

Performance-Based Eye-Tracking Analysis in a Dynamic Monitoring Task

Wei Du¹ and Jung Hyup Kim²(✉)

¹ Service Standards and Development Department,
Air China Cargo Co. Ltd., Beijing, China
uwei@airchinacargo.com

² Department of Industrial and Manufacturing Systems Engineering,
University of Missouri, Columbia, MO, USA
kijung@missouri.edu

Abstract. The goal of this study is to explore how ocular behavior is different in groups that possessed varying levels of performance in dynamic control tasks with complex visual components. Twenty two university students participated in this study by operating a human-in-the-loop (HITL) simulator. The participants were asked to identify unknown air track(s) and take proper actions to defend a battleship. During the experiment, a head-mounted eye-tracking device was used continuously to record participants' visual attention span regarding the normalized coordinates of their gaze points. In the current study, fixation duration was the main eye-tracking metrics. Air track identification accuracy and the NASA Task Load Index (NASA-TLX) were also used to measure participants' task performance and overall subjective mental workload.

Keywords: Mental workload · Cognitive modeling · Eye tracking analysis · Task performance · Human-in-the-loop simulation

1 Introduction

Nowadays, many complex tasks, including tasks related to monitoring the refinery process, as well as military command and control tasks, involve the use of visual information, such as display gauges and computer monitors. To carry out the responsibilities of these tasks, individuals need to conduct efficient visual searches and locate the relevant information quickly and precisely. Due to the characteristics of visual searching tasks, knowing individuals' visual attention span and their cognitive abilities is essential in evaluating task performance and display interface usability. Thus, the eye movement has become the focus of more and more studies in the last few decades. The eye-tracking technology has been successfully applied in many research areas to examine human visual attention and physiological change. The main reason that the eye-tracking method has been so widely adopted is that it provides evidence on human information processes in several aspects by using different measurements, such as fixations, saccades, scan path, pupil diameters, and so on. In our study, an experiment with the eye-tracking technology was conducted to collect the participants' eye-tracking data including eye fixation (when and where a person is looking at) and

eye fixation transition (the sequence in which their eyes are shifting from one location to another). The purpose of this study was to advance our understanding of eye movement and to explore the feasibility of various eye-tracking metrics in the visual searching task to identify differences in eye-tracking data across different performance-based groups. These quantifiable differences may be used in several ways. First, they help to further our understanding of the participants' problem-solving behaviors. Second, ocular behaviors may yield some insights into the perceptual process patterns of the participants with a high-performance level. In other words, using eye-tracking devices to identify the differences in perceptual process patterns would be advantageous in developing the appropriate manner to solve problems in some cases. Besides, the level of the participants' subjective mental workload during the task procedure was measured by using the NASA-TLX questionnaire as well. The workload result was used to develop the relationship between the participants' performance and the mental demands placed on the participants by a task.

2 Literature Review

2.1 Eye-Tracking

Eye-tracking technologies typically collect two types of eye movement information: location (where fixations tend to be directed) and duration (how long they typically remain there) within specific areas. Although it was found that human ocular behaviors showed various patterns when inspecting visual scenes, they can still be modulated, depending on cognitive demand and the characteristics of the scene (Yarbus 1967). Therefore, an eye's gaze can be considered an unbiased indicator of the focus of visual attention and the eye-tracking method has been successfully utilized to study human behavior in a wide range of research domains, such as reading (Hyönä and Niemi 1990; Rayner et al. 2006), program comprehension (Bednarik and Tukiainen 2006; Crosby and Stelovsky 1990), arithmetic problem solving (Andrá et al. 2013; Hegarty et al. 1992), multimedia learning (Tsai et al. 2012; van Gog and Scheiter 2010), and driving (Palinko et al. 2010), human-computer interaction (Granka et al. 2004; Jacob and Karn 2003). A review of the literature on eye-tracking revealed that researchers developed a wide variety of eye tracking metrics. However, one of the most commonly reported eye-tracking metrics in previous studies is fixation. There are also several derived metrics that stem from these basic measures, such as saccade, dwell (also known as gaze), scan path, and pupillary responses, such as the changes in pupil size (Jacob and Karn 2003; Poole and Ball 2005).

2.2 Fixation

Fixations, which are widely used eye tracking matrices, are commonly defined as moments in time when the eyes stay relatively stationary, taking in or "encoding" visual information. Fixation duration is one of the common metrics, both of which can be interpreted differently depending on the characteristics of the tasks. A longer duration indicates a greater difficulty in interpreting information from the object being

fixated, or it may also mean that the participant is more engaging (Just and Carpenter 1976). Also, fixation duration varies as a function of the particular task such as reading, visual search and typing (Rayner 1998). It can range from 60 to 500 ms being about 250 ms on average (Liversedge and Findlay 2000). Many previous eye-tracking studies employed fixation-derived metrics and provided interesting insights into the problem-solving process. Kun et al. (Guo et al. 2006) compared the cognitive processes for inspecting several visual scenes with different characteristics and found that the face and natural images attracted similar numbers of fixations while the viewing of faces was accompanied by longer fixation durations compared with natural scenes, which provided supportive evidence to the arguments that face perception is involved in a unique cognitive process compared with non-face object or scene perception. Bednarik et al. (2006) presented a study of the comprehension processes of programmers with the help of a remote eye-tracking device. A significant difference was found on the mean fixation durations over the main areas of interest, which, they believed, indicated levels of difficulty in comprehending different parts of a program. In their study, longer durations indicated more difficulties during cognitive processing. A more recent research done by Tsai et al. (2012) examined the students' visual attention when solving a multiple-choice problem. Based on the analysis, they suggested that the relevant information was fixated longer than the irrelevant one, and participants paid the most attention (longer total fixation durations) to their preferred answers before making decisions. Table 1 is a summary of the main fixation-derived metrics.

Table 1. Fixation-derived metrics with interpretations from literatures

Eye movement metric	Interpretation	Reference
Fixation duration mean	Longer durations indicate a greater difficulty in interpreting information from the object	Just and Carpenter (1976); Jacob and Karn (2003)
	Longer durations mean that the object is more engaging in some way	Just and Carpenter (1976)
Total fixation number	More overall fixations indicate more searching or less efficient search	Goldberg and Kotval (1999)
Fixation number per area of interest	A higher number of fixations indicates more importance or a more noticeable area in problem-solving tasks	Poole et al. (2005)
	A higher number of fixations means that an item is actually harder to recognize in reading tasks	Poole et al. (2005)

2.3 Workload Measurement

Currently, applied research has paid much attention to the human workload study to ensure high levels of safety, health, and comfort and the long-term productive efficiency of the operator. Eggemeier (1988) defined mental workload as “the degree of

processing capacity that is expanded during task performance”; in other words, the mental workload is a specification of the amount of information processing capacity that is used for task performance. There are several factors that are said to be related to mental workloads from the both task side (such as complexity, the difficulty of the task) and the individual side (such as effort expended by the operator) (De Waard and Studiecentrum 1996). Thus, a multidimensional measurement method is needed to scale the workload accurately. A number of tools for the evaluation and prediction of workload are available such as the National Aeronautics and Space Administration – Task Load Index (Hart and Staveland 1988), the Subjective Workload Assessment Technique (Reid and Nygren 1988), and the Workload Profile (Tsang and Velazquez 1996). Among these workload assessment instruments, NASA-TLX, which is a multifaceted tool for assessing subjective workload, has been applied in many domains and is widely accepted as one of the strongest tools for reporting perceptions of workload.

3 Method

In this study, the anti-air warfare coordinator (AAWC) human-in-the-loop test bed was used (Kim et al. 2015; Macht et al. 2014). Within this interactive simulator, participants handled identifying the unknown air tracks based on the engagement rules and the air tracks’ information that they gathered from the radar simulation system. The main reason for choosing this test platform was that it was a relatively complex task with dynamic visual components. Also, some studies have contributed to understanding the eye movement behaviors using this platform or other similar platforms. Thus, further evidence was needed to support the reliability of eye-tracking metrics in complex dynamic tasks such as the AAWC simulation (Fig. 1).



Fig. 1. Experimental environment

The twenty-two participants were divided into three groups based on their performance: high-accuracy group, medium-accuracy group, and low-accuracy group. Analysis of the data revealed some statistical changes in ocular behavior and mental workload among the three groups.

4 Data Collection

4.1 AAWC Simulation Task Performance

The accuracy of the identified unknown air tracks was calculated to determine the levels of the participants’ AAWC simulation task performance. As discussed above, we only focused on the performance of identification task in the current study. Thus, if and only if both the primary identification (such as friendly, hostile, and etc.) and the type identification (such as strike, commercial aircraft, helicopter and etc.) of an unknown air track were identified correctly before the end of a scenario, the result would be accounted as a right identification (Kim 2014; Kim et al. 2011).

$$\text{Accuracy} = \frac{\text{Total number of correctly identified air tracks}}{\text{Total number of unknown air tracks in the scenario}} \quad (1)$$

4.2 Area of Interest (AOI)

The AOI refers to the specified interested areas in the visual field in which the gaze point lies. Five AOIs were defined for further data analysis (See Fig. 2). They were radar screen AOI, menu bar AOI, data panel AOI, track profile AOI, and EWS AOI. Although being important components of the AAWC simulation task, some areas were defined as outside areas and were excluded from eye-tracking analysis for the following reasons. The AAWC simulator guide board was excluded due to the small amount of attention allocation. The air track behavior board area and the system response panel area were excluded due to the low eye-tracking accuracy on these two areas. Also, since the purpose of this experiment was to explore the differences on information process, the data input panel area was excluded as well as it did not contain much relative information that can be used to perceive the situation of the air space.

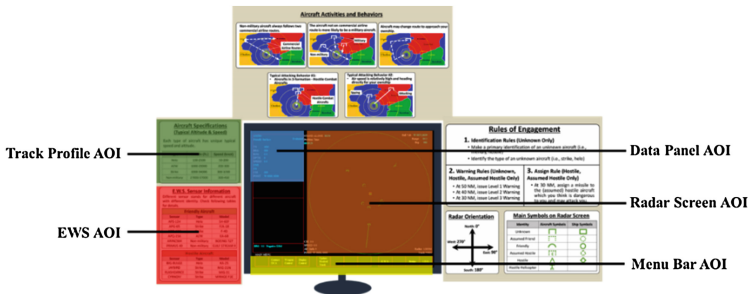


Fig. 2. Environment layout and areas of interest (AOIs)

4.3 Eye-Movement Data

The raw data recorded by the eye-tracking system were typically in terms of the normalized coordinates. From this raw data, fixations were calculated and derived. In this study, fixation occurred when a participant's gaze stabilized over at least 100 ms, and the visual angle degree was less than 1°. The fixation duration was calculated as the entire period of the fixation, in other words, the difference between stop and start time of the fixation.

5 Results

5.1 AAWC Simulation Task Performance

Although the participants went through the same experiment training phases, their mastering of the knowledge and skills were varied. The analysis of the participants' identification accuracy was performed to determine the level of performance. According to the identification result, twenty-two participants were divided into three groups with the different performance levels: the high-accuracy group (8 participants), the medium-accuracy group (7 participants), and the low-accuracy group (7 participants).

The followings are the criteria for grouping.

- High-Accuracy Group: The average accuracy of the two day's experiment was no less than 75 % and the accuracy on each day was no less than 70 % (This criteria filtered out about top 25 % participants with relatively high performance).
- Medium-Accuracy Group: The average accuracy of the two day's experiment was between 50 % and 65 % and none of the accuracy on each day was lower than 40 % or higher than 70 % (This criteria filtered out about medium 25 % participants with relatively medium performance).
- Low-Accuracy Group: The average accuracy of the two day's experiment was no more 45 % and the accuracy on each day was no more than 50 % (This criteria filtered out about bottom 25 % participants with relatively low performance).

5.2 Fixation Durations on AOIs

In order to determine the information processing time on the AOIs, we investigated the fixation duration time. The AAWC test platform consisted of five areas of interest (AOIs), which were radar screen AOI, menu bar AOI, data panel AOI, track profile AOI, and EWS AOI. Eye-tacking data was filtered out when eye gaze point fell outside these areas. Thus, fixation duration time was measured only for these five discrete AOIs. The results of fixation durations on each AOI are presented in Table 2.

Table 2. Mean fixation durations and standard deviations (in parentheses) over the main AOIs for high-accuracy, medium-accuracy and low-accuracy groups

Group	Radar screen	Data panel	Track profile	EWS	Menu bar
High-accuracy	0.35749 (0.38364)	0.30281 (0.25152)	0.23630 (0.13387)	0.21711 (0.11929)	0.29523 (0.20356)
Medium-accuracy	0.37653 (0.40092)	0.30621 (0.23886)	0.23693 (0.14868)	0.22769 (0.12502)	0.30204 (0.21927)
Low-accuracy	0.37107 (0.39057)	0.34800 (0.35655)	0.24653 (0.16285)	0.22862 (0.12380)	0.30735 (0.22311)

5.3 NASA-TLX

According to the assessment questionnaire, the mean of the overall weighted NASA-TLX score for the high-accuracy group was 42.77 (SD = 15.60), which was the lowest among all the three groups. The mean of the overall weighted NASA-TLX score for medium-accuracy group was of the medium level, which was 49.54 (SD = 16.55). However, as for the low-accuracy group, the overall weighted NASA-TLX score increased up to 56.59 (SD = 19.78) which was higher than any other groups. The ANOVA analysis showed that the three groups' NASA-TLX scores were significantly different ($F(2,129) = 7.24$, $p < 0.001$) (Table 3). The medium-accuracy group and the low-accuracy group's scores were significantly higher than the high-accuracy group's (all p -values < 0.05). However, no significant difference was found between medium-accuracy group and low-accuracy group ($p = 0.076$).

Table 3. NASA-TLX results of high-accuracy, medium-accuracy and low-accuracy groups

Group	Mean NASA TLX	Standard deviation	F value	P value
High-accuracy	42.77	15.60	$F(2,129) = 7.24$	0.001
Medium-accuracy	49.54	16.55		
Low-accuracy	56.69	19.78		

6 Discussion and Conclusion

By analyzing the result, we found that the fixation duration varied systematically as a function of the five different AOIs as expected. It means that the fixation duration is linked to the human's processing time. A longer duration indicates that more effort is expended in interpreting information during a problem-solving process. We discovered that a comparison of fixation duration data produced clear differences corresponding to the known levels of the performance in our study. It indicates that the processing time to comprehend information from different AOIs varied a lot and obviously followed the same trend across all three groups. The longest mean fixation duration belonged to the radar screen AOI, and the second-longest fixated AOI was the data panel AOI followed by the menu bar AOI, while the EWS AOI and the track profile AOI demanded the lowest mean fixation durations. This can be related to the different difficulty levels for

participants to encode the information from each AOI. Participants demanded the longest fixation duration on the radar screen AOI, which could be explained by the fact that this AOI displayed only graphic symbols, which indicated a lot of implicit information including current track identification, direction, speed, location, and behavior. It had a longer solution path and required more effort in order to extract the useful information from this AOI. As for the data panel AOI, which displayed the air track's parameters in digits, such as altitude and speed, the information on this AOI was much more straightforward compared with the information on the radar screen AOI. It saved participant effort in encoding implicit information from the symbols on the radar screen AOI. Additionally, the information on the EWS AOI and track profile AOI was the easiest to process due to the fact that these only displayed the direct correspondence between the track's real identification and the typical parameter/sensor code/track model, which required no further processing.

In addition, a significant difference on NASA-TLX scores was observed across the groups with different performance levels. It was found that participant performance had an obvious negative correlation with the NASA-TLX score. As the group's identification accuracy significantly increased, the NASA-TLX score decreased. This relationship revealed the fact that the high-accuracy group experienced a relatively low overall mental workload when performing the radar monitoring task. In other words, the mental demands expended during the task were lower for those with higher performances. The result can possibly be explained by the fact that if participants had a stably high performance, which means they could understand the task well, be aware of the situation correctly and interact with the radar simulator step by step in an orderly manner, they experienced less mental workload and might have felt more relaxed during the tasks. However, if participants had relatively worse performances, which means they experienced more difficulties when struggling in the task, they might be more stressful and expended higher mental workload. In a nutshell, the NASA-TLX could be used as a supportive source for measuring performance levels. However, there is a limitation on NASA-TLX. Though the NASA-TLX scores showed an apparent inverse pattern from the performance level, the difference in NASA-TLX scores between the medium-accuracy group and the high-accuracy group was not significant. It means that this subjective response might not precisely correspond to the participants' experience. In addition, NASA-TLX is a post-session subjective assessment. Hence, it could not account for rapid changes in mental workload.

In this research, we used the anti-air warfare coordinator (AAWC) radar simulator as the experiment test bed to explore how ocular behavior is different in high-accuracy, medium-accuracy, and low-accuracy groups. The findings suggested that eye tracking data may potentially be a reliable source to identify differences in problem-solving behaviors among performance-based groups in several aspects, and may provide insights into the cognitive process to interpret participant performance further.

References

- Andrá, C., Lindström, P., Arzarello, F., Holmqvist, K., Robutti, O., Sabena, C.: Reading mathematics representations: an eye-tracking study. *Int. J. Sci. Math. Educ.* **13**(2), 237–259 (2013)
- Bednarik, R., Tukiainen, M.: An eye-tracking methodology for characterizing program comprehension processes. Paper Presented at the Proceedings of the 2006 Symposium on Eye Tracking Research and Applications (2006)
- Crosby, M.E., Stelovsky, J.: How do we read algorithms? A case study. *Computer* **23**(1), 25–35 (1990)
- De Waard, D., Studiecentrum, V.: The measurement of drivers' mental workload. Groningen University, Traffic Research Center Netherlands (1996)
- Eggemeier, F.T.: Properties of workload assessment techniques. *Adv. Psychol.* **52**, 41–62 (1988)
- Granka, L.A., Joachims, T., Gay, G.: Eye-tracking analysis of user behavior in WWW search. Paper Presented at the Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (2004)
- Goldberg, J.H., Kotval, X.P.: Computer interface evaluation using eye movements: methods and constructs. *Int. J. Ind. Ergonomics*, **24**(6), 631–645 (1999)
- Guo, K., Mahmoodi, S., Robertson, R.G., Young, M.P.: Longer fixation duration while viewing face images. *Exp. Brain Res.* **171**(1), 91–98 (2006)
- Hart, S.G., Staveland, L.E.: Development of NASA-TLX (task load index): results of empirical and theoretical research. *Adv. Psychol.* **52**, 139–183 (1988)
- Hegarty, M., Mayer, R.E., Green, C.E.: Comprehension of arithmetic word problems: Evidence from students' eye fixations. *J. Educ. Psychol.* **84**(1), 76 (1992)
- Hyönä, J., Niemi, P.: Eye movements during repeated reading of a text. *Acta Psychol.* **73**(3), 259–280 (1990)
- Jacob, R., Karn, K.S.: Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind* **2**(3), 4 (2003)
- Just, M.A., Carpenter, P.A.: Eye fixations and cognitive processes. *Cogn. Psychol.* **8**(4), 441–480 (1976)
- Kim, J.H.: Simulation training in self-regulated learning: investigating the effects of dual feedback on dynamic decision-making tasks. In: Zaphiris, P., Ioannou, A. (eds.) *LCT 2014, Part I. LNCS*, vol. 8523, pp. 419–428. Springer, Heidelberg (2014)
- Kim, J.H., Chan, T., Du, W.: The learning effect of augmented reality training in a computer-based simulation environment. In: Zaphiris, P., Ioannou, A. (eds.) *LCT 2015. LNCS*, vol. 9192, pp. 406–414. Springer, Heidelberg (2015)
- Kim, J.H., Rothrock, L., Tharanathan, A., Thiruvengada, H.: Investigating the effects of metacognition in dynamic control tasks. In: Jacko, J.A. (ed.) *Human-Computer Interaction, Part I, HCII 2011. LNCS*, vol. 6761, pp. 378–387. Springer, Heidelberg (2011)
- Liversedge, S.P., Findlay, J.M.: Saccadic eye movements and cognition. *Trends Cogn. Sci.* **4**(1), 6–14 (2000)
- Macht, G.A., Nembhard, D.A., Kim, J.H., Rothrock, L.: Structural models of extraversion, communication, and team performance. *Int. J. Ind. Ergon.* **44**(1), 82–91 (2014)
- Palinko, O., Kun, A.L., Shyrovkov, A., Heeman, P.: Estimating cognitive load using remote eye tracking in a driving simulator. Paper Presented at the Proceedings of the 2010 Symposium on Eye-Tracking Research and Applications (2010)
- Poole, A., Ball, L.J.: Eye tracking in human-computer interaction and usability research: current status and future. Prospects. In: Ghaoui, C. (ed.) *Encyclopedia of Human-Computer Interaction*. Idea Group, Inc, Citeseer, Pennsylvania (2005)

- Rayner, K.: Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* **124**(3), 372 (1998)
- Rayner, K., Chace, K.H., Slattery, T.J., Ashby, J.: Eye movements as reflections of comprehension processes in reading. *Sci. Stud. Read.* **10**(3), 241–255 (2006)
- Reid, G.B., Nygren, T.E.: The subjective workload assessment technique: a scaling procedure for measuring mental workload. *Adv. Psychol.* **52**, 185–218 (1988)
- Tsai, M.-J., Hou, H.-T., Lai, M.-L., Liu, W.-Y., Yang, F.-Y.: Visual attention for solving multiple-choice science problem: an eye-tracking analysis. *Comput. Educ.* **58**(1), 375–385 (2012)
- Tsang, P.S., Velazquez, V.L.: Diagnosticity and multidimensional subjective workload ratings. *Ergonomics* **39**(3), 358–381 (1996)
- van Gog, T., Scheiter, K.: Eye tracking as a tool to study and enhance multimedia learning. *Learn. Instr.* **20**(2), 95–99 (2010)
- Yarbus, A.: *Eye Movements and Vision*. Plenum Press, New York (1967)