

Comparison of Three Eye Tracking Devices in Psychology of Programming Research

Seppo Nevalainen and Jorma Sajaniemi
University of Joensuu, Finland
seppo.nevalainen@cs.joensuu.fi
jorma.sajaniemi@cs.joensuu.fi

Keywords: POP-II.B. program comprehension, POP-V.B. eye tracking

Abstract

Eye tracking can be used in measuring point of gaze data that provides information concerning subject's focus of attention. The focus of subject's attention can be used as supportive evidence in studying cognitive processes. Despite the potential usefulness of eye tracking in psychology of programming research, there exists only few instances where eye tracking has actually been used. This paper presents an experiment in which we used three eye tracking devices to record subjects' points of gaze when they were studying short computer programs using a program animator. The results suggest that eye tracking can be used to collect relatively accurate data for the purposes of psychology of programming research. The results also revealed significant differences between the devices in the accuracy of the point of gaze data and in the times needed for setting up the monitoring process.

Introduction

In the focus of psychology of programming research is the understanding of cognitive processes of programmers when they are writing, reading and learning computer programs. These cognitive processes can't be observed directly. Instead, the researcher has to collect secondary data through which the processes can be inferred. One way to gather this secondary data is to make observations of the subject's actions. These observations can consist of for example errors the subject makes, time the subject uses or location of the point of gaze (POG) of the subject.

In the eye tracking process, the collection of POG data can be performed without the need of the subject performing any action. This can be seen as a benefit when studying cognitive processes that can be easily disturbed. The collected POG data provides information of subject's attention, even though the focus of attention is not necessarily at POG. Information about subject's attention can be used as supportive evidence when studying cognitive processes.

Eye tracking has been used in several usability studies (Goldberg & Kotval 1999, Sibert & Jacob 2000, Byrne, Anderson, Douglass & Matessa 1999) and cognitive psychology studies related to different search and reading strategies (Rayner 1992, 1998, Findlay 1992, Kennedy 1992). In psychology of programming research, eye tracking has been used by Crosby (1989), who studied the code viewing strategies of the subjects. Bednarik and Tukiainen (2004) have compared eye tracking with blurred display.

Despite the potential usefulness of eye tracking in psychology of programming research, there exists only few instances where eye tracking has actually been used. Therefore experience concerning the benefits, disadvantages and problems of eye tracking in psychology of programming research is needed. This paper reports an experiment in which we used three eye tracking devices to record subjects' POG when studying short computer programs using PlanAni animator (Sajaniemi & Kuittinen 2003). We studied the easiness of use and accuracy of the three devices. We also observed and estimated the amount of disturbance the devices caused to the subjects.

The rest of the paper is organized as follows. Next section gives an introduction to the eye tracking process and to the devices used in this experiment. Then the experiment is described and results are presented and discussed. The last section contains the conclusions.

Eye Tracking Methodology

Eye tracking process can be divided roughly into the following steps: subject set-up, adjustments, subject calibration, and monitoring.

In the subject set-up phase, the subject is seated and her location in relation to the eye-tracking device is adjusted. If head mounted optics is used, the eye tracking device is placed on subject's head and its position is adjusted.

The adjustments phase includes adjusting the settings of the eye tracking program; detecting and ensuring the recognition of the subject's eye(s); and opening the file used for the recording of the eye tracking data.

In the calibration phase, a calibration pattern consisting of a number of calibration points is shown to the subject. The subject is asked to direct her gaze to each of the calibration points and the location of the POG for each calibration point is recorded. The values from the calibration are used in calculating the locations of points of gaze from the values received from the eye tracking device. The calibration phase is repeated until satisfactory calibration values are recorded for each calibration point. One significant problem in eye tracking is the drift effect, which indicates a deterioration of the calibration over time (Tobii 2003). The drift effect can be reduced by ensuring the stability of the light conditions of the environment and the equal light intensity between calibration stimuli and the experiment stimuli.

The monitoring phase consists of viewing the status of the eye tracking and, if necessary, readjusting the settings during the tracking of the actual experiment tasks.

In the experiment we used the following three devices: Tobii 1750 from Tobii Technology, ASL 504 Pan/Tilt Optics from Applied Science Laboratories and ASL 501 Head Mounted Optics from Applied Science Laboratories. All three devices use video based combined pupil and corneal reflection eye tracking.

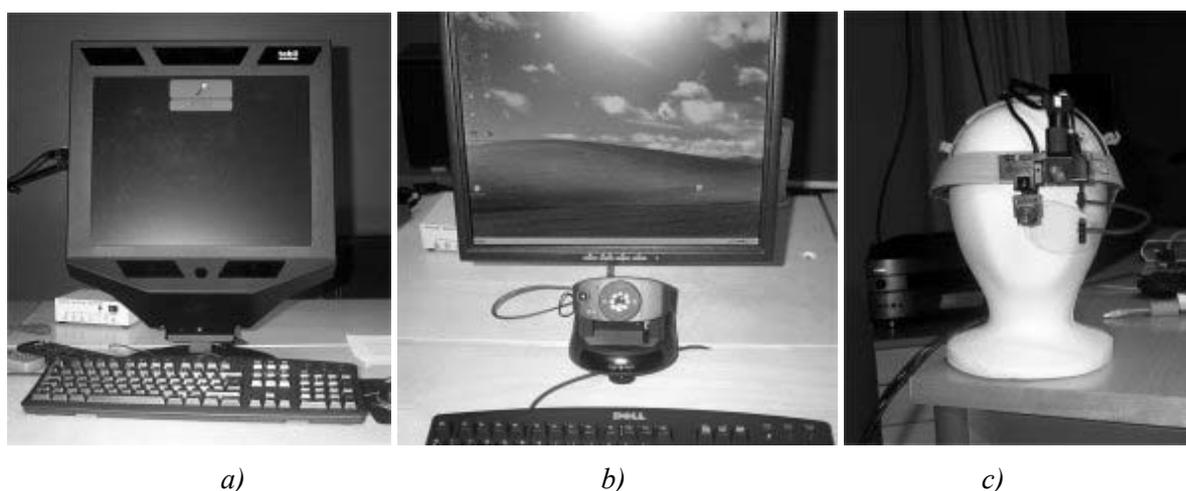


Figure 1. Eye tracking devices, a) Tobii 1750 b) ASL 504 Pan/Tilt Optics c) ASL 501 Head Mounted Optics.

In Tobii 1750 (Tobii 2003), the eye tracking device is embedded into the panels of the monitor that the subject is viewing (Figure 1a). The device uses a wide-angle camera to capture images of the subject and near infrared light emitting diodes for eye illumination. The device uses both eyes of the subject for tracking. Tobii 1750 records data at the rate of 30 Hz (30 gaze data points/second). When

the device does not detect the subject's eye(s), the recording rate is slowed down until proper detection is regained. The theoretical accuracy of POG coordinates provided by the device is 1 degree visual angle (approximately 1 cm error when the subject is seated at 50 cm distance from the display).

In ASL 504 pan/tilt optics (ASL 2003b), the eye tracking device is placed below the monitor the subject is viewing (Figure 1b). The device has an adjustable wide angle camera that repositions itself according to the movements of the subject. The device uses the wide angle camera to capture an image of the subject's eye and near infra-red light emitting diodes for eye illumination. The device uses one eye for tracking. ASL 504 pan/tilt optics records data at the rate of 50 or 60 Hz. The theoretical accuracy of POG coordinates provided by the device is 0.5 degree visual angle (approximately 0.5 cm error when the subject is seated at 50 cm distance from the display).

In ASL 501 head mounted optics (ASL 2003a), the optics device is placed on subject's head (Figure 1c). The device uses one wide angle camera to capture image of the subject's eye and another wide angle camera to capture the subject's field of view (the scene camera). The device uses near infra-red light emitting diodes for eye illumination. The device uses one eye of the subject for tracking. ASL 501 head mounted optics records data at the rate of 50 Hz. The theoretical accuracy of POG coordinates provided by the device is 0.5 degree visual angle.

Experiment

In our experiment, we studied the easiness of use of eye tracking devices by measuring the total amount of time needed for the preparations of the subject. The preparations consist of subject set-up, adjustments and calibration. We also observed and estimated the effort these activities required from the subject. The accuracy of the devices was measured by calculating mean distances between recorded points of gaze (in the data files) and requested points of gaze (measured with the eye tracking software). The experimenters were using eye-tracking devices for the first time.

Method

Design: A within-subject design was used with one independent variable (the eye tracking device used for collecting the data) and two dependent variables (the time needed for the preparation of the subject, and the accuracy of the device).

All subjects were measured using all three eye tracking devices (Tobii 1750, ASL 504 Pan/Tilt Optics, and ASL 501 Head Mounted Optics) and the order of the devices was counterbalanced. Each device occurred in each of the chronological position (1st, 2nd or 3rd measuring device) equal number of times. In the experiment we used two different versions of PlanAni. The order of the versions was varied so that with each tracking device and each of the viewed programs two of the four subjects used the animator with code view first and the other two used the animation view first.

Subjects: Twelve subjects, eight male and four female, participated in the experiment. The subjects were required to have at least basic programming skills and some experience in programming. They were recruited from third year courses in computer science and were given a coffee ticket for their participation.

Materials: For the purpose of the experiment, PlanAni was modified so that it showed either only the code-view that is located on the top left corner of the animator (Figure 2) or only variable animation-view that is located on the top center of the animator (Figure 3). All variables were depicted by the same neutral image. Both versions showed notifications for the subject and the input/output area. For the task of focusing at specific targets on the screen, screenshots of PlanAni were used. The PlanAni version was v0.53.

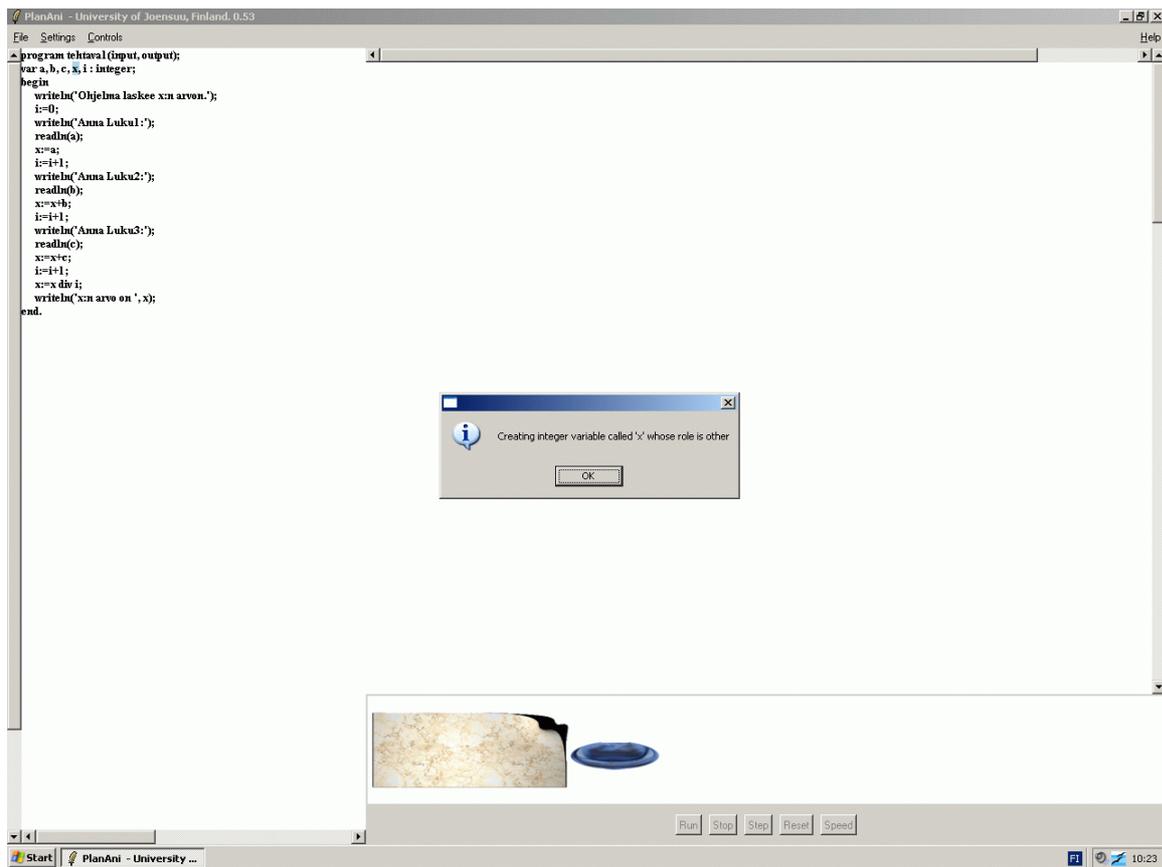


Figure 2. PlanAni with code view only

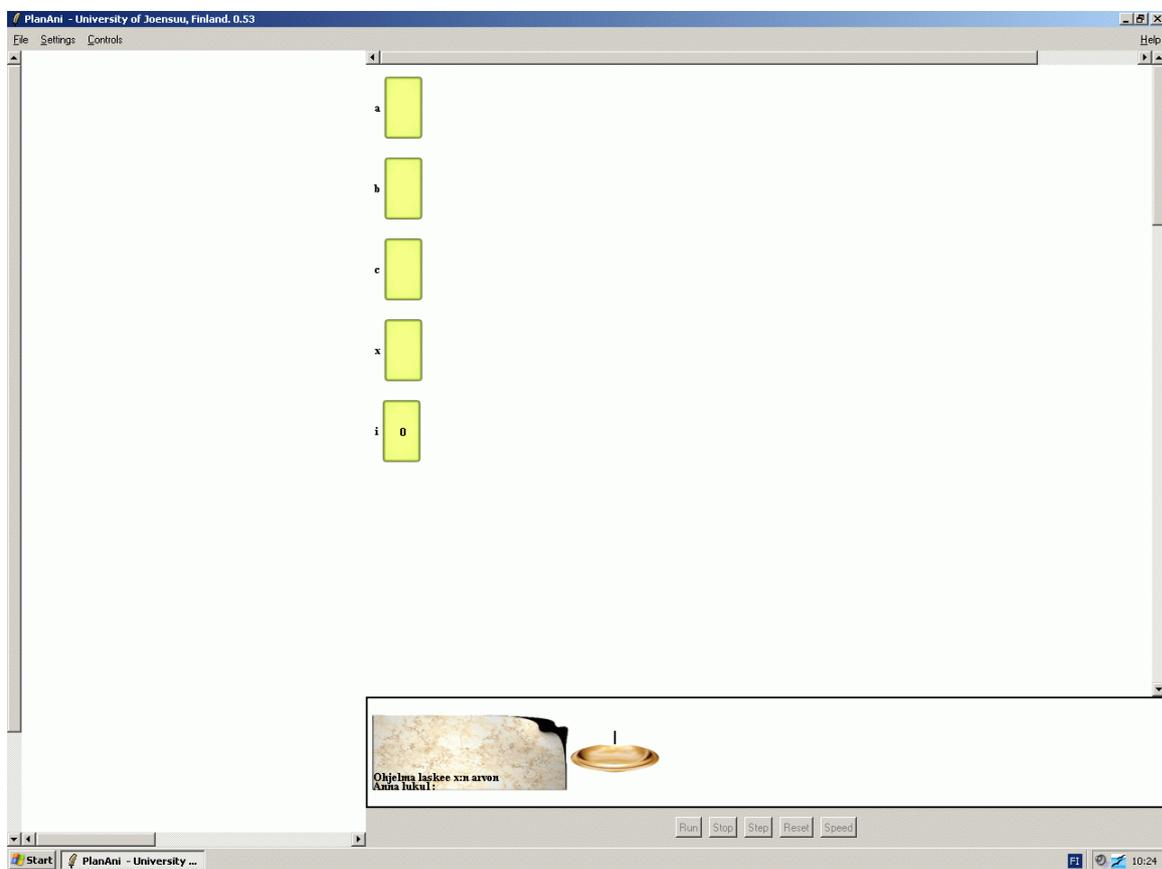


Figure 3. PlanAni with variable animation view only.

Procedure: The subjects used PlanAni to comprehend six short computer programs—two programs with each eye tracking device. They were allowed to view each program one time step by step. The POG of subjects during these tasks was measured. With each device, the subject was first seated and the eye tracking device's location in relation to the subject was adjusted (subject set-up phase). The movement of the subject was minimized by using a chair without wheels, by setting the chair close enough to the desk to minimize the horizontal rotation and advising the subject to avoid quick and large movements of her head. The subject was not explicitly demanded to stay perfectly immobile during the task. After set-up, the settings of the interface program were adjusted, detection of the subject's eyes was performed and the file used for storing the POG data was opened (adjustments phase). After this the calibration of the subject was performed (calibration phase). Time needed for these preparations was measured by the experimenter using a special program that required a single key-press to start and stop time measuring. With each device, the subject performed two program comprehension tasks so that she used both versions of PlanAni. After each viewing task, the subjects were asked to give a short program summary. The program summaries were collected for the purpose of motivating the subjects to study the program but they were not analyzed further in this experiment. After studying the programs, subjects were asked to look at eight specific targets on the screen before proceeding to the next eye tracking device.

Results

Table 1 gives the mean times (in seconds) needed for the preparation phase and the execution of the program comprehension tasks. The preparation time is measured from the beginning of the set-up to the end of calibration. The difference in preparation times between Tobii 1750 and ASL 501 (paired t test, $t = 8.187$, $df = 11$, $p < 0.0001$) as well as the difference between Tobii ASL 504 and ASL 501 (paired t test, $t = 6.417$, $df = 11$, $p < 0.0001$) are both statistically significant.

Table 1: Times (means in seconds with standard deviation in parentheses) needed for the preparation and execution of the tasks.

	Tobii 1750	ASL 504	ASL 501
Preparation	471.8 (128.9)	548.3 (126.8)	953.5(164.4)
Tasks	502.6 (112.1)	525.0 (122.9)	476.6 (68.4)

Table 2: Amount of valid, uncertain and invalid data from all collected gaze data and the percentage of invalid data from all data.

	Amount of valid data	Amount of uncertain data	Amount of invalid data	Percentage of invalid data
Tobii 1750	104101	16486	10629	8.1
ASL 504	182177	N/A	17722	8.7
ASL 501	198812	N/A	6099	3.0

Table 2 gives the amounts of valid, uncertain and invalid data as reported by the devices. Tobii 1750 provides validity codes 0-4 (0 = valid, 1-3 = uncertain and 4 invalid) for the data. For the ASL devices, the validity of the data is determined by value in the pupil size field (0 = invalid, otherwise valid).

Table 3: Distances (means in centimeters with standard deviation in parentheses) of measured points of gaze from the requested points of gaze, and the corresponding degrees of visual angle.

	Tobii1750	ASL 504	ASL 501
Distance	1.134 (0.203)	1.391 (0.351)	1.609 (0.314)
Visual angle	1.3	1.6	1.8

Table 3 gives the mean distances (in centimeters) of measured points of gaze from the requested points of gaze, and the corresponding visual angle when the subject is seated at 50 centimeters distance from the display. The distance was measured for points within a threshold of 2.5 cm from the center point of the target. The threshold was selected so that the theoretical accuracies of the devices and the microscopic movements of the eye fitted within the threshold. ASL 501 provides the POG coordinates on a plane that is in relation to the field of view of the subject, while the other two devices provide the coordinates on a fixed plane. With ASL 501, the location of the screen in the field of view shifts when the subject turns her head. This shift was visually detected and measured, and the corresponding corrections were calculated and applied to the coordinates before calculating the distances. The difference between Tobii 1750 and ASL 501 (paired t test, $t = 3.707$, $df = 8$, $p = 0.006$) is statistically significant.

Discussion

The time needed for preparation when using 501 was approximately twice as long as the time needed for Tobii 1750 or ASL 504 (see Table 1). In our experience, there are two main reasons that explain this difference. Firstly, the subject set-up phase with ASL 501 consisted of more individual steps and required more effort than with Tobii 1750 or ASL 504. One time consuming step was locating the image of the subject's eye through the visor so that it was in correct angle and the visor was not in front of the subject's field of view. Secondly, the calibration with the ASL 501 was more troublesome and needed to be repeated more often than with Tobii 1750 or ASL 504. This was mainly due to the fact that the subjects were required to keep their heads perfectly still during the calibration phase. One possibility to make the calibration easier and faster with the ASL 501 is to use a bite bar or chin rest during calibration. This may, however, cause discomfort to the subject and its applicability in a psychology of programming experiment is questionable.

Table 2 shows the amounts of collected valid, uncertain and invalid data. All the devices reported invalid data under 10 %. The difference between ASL 501 compared to Tobii 1750 and ASL 504 is most probably due to the fact that the two last mentioned devices lost the eye easily when the subject used the keyboard to provide input to the program and the eye moved out of the reach of the devices' cameras. ASL 504 also had difficulties in automatically relocating the eye and in some cases it needed to be aided by relocating the eye manually.

Table 3 shows the accuracies of the three devices. The values indicate that Tobii 1750 has the highest accuracy, ASL 504 provides second highest accuracy, and ASL 501 the lowest accuracy. Only the difference between Tobii 1750 and ASL 501 is, however, statistically significant. One factor in the low accuracy of the ASL 501 is probably an inaccuracy in the visually estimated correction due to head movements. The need for this correction can be removed by using magnetic head tracker with ASL 501. Tobii 1750's measured accuracy is quite near to the theoretical accuracy given in the Eye Tracking Methodology -section. The measured accuracy is 1 - 1.6 degrees for a subject sitting at 40 - 60 centimeters distance from the screen. ASL 504 and ASL 501 fall clearly behind the theoretical accuracy given in the Eye Tracking Methodology -section. For ASL 504 the measured accuracy is 1.3 - 2 degrees for a subject sitting at 40 - 60 centimeters distance from the screen. For ASL 501 the measured accuracy is 1.5 - 2.25 degrees for a subject sitting at 40 - 60 centimeters distance from the screen

Eye Tracking in Psychology of Programming Research

In psychology of programming research eye tracking can be used as an implication of the focus of subjects' attention. The POG is not, however, the same as the focus of attention. Attention is not necessarily always associated with the visual scene, even though POG is. The subject can also voluntarily target his attention slightly off the POG (Posner 1980).

The general unobtrusiveness of an eye tracking device can be seen as a factor when using this technology in psychology of programming research. Subjects' cognitive processes can be easily disturbed with objects in the field of view, sounds in the room, and extra activities required by the experimental settings. Some of the subjects commented that the scene camera of the ASL 501, positioned according to the manual, was disturbingly in their field of view. The visor of the ASL 501 remained in the lower part of the subject's field of view during the measuring. This did not, however, invoke any comments from the subjects. With ASL 504, the adjustable camera produced a buzzing sound when it repositioned itself, causing the subject to be aware of the device's existence. Tobii 1750 looks like a normal display device and makes no visible or audible interference.

When considering the required effort and caused disturbance, Tobii 1750 seemed to be the most unobtrusive for the subjects. With ASL 504, the subject was required to keep her head perfectly still during the detection of the eye, since the auto-follow property of the camera could be turned on only after the pupil and corneal reflection were found. The positioning of the optics device of ASL 501 on the subject's head was time consuming and caused physical discomfort to the subject.

Tobii 1750 enabled a subject to easily observe the tracking status before the calibration phase, and to take part in the detection of the eye. The calibration was not dictated by the operator but the tracking program performed the calibration by showing the subject calibration points in random locations at a slow pace.

In eye tracking, the quality and amount of recorded data is influenced by the amount of subjects' motions. The more immobile the subject is, the better data eye tracking devices usually record (Tobii 2003, ASL 2003a). When eye tracking is used in psychology of programming research, however, the immobilising of the subject can disturb the cognitive processes that are being studied. It seems that there is a trade off between the accuracy and the ecological validity of data. With the subject seating used in our experiment, we reached an accuracy that was quite near to the theoretical accuracy of Tobii 1750. With ASL devices, however, the measured accuracy was considerably behind the theoretical values.

Tobii 1750 and ASL 504 require the subject to be seated and tolerate limited movements of the head, only. ASL 501 allows the subject to move around—an activity needed in some experimental settings in psychology of programming.

Conclusions

We have conducted an experiment comparing the use of three eye tracking devices in a psychology of programming experiment in which subjects studied short computer programs using a program animator.

The results show that there are significant differences in the accuracy and easiness of use between the devices. The ASL 501 Head Mounted Optics required approximately twice as much time for the preparation than the other two devices. The ASL 501 was also the least accurate of the devices when it was used for the task in which the subject viewed a computer screen. This can be partly explained by inaccuracies in the manual correction of the shifting effect. This effect can be removed by using magnetic head tracker with ASL 501.

When considering the required effort and caused disturbance, Tobii 1750 seemed to be the most unobtrusive for the subjects. The device allowed the subject to take part in the detection of the eyes and the calibration process was performed without step-by-step dictation of the operator.

The monitoring process didn't contain any clear differences between the devices. The ASL 504 needed to be aided by relocating the eye manually in some cases.

Acknowledgements

This work was partially supported by the Academy of Finland under grant number 206574

References

- ASL (2003a). *Eye Tracking System Instruction Manual – Model 501 Head Mounted Optics*. Applied Science Laboratories
- ASL (2003b). *Eye Tracking System Instruction Manual – Model 504 Pan/Tilt Optics*. Applied Science Laboratories
- Bednarik, R. & Tukiainen, M. (2004). Visual attention and representation switching in Java program debugging: a study using eye movement tracking. In *the 16th Annual Workshop of the Psychology of Programming Interest Group (PPIG'04)*.
- Byrne, M. D., Anderson, J. R., Douglass, S. & Matessa, M. (1999). Eye tracking the visual search of click-down menus. In *Human Factors in Computing Systems: CHI '99 Conference Proceedings* (p.402-409). ACM Press.
- Crosby, M. & Stelovsky, J. (1989). The influence of user experience and presentation medium on strategies of viewing algorithms. In *Vol. II: Software Track, Proceedings of the Twenty-Second Annual Hawaii International Conference on System Sciences*, Kailua-Kona, HI USA, Jan. 1989.
- Findlay, J. M. (1992). Programming of stimulus-elicited saccadic eye movements. In K. Rayner (Ed.), *Eye Movements and Visual Cognition: Scene Perception and Reading* (p. 8-30). New York, NY: Springer-Verlag. (Springer Series in Neuropsychology)
- Goldberg, J. H. & Kotval, X. P. (1999). Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics*, 24, 631-645.
- Kennedy, A. (1992). The spatial coding hypothesis. In K. Rayner (Ed.), *Eye Movements and Visual Cognition: Scene Perception and Reading* (p. 379-396). New York, NY: Springer-Verlag. (Springer Series in Neuropsychology)
- Posner, M. I., Snyder, C. R. R. & Davidson, B. J. (1980). Attention and the detection of signals. *Experimental Psychology: General*, 109(2), 160-174.
- Rayner, K. (Ed.). (1992). *Eye Movements and Visual Cognition: Scene Perception and Reading*. New York, NY: Springer-Verlag. (Springer Series in Neuropsychology)
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422.
- Sajaniemi, J. & Kuittinen, M. (2003). Program animation based on the roles of variables. *Proceedings ACM 2003 Symposium on Software Visualization (SoftVis 2003)* (p. 7-16), San Diego, CA, June 2003. Association for Computing Machinery.
- Sibert, L. E. & Jacob, R. J. (2000). Evaluation of eye gaze interaction. In *Human Factors in Computing Systems: CHI 2000 Conference Proceedings*. ACM Press.
- Tobii (2003). *User Manual – Tobii Eye Tracker*, Clearview Analysis Software. Tobii Technology AB.