

# Sensitivity, Bias, and Mental Workload in a Multitasking Environment

Monika Putri, Xiaonan Yang, and Jung Hyup Kim<sup>(✉)</sup>

Department of Industrial and Manufacturing Systems Engineering,  
University of Missouri, Columbia, MO, USA  
{mapmtf, xyr29}@mail.missouri.edu, kijung@missouri.edu

**Abstract.** In this paper, we used signal detection theory (SDT) as a tool to evaluate human performance in a multitasking environment. The primary objective of using SDT is to assess an operator's sensitivity ( $d'$ ) and bias ( $\beta$ ). In addition, NASA-TLX was used to measure participants' workload under different complexity scenarios. During the experiment, participants were asked to detect abnormal and alarm signals on a gauge monitoring display. They also needed to perform multi-attribute task battery (MATB) tasks at the same time. The gauge-monitoring screen contains total 52 gauges (flow, level, temperature, and pressure). The MATB consists of system monitoring, target tracking, and dynamic resource management. The results of this study demonstrate that participants showed various levels of sensitivity ( $d'$ ) in the gauge-monitoring task based on the degree of task complexity.

**Keywords:** Signal detection theory · Human-in-the-loop simulation · Mental workload

## 1 Introduction

Operators who perform tasks under supervisory control continuously keep track of new and high priority events (Ratwani et al. 2010). The control room environment in oil and gas refineries is one example of this. The room consists of multiple, complex human-machine systems. The operator must observe numerous different control loops while concurrently performing other attention-demanding tasks (Kim et al. 2015; Noah et al. 2014). Although the systems generate lots of data, the amount of information transmitted to the operator is always smaller than the stimulus information. Hence, it is necessary to understand how the operators reform their detectability, under a multitasking environment, to improve control conditions and avoid mistakes caused by missing the critical information. According to the previous research done by Pashler (1994), multitasking is one of the main causes of increasing human errors and reaction time. The multitasking environment results in degradation of information to the operator. Swets and Biedsall's (1955) found that humans who are working with multiple displays are making their choices among a number of signal alternatives. In addition, a cooperative multitasking involves task switching activities, which causes a higher error rate and slower reaction time during the tasks (Spector and Biederman 1976).

In this study, we examined an impact of the complexity level in a multitasking environment (primary task: gauge monitoring task, secondary task: MATB). The signal rate of the gauge monitoring task was used to design the different levels of complexity. Each task requires a different cognitive resource, such as visual searching, target tracking, and diagnostic control. To conduct the experiment, a human-in-the-loop (HITL) simulation was used. A set of gauges shapes of flow, level, pressure, and temperature were used and underlay the design of the overview display (OD) that represents an actual refinery's operations (Bullemer et al. 2009). The HITL experiment allowed us to observe participants' direct responses and activities generated by characteristic functions of the given task. The activities comprised detecting abnormal events in a continuous gauge monitoring task. We applied SDT to measure human performance in a multitasking environment. SDT evaluates the psychometric function that describes how performance increases with stimulation degrees (García-Pérez and Alcalá-Quintana 2011). It measures an individual's ability to detect signals in a dual-task environment. The outcomes of SDT consist of Hit, Miss, False Alarm, and Correct Rejection. Using the outcome data, we will be able to calculate operator perceptual sensitivity ( $d'$ ) and operator bias, ( $\beta$ ) (Walker and Brewster 2000). Operator bias is defined as the likelihood that a participant will favor one direction as opposed to the other, whereas sensitivity measures operator accuracy in differentiating the signal from the noise (Lerman et al. 2010). In addition, NASA-TLX was used to measure participants' workload during the experiment. NASA-TLX is a multi-dimensional rating task that measures an overall workload score in a dynamic environment.

The goal of this study is to investigate operator sensitivity, bias, and workload in a multitasking environment. The findings are not only beneficial for understanding operator behavior but also developing an operator decision-making model in a multitasking environment.

## 2 Methods

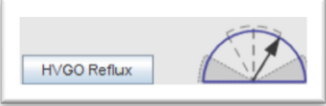
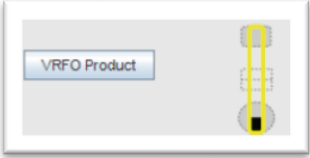
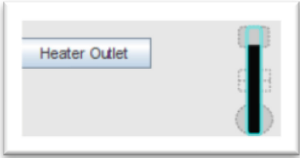
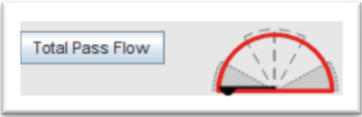
### 2.1 Participants

A total of 18 university students participated in this study. Participants included undergraduate and graduate students. The average age was 22 for male students and 23 for female students ( $M = 22.61$ ,  $SD = 4.36$ ). 45 % of participants were male and 55 % were female. Every participant had normal vision and no upper-body impairments that may have limited the use of a mouse as an interface.

### 2.2 Apparatus

To create a more realistic assessment in a multitasking environment, the gauge monitoring human-in-the-loop (HITL) simulation was used as a primary task (Kim et al. 2014). It contained five functional areas to represent all the major gauges (see Fig. 1). Each gauge showed different colored-outlines to indicate a normal, abnormal, or alarm state (see Table 1). We also used the multi-attribute task battery (MATB) as a second task. It consists of system monitoring, tracking, resource management, communication,

**Table 1.** Examples of different gauge states

Gauge Contour Color	Alarm Type
	Abnormal (dark blue colored-line)
	Low Alarm (yellow colored-line)
	High Alarm (blue colored-line)
	Hi-Hi Alarm & Lo-Lo Alarm (red colored-line)

and scheduling task, but the communication task was deactivated during the experiment (see Fig. 2). Every experimental scenario was developed based on the actual refinery's operation. Participants experienced two levels of complexity scenarios (low and high) during the experiment. The total number of events in the high complexity scenario category was twice as large as the low complexity scenario category.

There were 12 alarm events of low complexity and 24 alarm events of high complexity. Moreover, the MATB was used to keep the balance of complexity in each scenario. For example, when the number of alarm events decreased, the number of events in the MATB increased.

### 3 Procedure

The experiment was conducted as a two-factor experiment with repeated measures (within-subject factors: scenario complexity and day), and a multi-session study, which would be continued for five days. Day 1 was a training session, and Day 2 and 3 were

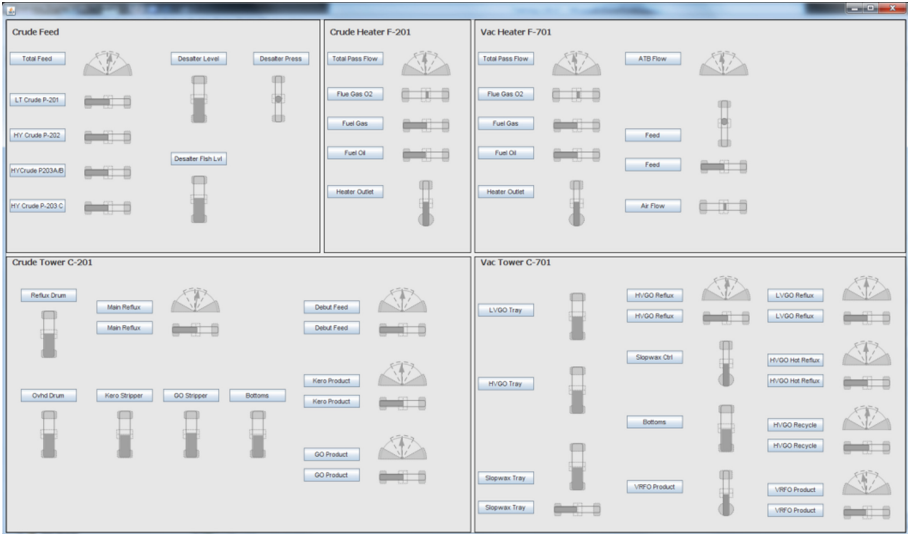


Fig. 1. The gauge monitoring HITL simulation

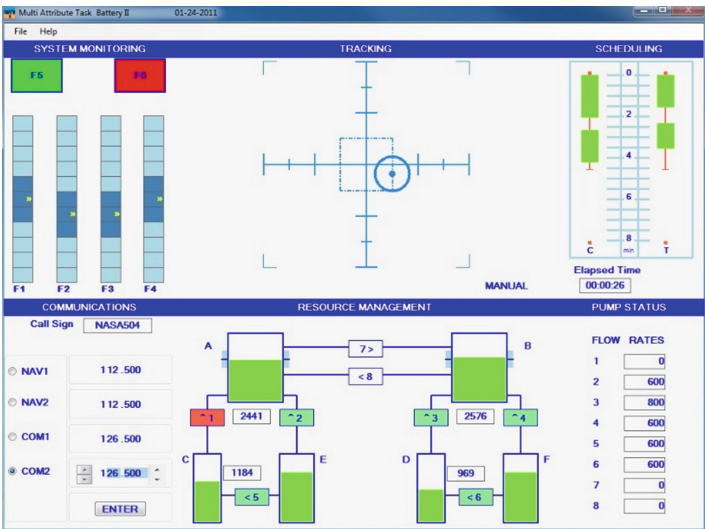


Fig. 2. MATB display

practice sessions. The data from Day 4 and 5 was used to analyze operator’s sensitivity and bias. A statistical model was developed to analyze the performance under different complexity scenarios. Day 1 required 60 min for an introduction of initial training trials. Day 2 and 3 involved additional practice tests to ensure that participants were

fully aware of all tasks. Day 4 and 5 were considered the actual experiment days in order to collect the performance data for the assessment of learning.

The first step of Day 1 was a general orientation about the primary and secondary task. The purpose of this training was to make the participants aware of the basic knowledge of all functions. During the practice sessions (Day 2 and 3), participants practiced both tasks multiple times. During the actual process monitoring sessions (Day 4 and 5), participants experienced six scenarios (3 low complexity and 3 high complexity). A scenario order was counterbalanced to eliminate the order effect. Participants were asked to answer NASA-TLX questionnaires after they completed each scenario.

## 4 Data Analysis

### 4.1 Gauge Monitoring Task

There were two measurements for the primary task: operator's sensitivity ( $d'$ ) and bias ( $\beta$ ). The sensitivity refers to how well an operator discriminates the signal from the noise (Lerman et al. 2010).  $d'$  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).

$$\text{Sensitivity } (d') = |Z(\text{Hit}) - Z(\text{False Alarm})| \quad (1)$$

The response bias ( $\beta$ ) is defined as the likelihood ratio of operator's response regarding the presence of signals.  $\beta$  was calculated by

$$\text{Bias } (\beta) = P(\text{ordinate of Hit}) / P(\text{ordinate of false alarms}) \quad (2)$$

### 4.2 Multi-Attribute Task Battery

There were three tasks in an MATB display: system monitoring, resource management, and tracking task. Each task requires a different cognitive resource to perform. Hence, the different metric was used to evaluate task performance on each task. For the tracking task, the performance is calculated by the root mean square deviation of the target center point from the center point in pixel units during a fixed time lapse (Santiago-Espada et al. 2011).

$$RMSD\_C = \text{Square Root } (SS/NUM) \quad (3)$$

The value range is from 0 for the center position to 300 for continuous maximum X and Y offset.  $NUM$  is the number of samples (15-s intervals), and  $SS$  is the sum of the squares of the X-offset plus the Y-offset. By using this, the tracking performance ( $A_T$ ) was calculated by

$$A_T = \frac{(\text{the highest } RMSD\_C - \text{current } tRMSD\_C)}{\text{the highest } RMSD\_C} \quad (4)$$

For the system-monitoring task, the performance ( $A_S$ ) was calculated by

$$A_S = \frac{H_s}{TA_s} \times 100 \quad (5)$$

$H_s$  = Total number of correct actions in system-monitoring task.

$TA_s$  = Total number of signal events.

For the resource management task, the performance ( $A_R$ ) was calculated by

$$A_R = \frac{\text{Total Normal} - \text{state Tanks}}{\text{Total Sampling Interval} \times 2 \text{ (for two tanks)}} \times 100 \quad (6)$$

$A_R$  is identified as the total normal-state tanks divided by the total sampling interval multiplied by two tanks. There are 18 sampling interval in total that is derived from each 9 min-length scenario.

### 4.3 NASA TLX

The NASA-TLX (NASA Task Load Index) is a multidimensional subjective workload rating technique. NASA TLX is commonly used to measure operators' workload (Cao et al. 2009; Hancock et al. 1995; Hart and Staveland 1988). It considers the magnitude of six possible load types: mental demand, physical demand, and temporal demand, own performance, effort, and frustration. It weighs the six types of load through a series of 15 combinations (close to 100 – high workload; close to 0 - low workload). In this study, the workload is defined as the expense caused by operators to achieve a certain performance level (Singh et al. 2008).

## 5 Results

We analyzed the participant's responses by using a two-way ANOVA. The dependent variables were sensitivity ( $d'$ ), operator's bias ( $\beta$ ), performance on tracking, system-monitoring, resource management, and NASA-TLX score. The independent variables were scenario complexity and day.

### 5.1 Gauge Monitoring Task

The ANOVA results showed the effect of scenario complexity and day for  $d'$  and  $\beta$ . For the sensitivity ( $d'$ ), it was significantly influenced by the scenario complexity,  $F_{1,17} = 4.25$ ,  $p < 0.05$ . However, for the operator bias ( $\beta$ ), there was a no significant effect on the scenario complexity and day (Table 2).

**Table 2.** Descriptive statistics for sensitivity and operator bias

Metrics	Complexity	Day 4		Day 5	
		M	SD	M	SD
$d'$	Low	4.1636	0.4121	4.254	0.3594
	High	3.912	0.846	4.007	0.715
$\beta$	Low	23.22	15.77	19.51	8.37
	High	19.89	12.47	22.61	14.08

## 5.2 MATB

For the tracking task, there was a significant effect on scenario complexity,  $F_{1,17} = 22.56$ ,  $p < 0.001$ . However, there was no significant effect for the system-monitoring and resource management task (Table 3).

**Table 3.** Descriptive statistics for system monitoring, resource management, and track

Metrics	Complexity	Day 4		Day 5	
		M	SD	M	SD
$A_S$	Low	80.45	20.51	84.17	23.88
	High	79.08	18.93	78.58	23.24
$A_R$	Low	93.83	9.85	92.28	14.67
	High	89.81	12.28	93.00	10.03
$A_T$	Low	49.41	12.04	49.08	11.61
	High	38.98	14.42	44.88	13.04

## 5.3 Mental Workload

The NASA-TLX results showed that participant's mental workload was significantly influenced by the scenario complexity ( $F_{1,17} = 11.43$ ;  $p < 0.001$ ) (Table 4).

**Table 4.** Descriptive statistics for NASA-Task Load Index

Metrics	Complexity	Day 4		Day 5	
		Mean	St. Dev	Mean	St. Dev
NASA-TLX	Low	50.37	15.93	48.15	18.21
	High	56.96	13.24	54.74	16.11

## 6 Discussion and Conclusion

In this research, we studied sensitivity ( $d'$ ) and operator's bias ( $\beta$ ), and mental workload of human operators in a multitasking environment by using the signal detection theory (SDT) and NASA-TLX.

The results showed that the operator's detectability regarding an alarm signal was significantly influenced by the number of visual stimuli from the primary task (the gauge monitoring task). When the scenario complexity was low, sensitivity ( $d'$ ) was better than during the high complexity condition. However, operator bias was not influenced by this, which means that the allocation of visual attention exists between the primary and secondary tasks. In this experiment, the participants had to pay attention to find abnormal situations in both displays. This divided the attention of participants. If they spent an equal amount of time detecting signals from both tasks, the number of missed signals on the gauge monitoring task during the high complexity condition would be larger than in the low complexity condition. Hence, the sensitivity was low when the participants executed high complexity scenarios, but this did not influence the performance on the system-monitoring task.

For the secondary task, the scenario complexity had a significant influence on the tracking task. The tracking task performance was better under the low complexity scenarios. However, the results of the system monitoring and resource management were not significantly influenced by the complexity. It means the complexity level of the gauge monitoring task could affect performance on the tracking task, although the tracking had more physical demands than mental demands. The performance on the tracking task was based on the deviation of the target point from the center point. A joystick controlled the target point. So, participants could control a joystick better in the low complexity condition. Tracking data showed that the switching time to the tracking task was longer than other tasks because of physical motions related to the joystick control. Therefore, we can conclude that the visual stimulus level in a multitasking environment could influence task performance related to physical activities.

For mental workload, we found a negative correlation between the sensitivity and NASA-TLX rating. With the increasing of the complexity level, the sensitivity worsened, and the NASA-TLX score increased.

Although we did not include any physiological measures to assess human performance in a multitasking environment, all findings suggest that multitasking performance is influenced by the visual stimulus level. In other words, the effect of the visual stimulus is significant on performing a visual searching task and physical control task in a multitasking environment. For future research, we will collect eye and head tracking data to investigate a relationship between the visual stimulus level and the multitasking performance. The results of this research will contribute to advance our understanding of behavior modification caused by visual stimuli and the operators' decision-making process in a multitasking environment.

## References

- Bullemer, P., Reising, D.V., Jones, M.: ASM® Consortium White Paper (2009)
- Cao, A., Chintamani, K.K., Pandya, A.K., Ellis, R.D.: NASA TLX: software for assessing subjective mental workload. *Behav. Res. Methods* **41**(1), 113–117 (2009)
- García-Pérez, M.A., Alcalá-Quintana, R.: Interval bias in 2AFC detection tasks: sorting out the artifacts. *Atten. Percept. Psychophys.* **73**(7), 2332–2352 (2011)



- Hancock, P., Williams, G., Manning, C.: Influence of task demand characteristics on workload and performance. *Int. J. Aviat. Psychol.* **5**(1), 63–86 (1995)
- 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)
- 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. *Applied ergonomics* **45**(3), 693–705 (2014)
- Kim, J.H., Rothrock, L., Laberge, J.: Arousal and performance in a process monitoring task using signal detection theory. In: Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting (2015)
- 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)
- Noah, B., Kim, J.-H., Rothrock, L., Tharanathan, A.: Evaluating alternate visualization techniques for overview displays in process control. *IIE Trans. Occup. Ergonomics Hum. Factors* **2**(3–4), 152–168 (2014)
- Pashler, H.: Dual-task interference in simple tasks: data and theory. *Psychol. Bull.* **116**(2), 220 (1994)
- Ratwani, R.M., McCurry, J.M., Trafton, J.G.: Single operator, multiple robots: an eye movement based theoretic model of operator situation awareness. In: Paper presented at the Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction (2010)
- Santiago-Espada, Y., Myer, R.R., Latorella, K. A., Comstock Jr., J.R.: The Multi-Attribute Task Battery II (MATB-II) software for human performance and workload research: a user's guide. NASA Tech Memorandum, 217164 (2011)
- Singh, I.L., Singh, A.L., Tiwari, T., Saha, P.K.: Multi-task performance in computer-aided systems: an appraisal. *J. Indian Acad. Appl. Psychol.* **34**(1), 115–125 (2008)
- Spector, A., Biederman, I.: Mental set and ental shift revisited. *Am. J. Psychol.* 669–679 (1976)
- Walker, A., Brewster, S.: Spatial audio in small screen device displays. *Pers. Technol.* **4**(2–3), 144–154 (2000)

Engineering Psychology and Cognitive Ergonomics  
13th International Conference, EPCE 2016, Held as Part  
of HCI International 2016, Toronto, ON, Canada, July  
17-22, 2016, Proceedings  
Harris, D. (Ed.)  
2016, XVII, 496 p. 176 illus., Softcover  
ISBN: 978-3-319-40029-7