National Technology Agency of Finland (Tekes). Correspondence concerning this article should be addressed to K. Lukander, Brain Work Research Center, Finnish Institute of Occupational Health, Topeliuksenkatu 41 a A, FIN-00250 Helsinki, Finland (e-mail: kristian.lukander@ttl.fi).

However, these systems require the use of a heavy helmet that carries the camera, and lack the capability to track moving areas of interest.

From a methodological point of view, some of the most similar approaches (in comparison with the system presented in this article) have been made in the research field of gaze tracking in virtual environments. These include the work of Duchowski et al. (2000) and Christou (2001). Also related is the work of Babcock and Pelz (2004) and Pelz and Canosa (2001), who have built a lightweight eye-tracking system to track eye movements in complex tasks and natural environments. However, their system records eye movements as an overlaid pointer on a scene video, and therefore provides data only for subjective evaluation.

Small interfaces have been described as the blind spot of academic research—likely, in part, due to the lack of proper research equipment (Kuutti, 1999). The typically small viewing area of the moving display, differences in input devices, and changing lighting conditions affect user performance and user experience on mobile devices. The rapid increase in the number of mobile devices with complex user interfaces—used both in work situations and during free time—makes further research in this area necessary. The proposed system is an attempt to develop appropriate equipment with which to carry out this research.

The present article describes a prototype of a novel research system developed for tracking the gaze point of a subject using a mobile, handheld device, without placing restrictions on the natural movements of the subject.

Structure of the System

The system is implemented as software, and integrates a head-mounted video-oculography device with a magnetic positional tracker. The current implementation uses a head-mounted video-oculography device (EyeLink I, SR Research, Mississauga, ON) for measuring binocular gaze direction, and a magnetic positional tracker (Polhemus FASTRAK, Polhemus, Colchester, VT) to measure the position and orientation of two magnetic sensors attached to the head frame and the screen frame (hereafter referred to as the “head sensor” and “screen sensor”).

The current system components were chosen from equipment readily available to the laboratory. Both components place technical limitations on the current prototype that could possibly be removed or alleviated by using different components. The EyeLink system requires the use of headgear, and the Polhemus tracker reports posi-

tional coordinates only within a given operational envelope. (The current transmitter works up to a distance of 75 cm.) Both components currently restrict user movement to a relatively small area. Although this is adequate for the present purposes, the need for possible modifications has been taken into consideration in the design of the system: Both trackers have been abstracted in the software, so that changing the underlying equipment has been made as easy as possible.

The system communicates with the EyeLink tracker through the programming interface (API) provided by the manufacturer, and with the Polhemus tracker through a standard RS-232 serial interface.

The system is implemented as an object-oriented program, with Borland Delphi running on a Windows platform. A simplified schematic of the system is presented in Figure 1. The tracker components of the software (i.e., Eye Tracker and Pos Tracker) communicate with the actual trackers, abstracting their functionality, and provide data to the Data Processor. The Data Processor takes care of the calibrations and calculations during the measurement. The Data Processor feeds the processed data to the Visualizer for drawing an onscreen representation of the measurement, and to the File I/O component for data storage. The Playback component deals with saved recordings, feeding the data to the Visualizer for later review.

METHOD

3-D Gaze Points

Calculating the binocular gaze points in 3-D space requires knowledge of (1) the position of the eyes, (2) the direction of gaze for both eyes, and (3) the position and orientation of the gaze target (see Figure 2). A gaze point is defined, then, as the intersection point of the gaze vector and the gaze target surface (the screen).

The positions of the eyes are defined as the positions of the eye sighting centers (see definition below). The gaze vectors have their origin in the sighting center of the corresponding eye, and their direction is determined with data from the eyetracker. A positional tracker is used to obtain the position of the head and the eyes (i.e., the sighting centers of the eyes) of the subject, and the position of the target surface.

Calibration Process

To provide accurate data, the system variables must be calibrated to the individual measurement setup. The calibration has three stages. First, the EyeLink tracker is calibrated through the calibration procedure provided with the device. Next, the eye sighting centers are calibrated as described below. Finally, a transform array of gaze reference vectors is formed for transforming the gaze angle values given in eyetracker coordinates to gaze vectors in world coordinates.

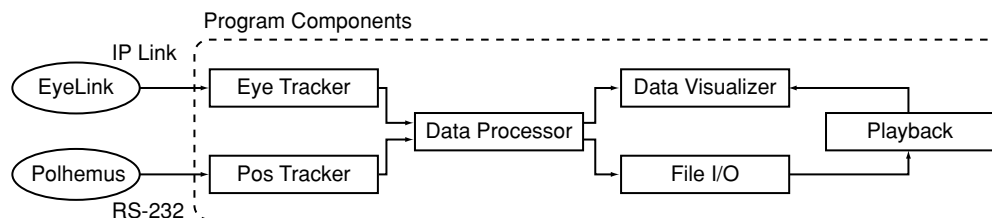


Figure 1. A simplified schematic of the system.

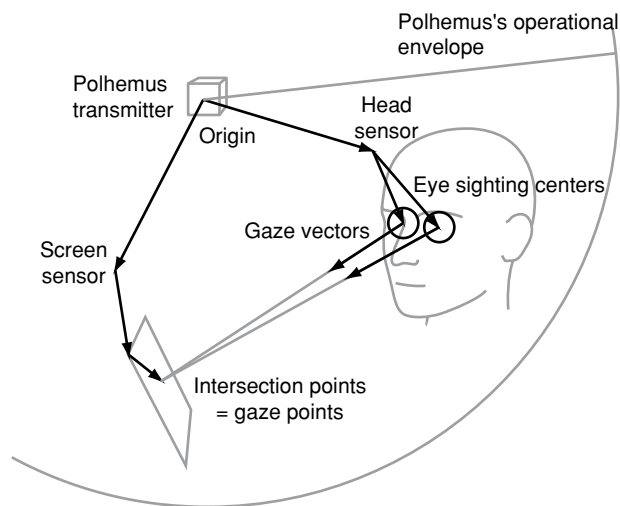


Figure 2. Definition of the binocular gaze points.

Sighting Center

The sighting center is defined as the point within the eye through which all gaze vectors, or lines of sight, pass. The concept has been described in previous work (Epelboim et al., 1995; Park & Park, 1933). In Park and Park, this point was found to be approximately 13.5 mm behind the front surface of the cornea, along the line of sight. Epelboim et al. used a pointing procedure: A pointer was positioned to touch the surface of the subject's eyelid while his or her head was stabilized on a bite bar. The subject's gaze vector was then aligned with the centerline of the pointer.

Here, the sighting center is defined for each eye by using the pointing procedure presented in Figure 3, which requires the subject to look through an aiming apparatus at different angles. The sighting centers for each eye are then calculated using the intersection or near-intersection point (see definition below) of the vectors defined by the centerline of the aiming apparatus in the acquired positions.

To calibrate the sighting centers, the subject points a narrow tube (AIM, in Figure 3) attached to a magnetic sensor (AIMREF) toward his or her eye so that the vector (AV1, AV2), defined as the centerline of the tube, effectively points at the sighting center (that is, so that the subject can see the background light through the tube). The centerline vector of the tube is known, in relation to the magnetic sensor. This position is then recorded in relation to the head sensor (HEADREF). Another line-of-sight vector in a similar manner after moving the tube approximately 30°–40° of visual angle in relation to the eye. The procedure is repeated for both eyes.

The sighting center, then, is defined as the intersection or near-intersection point of the two line-of-sight vectors for each eye. Because the two vectors seldom actually intersect, the near-intersection point is typically used. The near-intersection point is the middle point of the shortest distance vector (SDL, orthogonal to both AV1 and AV2) between two vectors. In Figure 3, the two aim vectors do not intersect; the gray field shows the projection of AV1 on AV2. The goodness/validity of the calculated eye sighting center can be evaluated with the length of the shortest distance vector, representing the uncertainty in pinpointing one exact location in space. The sighting centers are calibrated relative to the head sensor so that when the subject moves, the sighting center positions move, respectively.

Gaze Vectors

The gaze vectors have their origins in the sighting centers of the corresponding eyes. The direction of the gaze vector is determined with the eyetracker. The current system, using EyeLink I, uses a set

of calibrated reference vectors to resolve the direction of gaze in the world coordinate system through a linear transformation from the proprietary coordinate system provided by EyeLink.

Headgear

The EyeLink headband proved to be relatively unstable, moving around on the subject's head when the subject turned his or her head. The problem was probably largely caused by the tug of the heavy wire leading from the back of the unit. This necessitated the development of a more stable headset, which was then constructed from modified scuba-diving glasses. The headgear supplies a rigid frame for holding the EyeLink cameras and the Polhemus sensor steady, in relation to the eyes of the subject. (See Figure 4.)

The current headgear, however, makes the system quite uncomfortable for the user. Because this significantly limits the use of the system, a better solution is now under development.

Measurement Cycle

For each gaze point sample, the system goes through the measurement cycle presented in Table 1.

First, the system samples the current head and screen sensor positions from the positional tracker. Then, using calibration information, the system calculates the eye sighting center positions, and the position and orientation of the screen surface. Next, the system polls EyeLink for the current head-referenced gaze angles, and transforms those values to world coordinate gaze vectors. Then, the gaze point is calculated as the intersection point of the gaze vectors and the screen surface. Finally, that position is projected to the screen coordinate system, resulting in an (x, y) gaze position on the screen.

RESULTS

Performance and Evaluation

The system provides gaze point data. The actual output parameters are user configurable and can include a time stamp, the onscreen gaze point (x, y) , the 3-D eye position, the 3-D gaze vector, and the 3-D screen position and orientation values. The system is capable of presenting recorded data in Playback mode. However, the system does not include a data analysis module for gaze data, so

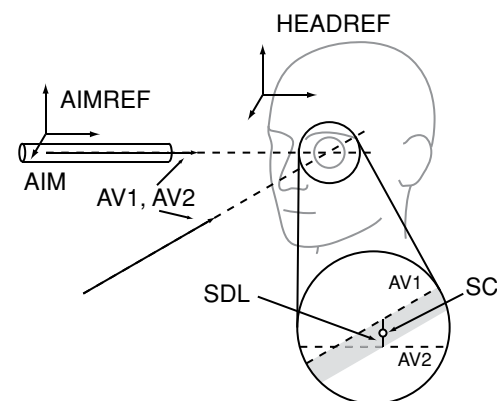


Figure 3. Defining the sighting center of the eye. AIM, aiming apparatus; AIMREF, aim reference coordinate system; HEADREF, head reference coordinate system; AV1, AV2, aim vectors; SDL, shortest distance line between the two aiming vectors; SC, midpoint of the SDL, which defines the sighting center.

Table 1
Measurement Cycle

Stage	Description	Device
1	Get eye positions (sighting centers)	Polhemus
2	Get screen position and orientation	
3	Get EyeLink head-referenced coordinate values	EyeLink
4	Transform EyeLink coordinate values to world coordinate gaze vectors	System
5	Calculate gaze vector (which has its origin in the sighting center) intersection with the screen surface	
6	Project the screen gaze vector to the screen coordinate system	

the data should be subjected to a separate data analysis method if fixation data are required.

Because the system is implemented as software, the temporal resolution is largely affected by the underlying equipment. Currently, when using a 1.6-GHz Pentium 4 processor, the time required by the sampling from the trackers and the calculations in the measurement cycle is less than 4 msec, enabling a sampling frequency of over 200 Hz. However, the test measurements for this work were made with a more moderate sampling frequency of 30 Hz, because the visualizing component had not been thoroughly optimized (but was essential for observing the operation of the system in early development).

The system was evaluated using two pilot experiments. The first experiment evaluated the accuracy of the data produced by the system. The subjects were asked to fixate back and forth between two diagonal fixation points while changing their position freely, moving both their head and the handheld device.

The second experiment was a standard reading task, in which the subjects were asked to read a passage of text from the handheld device, which allowed them to change their position freely. The data were then used to subjectively evaluate the validity of the data against typical reading data from the literature.

In both tasks, the user sat in a comfortable position in a typical office chair. The handheld device was held at an average viewing distance of 60 cm during the test, the

typical posture being with the user's hand resting on his or her lap and the head tilted forward so that the device in the user's hand was within the central viewing area. While the subject moved the device, it was brought to about 20 cm from the nose, at the closest, and was at an arm's length, at the farthest. During the task, the user was also instructed to turn around and move the device horizontally and vertically so that the test data would cover most of the space surrounding the user.

The users moved much more extensively in the first task; in the second task, they concentrated more on reading. The pilot tests were performed with 3 subjects, and all of these produced similar data.

Experimental Results

Example data from Scenario 1 are presented in Figure 5. The figure shows the fixation targets and the superimposed gaze samples on the screen of the handheld device. The histograms on the left-hand and upper side show the horizontal and vertical distribution of the gaze samples. The bar beneath the histogram shows the span of 1° of visual angle in the average viewing distance (60 cm) during the test. In the example data, the user made 22 back-and-forth cycles between the fixation targets. The fixation targets were located at points (70, 235) and (160, 80), in screen pixels from the lower left corner. The mean fixation locations and their standard deviation were $(71 \pm 12; 232 \pm 17)$ and $(160 \pm 21; 100 \pm 20)$, respectively.

The data show that 80% of the samples (74% vertical, 82% horizontal) were within 1° of the fixation targets. It should be noted that these were raw data and had not been subjected to a fixation analysis. Therefore, some of the gaze points were sampled during a fixation on a target and some during a saccade between the targets. Blinking, and the resulting loss of eye data from EyeLink, currently results in the gaze vector pointing at the coordinate system origin, which explains the data momentarily drifting outside the screen. Subjecting the data to a fixation analysis algorithm would remove the individual saccade points, at least, thus improving the percentages noted above and decreasing the amount of dispersion of the data points. One possible reason for the larger vertical deviation in the

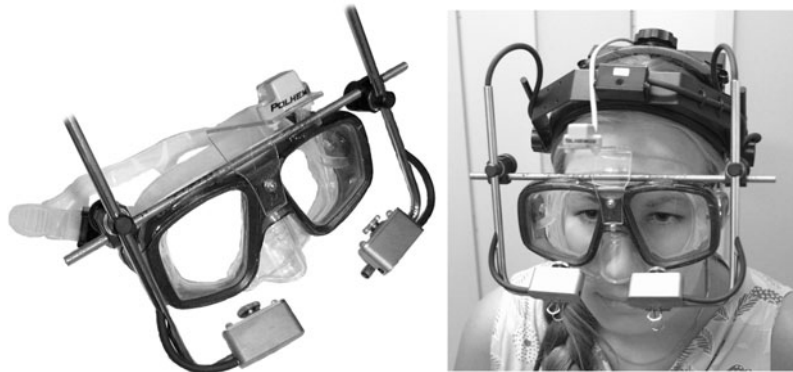


Figure 4. Headgear.

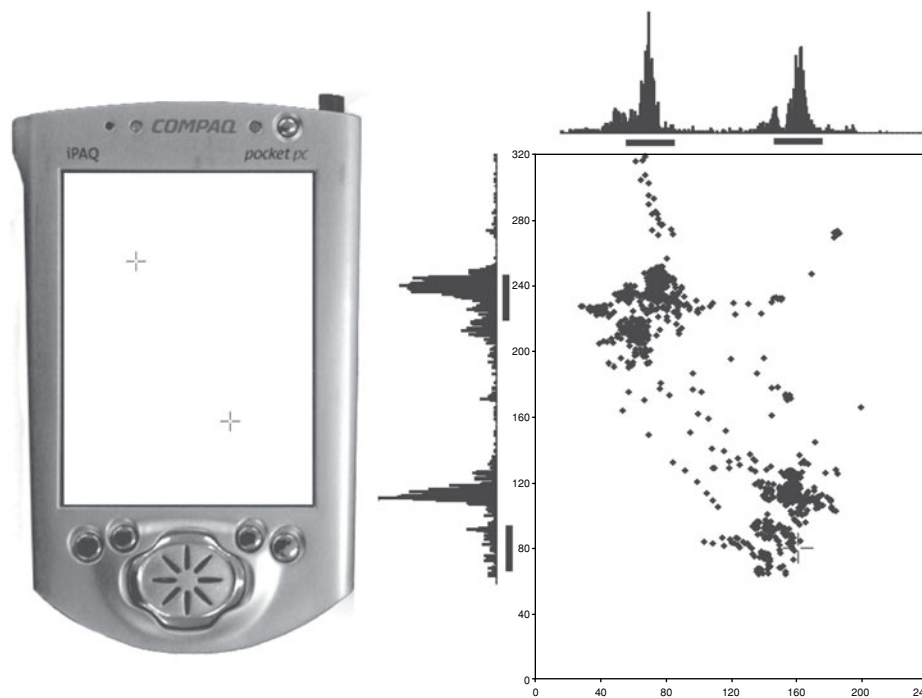


Figure 5. Example data for the fixation test.

data for the lower right fixation point could be the narrow linear range of the EyeLink tracker: If the lower part of the screen had been held more than 20° below the primary line of gaze, the data probably would have been “squeezed,” as in the example data.

Figure 6 shows example data from Experiment 2. The picture on the left-hand side shows the PDA screen as the user saw it. The passage of text was displayed with a resolution of 240×320 pixels and a letter height of 3 mm. The figure shows a typical gaze path in a reading task, with the text lines being fixated relatively accurately. When the data are compared with characteristics of reading provided in Rayner and Pollatsek (1989), it seems that the system produces valid data.

The amount of scatter in the data from the first experiment, when compared with the relatively steady gaze path of the second experiment, could be the result of the user having moved quite a bit more during the first experiment. The larger range of movements in the first experiment took the screen beyond the linear range of the EyeLink tracker.

DISCUSSION

A prototype system for tracking gaze on a handheld device has been presented. Evaluation shows that the system is capable of producing valid data.

The static accuracy of the currently used Polhemus FASTRAK system is reported as 0.8-mm RMS for the sensor position (x, y, z) and 0.15° RMS for receiver orientation (Polhemus, 2000). It should be noted that these values are valid only within the operational envelope of the

tracker, which, with the current transmitter, has a radius of 75 cm. By using different types of transmitters available from Polhemus, this range can be extended to up to 5 m.

With a trained subject, the definition of the sighting centers gives submillimeter accuracy values for pinpointing the sighting centers as single points in space. The accuracy of the total system depends largely on the successful calibration of the EyeLink system. Even with good calibration, the output of EyeLink seems to be linear on a relatively small area, extending to about 30° horizontally and about 20° vertically, effectively limiting the usable area of the system.

Because the system uses EyeLink for tracking the eyes, it produces binocular data. Although binocular data are not essential for tracking gaze (because the eyes usually fixate on approximately the same spot), these data could be used to study convergence, when looking at targets at different depths. Also, eye dominance could be of significance in certain types of tests, and measuring binocular data gives the option to select the eye used as the data source if monocular data are needed.

The present approach could also benefit the field of research on augmented reality (AR). In addition to a detailed description of the scene around the user, the AR systems need an accurate description of the location and optical properties of the viewer (in this case, the user's eyes) for realistic merging of virtual objects with the real scene. These should also be tracked over time when the user moves and interacts with the scene (Azuma et al., 2001).

As noted above, the system is designed so that either of the trackers can be replaced with a substitute. Using an

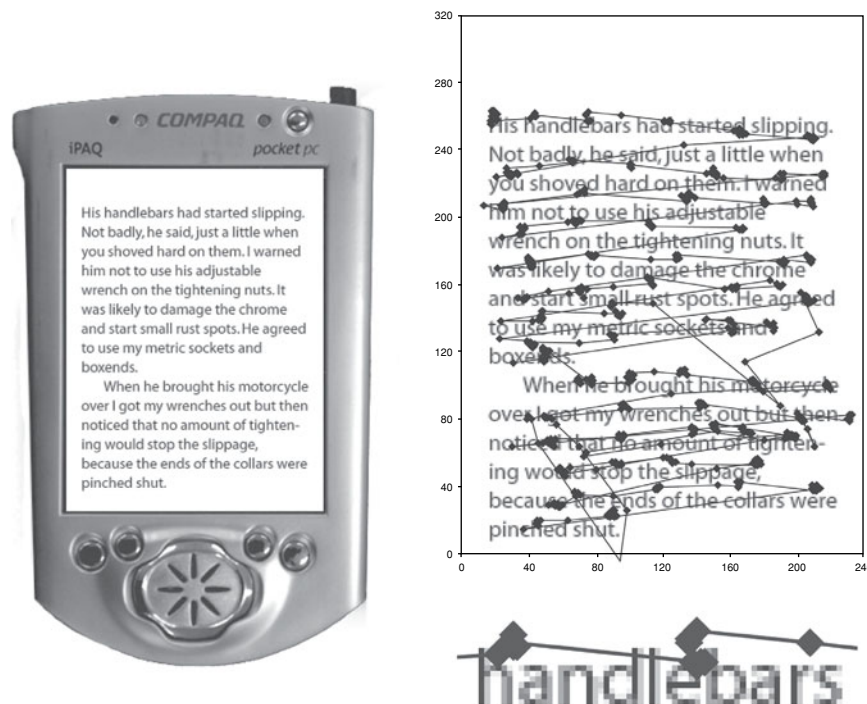


Figure 6. Example data for the reading test.

optical tracker, rather than the Polhemus tracker, would free the system from wires, and possibly extend the operational envelope. Also, using a remote system for eye tracking would free the user from wearing headgear, but this would result in reduced freedom of movement, because the user would have to stay in view of the gaze-tracking cameras. In either case, the basic operational principle of the system would not be affected by the use of other types of trackers.

Although the present system operates only within an area defined by the operational envelope of the positional tracker and the wiring of the trackers, it makes new kinds of studies possible. These studies could also be extended—for example, studying mobile use while walking could be accomplished using a treadmill, which would allow the user to walk without drifting outside the operational range.

The future development of the system will concentrate on three things: (1) improving (possibly removing) the headgear worn by the subject, reducing stress and discomfort without compromising stability, (2) improving the calibration, with the objective of a quicker and easier method, and (3) extending the use of the system from a moving-screen-based to a 3-D-model-based environment.

REFERENCES

- AZUMA, R., BAILLOT, Y., BEHRINGER, R., FEINER, S., JULIER, S., & MACINTYRE, B. (2001). Recent advances in augmented reality. *IEEE Computer Graphics & Applications*, *21*, 34-47.
- BABCOCK, J. S., & PELZ, J. B. (2004). Building a lightweight eyetracking headgear. *Proceedings of the 2004 Symposium on Eye Tracking Research and Applications* (pp. 109-114). New York: ACM Press.
- BAREA, R., BOQUETE, L., MAZO, M., LÓPEZ, E., & BERGASA, L. M. (2000). E.O.G. guidance of a wheelchair using spiking neural networks. *Proceedings of European Symposium on Artificial Neural Networks* (pp. 233-238). Evre, Belgium: D-side.
- CARPENTER, R. H. S., & ROBSON, J. G. (Eds.) (1999). *Vision research: A practical guide to laboratory methods*. Oxford: Oxford University Press.
- CHRISTOU, C. (2001, August). *iTrack: Integrated eye, head and finger tracking in virtual environments*. Paper presented at the 12th European Conference on Eye Movements, Dundee, Scotland.
- CROSBY, M. E., & PETERSON, W. W. (1991). Using eye movements to classify search strategies. *Proceedings of the Human Factors Society 35th Annual Meeting: Vol. 2* (pp. 1476-1480). Santa Monica, CA: Human Factors and Ergonomics Society.
- DUCHOWSKI, A. T., SHIVASHANKARAIHAH, V., RAWLS, T., GRAMOPADHYE, A. K., MELLO, B. J., & KANKI, B. (2000). Binocular eye tracking in virtual reality for inspection training. *Proceedings of the 2000 Symposium on Eye Tracking Research and Applications* (pp. 89-96). New York: ACM Press.
- EDEN, G. F., STEIN, J. F., WOOD, H. M., & WOOD, F. B. (1994). Differences in eye movements and reading problems in dyslexic and normal children. *Vision Research*, *34*, 1345-1358.
- EPELBOIM, J. L., STEINMAN, R. M., KOWLER, E., EDWARDS, M., PIZLO, Z., ERKELENS, C. J., & COLLEWIN, H. (1995). The function of visual search and memory in sequential looking tasks. *Vision Research*, *35*, 3401-3422.
- HYÖNÄ, J., RADACH, R., & DEUBEL, H. (Eds.) (2003). *The mind's eye: Cognitive and applied aspects of eye movement research*. Amsterdam: Elsevier.
- KUUTTI, K. (1999). Small interfaces: A blind spot of the academical HCI community? In H.-J. Bullinger & J. Ziegler (Eds.), *Human-Computer Interaction: Vol. 1. Ergonomics and user interfaces*. (pp. 710-714). Mahwah, NJ: Erlbaum.
- MAJARANTA, P., & RÄIHÄ, K.-J. (2002). Twenty years of eye typing: Systems and design issues. *Proceedings of the 2002 Symposium on Eye Tracking Research and Applications* (pp. 15-22). New York: ACM Press.
- NÄSÄNEN, R., OJANPÄÄ, H., & KOJO, I. (2001). Effect of stimulus con-

- trast on performance and eye movements in visual search. *Vision Research*, **41**, 1817-1824.
- OYEKOYA, O. K. (2004). *Overview of available commercial eye trackers*. Retrieved November 8, 2004, from University College London, Electronic and Electrical Engineering Web site: <http://www.ee.ucl.ac.uk/~ooyekoya/CommercialEyeTrackers.pdf>.
- PARK, R. S., & PARK, G. E. (1933). The center of ocular rotation in the horizontal plane. *American Journal of Physiology*, **104**, 545-552.
- PELZ, J. B., & CANOSA, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research*, **41**, 3587-3596.
- POLHEMUS, INC. (2000). 3SPACE FASTRAK user's manual. Colchester, VT: Author.
- RAYNER, K., & POLLATSEK, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.
- STANFORD POYNTER PROJECT (2000). Eye tracking online news [Electronic version]. Retrieved November 8, 2004, from www.poynterextra.org/et/i.htm.
- YOUNG, L. R., & SHEENA, D. (1975). Survey of eye movement recording methods. *Behavior Research Methods & Instrumentation*, **7**, 397-429.

(Manuscript received December 9, 2004;
revision accepted for publication September 1, 2005.)