# Evaluating Eye Tracking with ISO 9241 - Part 9

Xuan Zhang and I. Scott MacKenzie

Department of Computer Science and Engineering
York University
Toronto, Ontario, Canada M3J 1P3
{xuan,mack}@cse.yorku.ca

**Abstract.** The ISO 9241-9 standard for computer pointing devices proposes an evaluation of performance and comfort [4]. This paper is the first eye tracking evaluation conforming to ISO 9241-9. We evaluated three techniques and compared them with a standard mouse. The evaluation used throughput (in bits/s) as a measurement of user performance in a multi-directional point-select task. The "Eye Tracking Long" technique required participants to look at an on-screen target and dwell on it for 750 ms for selection. Results revealed a lower throughput than for the "Eye Tracking Short" technique with a 500 ms dwell time. The "Eye+Spacebar" technique allowed participants to "point" with the eye and "select" by pressing the spacebar upon fixation. This eliminated the need to wait for selection. It was the best among the three eye tracking techniques with a throughput of 3.78 bits/s, which was close to the 4.68 bits/s for the mouse.

**Keywords:** Pointing devices, ISO 9241, Fitts' law, performance evaluation, eye movement, eye tracking.

## 1 Introduction

### ISO 9241 - Part 9

Beginning with the Apple *Macintosh* in 1984, graphical user interfaces (GUIs) have evolved and matured. The key feature of modern GUIs is the ability for users to interact with simple point-and-select operations. The most common pointing device in desktop systems is the mouse. To select an on-screen target with a mouse, a user manipulates the mouse to maneuver the cursor to a target, then selects the target by pressing and releasing a button. Simple as this seems, the interaction is even simpler with an eye tracker. The user locates the target by looking at it and follows immediately with selection [10].

Although considerable research exists in eye tracking [3], [6], [9], [10], none has evaluated eye tracking with the *ISO 9241 Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices*. ISO 9241-9 establishes uniform guidelines and testing procedures for evaluating computer pointing devices. The metric for comparison is *Throughput*, in bits per second (bits/s), which includes both the speed and accuracy of users'

performance. The equation for throughput is Fitts' Index of Performance except using an *effective* index of difficulty ($ID_e$). Specifically,

$$Throughput = ID_e / MT .\tag{1}$$

where *MT* is the mean movement time, in seconds, for all trials within the same condition, and

$$ID_e = log_2(D / W_e + 1) .\tag{2}$$

$ID_e$, in bits, is calculated from *D*, the distance to the target, and $W_e$, the effective width of the target. $W_e$ is calculated as

$$W_e = 4.133 \times SD .\tag{3}$$

where *SD* is the standard deviation in the selection coordinates measured along the line from the center of the home square to the center of a target. Using effective width allows throughput to incorporate the spatial variability in human performance. It includes both speed and accuracy [5].

**Prior Evaluations**
ISO 9241-9 was in Draft International Standard form in 1998 and became an International Standard in 2000. If one considers mouse evaluations in research not following the standard, throughput ranged from about 2.6 bits/s to 12.5 bits/s. On the contrary, studies conforming to the standard reported throughput from about 3.7 bits/s to 4.9 bits/s [8]. The data appear much more uniform and consistent. In short, ISO 9241-9 improves the quality and comparability of device evaluations.

Although several papers follow ISO 9241-9 and dozens of others use Fitts' law to evaluate non-keyboard input devices, Ware and Mikaelian published in 1987 what remains the only Fitts' law evaluation of an eye tracking system [10]. They used a serial Fitts' law task to test three eye tracking techniques. Task completion time was the only performance measure used. They compared eye tracking with the mouse but did not calculate or report on throughput as a performance measure. No eye tracking evaluation paper has ever been published since then using Fitts' law (or ISO 9241-9).

By following the standard and comparing throughput for eye tracking with a baseline technique (i.e., a mouse), we can determine how good an eye tracking system is. This paper is the first eye tracking evaluation conforming to ISO 9241-9.

The rest of this paper is organized as follows. In section 2, we described the methodology of our experiment. In section 3, the results are presented and discussed. Finally, we present our conclusions in section 4.

## 2   Methodology

An experiment was designed to implement the performance and comfort elements of ISO 9241-9. Effort was not tested since we do not have the sophisticated equipment necessary for measuring biomechanical load.

Performance testing was limited to pointing and selecting using multi-directional point-and-select tasks following ISO 9241-9 [2]. The testing environment was modeled on Annex B in the ISO standard [4]. Comfort was evaluated using the ISO

"Independent Rating Scale". The design followed, as reasonably as possible, the description in Annex C [4].

**Participants**
Sixteen paid volunteer participants (11 male, 5 female) were recruited from the local university campus. Participants ranged from 22 to 33 years (*mean* = 25). All were daily users of computers, reporting 4 to 12 hours usage per day (*mean* = 7). None had prior experience with eye tracking. All participants had normal vision, except one who wore contact lenses. Nine participants were right-eye dominant, seven left-eye dominant, as determined using the eye dominance test described by Collins and Blackwell [1].

**Apparatus**
A head-fixed eye tracking system, *ViewPoint*™ from ArringtonResearch, served as the input device (Fig. 1). The measurement method was Pupil and Corneal Reflection for greater tolerance to head movements. The infrared camera was set to focus on a participant's dominant eye. The monitor was a 19-inch 1280 x 1024 pixel LCD. Participants sat at a viewing distance of approximately 60 cm. The eye tracker sampled at 30 Hz with an accuracy of 0.25° - 1.0° visual arc, or about 10 – 40 pixels with our configuration. Calibration was performed before the first technique involved with the eye, with re-calibration as needed. Raw eye data and event data were collected and calculated using experimental software developed in our laboratory.



**Fig. 1.** Eye Tracking System

**Procedure**
The main independent variable was Interaction Technique with four levels:

- ETL    - Eye Tracking Long
- ETS    - Eye Tracking Short
- ESK    - Eye+Spacebar
- M      - Mouse

The ETL technique required participants to look at an on-screen target and dwell on it for 750 ms for selection. The dwell time was 500 ms for the ETS technique. The ESK technique allowed participants to "point" with the eye and "select" by pressing the spacebar upon fixation. To minimize asymmetric learning effects, the four interaction techniques were counterbalanced using a 4 × 4 balanced Latin square [7].

There were three additional independent variables. These were included to ensure that the trials covered a reasonable range of difficulties and to collect multiple sample points for each condition:

- Target Width:    75 pixels, 100 pixels
- Distance:        275 pixels, 350 pixels
- Trial:           1 to 16

Target width was the diameter of the circle target. Distance was the radius of the big circle, which was the distance from the center of the home square to the center of the circle target. These four conditions and the desired target were randomized. For each of the four conditions, the task involved 16 circle targets (Fig. 2). The total number of trials was 4096 (16 participants × 4 interaction techniques × 2 distances × 2 widths × 16 trials).

At the onset of each trial, a home square appeared on the screen. The home square allowed the distance of eye movement for each trial to be approximately the same. The home square disappeared after participants dwelled on it, pressed the spacebar, or clicked the left mouse button depending on the interaction technique. To exclude physical reaction time, positioning time started as soon as the eye or mouse moved after the home square disappeared. A window of 2.5 seconds was given to complete a trial after the home square disappeared. If no target selection occurred within 2.5 seconds, a time-out error was recorded. Then, the next trial followed.
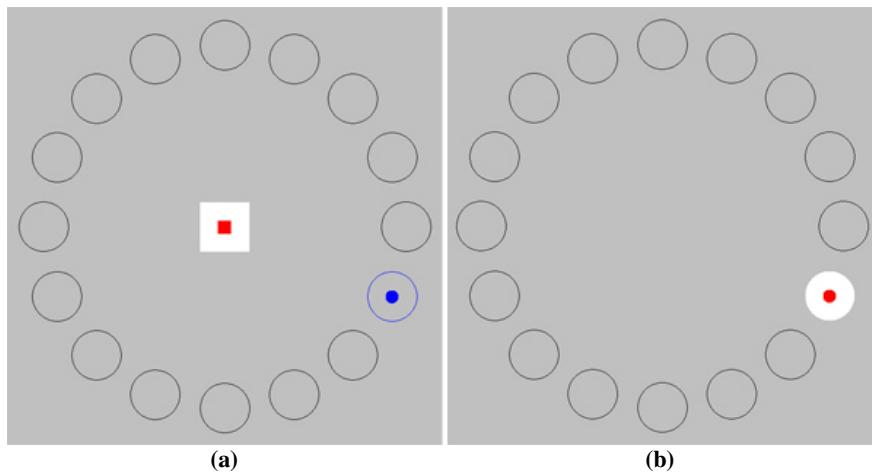


**(a)**                **(b)**

**Fig. 2.** Multi-directional Fitts' law task (2D Fitts discrete task). (a) Home square in focus (red dot with white background). Current target not in focus (blue dot with blue outline). (b) Home square disappeared (time started). Current target in focus (red dot with white background).

To minimize visual reaction time, the desired target was highlighted as soon as the participant fixated on the home square (Fig. 2a). The current target showed a blue dot when not in focus and a red dot when in focus (Fig. 2b). The dot helped participants fixate at the center of the target. The gray background was designed to reduce the eye stress caused by a bright color, such as a white background. For all three eye techniques, the mouse pointer was hidden to reduce visual distraction.

Participants were instructed to point to the target as quickly as possible (look at the target or move the mouse depending on the interaction technique), and select the target as quickly as possible (dwell on the target, press spacebar, or click the left mouse button depending on the interaction technique). After finishing the trials, we interviewed each user and gave a questionnaire.

## 3   Results and Discussion

**Throughput**

As evident in Fig. 3, there was a significant effect of interaction technique on throughput ($F_{3,45} = 47.46$, $p < .0001$). The 500 ms dwell time of the ETS technique seemed just right. Too short and participants accidentally selected the wrong target; too long and participants became impatient while waiting for selection. ETL had a lower throughput than ETS due to this. ESK was the best among the three eye tracking techniques. We attribute this to participants effectively pressing the spacebar immediately upon fixation on the target. This eliminated the need to wait for selection. The throughput of the ESK technique was 3.78 bits/s, which was close to the 4.68 bits/s for the Mouse. Considering the mouse has the best performance among non-keyboard input devices [8], the ESK technique is very promising. As the user
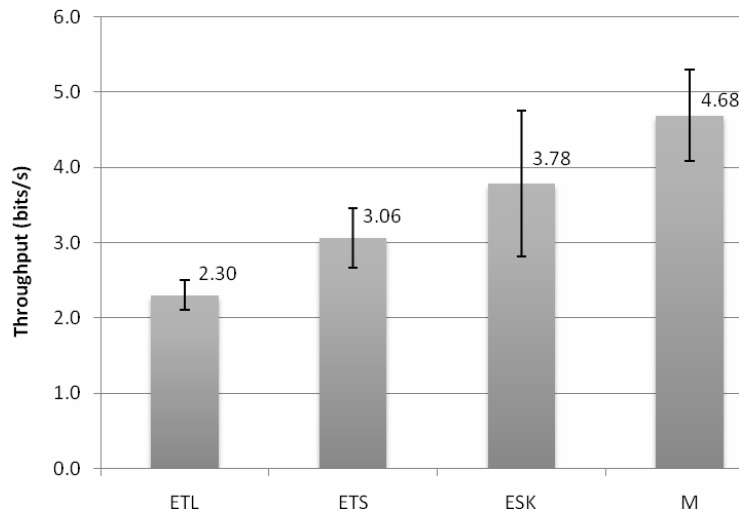


**Fig. 3.** Throughput as a function of interaction technique

must press the spacebar (or other key), this observation is qualified by noting that the ESK technique is only appropriate where an additional key press is possible and practical.

**Point-select Time**

Point-select time is the sum of the positioning time and the selection time. As shown in Fig. 4, the point-select time of the ESK technique was significantly lower than for the other interaction techniques ($F_{3,45} = 60.82$, $p < .0001$). A post hoc multiple comparisons test was performed using the Student-Newman-Keuls method. This revealed significance at the $p = .05$ level for all six comparisons except ETS vs. Mouse.
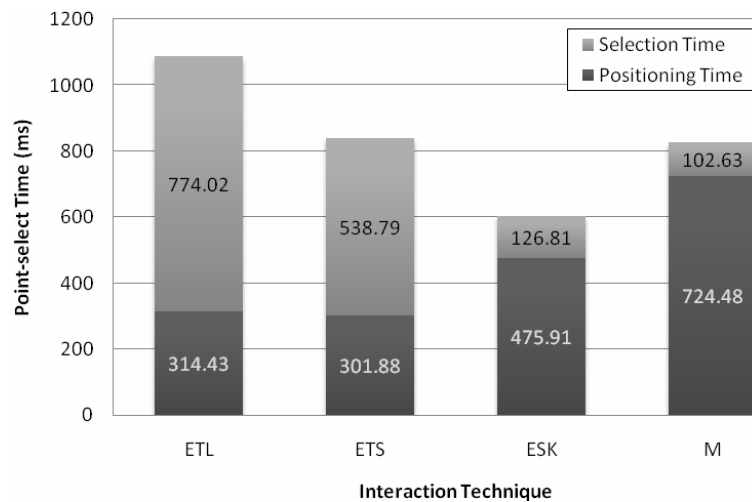


**Fig. 4.** Point-select Time as a function of interaction technique

**Error Rates and Time-out Error**

For the ETL and ETS techniques, participants selected the target by dwelling on it. Thus, the outcome was either a selection or a time-out error. Therefore, the error rate for ETL and ETS were both zero, as shown in Fig. 5. Time-out errors for the ETL, ETS and ESK techniques were mainly caused by eye jitter and eye tracker accuracy. The longer time needed to perform a selection, the higher chance there would be a time-out error. ESK had 2.89% time-out errors, which much closer to the 1.07% time-out errors for the mouse, compared with the other eye tracking techniques.

Although ESK yielded the fastest point-select time, as aforementioned, it suffered from a high error rate. This is a classic speed-accuracy tradeoff and we attributed it to participants pressing the spacebar slightly before fixating on the target, or slightly after the eye moved out of the target. Because no participant had prior experience with eye tracking, few could do the coordinated work of eye pointing and hand pressing of the spacebar very well. The error rate for the ESK technique varied a lot across participants (*standard deviation* = 11.43, *max* = 35.59, *minimum* = 3.13). We
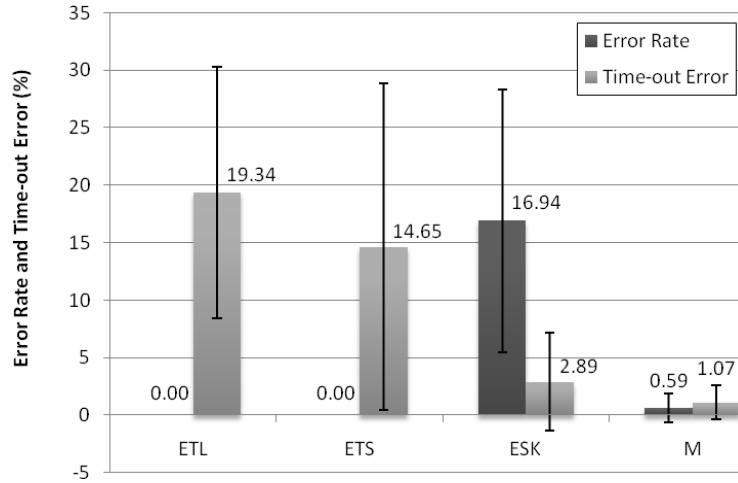
**Fig. 5.** Error Rate and Time-out Error as a function of interaction technique

believe participants could have much lower error rates if further training was provided and improved feedback mechanisms were considered and tested.

**Target Width**

As we analyzed all data, an interesting finding surfaced. The width of a target can affect the error rate and time-out error. For the ETL, ETS and ESK techniques shown in Fig. 6, the time-out error of the large-width target was generally better than for the small-width target. For the ETS and ESK techniques, the difference was substantial, with
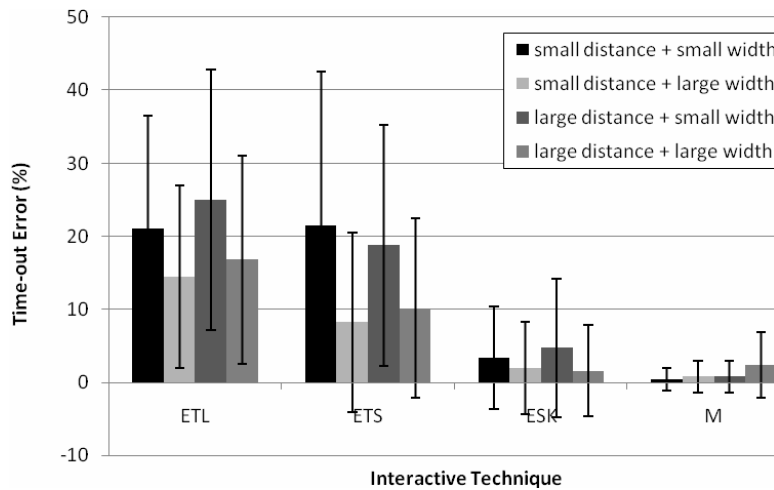


**Fig. 6.** Time-out Error as a function of interaction technique, target width, and distance

about 50% fewer time-out errors for the large-width targets than for the small-width targets. We observed a similar pattern as in error rate. We also found that although a larger target width can help reduce errors, it had little impact on throughput or point-select time.

**Questionnaire**

The device assessment questionnaire consisted of 12 questions. The questions pertained to eye tracking in general, as opposed to a particular eye tracking interaction technique. Each response was rated on a seven-point scale, with 7 as the most favorable response, 4 the mid-point, and 1 the least favorable response. Results are shown in Fig. 7.
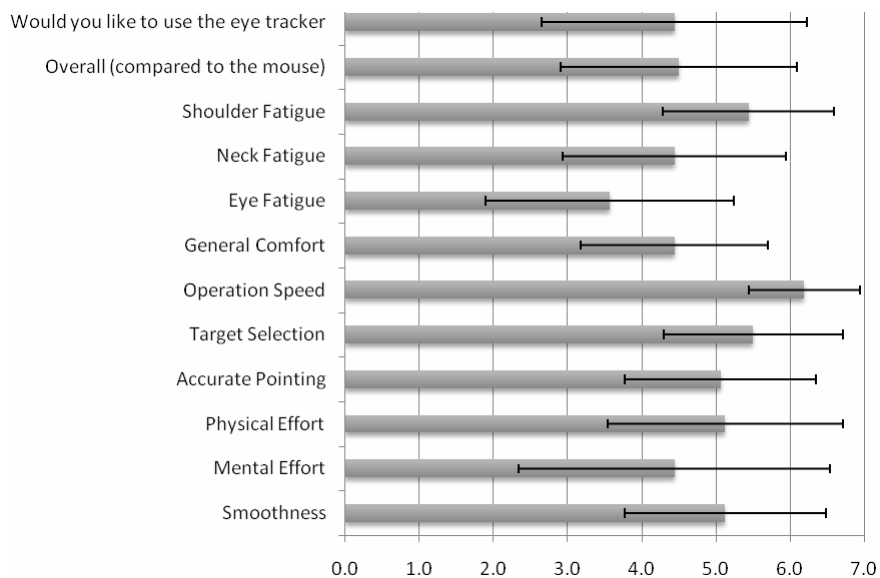


**Fig. 7.** Eye tracker device assessment questionnaire. Response 7 was the most favorable, response 1 the least favorable.

As seen, participants generally liked the fast positioning time of the eye tracker. On Operation Speed, the mean score was high at 6.2. However, Eye Fatigue was a concern. Participants complained that staring at so many targets made their eyes dry and uncomfortable. Eye Fatigue scored lowest among all the questions. Neck Fatigue and Shoulder Fatigue were also an issue, since the eye tracking system we tested was head-fixed. Participants gave eye tracking a favorable response overall of 4.5, just slightly higher than the mid-point (see top two entries in Fig. 7). Discussions following the experiment revealed that participants liked to use eye tracking and believed it could perform similar to the mouse. Of the three eye tracking techniques,

participants expressed a preference for the Eye+Spacebar technique. Concerns were voiced, however, on the likely expense of eye tracking system, the troublesome calibration procedure, and uncomfortable need to maintain a fixed head position.

## 4 Conclusion

This paper is the first eye tracking evaluation conforming to ISO 9241-9. Four point-select interaction techniques were evaluated, three involving eye tracking and one using a standard mouse. The Eye Tracking Long technique yielded a lower throughput than the Eye Tracking Short technique. The Eye+Spacebar technique was the best among the three eye tracking interaction techniques. It had a throughput of 3.78 bits/s, which was close to the 4.68 bits/s for the Mouse. Participants generally liked the Eye+Spacebar technique.

More work is planned to determine the best settings for eye tracking, for example, the optimal target size and color highlighting. In the future, we intend to evaluate eye tracking in a longitudinal study and in text entry applications.

## References

1. Collins, J.F., Blackwell, L.K.: Effects of eye dominance and retinal distance on binocular rivalry. Perceptual and Motor Skills 39, 747–754 (1974)
2. Douglas, S.A., Kirkpatrick, A.E., MacKenzie, I.S.: Testing pointing device performance and user assessment with the ISO 9241, Part 9 standard. In: Proceedings of the ACM Conference on Human Factors in Computing Systems - CHI '99, New York, ACM, pp. 215–222 (1999)
3. Hennessey, C., Noureddin, B., Lawrence, P.A: single camera eye-gaze tracking system with free head motion. In: Proceedings of the Symposium on Eye Tracking Research and Applications – ETRA 2006 New York, ACM, pp. 87–94 (2006)
4. Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices. International Standard, International Organization for Standardization (2000) ISO. ISO/DIS 9241-9
5. MacKenzie, I.S.: Fitts' law as a research and design tool in human-computer interaction. Human-Computer Interaction 7, 91–139 (1992)
6. Majaranta, P., MacKenzie, I.S., Aula, A., Räihä, K.-J.: Effects of feedback and dwell time on eye typing speed and accuracy. Universal Access in the Information Society (UAIS) 5, 199–208 (2006)
7. Martin, D.W.: Doing psychology experiments, 6th edn. Wadsworth Publishing, Belmont, CA (2004)
8. Soukoreff, R.W., MacKenzie, I.S.: Towards a standard for pointing device evaluation: Perspectives on 27 years of Fitts' law research in HCI. International Journal of Human-Computer Studies 61, 751–789 (2004)

9. Wagner, P., Bartl, K., Günthner, W., Schneider, E., Brandt, T., Ulbrich, H.: A pivotable head mounted camera system that is aligned by three-dimensional eye movements. In: Proceedings of the Symposium on Eye Tracking Research and Applications – ETRA 2006, New York, ACM, pp. 117–124 (2006)
10. Ware, C., Mikaelian, H.H.: An evaluation of an eye tracker as a device for computer input. In: Proceedings of the ACM Conference on Human Factors in Computing Systems – CHI+GI '87 New York, ACM, pp. 183–188 (1987)