

Comparing Pointing Performance of Mouse and Eye-Gaze Input System

Wenbin Guo^(✉) and Jung Hyup Kim

Department of Industrial and Manufacturing Systems Engineering,
University of Missouri, Columbia, USA
wgk95@mail.missouri.edu, kijung@missouri.edu

Abstract. The purpose of this study is to identify strengths and weaknesses of an eye-gaze input system. The pointing performance of an eye-gaze input, such as reaction time and accuracy, were compared to the performance of a mouse input. Operator's sensitivity (d'), bias (β) and NASA-TLX were used as metrics for performance assessment during a gauge monitoring task. The results indicated that there was a significant difference in reaction time between a mouse input and eye-gaze input system. The participants who used a mouse as their input device showed more miss responses. However, the eye-gaze device users had more false-alarm responses. For workload, there was no significant difference between the eye-gaze input system and the mouse input system. The findings from this study showed that an eye-gaze system is favorable to the operations requiring fast human reactions and less sensitivity on false. The study provided deeper insight into user behavior regarding an eye-gaze input system.

Keywords: Eye-gaze input system · Visual search task · Human-in-the-loop simulation · Human performance · And workload

1 Introduction

There are numerous input systems, such as a mouse, multi-touch, and voice, in human-computer interaction (Bai et al. 2014). Among them, an eye-gaze system is one of the advanced systems (Chandra et al. 2015) and become more natural and non-intrusive recently (Chen and Ji 2015). One of the benefits of using an eye-gaze input system is that the system can quickly reflect human cognitive processes. For that reason, many researchers have studied an eye-gaze input system and found the impacts on pointing performance using eye-gaze input system (Jacob and Karn 2003; Murata et al. 2014; Murata et al. 2012; Sibert et al. 2001). However, they have not considered the practical usability of the eye-gaze input system. For example, reaction time, accuracy and mental workload in a dynamic control environment have not yet been examined until now.

In this study, the pointing performance of an eye-gaze input system was compared to a mouse input system. The time window-based human-in-the-loop (TWHITL) simulation (Kim et al. 2014) was used as a tool to collect accurate participant responses corresponding to the simulation events. Operator bias β and sensitivity d' (Lynn and Barrett 2014; Maniscalco and Lau 2012) were calculated by using the participant's

outcomes to compare accuracy between an eye-gaze input and mouse input system. During the experiment, the participants' mental workload was also measured by NASA-TLX. It is one of the well-known metrics for the mental workload in a dynamic task environment (Bodala et al. 2014; De la Torre et al. 2014).

The key research question for this study was: "how do the participants perform the TWHITL simulation differently when they use an eye-gaze input system compared to a mouse input system?"

We hypothesized that there was no difference in reaction time between a mouse input and an eye-gaze input. The response times from both input systems were expected to be similar when the participants experienced same stimuli during the task. Second, there was no significant difference in accuracy between a mouse input and an eye-gaze input. Finally, there was a significant difference in mental workload between a mouse input and eye-gaze input system. The mental workload of using an eye-gaze input system is expected to be higher than the workload of using a mouse input system.

2 Method

2.1 Apparatus













For the experiment, we selected a desk-mounted eye tracker Tobii EyeX Dev Kit (see Fig. 1). There were three synced infrared sensors in the Tobii EyeX sensor. Based on the angle and glint of the operator's eye, Tobii Eye X device calculated where he or she is looking at during the experiment.



Fig. 1. Tobii EyeX Dev Kit

During the experiment, the visual image of flow, level, pressure, and temperature gauge shapes from the Visual Thesaurus as the underlying design of the gauge (Noah et al. 2014; Tharanathan et al. 2010). The participants were asked to acknowledge changes in the current state of all gauges by clicking the button corresponding to the gauge. Table 1 shows the more details of flow, level, pressure, and temperature gauge shape.

Table 1. Gauge shape and region

Event	Flow	Level	Pressure	Temperature
NORMAL				
ABNORMAL				
ALARM				

2.2 Participants

A total of twenty-four participants were recruited from the University of Missouri. Participants' age ranged from 18 to 28 (Mean = 21.4, SD = 1.92). First of all, participants were required to fill out a consent form and a demographic questionnaire. 90% of participants were right-handed, which mean that the majority of the participants used their right hand to use a mouse or keyboard. The participants were asked to use their input device to point and click alternately between the two presented target gauges as rapidly and accurately as they could.

2.3 Experimental Setup

There were two groups in this experiment (Group A: mouse input system and Group B: eye-gaze input system). For both groups, the subjects were seated roughly 60 cm in front of a 21.5" LCD monitor screen with a resolution of 1024×768 . The experimental setup is illustrated in Fig. 2.

The participants took a pre-test before the experiment to determine their computer experience level and verify their eyesight. After that, a nine-points calibration was performed for the participants who assigned in the Group B. So that, they could control the eye-gaze input system sit without discomfort. During the calibration, they should not move their head. After the calibration, they can move their heads gently. This process is good for the accuracy of the gaze and fixation.



Fig. 2. Experimental setup

2.4 Study Design

The TWHITL gauge monitoring simulation is an interactive, real-time system (Kim et al. 2014). During the test, if a gauge value went beyond the normal range, the user needed to respond the change by clicking the button corresponding to the gauge. For the Group A, they could click the button as same as when they used a normal mouse device. For the Group B, however, the participants must look at the gauge first and click “Alt” button on the keyboard. The green box (see Fig. 3) represented the current location of the eye-gaze cursor. The simulation began with a normal condition for all 44 gauges, which include 11 flow gauges, 11 level gauges, 11 temperature gauges and 11 pressure gauges. The width of each gauge was 200 pixels, and the amplitude between abnormal gauges was different based on the given scenario conditions. Multiple events were scheduled in each scenario. Each event was designed based on the index of difficulty (ID) ranges from 1.0 to 4.0. By using the simulation, we collected the reaction time, task performance and cognitive workload of both groups. The data was used to evaluate the participants’ ability to detect abnormal events in a continuous monitoring task.

As seen in Fig. 4, if a participant detected the abnormal condition gauge within the event duration time, the participant action would change the status of the gauge from an abnormal condition to a normal condition. This action was recorded as “Hit.” If a participant failed to detect the abnormal status of this gauge, it would move to the alarm condition. If the participant also failed to detect the alarm condition within the full event duration time, then this one would be recorded as “Miss.”

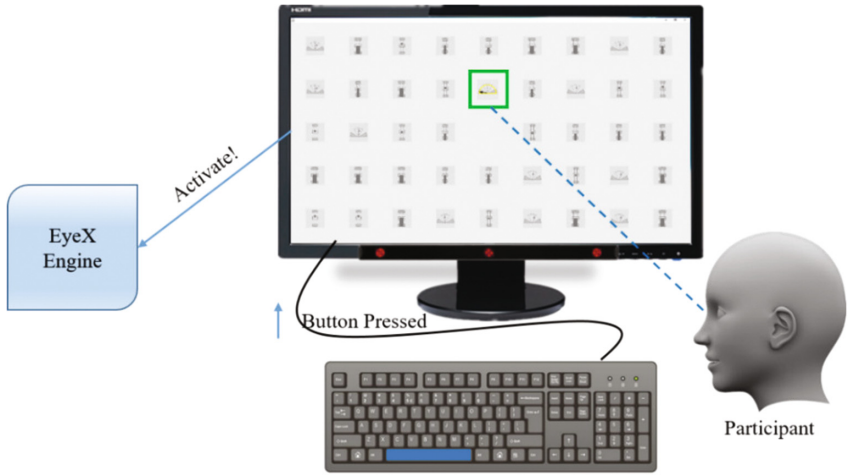


Fig. 3. Eye-gaze input human-in-the-loop simulation

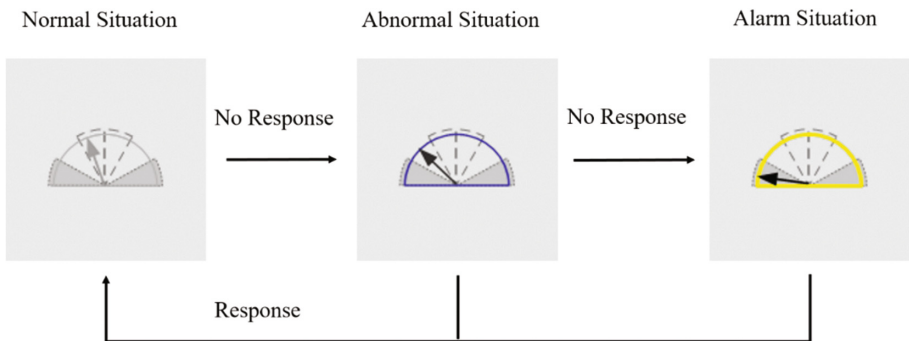


Fig. 4. Participant response

3 Procedure

The experiment began with a briefing on the TWHITL simulation. The purpose of this presentation was to explain the primary interface to the participants. During the presentation, which lasted approximately 10 min, they were allowed to ask any questions. Then, the participants performed a practice session. The training course was designed to teach participants how to respond to abnormal or alarm events. For the Group B, the participants needed to be trained how to control the eye-gaze input system.

The primary goal of this task was to compare the ability to detect abnormal situation between a mouse and eye-gaze input system and understand how participants used the two systems differently. The participants were free to ask questions throughout the practice tests. This training session lasted five minutes. The final step of the experiment was data collection. Each participant engaged in six scenarios. All participants were

randomly assigned to six different scenario orders. Each test scenario lasted approximately eight minutes. After they had finished one scenario, the participants were asked to answer a NASA-TLX questionnaire which consisted of six questions concerning the more important contributor to the workload for this task. The participants were not allowed to ask any questions during the data collection. The total experiment time was approximately 60 min.

4 Data Analysis

4.1 Fitts' Law and Reaction Time

Fitts' law (1954) has been used widely to understand human responses related to different input systems (Burstyn et al. 2016; Forlines et al. 2007). It is primarily used in human-computer interaction and human factors as a descriptive model of human movement. This scientific law predicts that the time required to move the object to a target area is a function of the relation between the distance from the start to the target and the width of the target. The Eq. 1 shows how to calculate a difficulty level of a given task, called index of difficulty (ID). A is the distance or amplitude between the targets, and W is the width of the target. The bigger ID value represents the more difficult task to perform. Also, we analyzed the relation between the reaction time and ID by using the Eq. 2, where a and b are empirically derived constants.

$$ID = \log_2(2A/W) \quad (1)$$

$$T(\text{milliseconds}) = a + b \log_2(2A/W) \quad (2)$$

4.2 Sensitivity (D') and Operator Bias (β)

Accuracy was measured by operator's sensitivity (d') and bias (β). Estimates of the operator's sensitivity and bias were collected for the participant performance during the TWHITL simulation (Kim et al. 2015). The sensitivity refers to how well the operator discriminates the signal from the noise (Lerman et al. 2010). Thus, d' was calculated by subtracting the z-score that corresponds to the false alarm rate from the z-score that corresponds to the hit rate (Macmillan 1993).

$$\text{Sensitivity}(d') = |Z(\text{Hit}) - Z(\text{False Alarm})| \quad (3)$$

The operator's bias is the likelihood ratio of human response related to the presence of a signal. Z (Hit) is the z-score which corresponds to a hit rate, whereas Z (False Alarm) is the score which corresponds to a false alarm rate (Stanislaw and Todorov 1999).

$$\text{Bias}(\beta) = P(\text{ordinate of Hit})/P(\text{ordinate of false alarms}) \quad (4)$$

4.3 Mental Workload

NASA-TLX is a six-dimensional scale designed to obtain mental workload during a task or immediately afterward (Hart 2006). It has been applied within numerous domains, including civil and military aviation, driving, nuclear, power plant control room operation and air traffic control (Caldwell 2005; Erzberger 2005; Hwang et al. 2008; Kim et al. 2016; Lauer et al. 2007; Palinko et al. 2010; Wiegmann and Shappell 2001). Over the past 30 years, the research has revealed the NASA-TLX approach to be a tool which is easy to use and reliable for experimental manipulations.

5 Results

Figure 5 showed the reaction time comparisons between a mouse input and an eye-gaze input system. The horizontal axis displays two input systems and each index of difficulty (ID). Seven levels of ID were tested during the experiment (ID = {1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0}). The vertical axis presents the reaction time. Reaction time was charted as a function of different input systems and index of difficulty based on Fitts' law model.

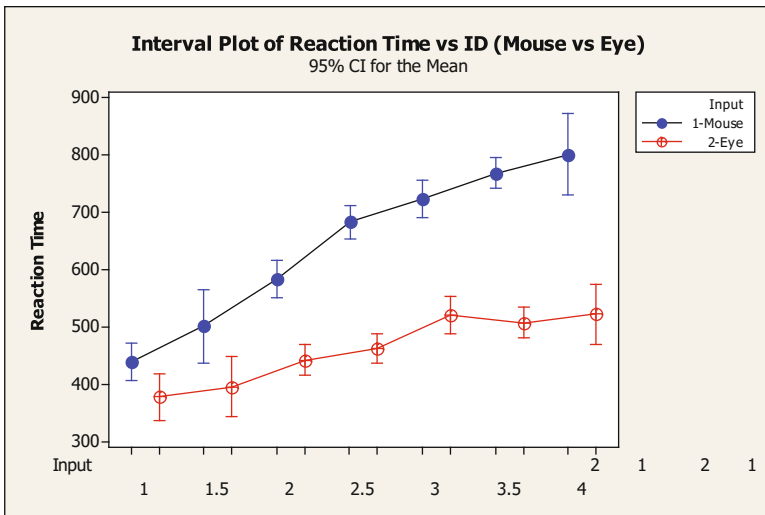


Fig. 5. Interval plot of reaction time vs ID (Mouse vs Eye)

Mouse input system:

$$RT = 327.4 + 130.1 ID \quad (R^2 = 40.2\% \quad P < 0.001) \quad (5)$$

Eye-gaze input system:

$$RT = 325.2 + 55.86 ID (R^2 = 9.4\% P < 0.001) \tag{6}$$

A two-way ANOVA general linear model was used (factor A: input systems, factor B: index of difficulty), which revealed that there were significant differences for reaction times between the input systems ($F(1,11) = 10.11, p = 0.002$), and index of difficulty ($F(6,66) = 5.89, p < 0.001$).

In Table 2, the mean reaction time is illustrated as a function of different IDs for the mouse and eye-gaze input systems, respectively. The smallest reaction times (ID 1.0) were 284 ms (milliseconds) and 152 ms for the mouse and eye-gaze input system, respectively. Additionally, the smallest reaction times (ID 4.0) were 579 ms and 267 ms for the mouse and eye-gaze input system, respectively. The average reaction time difference between the mouse and eye-gaze input systems is 193.76 ms.

Table 2. Reaction time (milliseconds) vs index of difficulty (Mouse vs Eye-gaze)

ID	Mouse			Eye-gaze		
	Mean	STD	MIN	Mean	STD	MIN
1.0	438.88	108.14	284	376.92	176.29	152
1.5	501.32	132.73	317	395.22	159.22	175
2.0	584.36	137.31	341	441.41	140.35	187
2.5	683.93	136.20	452	462.55	145.00	170
3.0	723.14	138.57	478	520.44	149.95	171
3.5	769.05	121.24	510	506.98	128.71	188
4.0	881.53	168.21	579	522.35	102.88	267

In Table 3, a one-way (input system) ANOVA was conducted on the reaction time and revealed no significant accuracy ($p = 0.992, F(1, 23) = 0$) between the different input systems. However, the result revealed significant differences in operator bias ($p < 0.001, F(1, 23) = 40.69$) between the mouse and eye-gaze input systems.

Table 3. Sensitivity (d') and bias (β) detection (Mouse vs Eye-gaze)

	Mouse		Eye-gaze	
	Mean	STD	Mean	STD
d'	3.3922	0.6366	3.3943	0.7009
β	69.67	29.38	21.13	22.94

In Table 4, a one-way (input system) ANOVA was conducted on the different input systems' mental workload and no significant differences ($F(1, 23) = 0.30, p = 0.587$) in the mental workload index between mouse and eye-gaze input systems were detected. The mean of the mental workload indices are 38.64 and 35.43 for mouse and eye-gaze input systems, respectively.

Table 4. Mental workload rating (Mouse vs Eye-gaze)

Group	Index	STD	F Value	P Value
Mouse	38.64	20.85	F(1, 23) = 0.30	0.587
Eye-gaze	35.43	19.82		

6 Discussion

The objective of this study was to compare the pointing performance between a mouse input and an eye-gaze input. By using reaction time (millisecond), sensitivity (d'), operator's bias (β), and NASA-TLX, we were able to analyze participants' ability to detect abnormal events during the experiment.

6.1 Reaction Time

The results (see Fig. 5) showed that the index of difficulty (ID) influenced the reaction time of both input systems. However, according to the regression analysis, the model from the eye-gaze system was significantly different than the model from the mouse input. The Fitts' law could explain the regression model of the mouse input. The reaction time increased proportionally as ID increased. However, the model from the eye-gaze system was not. The influence caused by ID was compared to the mouse input. We also compared the reaction time difference between the mouse and eye-gaze system times for each ID level in Table 2. The results showed that the reaction time of the eye-gaze system was always faster than the mouse input system. The average reaction time difference from the easiest event (ID = 1.0) to the hardest event (ID = 4.0) was only 146 ms for the eye-gaze system. All results pointed out that there was a significant difference between the eye-gaze input and traditional mouse input system in reaction time. Thus, one of our hypotheses: there is no significant difference in reaction time between mouse and eye-gaze systems should be rejected.

6.2 Accuracy

The results (see Table 3) showed that there was no significant difference in sensitivity between the mouse and eye-gaze system. Thus, the another hypothesis: there is no difference in accuracy between the mouse and eye-gaze system is proven. However, the results also revealed different operator biases between the mouse and eye-gaze system. The high β average represents that participants allowed more misses to avoid false alarms. This means that the mouse input system has more miss responses in comparison to the eye-gaze system, and the participants were more conservative when they experienced abnormal events. However, the users of the eye-gaze system made false alarm responses more often compared to the mouse input system, which means that the participants were more liberal when they made decisions about the abnormal events.

As illustrated in Fig. 6, the process concerning a stimulus response was faster, and the user response path was relatively short for the eye-gaze input system compared to

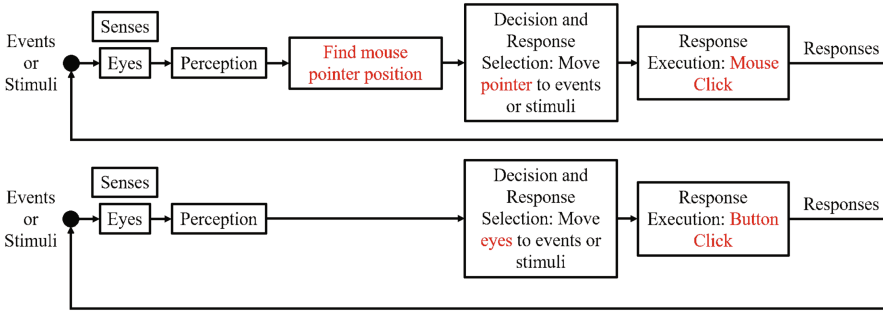


Fig. 6. Different responses procedure (Mouse vs Eye-gaze)

the mouse input system. The findings indicated that participants in the Group B did not need to navigate the input device to perform the given task. They only needed to pay attention to the abnormal gauges to reach out the targets and click them. However, the participants in the Group A needed to find the current position of the mouse after they discovered the abnormal gauges on the screen, and moved to the target location. Therefore, the probability of judging the stimulus as a signal was high when the participants used the eye-gaze input system.

6.3 Mental Workload

Although we hypothesized that the eye-gaze system would increase the user’s mental workload compared to the mouse input system, it turned out that there was no workload significance between the mouse and eye-gaze system (see Table 4). Hence, our hypothesis regarding the workload should be rejected.

7 Conclusion and Future Work

7.1 Conclusion

In this research, we verified and confirmed the following three hypotheses. Firstly, there was a significant difference in reaction time between mouse and eye-gaze input system. The reaction time of the eye-gaze input system was faster than that of the mouse input, and was less influenced by the ID. According to the regression model, the mouse input system was more predictable as a linear pattern, but the eye-gaze input system was not. Secondly, there was no difference in accuracy between the mouse system and the eye-gaze system. The participants in both groups had a similar ability to detect the abnormal events during the experiment. It means that the eye-gaze input system could perform faster at the same accuracy level. Moreover, the probability of judging the stimulus as a signal was high for the eye-gaze input system, and its operator bias β values were lower than the mouse input system. This means that the eye tracking system users might be more liberal to make their decision compared to the mouse

system users. However, the eye-gaze input system was easier to make false alarm responses. Thus, the findings recommend us to use either input system based on the cost of human errors: miss and false alarm. Finally, there was no significant difference in mental workload between the mouse and eye-gaze input system based on NASA-TLX results. The workload caused by using the eye-gaze system was not different than the mouse input system.

In conclusion, the eye-gaze input system can be regarded as a better input method for human computer interactions. It is confirmed that the eye-gaze system users' performance accuracy and mental workload were similar to the traditional mouse input system. However, the eye-gaze input system was better than the mouse input system with regard to the reaction time. Therefore, the eye-gaze system is favorable to the tasks, which require fast human responses and charge lower false alarm cost compared to the cost of miss.

7.2 Future Work

For future research, it is beneficial to explore how exactly the index of difficulty (ID) influences the response of the eye-gaze input system. Developing a generalized and extended model which can describe the relationship between reaction time and the ID for the eye-gaze system. Furthermore, it is necessary to study how to reduce the false alarm rate when the users use the eye-gaze system. Finally, for mental workload, different methods, such as pupil dilation, electroencephalogram, and functional magnetic resonance imaging, could be used as tools to measure the workload caused by various input systems.

References

- Bai, H., Lee, G., Billingham, M.: Using 3d hand gestures and touch input for wearable ar interaction. Paper Presented at the CHI 2014 Extended Abstracts on Human Factors in Computing Systems (2014)
- Bodala, I.P., Ke, Y., Mir, H., Thakor, N.V., Al-Nashash, H.: Cognitive workload estimation due to vague visual stimuli using saccadic eye movements. Paper Presented at the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (2014)
- Burstyn, J., Carrascal, J.P., Vertegaal, R.: Fitts' law and the effects of input mapping and stiffness on flexible display interactions. Paper Presented at the Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (2016)
- Caldwell, J.A.: Fatigue in aviation. *Travel Med. Infect. Dis.* **3**(2), 85–96 (2005)
- Chandra, S., Sharma, G., Malhotra, S., Jha, D., Mittal, A.P.: Eye tracking based human computer interaction: applications and their uses. Paper Presented at the 2015 International Conference on Man and Machine Interfacing (MAMI) (2015)
- Chen, J., Ji, Q.: A probabilistic approach to online eye gaze tracking without explicit personal calibration. *IEEE Trans. Image Process.* **24**(3), 1076–1086 (2015)
- De la Torre, V., Hernandez, J.L., García, J.L., Mejia, G.I.: Workload Assessment in Industrial Settings: A Proposal Applying the Analytic Hierarchy Process (2014)

- Erzberger, H.: Automated conflict resolution for air traffic control (2005)
- Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. *J. Exp. Psychol.* **47**(6), 381 (1954)
- Forlines, C., Wigdor, D., Shen, C., Balakrishnan, R.: Direct-touch vs. mouse input for tabletop displays. Paper Presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2007)
- Hart, S.G.: NASA-task load index (NASA-TLX); 20 years later. Paper Presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting (2006)
- Hwang, S.-L., Lin, J.-T., Liang, G.-F., Yau, Y.-J., Yenn, T.-C., Hsu, C.-C.: Application control chart concepts of designing a pre-alarm system in the nuclear power plant control room. *Nucl. Eng. Des.* **238**(12), 3522–3527 (2008)
- Jacob, R., Karn, K.S.: Eye tracking in human-computer interaction and usability research: ready to deliver the promises. *Mind* **2**(3), 4 (2003)
- Kim, J.H., Rothrock, L., Laberge, J.: Using Signal Detection Theory and Time Window-based Human-In-The-Loop simulation as a tool for assessing the effectiveness of different qualitative shapes in continuous monitoring tasks. *Appl. Ergon.* **45**(3), 693–705 (2014)
- Kim, J.H., Rothrock, L., Laberge, J.: Arousal and performance in a process monitoring task using signal detection theory. Paper Presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting (2015)
- Kim, J.H., Yang, X., Putri, M.: Multitasking performance and workload during a continuous monitoring task. Paper Presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting (2016)
- Lauer, T.R., Faber, S., Richstone, D., Gebhardt, K., Tremaine, S., Postman, M., Green, R.: The Masses of Nuclear Black Holes in Luminous Elliptical Galaxies and Implications for the Space Density of the Most Massive Black Holes Based on observations made with the NASA/ESA Hubble Space Telescope, obtained at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. These observations are associated with GO and GTO proposals 5236, 5446, 5454, 5512, 5943, 5990, 5999, 6099, 6386, 6554, 6587, 6633, 7468, 8683, and 91. *Astrophys. J.* **662**(2), 808 (2007)
- Lerman, D.C., Tetreault, A., Hovanetz, A., Bellaci, E., Miller, J., Karp, H., Keyl, A.: Applying signal-detection theory to the study of observer accuracy and bias in behavioral assessment. *J. Appl. Behav. Anal.* **43**(2), 195–213 (2010)
- Lynn, S.K., Barrett, L.F.: “Utilizing” signal detection theory. *Psychol. Sci.* **25**(9), 1663–1673 (2014)
- Macmillan, N.A.: Signal detection theory as data analysis method and psychological decision model (1993)
- Maniscalco, B., Lau, H.: A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Conscious. Cogn.* **21**(1), 422–430 (2012)
- Murata, A., Uetsugi, R., Fukunaga, D.: Effects of target shape and display location on pointing performance by eye-gaze input system. In: 2014 Proceedings of the Paper Presented at the SICE Annual Conference (SICE) (2014)
- Murata, A., Uetsugi, R., Hayami, T.: Study on cursor shape suitable for eye-gaze input system. In: 2012 Proceedings of the Paper Presented at the SICE Annual Conference (SICE) (2012)
- Noah, B., Kim, J.-H., Rothrock, L., Tharanathan, A.: Evaluating alternate visualization techniques for overview displays in process control. *IIE Trans. Occup. Ergon. Hum. Factors* **2**(3–4), 152–168 (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 & Applications (2010)

- Sibert, L.E., Templeman, J.N., Jacob, R.J.: Evaluation and analysis of eye gaze interaction (2001). Accessed
- Stanislaw, H., Todorov, N.: Calculation of signal detection theory measures. *Behav. Res. Methods Instrum. Comput.* **31**(1), 137–149 (1999)
- Tharanathan, A., Laberge, J., Bullemer, P., Reising, D.V., McLain, R.: Functional versus schematic overview displays: Impact on operator situation awareness in process monitoring. Paper Presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting (2010)
- Wiegmann, D.A., Shappell, S.A.: Human error analysis of commercial aviation accidents: Application of the Human Factors Analysis and Classification System (HFACS). *Aviat. Space Environ. Med.* **72**(11), 1006–1016 (2001)