Using Eye-tracking to Acknowledge Attended Alarms

Katherine E. Herdt

Thesis submitted to the faculty of the Virginia Polytechnic and State University in fulfillment of the requirements for the degree of

Master of Science

In

Industrial and Systems Engineering

Nathan K. Lau, Chair

Katya L. LeBlanc

Sheila G. Klauer

December 14, 2021

Blacksburg, Virginia

Keywords: alarm management, gaze-based interaction, industrial control systems

Using Eye-tracking to Acknowledge Attended Alarms

Katherine E. Herdt

Abstract

A lack of alarm management for industrial control rooms has led to frequent alarm floods that have the potential to overwhelm operators within minutes. One approach to managing alarm floods would be altering the salience of alarms that operators might already notice, thereby reducing the disruption on workflow and attention for managing uninformative alarms. This research investigated the central hypothesis that eye fixations could supply passive input to acknowledge alarms anticipated by the operators and thereby improve their overall task performance. A dual-task experiment recruiting 24 participants was conducted to compare three gaze-based alarm acknowledgement methods –Proximity, Prediction, and Entropy- against no acknowledgement across three types of scenarios – Near-threshold, Trending, and Fluctuation. The gaze-based acknowledgement methods reduced visual and auditory salience of alarms as a function of the number of fixations on parameters as well as characteristics of the parameter known to influence operator monitoring behaviors. The participants performed an alarm monitoring task while controlling a continuous parameter within an acceptable range. While participants showed a preference for all of three gaze-based acknowledgment methods, performance of the parameter control task did not improve with gaze-based acknowledgement. Scenario types, as defined by the behavior of the parameters, exhibited a significant effect on the performance of the parameter control task, suggesting a greater influence on participant attention than the reduced salience associated with the gaze-based acknowledgments. Additional analysis revealed that gaze-acknowledgements are higher in scenarios with the most suitable for the gaze-based acknowledgement methods, although the participants did not show any gaze-based acknowledgements and did not make a prediction of an alarm for a significant portion of the trials, suggesting a lack of resource allocation to the alarm monitoring task. This result suggests that the effectiveness of gaze-based acknowledgement may depend on the combination of on-going tasks. Taken together, the experimental results showed some utility of user gaze in managing alarms given how acknowledgement occurred more often when the acknowledgement methods and parameters matched; however, further design research is necessary to translate the utility into clear performance or productivity benefits.

Using Eye-tracking to Acknowledge Attended Alarms

Katherine E. Herdt

General Audience Abstract

Industrial control rooms are notorious for having too many alarms triggered within minutes and operators are hindered by responding to these alarms as opposed to the actual process faults. Existing alarm management research and applications have already reduced nuisance alarms by filtering out those correlated to one another according to historical data or plant models. However, existing approaches have not eliminated the process parameters that operators already expect to reach alarm thresholds. In other words, current alarm management has not adapted for operator awareness of impending alarms. This study explored how eye-tracking might be used to acknowledge alarms anticipated by operators, thereby reducing uninformative alarms and interruption to operator work. The participants performed an alarm monitoring task while trying to maintain a fluctuating parameter within an acceptable range. While participants liked the gaze-based acknowledgement methods, their performance on the parameter control task did not improve over conditions without any alarm acknowledgement. The alarm monitoring task may not have received sufficient attention to induce an observable benefit. The characteristics of the parameter seemed to have a larger effect on participants’ attention than the muted alarm presentation associated with the gaze-based acknowledgment. Further research is necessary to refine the current design to induce the postulated attention and performance benefits with gaze-based acknowledgement.

Acknowledgements

I would like to acknowledge the support of Idaho National Laboratory in providing the process simulator for developing the alarm monitoring task. I also recognize Michael Hildebrandt and Tai Le for their generous technical support in developing and providing Synquesticon for conducting the experiment. Lastly, many thanks to Drs. Nathan Lau, Katya LeBlanc, and Sheila Klauer for their service in my thesis committee and their mentorship throughout this process.

Contents

List of Figures

List of Tables

List of Equations

1 Introduction

Control room operators depend on alarms as one of the many mechanisms to facilitate immediate response to abnormal events, thereby preventing accidents and plant shutdowns (Goel et al., 2017). According to ISA-18.2, an industrial control room alarm serves as “an audible and/or visible means of indicating to the operator an equipment malfunction, process deviation, or abnormal condition requiring a response” (p.14, 2009). Operators require monitoring assistance from an alarm system because the amount of information displayed within the control room is more than can be surveyed at once. Instead, operators survey parameters in a top-down approach to compare their expectation of the plant state with actual process indicators, making the task an active form of problem-solving (Lau, Jamieson, and Skraaning, 2016). They continuously search for abnormal indications amongst all the ‘noise’ (Munaw et al., 2000). Alarms thus represent a monitoring tool that shifts the operator’s attention towards parameters not being surveyed when those parameters may need immediate attention (Woods, 1995).

Given the goal of redirecting attention, alarms are designed to be very salient to human perception and difficult to disregard (i.e., perceptual inhibition). For example, humans are very capable of detecting intermittent stimulus of light and motion in their peripheral vision due to the higher flicker fusion frequency afforded by the number of rods and the magnocellular cells (Simonson & Brozek, 1952; Carrasco et al., 2003). While vision is constrained to the person’s field of view, sound is omnidirectional, capturing attention when the visual cue cannot (Stanton & Edworthy, 1999). An auditory stimulus can also readily override concentration of ongoing visual tasks (Wickens et al., 2005; Wickens & Colcombe, 2007). Consequently, visual and auditory cues together ensure that the operator is alerted and directed towards the alarm.

1.1 False and Nuisance Alarms

When alarms are uninformative, the selective attention of the operator is compromised because alarms are difficult to ignore (Carrasco, 2011). Uninformative alarms can disturb or override the operator’s focused attention. *False alarms* are warnings where the emergency does not happen, while *nuisance alarms* are a type of false alarm that is uninformative due to context insensitivity (Xiao & Seagull, 1999). Both false and nuisance alarms are a common hinderance because of their uninformative nature and contribution to operator workload in alarm management (Goel et al., 2017).

False alarms are largely the result of conservative threshold settings to ensure safety because any missed event is commonly thought to result in serious physical and financial consequences (Getty et al., 1995). However, excessive false alarms can result in more non-responses and decreased operator compliance towards alarms (Breznitz, 1983; Meyer, 2004; Wickens et al., 2009). Conservative alarm settings have caused many alarms to interrupt operators unnecessarily and waste their time cross-checking parameters (Stanton & Babar, 1995; Dixon, Wickens, & McCarley, 2007; Maltz & Shinar, 2003). An operator can be pulled away from their primary task to manage these alarms that does contribute to the overall operational goals. One investigation in the Aviation Safety Reporting System (ASRS) database cited the cause of one accident was that a pilot became so preoccupied with clearing alarms that they failed to observe their altitude change (Bliss, 2003).

Many nuisance alarms occur because of the frequent correlation between alarms that is intrinsic to highly coupled equipment and process parameters (Moray, 2006). A single fault can generate many alarms simultaneously, and thereby reduce the informational value of an individual alarm to an operator. Compounding this issue, most control room designs still adopt the single-sensor single-alarm philosophy that results in multiple alarms for each process parameter (Vicente et al., 1996). Consequently, operators sometimes need to manage many simultaneous alarms that convey duplicate information.

Chattering alarms are also considered a nuisance because they occur due to small signal noises or fluctuations that happen around the alarm setpoints but do not indicate true disturbances. The International Society of Automation (ISA-18.2) defines a chattering alarm as one that repeatedly transitions between the alarm state and the normal state in a brief period (p.16, 2009). These alarms can comprise up to 60% of alarm annunciations (Rothenburg, 2009).

Alarms also become nuisance if they do not require operator action (US NRC, 2019). Some alarms only provide status messages (e.g., standing alarms) that do not link to any operational safety issues or prompt any control actions. Kragt and Bonten (1983) found that only 7% of alarms corresponded to operator action in a fertilizer production plant. Alarms are defined by their need for a response (ISA, 2009). If no response is necessary, then the alarm communicates a misguided urgency to attend an issue.

Alarm setpoints configured for one operating mode may become false in another (Hollifield & Habibi, 2011; US NRC, 2019). For example, during startup of a nuclear plant, the control room commonly displays a myriad of alarms because parameter values are below their lower bound setpoints (i.e., as water temperature of the steam generator increases). However, these alarms are unnecessary because the parameter values are normal within that context. This lack of context sensitivity can cause alarms to indicate normal rather than abnormal operation, which is confusing and a timewaster for operators to acknowledge.

Having these various nuisance alarms is problematic because it increases the chance of an alarm flood. Alarm floods occur when numerous alarms are triggered within a brief period exceeding an operator’s ability to resolve them promptly (Bransbury, 2001; Bullemer et al., 2011). Alarm flooding can result in the operator missing/neglecting relevant alarms, slowing down in their response, or even actively disabling informative alarms to reduce workload (Sorkin, 1988; Getty et al. 1995; Rayo & Moffat-Bruce, 2015). The sheer number of alarms can weaken a person’s ability to suppress distractor stimuli because of reduced signal to noise ratio (Carrasco, 2011). Collectively, the influence of nuisance alarms may overwhelm operators because “many nearly simultaneous alarms [have] varying degrees of relevance to the operators’ tasks” (p. 4-2 US NRC, 2019).

1.2 Advanced Alarm Management

The observation that traditional alarm systems can sometimes hinder rather than aid operators has led to substantial research adopting a process-centric approach to reduce nuisance and false alarms. Many of these techniques decrease correlated alarms. Multivariate process monitoring employs statistical analysis to test combinations of variables that can explain the process outcomes (Izadi et al., 2009a; Chen & Wang, 2017; Yu et al, 2017). Noda et al. (2011) applied event correlation analysis which groups similar process events together and identified 978 groups from a month of plant data.

Other research has focused on finding the root causes for alarms to minimize correlation and redundancy between alarms. The theory is that operators may only need the causal alarm presented rather than all the subsequent alarms to diagnose the incident. Rodrigo et al. (2016) used a similarity index and transfer entropy to identify the causal alarm during an alarm flood.

Similarly, Hu et al. (2017) also measured the amount of transfer entropy, which calculates the amount of independent information alarms have from one another. Their study was tested on 6 variables that caused 282 alarms in a 3-day period in which they were able to identify the two main variables causing the consequential alarms. Using dynamic fault trees to analyze the consequential effects of fault events, 628 false alarms were filtered out during 18 helicopter flights, with 12 flights having false alarms reduced by 80% or more (Simeu-Abazi, Lefebrvre, & Derain, 2011). Through identifying the temporal relationships between alarms, other researchers calculated the time delay between events and applied a multi-temporal sequence algorithm to uncovered causality (Bauer and Thornhill, 2008; Dorgo & Abonyi, 2018). The former was tested on a large chemical plant incident while the latter was tested on data from a laboratory-scale water treatment testbed.

Time-delays, filters, and dead-bands are commonly used to eliminate chattering alarms (Goel et al. 2017). A *time-delay* waits to trigger an alarm until multiple samples of the parameter exceed the alarm threshold, while a *dead-band* sets a higher limit to clear the alarm after it has been raised. *Filters* smooth out process data and remove noise that can cause nuisance alarms. Izadi et al. (2009b) showed that the combination of delays, filters, and dead bands could be used to optimize tradeoffs between false alarm rates and missed alarm rates based on the Receiver Operating Characteristic (ROC) curve. Wang and Chen (2014) employed timing intervals which is a modification of alarm delays that reduced chattering alarms by 88%. Additionally, the US Nuclear Regulatory Commission (US NRC) recommends separating status annunciators from alarms because they do not require operator action (2019).

Some approaches of alarm management have focused on reducing nuisance alarms by characterizing true abnormal circumstances carefully. For example, state-based alarming detects and gives separate alarm limits for each identified mode of operation (Hollifield & Hababi, 2011; Wang & Chen, 2016; Hu, Chen, & Shah, 2017). Additionally, model-based plant monitoring compares the modeled outcome of a plant's response versus its actual behavior to identify true abnormalities and subsequently true alarms quickly (Kim, Modaress, & Hunt, 1990; Izadi et al., 2009a).

Collectively, these alarm management techniques have worked to reduce the number of false and nuisance alarms. However, process-centric techniques alone may be insufficient. Despite the many techniques presented to reduce alarms, much of this research has not been implemented in practice, based on the observations of 15 various industrial control rooms in Germany (Bockelman et al., 2018). Furthermore, one study calculated that when an operator receives 20 simultaneous alarms, typical response to all of them takes approximately 88 minutes (Reising et al., 2004). Based on those results, the number of alarms experienced in control rooms seems inappropriate and ineffective.

Another concern is that system designers cannot feasibly anticipate all possible plant operations and thus all nuisance alarms a priori (Smith, Mccoy, & Layton, 1997; Wu & Li, 2018). The informativeness of an alarm also changes with respect to the operators' monitoring behavior, which evolves in real-time. The Positive Predictive Value (PPV) of an alarm system, a measure of informativeness, decreases as a function of operator expertise or vigilance (Seagull et al., 2000; Meyer & Bitan, 2002). If an operator is proactive in monitoring and can anticipate the upcoming event, some alarms would be considered premature and nuisance rather than helpful (Lee & Lees, 2007; Wickens et al., 2009). Alarms expected by the operators do not communicate new information by definition and turn into a distraction from other activities. In essence, an operator who is monitoring well would find more alarms to be distractors. As alarms are configured now, there are multiple situations in which alarms become non-informative due to operator expectancy (Xiao & Seagull, 1999). Thus process-centric techniques have intrinsic limitations in eliminating all false or nuisance alarms.

A complimentary approach of user-centric alarm management may help resolve the remaining nuisance alarms. Vicente (1999) advocates allowing end users to finish designs for flexibility during unanticipated situations. Recent alarm systems have advocated for more adaptive automation that support operator inputs (Miller & Parasuraman, 2007). These alarm system features enable users to define temporary alarms, adjust setpoints, configure filtering options, and sort alarms (Guerlain & Bullemer, 1996; Lehto, Papastavrou, & Giffen, 1998; Hollifield & Habibi, 2011; NRC, 2019). However, allowing users to set their own alarm thresholds can increase the chance of false negatives if they do not have the knowledge to make an informed alteration (Watson, Sanderson, & Russell, 2004; Botzer et al., 2010; Meyer & Sheridan, 2017). Also, configuring alarms represents a secondary task and imposes more workload on users. The exception to these limitations is a study that developed a human factor index that adapts alarm thresholds based on perceptual capacity and alarm treatment rate (Li, Wang, & Yan, 2019). But it remains that current user-centric alarm management cannot fully circumvent the issue of operator anticipation causing additional nuisance alarms.

Given the limitations of current adaptable alarm systems, the user-centric approach to alarm management could be supplemented with techniques that accommodate additional user requirements for handling nuisance alarms. The user-centric approach could include techniques of gathering feedback passively without introducing workload on the operator and adapting to the operator monitoring performance in relation to the informativeness of alarms. Past studies have demonstrated that eye tracking can estimate visual attention passively, and thus gauge their awareness of process parameters. For this reason, eye tracking represents a potentially useful technique for user-centric alarm management.

1.3 Eye Movements and Parameter Behaviors

Research has shown that operators change their visual sampling as a function of parameter behavior, which can indicate operator expectation and awareness. For instance, operators increase their visual sampling rate on parameters near the control limit (Crossman, Cooke, & Beishan, 1974; Kvalseth, 1979). In some cases, operators set mental targets for when they need to watch a parameter value more closely, which are more conservative than the original alarm settings (de Jong & Koster, 1971; Patternote, 1978). Similarly, Vicente et al. (2001) observed that operators tightened alarm limits or created pre-alarms, if they suspected that parameters would drift. This implies that when a parameter is at risk of reaching abnormal values, operators will be monitoring the parameter more closely to make sure those parameters do not alarm. These findings suggest that frequent sampling of a parameter near its control limit is likely correlated with the expectation of an alarm.

Operators also rely on trends of parameter values to inform their expectancy of alarms. This is because trend graphs reveal patterns of past data, allowing operators to predict future behavior (Burns & Hajdukiewicz, 2004). When watching random step changes of a dial, Robinson (1967) observed that a person’s estimation of the next value was the difference between their current observation and the average change of their earlier observations. This shows that human estimation is systematic and relies upon what they have seen in the past. Similarly, an operator’s estimation of parameter values has been modelled as calculating the likelihood of a parameter reaching a self-set limit where each visual sample resets their projection of the probability distribution curve (de Jong and Koster, 1971). Therefore, the more consistent the trendline has been in the past, the smaller the range of the probability distribution curve. More recently, studies have shown that fixation frequency is positively correlated to parameter's rate of change (Bitan and Meyer, 2007; Eisma, Cabrall, & de Winter, 2018). Together these studies suggest that an operator has greater expectancy of an impending alarm when the parameter shows a steady slope.

Monitoring behavior is also influenced by the entropy or fluctuation level of the parameters. Information theory defines entropy as an estimate of information value based on predictability (Shannon, 1948). Therefore, parameters which fluctuate are less predictable and provide more information per visual sample. Models of monitoring behavior have emphasized that operators sample parameters to maximize the return value for their effort rather than equally across parameters (Crossman, Cooke, & Beishon, 1974; Sheridan, 1970). Senders (1964) found that there were higher gaze rates and durations on parameters with higher entropy. Another study confirmed this monitoring behavior, revealing that people would not visually sample a parameter until the entropy was sufficiently high (83-100% of asymptotic value; Kvalseth, 1978). Bitan and Meyer (2007) found that slope variability of a parameter hindered a person's ability to predict future values, and therefore induced higher sampling rates compared to parameters changing at constant rates. These monitoring behaviors suggest that operators sample parameters with high entropy more often compared to other others. As a result of frequently monitoring a parameter because of the uncertainty and constant risk of an alarm, the operator will likely anticipate when an alarm occurs.

Current research reveals that monitoring behaviors or eye movements change based on actual and expected parameter behaviors. When a parameter is near the control limit, trending towards abnormal states, or fluctuating drastically, the operator will sample the parameter more frequently, suggesting anticipation of an alarm. In such circumstances, the operator anticipation could render those alarms less informative, if not a nuisance. Thus, the saliency of those types of alarms should be minimized so that operator's attention can be allocated towards more informative alarms and their current tasks.

My thesis research investigates eye-tracking to assist with acknowledging alarms by subduing auditory and flashing visual cues when eye-gaze fixates on a parameter shortly before crossing the alarm threshold. Three types of gaze-based acknowledgement methods were designed corresponding to the situations when operators fixate on parameters that are near the alarm thresholds, trending towards alarm thresholds, and fluctuating dramatically, respectively. A dual-task experiment involving alarm monitoring and parameter control was conducted to evaluate the three gaze-based acknowledgment methods against the normal alarm presentation method (i.e., control/no gaze-based acknowledgement) in three different scenario types. Given the potential to reduce alarm distraction, the central experimental hypothesis was a main effect of gaze-based acknowledgement methods, in that the three gaze-based acknowledgement methods will outperform having no gaze-based acknowledgement in performing a parameter control task while also performing a monitoring task. Furthermore, an interaction effect between gaze-based acknowledgement methods and scenario types was hypothesized in that the gaze-based acknowledgement methods yielded different parameter control task performance depending on the scenario types.

2 Methods

2.1 Participants

Twenty-four participants were recruited (14 males and 10 females between the ages of 18 to 45 years old) via emails to various listservs and postings on physical bulletin boards on a university campus. Prerequisite knowledge of plant operation was not an inclusion criterion. The gaze-based acknowledgement methods proposed do not yet account for the unique mental models and strategies of professional operators. Therefore, this research began investigation on the psychophysical aspects of the gaze interaction that can be tested without prior operational experience. The inclusion criteria for this study were the non-vulnerable population that also had normal or corrected-to-normal vision and no history of photosensitive epilepsy.

2.2 Experimental Apparatus

The experimental apparatus consisted of two LCD monitors that were operated using a keyboard and mouse (Figure 1). The bottom monitor was dedicated to the alarm monitoring task operated using a mouse (see section 2.3.1). Tobii X3-120 remote eye tracker was attached to the bottom monitor to collect eye gaze data on that monitor. The top monitor was dedicated to the parameter control task operated using a keyboard (see section 2.3.2).

Figure 1. Participant (left) views tasks from two monitors while facilitator (right) observes and sets up each trial

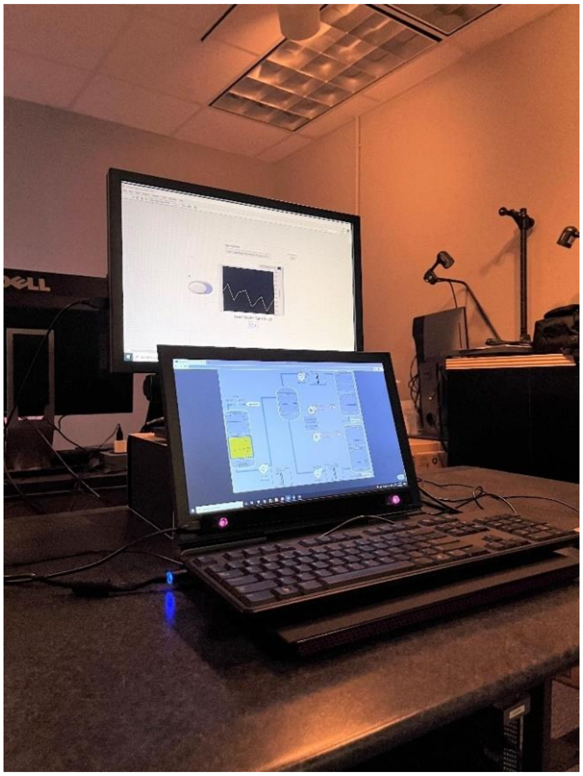


Figure 2. Parameter control task (top screen) and alarm monitoring task (bottom screen) with the remote eye tracker

2.3 Experimental Tasks

2.3.1 Alarm Monitoring Task

On one LCD monitor (top one in Figure 2), participants watched alarms in videos of simulated chemical plant operations. These videos were displayed through Synquesticon, a software designed for collecting eye-tracking data and user inputs via mouse and keyboards to evaluate display design and human performance in the process control domain (Hildebrandt, Langstrand, & Nguyen, 2019).

The videos for the alarm monitoring task were created with a chemical plant simulator developed by the Idaho National Laboratory (INL; LeBlanc et al., 2015). The chemical plant moves water into storage tanks, heats the water, and mixes the water with the "Pantheradine" chemical into two different concentrations (Figure 3). The simulator includes seven major process parameters: heater tank fill level, heater tank water temperature, accumulation tank fill level, mixing tank A fill level, mixing tank A concentration level, mixing tank B fill level, and mixing tank B concentration level. Each parameter is depicted with numerical values along with a trend graph that accounts for approximately 15-second of operation history. The trend graph shows the allowable operating range with two sets of tick marks. The inner tick marks trigger the high alarm, and the outer ones trigger the high-high alarm that shuts down the plant. The high alarm is represented by highlighting the parameter yellow and a repeating chime. This display design represents a modern visual display of process plants where alarm signals are embedding into the mimic display.

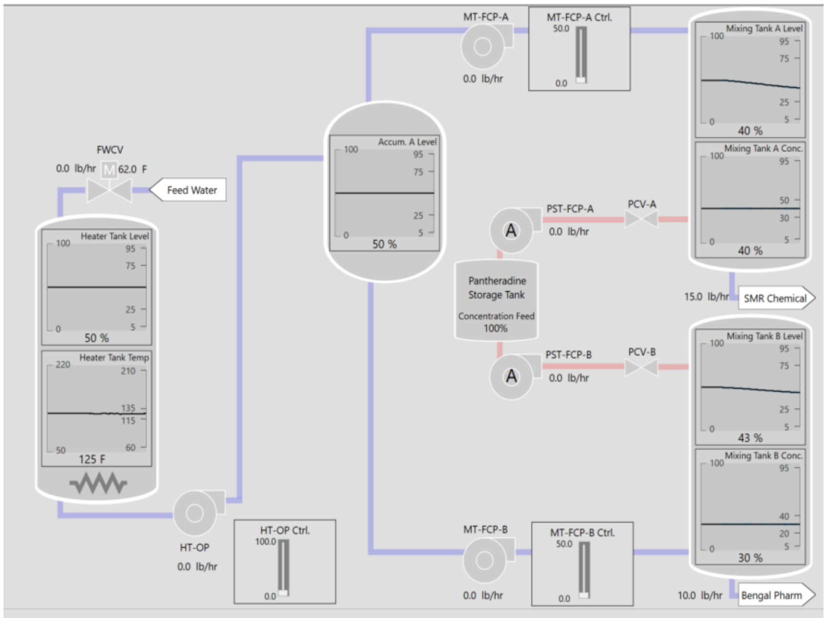


Figure 3. Chemical plant simulator

Participants were tasked to click with their mouse on any parameter that they predicted would reach its alarm threshold within four seconds. A blue dotted line around the parameter marked their prediction of an impending alarm, which would remain for three seconds (Figure 4).

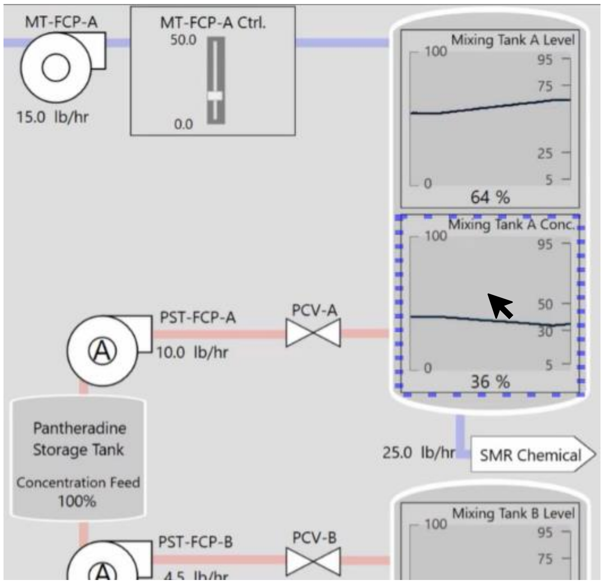


Figure 4. Blue dotted outline signifies participant prediction of impending alarm

Simultaneously, the participants' eye gaze was tracked so that alarms could be acknowledged according to the gaze-based method being tested for that block of trials. If acknowledged, the normal alarm presentation of the yellow highlight visual cue and auditory alert is replaced with a yellow outline and no auditory alert (see Figure 5). This presentation design was intended to reduce the salience of the alarm when the participant has gazed at the parameter earlier and to provide feedback to the participant of the gaze-based acknowledgement.

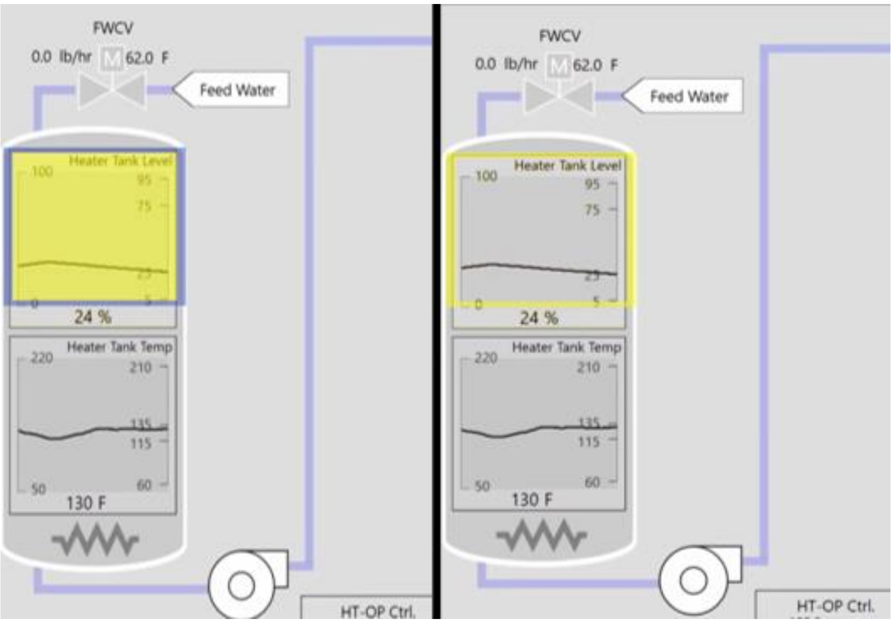


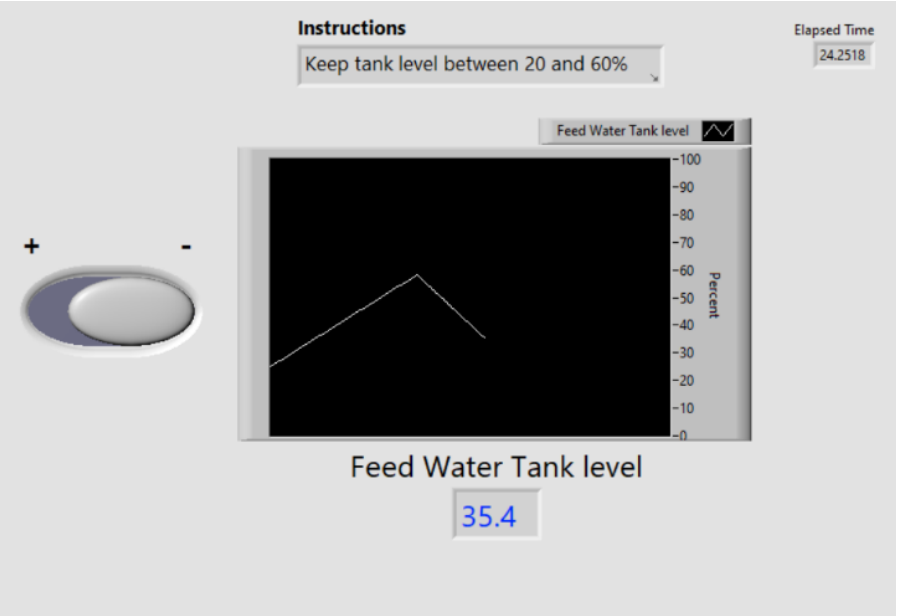
Figure 5. Alarm (left) versus alarm during gaze-based acknowledgement (right)

2.3.2 Parameter Control Task

On another LCD monitor (bottom in Figure 2), participants performed a parameter control task of maintaining process flow of the feed water tank between 20-60% in a LabView-based software application. Participants were able to open a drain, represented by a toggle on switch on the user interface (Figure 6). If the level got too high, the participant should press the 'Enter' button on the keyboard to open the drain. Activating the button caused the current slope to decrease by 1.5 every 100ms. Equation 1 defines the parameter behaviors of the parameter control task. The participant began this task ten seconds prior to the alarm monitoring task.

$$Y = \text{random}(t_1) + C - 20(t_2) \qquad \textbf{(1)}$$

where Y is the current parameter value; random is the slope value; t_1 is time since start of task; C is the starting value of the parameter, and t_2 is the duration the toggle is negative.



Instructions

Keep tank level between 20 and 60%

Elapsed Time 24.2518

Figure 6. Parameter control task

There were three different versions of the parameter control task being randomly administered across trials for all participants. The three versions varied in the starting value of the line (25, 40, and 55%) and the slope values (between .1 to .9 every 6 seconds) to limit practice effect of anticipating when to open the drain from trial to trial.

2.4 Experimental Manipulations

For this study, there were two experimental manipulations: (1) gaze-based alarm acknowledgement method, and (2) type of operational scenarios.

2.4.1 Gaze Acknowledgement Methods

This study investigated three gaze-based alarm acknowledgement methods and a control condition (which did not have any alarm acknowledgement). The gaze-based acknowledgement methods are the different requirements on the gaze behaviors for changing alarm presentation or salience. The three gaze-based acknowledgement methods were: proximity-based, prediction-based, and entropy-based. The seven parameters were marked as Areas of Interest (AOIs) where fixations were computed in real time during data collection (Figure 7). The parameter predetermined to alarm in each video was labeled the target AOI. Fixations were computed with the Velocity-Threshold Identification (I-VT) algorithm, which distinguishes fixations from saccades based on the difference in angular speed (Salvucci & Goldberg, 2000). This algorithm has the fastest computation time and was thus deemed most appropriate for our real-time application.

The gaze-based acknowledgement methods functioned by counting fixations on the target AOI within a time window of 1 to 4 seconds prior to the alarm. This time window is selected based on intention-based research in the automotive industry which has shown that a three to five second window was the most accurate for predicting intention (Lethaus et al., 2013). In this study, all methods disregard fixations on target AOIs that are one second or less before an alarm occurs because the time interval between the glance and alarm would likely be too short for a human to anticipate the alarm in an endogenous manner.

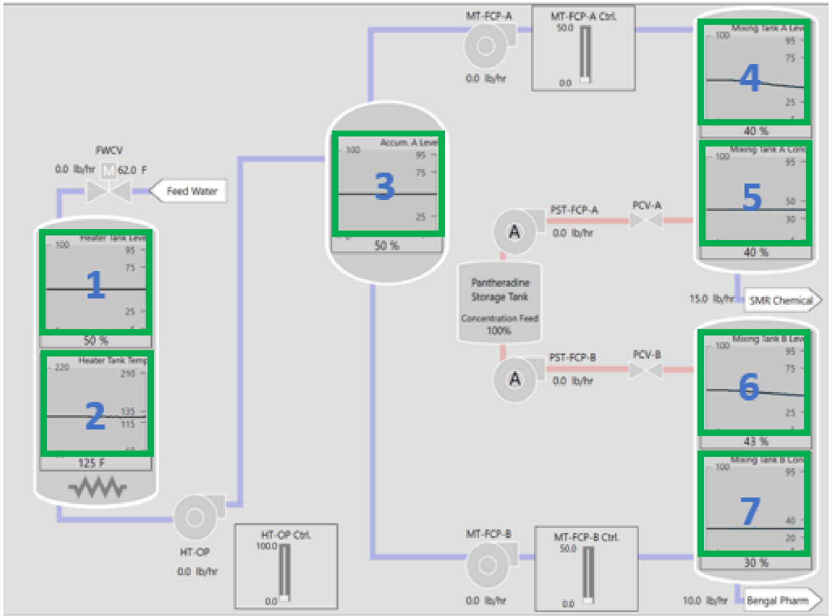


Figure 7. AOI locations and parameter identifiers

2.4.1.1 Proximity-based Acknowledgement Method

The proximity-based method acknowledged an alarm when (1) the parameter value *exceeded the acknowledgement thresholds*, and (2) the person had *two or more fixations on the target AOI* within the time window between one and four seconds before the alarm. The gaze acknowledgement thresholds are set at 10% and 90% of the range between the high and low alarms of the parameter. Table 1 presents the acknowledgment and alarm thresholds for each parameter. Figure 8 presents instances of a parameter activating (black line) and not activating (red line) the gaze-based alarm acknowledgement by the proximity-based method, respectively. The black line activates the gaze-based acknowledgement because the value of the parameter is above the high acknowledgement threshold, assuming the person also has two or more fixations on the parameter/target AOI. The red line does not activate alarm acknowledgement because the parameter value is not surpassing the acknowledgement thresholds irrespective of number of fixations on the parameter.

Parameter	High alarm/Gaze Acknowledgement threshold	Low alarm/Gaze Acknowledgement threshold
Heater tank level	75/70	25/30
Heater tank temperature	135/133	115/117
Accumulator A level	75/70	25/30
Mixing tank A level	75/70	25/30
Mixing tank A concentration	50/48	30/32
Mixing tank B level	75/70	25/30
Mixing tank B concentration	40/38	20/22

Table 1. Proximity-based thresholds compared to original alarm

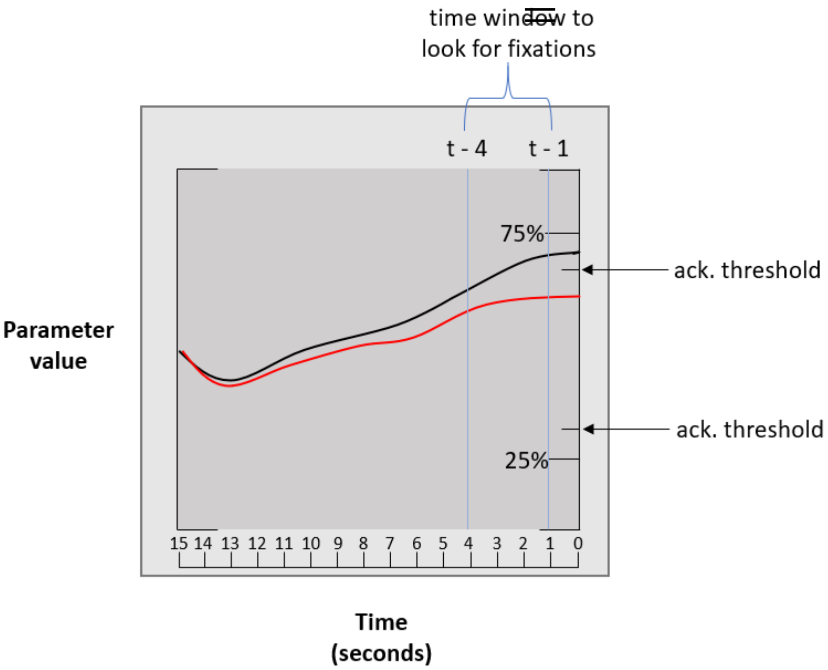


Figure 8. An instance where the proximity-based acknowledgement method would activate (black trendline) and an instance where it would not (red trendline). t denotes time.

2.4.1.2 Prediction-based Acknowledgement Method

The prediction-based method acknowledged an alarm when (1) the parameter value was *predicted to exceed the alarm thresholds* using an equation that forecasts three seconds ahead, and (2) the person had *two or more fixations on the target AOI* within the time window between one and four seconds before the alarm. To predict whether the parameter would exceed the alarm threshold, a second-degree polynomial curve was fitted based on parameter's value over the past 15 seconds. Figure 9 shows a graphical representation of the prediction-based method. The black line depicts a parameter that is projected to reach the high alarm threshold based on the end value of the look-ahead time window (extended dotted line). The gaze-based acknowledgement method would activate in this instance if the person had two or more fixations on the parameter between one and four seconds prior to the alarm. The red line is not projected to reach the high alarm threshold and therefore, would not activate the gaze-based acknowledgment irrespective of the number of fixations on the parameter.

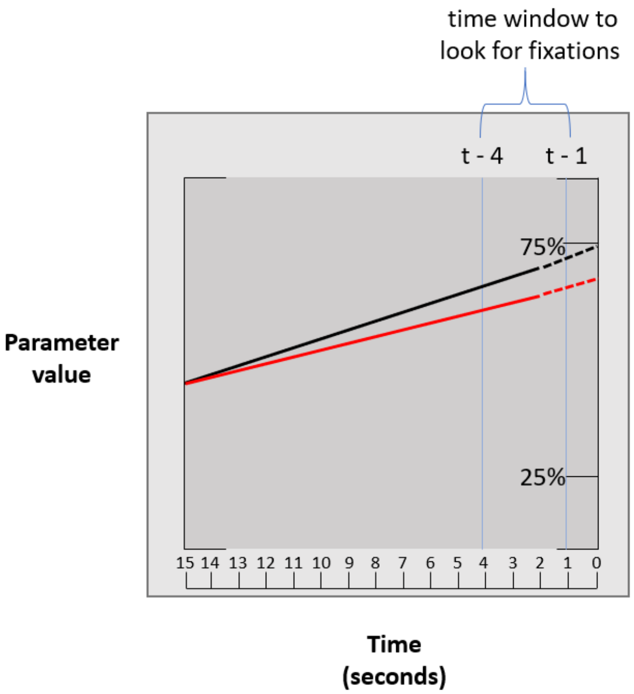


Figure 9. An instance where the prediction-based acknowledgement method would activate (black trendline) and an instance where it would not (red trendline). t denotes time.

2.4.1.3 Entropy-based Acknowledgement Method

The entropy-based method acknowledged an alarm when (1) the parameter values of the past 15-second time window *exceeded an entropy value of 1.2*, and (2) the person had *two or more fixations on the target AOI* within the time window between one and four seconds before the alarm. Entropy was calculated using a histogram estimator of the parameter value frequencies during the past 15 seconds (Hall & Morton, 1993; Wallis, 2006). Figure 10 presents a graphical representation of this entropy-based method. The black line shows high fluctuation within the last 15 seconds that has exceeded an entropy value of 1.2; thus, the gaze-based alarm acknowledgement would occur if the person also had at least two fixations on the parameter between one and four seconds prior to the alarm. For the red line, although the parameter drops below the low alarm threshold, the entropy-based method would not acknowledge the alarms irrespective of the fixations on the parameter because of the low entropy value for that parameter.

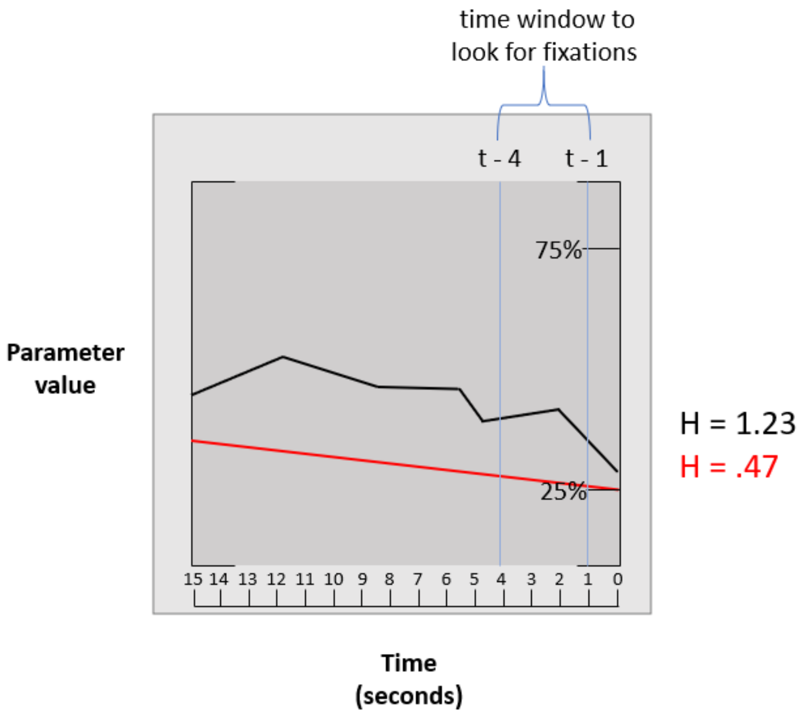


Figure 10. An instance where the entropy-based acknowledgement method would activate (black trendline) and an instance where it would not (red trendline). t denotes time.

2.4.2 Scenario Types

The gaze-based acknowledgement methods were tested on three scenario types. A scenario type is a video of the chemical simulator that is labeled based on the behavior of the target parameter or AOI right before the alarm triggers. Each video was a minute long. A scenario contained one alarm while other parameters served as distractors. Distractors were defined as parameters behaving in a way that would fulfill the requirements of one of the three gaze-based acknowledgement methods, but the parameters never reached their alarm thresholds. The parameter chosen to alarm varied evenly by the location on the LCD monitor (i.e., left, middle, right). The three scenario types were:

1. *Near-threshold scenarios* were characterized by a parameter stabilizing near the alarm threshold and thereby presenting an imminent risk of crossing the threshold.
2. *Trending scenarios* were characterized by a parameter increasing or decreasing towards the alarm thresholds continuously for four or more seconds and thereby presenting an increasing risk of crossing the alarm thresholds.
3. *Fluctuation scenarios* were characterized by a parameter having entropy greater than one based on past 15 seconds of parameter values and thereby presenting risk of crossing the threshold in an unexpected manner.

2.5 Experimental Design

This study employed a 4x3 fully crossed within-subject design with a treatment of gaze acknowledgement method of four levels (no gaze acknowledgement/control, proximity-based, prediction-based, and entropy-based), and a treatment of scenario type of three levels (near-threshold, trending, and fluctuation). The control condition had no gaze alarm acknowledgement (i.e., all alarms occur with the visual and auditory cues regardless of the participant eye gaze).

The participants experienced each gaze-based acknowledgement method in a block of six trials. The presentation order of alarm acknowledgement method was fully counterbalanced given the four conditions distributed over 24 participants (i.e., 4! =24).

The experiment included six unique scenarios, two of each scenario type. The participants experienced all six scenarios, once for every gaze-based acknowledgement method, resulting in 24 trials in total for each participant. The presentation order of the scenarios was randomized with no repeating permutations (see Table 2 for an example of trial ordering). The total number of trials is N=576 (i.e., 24 participants x 4 acknowledgement methods x 3 scenario types x 2 scenarios of each type).

Block 1: Control	SCN 1	SCN 2	SCN 3	SCN 4	SCN 5	SCN 6	questionnaire/ BREAK
Block 2: Proximity-based	SCN 2	SCN 3	SCN 4	SCN 1	SCN 6	SCN 5	questionnaire/ BREAK
Block 3: Prediction-based	SCN 3	SCN 4	SCN 6	SCN 2	SCN 1	SCN 5	questionnaire/ BREAK
Block 4: Entropy-based	SCN 5	SCN 6	SCN 3	SCN 1	SCN 2	SCN 4	questionnaire/ END

Table 2. Example of scenario order for one participant

2.6 Procedure

This experiment took place at the Virginia Cognitive System Engineering (VACSE) Laboratory at Virginia Polytechnical and State University. The experimenter gave participants an information sheet about the experiment. Obtaining written consent was waived by the IRB given the minimal risks imposed by the study. A demographic questionnaire was administered to collect data on age, gender, degree, and any relevant experience. Participants received training on how to complete both alarm monitoring and parameter control tasks through a PowerPoint presentation.

After the PowerPoint presentation, the experimenter performed eye-tracker calibration on the participants. Then, they practiced each task twice and had three practice trials of the dual-task, one for each scenario type. Following the three practice trials, the participants completed four blocks of six video trials. Each block administered only one gaze-based acknowledgement method and took approximately ten minutes. After each block, the participants answered a short questionnaire inquiring about their awareness of parameters being monitored for alarms and perception of usability. Then they were given a two-minute rest break. Re-calibration of the eye-trackers occurred after each break. After completing four blocks, the participants were debriefed on the experiment and compensated for their time. The experiment lasted approximately 90 minutes per participant.

2.7 Measures

The measurements included: gaze-based acknowledgement occurred for each trial, prediction of impending alarms by clicking on parameters/target AOI, the amount of time of parameter out-of-range for the parameter control task, and subjective ratings on parameter awareness of parameters in the alarm monitoring task and usability on alarm acknowledgement methods (Table 3).

Category	Metric
Eye Movement	Gaze-based acknowledgement occurred/did not occur
Prediction Performance	Number of hits
	Number of misses
Parameter Control Task Performance	Time out of permissible parameter range
Perception Questionnaire	Usability (3 items)
	Awareness (3 items)

Table 3. Collected metrics

2.7.1 Eye Movement

Measuring whether gaze-based acknowledgement occurred during a trial indicates the usage level of each gaze-based acknowledgement method. Low acknowledgement numbers imply that the method is not being passively used by the participants.

2.7.2 Prediction Performance

Performance of monitoring alarms were based on number of hits and misses computed by the mouse clicks on the parameters to indicate impending alarms in each trial. Correct rejection was omitted because all trials contain an alarm. False alarm was also omitted because multiple false alarms could occur within a trial and would skew the results without correct rejection. Table 4 summarizes the prediction performance indicators. These prediction performance indicators provided an explicit verification on whether the gaze-based acknowledgement methods match participant anticipation of alarms. Eye tracking alone cannot serve as both the means of interaction and verification of awareness for the experiment.

Miss	No click on target AOI during fixation window within $t < 4$ seconds and $t > 1$ second to alarm
Hit	Click on target AOI with alarm incurring within $t < 4$ seconds and $t > 1$ second

Table 4. Definition of hits and misses for prediction performance

2.7.3 Parameter Control Task Performance

Parameter control task performance was assessed by the amount of time that the parameter falls out of range within the scenario. A large amount of time outside the designated range of the parameter was interpreted as poor performance.

2.7.4 Subjective Ratings

A questionnaire was administered to collect participant ratings on usability (3 items) and awareness (3 items) after every block. Table 5 presents the questionnaire items. A high usability score implies that the gaze-based acknowledgement method was deemed helpful by participants. A high awareness score would suggest that the participant was confident about changes in parameter values of the chemical plant simulator during the block of trials.

Construct	Question (1-strongly disagree, 5- strongly agree)

Usability	The gaze-based interaction anticipated my needs.
	The gaze-based interaction was correct in assuming I anticipated the alarm.
	The gaze-based interaction was not effective in minimizing alarms I was aware of.
Awareness (Monitoring task only)	I was aware of the parameters' general value most of the time.
	I did not anticipate when parameter values would change.
	I easily detected when parameters were reaching their thresholds.

Table 5. Questionnaire items

3 Data Analysis and Results

3.1 Parameter Control Task Performance

To determine whether disruption to the parameter control task was reduced with the gaze-based acknowledgement, the mean time out of range for the parameter control task for each trial (N=576, also see experimental design in Section 2.5) were analyzed with a repeated measures multivariate analysis of variance (MANOVA) with an experimental design of four acknowledgement methods full crossed with three scenario types. Instead of treating each response as an observation and blocking for subjects in a conventional repeated measures ANOVA, each repeated factor combination is treated as a separate variable in a MANOVA. The multivariate approach was chosen because it is a more conservative test which reduces the likelihood of a Type I error (Lehman et al., 2013). MANOVA is also generally more appropriate for small sample sizes because the sphericity assumption needed for repeated measures ANOVA may not have enough power to represent an accurate result. The multivariate normality assumption was not satisfied but MANOVA is generally robust to this assumption (Lehman et al., 2013). The data distribution suggested a moderate ceiling effect. The MANOVA (Table 6) revealed a significant main effect for scenario type, F(2,22) = 3.53, p=0.047 (Table 6). Fluctuation scenarios had the largest mean time out of range compared to near-threshold or trending (Figure 11). Contrary to what was hypothesized, there was no main effect of acknowledgment method (p = 0.176; Table 6).

Source	NumDF	DenDF	F-Value	P-Value
Acknowledgment Method	3	21	1.81	0.176
Scenario Type	2	22	3.53	0.047 *
Ack Method* Scenario Type	6	18	2.55	0.058 *

Table 6. MANOVA of parameter control task performance

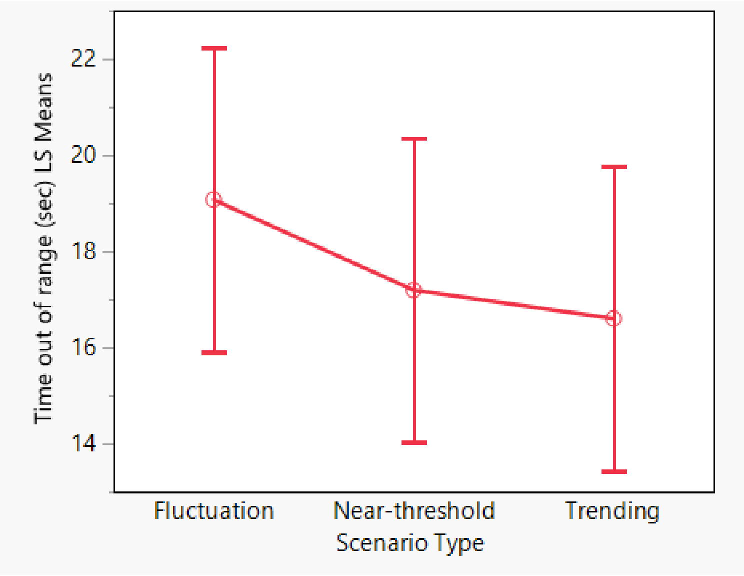


Figure 11. Means plot of parameter control task performance by scenario type.

	Near threshold	Trending	Fluctuation
Proximity-based	Proximity 🟥 + Near threshold	Proximity + Trending	Proximity + Fluctuation
Prediction-based	Prediction + Near threshold	Prediction + Trending	Prediction + Fluctuation
Entropy-based	Entropy + Near threshold	Entropy + Trending	Entropy + Fluctuation
Control	Control + Near threshold	Control + ◆ Trending	Control + 🔵 Fluctuation

Table 7. Predicted interaction (highlighted) versus actual best performance (shapes)

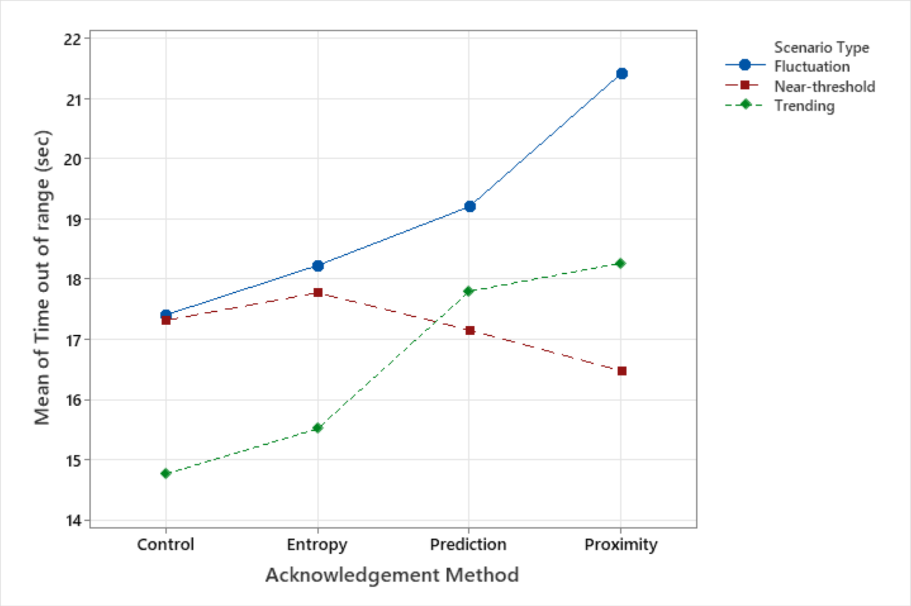


Figure 12. Interaction effect between acknowledgement method and scenario type on parameter control task performance

The MANOVA also revealed a marginal interaction effect, F(6,18) = 2.55, p=0.058, but the performance did not vary between gaze-based acknowledgement methods and scenario types as hypothesized (**Error! Reference source not found.**). Results indicate that fluctuation scenarios always had the highest time out of range regardless of acknowledgement methods (Figure 12). The best performance (lowest time out of range) for fluctuation and trending scenarios was during the control condition. Only the proximity-based acknowledgement method yielded the hypothesized parameter control performance result, showing the least time out of range for the near-threshold scenarios.

3.2 Calibration of Acknowledgement and Prediction Performance

Given that gaze-based acknowledgement should represent a person's anticipation of an alarm, an analysis was conducted that compared whether gaze acknowledgement would differentiate prediction outcome. For the proximity-based, prediction-based, and entropy-based acknowledgement conditions (N=432=3 acknowledgement method x 3 scenario types x 2 scenario/type x 24 participants), a contingency table is generated by tallying the trials having a hit or a miss for the alarm prediction decision, and an acknowledgement to have occurred or not occurred (Table 8). This statistical test omitted all 144 trials of no gaze acknowledgement condition as acknowledgement was not possible. The Chi squared test of independence showed a difference in the number of gaze-based acknowledgements between whether the participants correctly predicted the alarm by clicking on the target AOI/impending alarm (i.e., hit; X^2 (1, N = 432) = 23.802, p < 0.0001, V=0.235). Prediction hit rate was more frequent when acknowledgements did occur than did not occur (Table 8). The largest cell value was when participants neither clicked nor gaze acknowledged the target AOI/impending alarm (i.e., miss and no gaze acknowledgement). Additionally, participant gaze acknowledgements of an alarm also occurred more frequently when participants did not indicate a prediction of an alarm (i.e., miss) than when they did indicate a prediction (i.e., hit).

		Acknowledgement Trial count (% of total trials)		
		Not Occurred	Occurred	Total
Prediction Performance (Clicked on target AOI/impending alarm)	Miss	378 (64.8)	89 (20.6)	367
	Hit	29 (6.7)	36 (8.3)	65
	Total	307	125	432

Table 8. Contingency of acknowledgements and number of hits

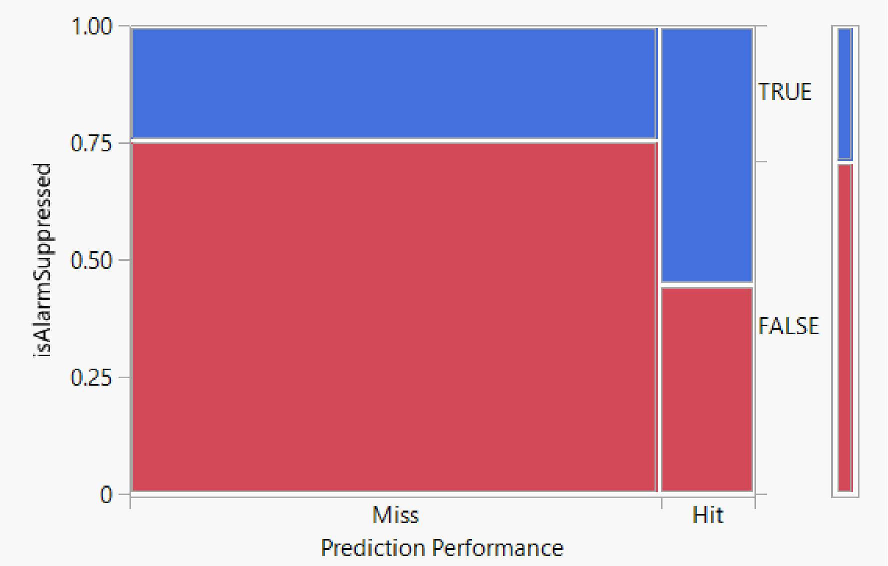


Figure 13. Proportion of acknowledgements with prediction performance.

Two separate Chi squared tests were conducted to investigate how scenario types affected the number of acknowledgements and prediction performance. A contingency table was generated by tallying the trials having an acknowledgement to have occurred or not occurred by scenario types over all trials except the no gaze acknowledgement condition (N=432; Table 9). This statistical test omitted all 144 trials of no gaze acknowledgement condition as acknowledgement was not possible. The Chi squared test of independence showed a significant difference between the number of gaze-based acknowledgement that occurred across the scenario types (X^2 (2, N = 432) = 30.147, p < 0.0001, V=0.264).

Another contingency table is generated by tallying the trials having a hit or a miss in alarm prediction by scenario types over all trials (N=576; Table 10). The Chi squared test of independence showed a significant difference in the number of hits (or correctly predicting alarms) across scenario types (X^2 (2, N = 576) = 10.567, p = 0.005, V=0.135). The two statistically significant results indicate that scenario type impacted both gaze-based acknowledgements and prediction hit rate. re appears to be no correspondence between the number of acknowledgements and correct alarm prediction (i.e., hit) with respect to scenario types. Most gaze acknowledgements occurred in the fluctuation scenarios (Table 9) but the most correct predictions occurred in the near-threshold scenarios (Table 10).

		Acknowledgement Count (% of total trials)		
		Not Occurred	Occurred	Total
Scenario Type	Fluctuation	78 (18.1)	66 (15.3)	144
	Trending	116 (26.9)	31 (7.1)	144
	Near-threshold	113 (26.1)	28 (6.5)	144
	Total	307	125	432

Table 9. Contingency of acknowledgements and scenario type

		Clicked on Target AOI/Impending Alarm Count (% of total trials)		
		Miss	Hit	Total
Scenario Type	Fluctuation	164 (28.5)	28 (4.9)	192
	Trending	172 (29.9)	20(3.7)	192
	Near-threshold	149 (25.9)	43 (7.3)	192
	Total	485	91	576

Table 10. Contingency of the number of hits by scenario type

3.3 Usage of Gaze-based Acknowledgement

To analyze usage of an acknowledgement methods by scenario types, a contingency table is generated by tallying the trials having an acknowledgement to have occurred for each acknowledgement method by scenario types (Table 11). This statistical test omitted all trials of no gaze acknowledgement condition as acknowledgement was not possible. The Chi squared test showed a significant difference between number of acknowledgements in acknowledge methods across the three scenario types (X^2 (4, $N = 125$) = 26.55, $p < 0.0001$, V=0.325; Table 11). The trials employing the entropy-based acknowledgement method in fluctuation scenarios resulted in the most acknowledgements. The fluctuation scenarios also had the most acknowledgements across all acknowledgement methods (Figure 14).

		Scenario Types Count of Acknowledgement (% of total trials)			
		Fluctuation	Near-threshold	Trending	Total
Acknowledgement Method	Entropy	28 (22.4)	0 (0)	5 (4)	33
	Prediction	18 (14.4)	8 (6.4)	15 (12)	41
	Proximity	20 (16)	20 (16)	11 (8.8)	51
	Total	66	28	31	125

Table 11. Contingency of acknowledgements sorted by acknowledgement method and scenario type

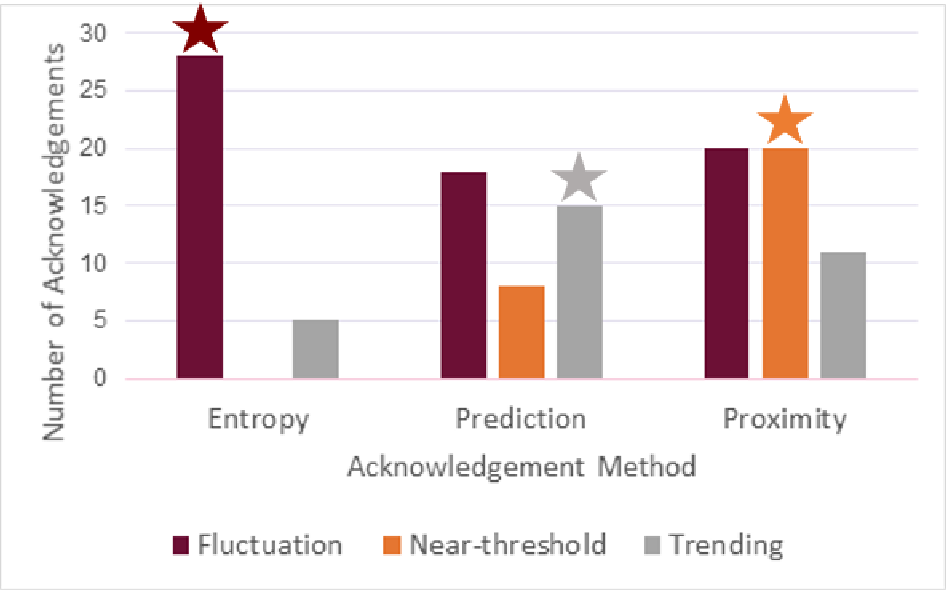


Figure 14. Number of acknowledgements across scenario type (bars) and acknowledgement method. The stars denote the expected interaction.

3.4 Subjective Ratings

Lastly, the usability questionnaire administered at the end of every block of trials was used to determine participants' opinion of the usefulness of each gaze-based acknowledgement. A one-way repeated measure Analysis of Variance (ANOVA) with a fixed factor of alarm acknowledgement method and a random factor of participant was conducted analyzed usability ratings that was averaged across the three usability questionnaire items for every block of trials per participant. However, the analysis omitted usability ratings from two participants who did not complete the questionnaire properly (thus, N=22x4). The usability scale showed good internal consistency across the three items (i.e., $\alpha \geq 0.8$), as determined by Cronbach's alpha (α= 0.829). The usability ratings were not normally distributed (Shapiro-Wilk test W=.0167 p = 0.017). The data distribution has kurtosis of -0.389 and skewness of -0.305. The type I error rate for non-normal data for ANOVA should remain within the bounds of Bradley's criterion (i.e., .025 and .075) if the skewness or kurtosis is between -1 to 1 (Bradley, 1978; Blanca Mena et al., 2017). The sphericity assumption was satisfied according to the Mauchly criterion (X^2 (3) = 3.836, p = 0.573). Table 12 showed significant difference in usability ratings between acknowledgement methods (F(3, 63) =3.744, p = 0.015, η^2=0.151). Proximity had the best usability score (M=3.59, SE=0.19), followed by prediction (M=3.50, SE=0.19), entropy (M=3.30, SE=0.19), and then control (M=2.92, SE=0.19; see Figure 15). Tukey's HSD Test found that the mean usability rating was significantly different between Control and Proximity ($p < 0.05$, 95% C.I. = [0.096, 1.237]) as well as Control and Prediction ($p < 0.05$, 95% C.I. = [0.005, 1.146]).

The analysis and results on the awareness ratings were omitted in thesis because the awareness scale showed questionable internal consistency across the three items given a Cronbach's alpha of 0.667, which is less the generally accepted criterion of 0.8. Participants also verbally indicated confusion about the questionnaire items.

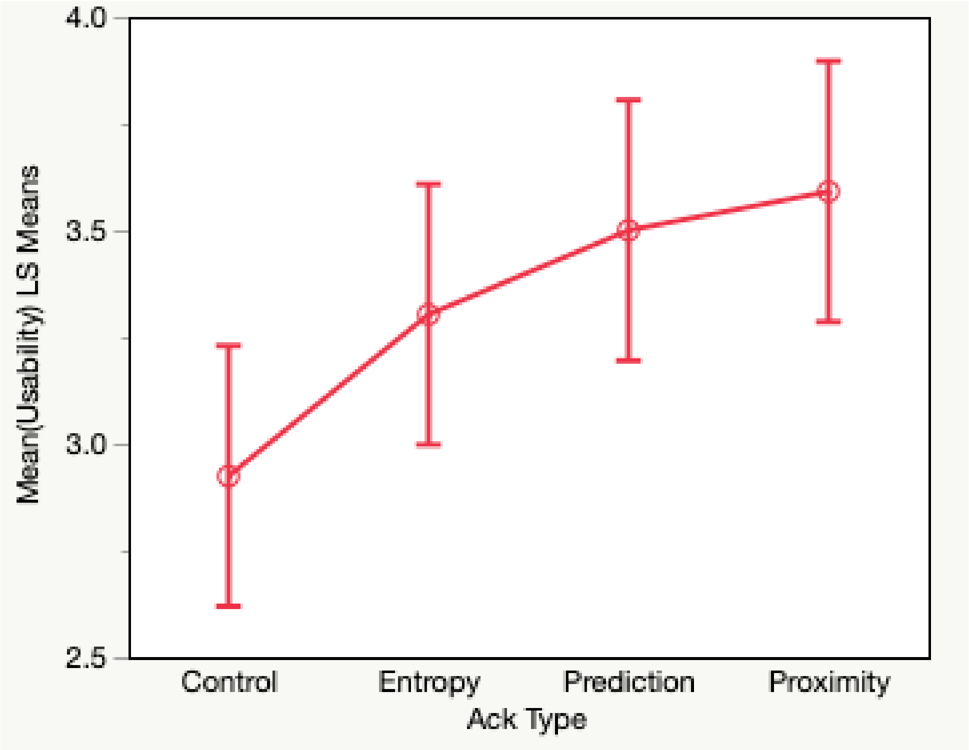


Figure 15. Means plot of usability ratings across acknowledgement method

Source	DF	SS	MS	F-Value	P-Value
Acknowledgement Method	3	5.771	1.923	3.7445	.0153*
Error	63	32.367	0.514		
Total	66	38.138			

Table 12. ANOVA of Usability ratings

4 Discussion

The design of gaze-based acknowledgement is intended to reduce the salience of alarm presentation and thereby improve focused attention and performance of other on-going tasks. The central study hypothesis is that providing gaze-based acknowledgement methods would result in less time out of range in the parameter control task because of reduced disruption on focused attention due to nuisance alarms. This main hypothesis was not well-supported in that there was no significant main effect of alarm acknowledgement on the parameter control task performance (i.e., time out of range) in the MANOVA results, although participants gave higher usability ratings for the gaze-based acknowledgement methods compared to no gaze-based acknowledgement, suggesting some potential merits for the alarm management approach.

One explanation for these results could be that participants may not have given sufficient attention to the alarm monitoring task. There was a minor ceiling effect in the control task performance. The number of misses without gaze-based acknowledgement were also disproportionally high (Table 8) indicating a lack of visual sampling as well as poor monitoring performance. These results support the interpretation that the participants did not allocate sufficient attention to the alarm monitoring task. Additionally, more than 50% of participants rated monitoring alarms as difficult in the debrief questionnaire (Appendix C), confirming insufficient resource allocation for adequate performance for alarm monitoring. Due to the parameter control task already taking most of participants' attention, the gaze-based acknowledgement might have limited impact on re-directing attention.

This explanation indicates the importance of task characteristics on whether gaze-based acknowledgement could reduce disruption to other tasks. The alarm monitoring task had multiple parameters to watch, with a single alarm occurring once every 70 seconds. Participants had no direct influence to change plant operations but marked their predictions for impending alarms. In contrast, the parameter control task required participants to monitor one parameter that would require corrective action approximately every 10 seconds. The aspect of the task which seems to influence allocation of attention might be the amount of interaction demanded by the task and the timing of the feedback. If the parameter control task reached the limit of the desired range, participants could toggle the switch and immediately alter the outcome of that task, whereas the alarm monitoring task did not set any specific requirements as the participants could predict alarms whenever and however many times they wanted. Participants did not know whether they predicted accurately unless the alarm occurred subsequently. This contributed to uneven allocation of resources between the two parallel tasks. Consequently, the attention given to the alarm monitoring task may be whatever is remaining after sufficiently managing the parameter control task, although the participants were instructed that the parameter control task could be sacrificed to meet the objectives of alarm monitoring. From this perspective, the design implication is that gaze-based acknowledgments might only be effective at reducing disruption for on-going tasks that require similar cognitive processing and physical interactions. Perhaps future research might examine two inter-related tasks or downstream impacts of series of similar tasks. In such cases, impacts of alarms would translate to all task performance.

An alternative explanation for the lack of improvement on the parameter control task is that the participants relied on the alarms for feedback on their prediction. If so, the higher alarm salience may provide a facilitation effect in confirming their prediction of parameter behaviors. Their task objective was framed such that the focus was on predicting alarms through clicking on the parameter. There was no immediate indication of how well they did until the gaze-based acknowledgement or alarm occurred. Given that the gaze-based acknowledgement was silent and had reduced color compared to the normal alarm presentation, the participants would need to wait and continue to look at the parameter until the alarm triggered. The reduced salience of the alarm due to gaze-based acknowledgment, therefore, might have left participants with higher uncertainty, redirecting them to monitor the alarms more frequently than otherwise. The fact that trials in the control condition yielded the best parameter control performance lends support to this explanation. Woods (1995) observed a similar phenomenon, in which removing nuisance alarms proved counterproductive to operator performance due to loss of validating information. Additional analysis of the raw eye gaze data is necessary to substantiate this explanation further, particularly in comparing the amount of gaze between the alarm monitoring and parameter control tasks. If future analysis and research support this justification, then gaze-based acknowledgement may consider altering the presentation to have audio cues with shorter duration and unique tone.

The gaze-based acknowledgement methods were also designed to yield different levels of benefit depending on scenario types or the parameter characteristics for the impending alarm. That is, one gaze-based acknowledgement method is expected to support the dual task in one scenario type better than other ones. This study provides very complex results speaking to this interaction effect. While there was an interaction effect of marginal significance on the parameter control performance (Table 6), the hypothesized performance benefits when acknowledgement methods matched the scenario types (Table 7) were not observed. Thus, the marginal significant effect cannot be interpreted as supporting evidence.

Despite lacking the hypothesized interaction effect on the parameter control task performance, a Chi squared test did reveal a significant difference in the number of gaze-based acknowledgements across acknowledgement methods and scenario types. The contingency table (Table 11) indicated that number of acknowledgements tended to be higher when the acknowledgement methods were matched with scenario types as hypothesized. In other words, the usage level of the acknowledgement methods did cater to scenario types, providing modest support for the hypothesized interaction. More importantly, this finding can inform future design of gaze-based acknowledgement methods in that the types of parameter behavior prior to reaching the alarm threshold should be a critical factor in design.

Results on alarm prediction decisions by the participants provide further insights into how eye gaze might reflect cognitive processing of parameter and alarm behaviors. The participants predicted an alarm more accurately with an occurrence of gaze acknowledgement (see Table 8and X^2 statistics in Section 3.2). This is a positive indication that gaze-based acknowledgement, or eye fixations, is a good estimation of conscious anticipation of alarms. However, the number of gaze-based acknowledgements (125) greatly exceeded than the number of predictions of an alarm (65; see Table 8), suggesting that gaze-based acknowledgements likely reflected more than just the participant prediction decision or conscious anticipation of alarms. This interpretation of gaze-based acknowledgements is further supported by the finding that near-threshold scenarios resulted in the most accurate alarm prediction (see Table 9 and Section 3.2) but fluctuation scenarios induced the most gaze acknowledgements (see Table 9 and Section 3.2). Note that fluctuation scenarios also resulted in worst parameter control performance (most time out of range), suggesting greater allocation of visual attention towards the alarm monitoring task (see Table 6 and Figure 11). These findings collectively suggest that fixations on parameters prior to alarm threshold can indicate *either* a conscious prediction of an alarm *or* perceived uncertainty surrounding a parameter. While the former interpretation has been discussed in the introduction as the motivation of this research study, the latter interpretation has been adopted by some prior research that argues high fixations as an indicator of uncertainty (Holmqvist et al., 2011).

The fact that fixations on parameters prior to alarm threshold could be interpreted as *either* conscious prediction of an alarm *or* perceived uncertainty surrounding a parameter could have major implications for designing gaze-based alarm acknowledgement. First, gaze-based acknowledgement might be more effective if the fixation requirement differs by types of parameter behaviors that inherently induce different perceived uncertainty. For example, fluctuating parameters might need a higher number of fixations to indicate conscious anticipation of an alarm. A single threshold of fixations would not adequately reflect anticipation of alarms driven by past behaviors of the parameter. Second, a more nuanced approach might be necessary by incorporating additional eye-gaze metrics, such as stationary gaze entropy, to estimate perceived uncertainty or confusion of the user. When uncertainty of the parameter behavior rather than anticipation of an alarm is determined, alarm acknowledgement would not occur.

This study has contributed a unique design approach towards managing alarms and empirical evidence of the effectiveness of gaze-based acknowledgement. The study showed that gaze-based acknowledgement in the parameter monitoring task did not seem to minimize disruption for parameter control task. Our results suggested that task characterization and scenario type (i.e., parameter behaviors) might have greater attentional importance than salience.

Therefore, to minimize alarm disruption to other tasks, the parallel tasks need to be more comparable and perhaps even influence one another.

The design of gaze-based acknowledgement was based on parameter behaviors and eye fixation requirements; thus, the study results provided new insights on how to jointly monitor parameter and gaze behaviors for inferring alarm prediction/awareness by the user. The results indicated that the scenario type affects the amount of attention given to a parameter. Fluctuating parameters drew the most attention when compared to near-threshold, and trending behavior. This is a novel finding regarding how visual attention differs between different parameter behaviors as prior studies have not directly compared the impact of near-threshold, trending, and fluctuating parameter characteristics on visual sampling. However, as indicated by the discrepancy in prediction versus gaze acknowledgements, many fixations do not always indicate anticipation of an alarm. The results indicate that the fixation requirement needs greater precision and may differ between scenario types.

4.1 Limitations

Only twenty-four participants were included in the study A previous power analysis conducted with G*Power 3.1 suggested that for a MANOVA with repeated measures, a medium effect size (f = .20), and an alpha of .05, 52 participants were needed to reach a power of .80 (Faul et al., 2007). A sample size of 52 could reveal effect sizes comparable to the findings of Meyers and Bitan (2007) whose study investigated the number of inspections a person took to prevent a parameter from exceeding a value and how those change with rate of change and slope variability. Given this power analysis, the study results must be interpreted with caution and not definitively dismiss gaze-based acknowledgement despite modest supporting results found in this study.

This study recruited students, which puts a limitation on the generalizability of these findings to experienced operators. Industrial operators may have a unique mental model of alarms, given their knowledge of alarm interconnectivity, that may affect their visual sampling and anticipation. Additionally, participants were only tasked with predicting alarms. However, their behavior and interaction with gaze-based acknowledgement may alter when charged with operating and resolving alarms as opposed to only monitoring for them.

4.2 Future Work

Future work should conduct exploratory data analysis on the number of eye fixations, fixation timestamps, and fixation duration of the target AOIs in relation to prediction performance, scenario type, and alarm timestamps. The current results suggest that fixations may be insufficient at inferring participant prediction decisions. Identifying additional and combination of eye-gaze metrics that is indicative of the prediction decision of the participants could be very informative at designing when and how to acknowledge alarms with eye-tracking. The exploratory analysis could reveal different visual scanning patterns by scenario type and/or prediction decisions.

Further task characterization is necessary to determine how participants allocate attention and to ensure that subsequent experiments can adequately test the gaze-based alarm acknowledgement methods in moderating focused attention. The analysis could examine the amount of time spent looking at the alarm monitoring task versus the parameter control task. A comparison of different combination of parallel tasks would provide greater insights as to when gaze-based acknowledgement would be particularly useful. The results indicated that gaze-based alarm acknowledgement might not induce a significant impact on attention allocation for the concurrent tasks in this study. However, gaze-based acknowledgement functions could be more effective with other combinations of task designs.

5 Conclusion

This study investigated whether measuring gaze behaviors is a viable way of passively reducing nuisance alarms. Studies show that operators who monitor vigilantly find that alarm informativeness decreases because they foresee the alarms. Thus, manually acknowledging alarms that operators are aware of may disrupt attention and time spent on more critical tasks. The gaze-based alarm acknowledgement methods uniquely rely on eye-tracking to provide non-intrusive and passive measurement of operator awareness of alarms, and in turn, adapts the presentation or salience to minimize disruption of focused attention.

Three types of gaze-based acknowledgement methods were designed based on parameter characteristics that correspond to operators' monitoring behavior and were evaluated with an experiment recruiting participants to monitor and predict impending alarms while maintaining a single parameter within range. The performance of the two tasks was analyzed to compare the three gaze-based acknowledgment methods against the normal alarm presentation method (i.e., control/no gaze-based acknowledgement) across three different scenario types, which were defined by different parameter behaviors prior to passing the alarm thresholds.

The results only provide modest support for gaze-based acknowledgement's ability to minimize disruption of other tasks. Results demonstrated that there was no significant effect of alarm acknowledgement on minimizing time out of range for the parameter control task. However, the alarm monitoring task did not receive sufficient attention which may have led to a lack of observable effects, which suggests revision to the task design. While there was a marginal interaction effect on the parameter control performance, the hypothesized performance benefits expected were not observed. Nevertheless, the usage level of the acknowledgement methods did cater to scenario types, and participants also gave higher usability ratings for the gaze-based acknowledgement methods compared to no gaze-based acknowledgement. Regarding the accuracy of measuring operator's anticipation of alarms, gaze-based acknowledgement seemed to depict instances of uncertainty and prediction, which also suggests reiteration of eye metric requirements. Further research is necessary to understand the utility of eye-tracking in alarm management.

In closing, this research produced novel empirical evidence on how eye-tracking could support alarm management based on operator anticipation of alarms. Our results offer greater understanding of the nuances in eye-fixations for inferring operator conscious anticipation or prediction of alarms. While the designs of gaze-based acknowledgement in this study require more iterations, more investigation of alarm management approaches of this nature should continue to improve the informativeness of alarms received by operators.

References

Bauer, M. & Thornhill, N.F. (2008) A practical method for identifying the propagation path of plant-wide disturbances. *Journal of Process Control, 18* (7-8), 707-719.

Bitan, Y., & Meyer, J. (2007). Self-initiated and respondent actions in a simulated control task. *Ergonomics, 50*(5), 763-788.

Blanca Mena, M. J., Alarcón Postigo, R., Arnau Gras, J., Bono Cabré, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema, 29*(4), 552-557.

Bliss, J. P. (2003). Investigation of alarm-related accidents and incidents in aviation. *The International Journal of Aviation Psychology, 13*(3), 249-268.

Bockelmann M., Nickel P., Nachreiner F. (2019) *Ergonomics Analysis of Alarm Systems and Alarm Management in Process Industries.* In: Bagnara S., Tartaglia R., Albolino S., Alexander T., Fujita Y. (eds) Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018). IEA 2018. Advances in Intelligent Systems and Computing, vol 822. Springer, Cham. https://doi.org/10.1007/978-3-319-96077-7_79

Botzer, A., Meyer, J., Bak, P., & Parmet, Y. (2010). User settings of cue thresholds for binary categorization decisions. *Journal of Experimental Psychology: Applied*, *16*(1), 1-15. https://doi.org/10.1037/a0018758

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*(2), 144-152. https://doi.org/10.1111/j.2044-8317.1978.tb00581.x

Bransby, M. (2001). Design of alarm systems. In J. Noyes & M. Bransby (Eds.), *People in control* (pp. 207-221). Institution of Engineering and Technology.

Breznitz, S. (1983). *Cry-wolf: The psychology of false alarms.* Hillsdale, NJ: Lawrence Erlbaum.

Bullemer, P. T., Tolsma, M., Reising, D. V. C., & Laberge, J. C. (2011). Towards improving operator alarm flood responses: Alternative alarm presentation techniques. *Abnormal Situation Management Consortium*.

Burns, C. M., & Hajdukiewicz, J. (2004). *Ecological interface design*. CRC Press.

Carrasco, M., McElree, B., Denisova, K., & Giordano, A. M. (2003). Speed of visual processing increases with eccentricity. *Nature neuroscience*, *6*(7), 699-700.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision research*, *51*(13), 1484-1525.

Chen, K., & Wang, J. (2017). Design of multivariate alarm systems based on online calculation of variational directions. *Chemical Engineering Research and Design*, *122*, 11-21.

Crossman, E. R. F. W., Cooke, I. E., & Beishan, R. J. (1974) Visual attention and the sampling of displayed information in process control. In *The human operator in process control.* New York: Taylor and Francis.

de Jong, J., & Koster, E. P. (1971, January). The human operator in the computer-controlled refinery. In *8th World Petroleum Congress*. Moscow, USSR.

Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007). On the independence of compliance and reliance: Are automation false alarms worse than misses? *Human factors, 49*(4), 564-572.

Dorgo, G., & Abonyi, J. (2018). Sequence mining based alarm suppression. *IEEE Access*, *6*, 15365-15379.

Eisma, Y. B., Cabrall, C. D., & de Winter, J. C. (2018). Visual sampling processes revisited: replicating and extending Senders (1983) using modern eye-tracking equipment. *IEEE Transactions on Human-Machine Systems, 48*(5), 526-540.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175-191.

Getty, D. J., Swets, J. A., Pickett, R. M., & Gonthier, D. (1995). System operator response to warnings of danger: A laboratory investigation of the effects of the predictive value of a warning on human response time. *Journal of experimental psychology: Applied, 1*(1), 19-33. https://doi.org/10.1037/1076-898X.1.1.19

Goel, P., Datta, A., & Mannan, M. S. (2017). Industrial alarm systems: Challenges and opportunities. *Journal of Loss Prevention in the Process Industries, 50*, 23-36.

Guerlain, S., & Bullemer, P. (1996, October). User-initiated notification: A concept for aiding the monitoring activities of process control operators. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 40*(4), 283-287. Sage CA: Los Angeles, CA: SAGE Publications.

Hall, P., & Morton, S. C. (1993). On the estimation of entropy. *Annals of the Institute of Statistical Mathematics*, *45*(1), 69-88.

Hildebrandt, M., Langstrand, J. P., & Nguyen, H. (2019). *Report for 2.2. 1 Task 4: Software-Based Tools to Support Human-System Evaluation Studies* (No. INL/EXT-19-55789-Rev000). Idaho National Lab. (INL), Idaho Falls, ID (United States).

Hollifield, B. R., & Habibi, E. (2011). Alarm management: A comprehensive guide: Practical and proven methods to optimize the performance of alarm management systems. International Society of Automation.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Hu, W., Chen, T., & Shah, S. L. (2017). Discovering association rules of mode-dependent alarms from alarm and event logs. *IEEE Transactions on Control Systems Technology*, *26*(3), 971-983.

Hu, W., Wang, J., Chen, T., & Shah, S. L. (2017). Cause-effect analysis of industrial alarm variables using transfer entropies. *Control Engineering Practice*, *64*, 205-214.

ISA (2009). Isa-18.2: Management of alarm systems for the process industries. *International Society of Automation. Durham, NC, USA.*

Izadi, I., Shah, S. L., Shook, D. S., & Chen, T. (2009). An introduction to alarm analysis and design. *IFAC Proceedings Volumes, 42*(8), 645-650.

Izadi, I., Shah, S. L., Shook, D. S., Kondaveeti, S. R., & Chen, T. (2009). A framework for optimal design of alarm systems. *IFAC Proceedings Volumes*, *42*(8), 651-656.

Kim, I. S., Modarres, M., & Hunt, R. N. M. (1990). A model-based approach to on-line process disturbance management: the models. *Reliability Engineering & System Safety*, *28*(3), 265-305.

Kragt, H., & Bonten, J. (1983). Evaluation of a conventional process-alarm system in a fertilizer plant. *IEEE Transactions on Systems, Man, and Cybernetics, 13*(4), 586-600. doi.org/10.1109/TSMC.1983.6313147.

Kvålseth, T. O. (1978). Human information processing in visual sampling. *Ergonomics*, *21*(6), 439-454.

Kvålseth, T. O. (1979). A Decision-Theoretic Model of the Sampling Behavior of the Human Process Monitor: Experimental Evaluation. *Human factors*, *21*(6), 671-686.

Lau, N., Jamieson, G. A., & Skraaning Jr, G. (2016). Situation awareness acquired from monitoring process plants–the Process Overview concept and measure. *Ergonomics*, *59*(7), 976-988.

LeBlanc, K., Spielman, Z., Bower, G., Oxstrand, J., & Bly, A. (2015). *Enabling Situation Awareness Under High Levels of Automation: Results from an Experimental Study* (No. INL/EXT-15-35791-Rev002). Idaho National Lab. (INL), Idaho Falls, ID (United States).

Lee, M. N., & Lees, J. D. (2007). The influence of distraction and driving context on driver response to imperfect collision warning systems. *Ergonomics, 50*(8), 1264-1286.

Lehman, A., O'Rourke, N., Hatcher, L., & Stepanski, E. (2013). *JMP for basic univariate and multivariate statistics: methods for researchers and social scientists*. SAS Institute.

Lehto, M. R., Papastavrou, J. D., & Giffen, W. J. (1998). An empirical study of adaptive warnings: Human-versus computer-adjusted warning thresholds. *International Journal of Cognitive Ergonomics*.

Lethaus F., Harris R.M., Baumann M.R.K., Köster F., Lemmer K. (2013) Windows of Driver Gaze Data: How Early and How Much for Robust Predictions of Driver Intent?. In: Tomassini M., Antonioni A., Daolio F., Buesser P. (eds) *Adaptive and Natural Computing Algorithms. ICANNGA 2013. Lecture Notes in Computer Science,* 7824. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-37213-1_46

Li, J., Wang, Y., Li, H., & Yang, B. (2019, June). An Adaptive Optimization Approach Based on the Human factors and Its Application to Process Alarm Thresholds. In *2019 12th Asian Control Conference (ASCC)* (pp. 823-828). IEEE.

Maltz, M., & Shinar, D. (2003). New alternative methods of analyzing human behavior in cued target acquisition. *Human factors, 45*(2), 281-295.

Mumaw, R. J., Roth, E. M., Vicente, K. J., & Burns, C. M. (2000). There is more to monitoring a nuclear power plant than meets the eye. *Human factors, 42*(1), 36-55.

Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human factors, 46*(2), 196-204.

Meyer, J., & Bitan, Y. (2002). Why better operators receive worse warnings. *Human Factors, 44*(3), 343-353.

Meyer, J., & Sheridan, T. B. (2017). The intricacies of user adjustments of alerting thresholds. *Human factors*, *59*(6), 901-910.

Miller, C. A., & Parasuraman, R. (2007). Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. *Human factors*, *49*(1), 57-75.

Moray, N. (2006) Process control. In W., Karwowski (Ed.). *International Encyclopedia of Ergonomics and Human Factors*. 3, 1148-1150. USA: Taylor and Francis Inc.

Noda, M., Higuchi, F., Takai, T., & Nishitani, H. (2011). Event correlation analysis for alarm system rationalization. *Asia‐Pacific Journal of Chemical Engineering*, *6*(3), 497-502.

NRC (2019). Human-System Interface Design Review Guidelines (NUREG-0700, Rev 3). Washington, D.C.: U.S. Nuclear Regulatory Commission. Retrieved from: https://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr0700/

Patternote, P. H. (1978). The control performance of operators controlling a continuous distillation process. *Ergonomics*, *21*(9), 671-679.

Rayo, M. F., & Moffatt-Bruce, S. D. (2015). Alarm system management: evidence-based guidance encouraging direct measurement of informativeness to improve alarm response. *BMJ quality & safety, 24*(4), 282-286.

Reising, D. V. C., Downs, J. L., & Bayn, D. (2004, September). Human performance models for response to alarm notifications in the process industries: An industrial case study. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 48(10), 1189-1193. Los Angeles, CA: SAGE Publications.

Robinson, D. N. (1967). Visual discrimination of temporal order. *Science*, *156*(3779), 1263-1264. http://doi.org/10.1126/science.156.3779.1263

Rodrigo, V., Chioua, M., Hagglund, T., & Hollender, M. (2016). Causal analysis for alarm flood reduction. *IFAC-PapersOnLine*, *49*(7), 723-728.

Rothenberg, D. H. (2009). *Alarm management for process control: a best-practice guide for design, implementation, and use of industrial alarm systems*. Momentum Press.

Salvucci, D. D., & Goldberg, J. H. (2000, November). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, 71-78. https://doi.org/10.1145/355017.355028

Seagull, F. J., Xiao, Y., Mackenzie, C. F., & Wickens, C. D. (2000, July). Auditory alarms: from alerting to informing. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *44*(1), 223-226. Los Angeles, CA: SAGE Publications.

Senders, J. W. (1964). The human operator as a monitor and controller of multidegree of freedom systems. *IEEE Transactions on Human Factors in Electronics*, 5(1), 2-5.

Shannon, C.E. (1948) A mathematical theory of communication. *The Bell system technical journal*, *27*(3), 379-423.

Sheridan, T. B. (1970). On how often the supervisor should sample. *IEEE Transactions on Systems Science and Cybernetics*, *6*(2), 140-145.

Simeu-Abazi, Z., Lefebvre, A., and Derain, J.P. (2011) A methodology of alarm filtering using dynamic fault tree. *Rel. Eng. Syst. Safety, 96*(2), 257–266

Simonson, E., & Brozek, J. (1952). Flicker fusion frequency: background and applications. *Physiological reviews*, *32*(3), 349-378.

Smith, P. J., McCoy, C. E., & Layton, C. (1997). Brittleness in the design of cooperative problem-solving systems: The effects on user performance. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *27*(3), 360-371.

Sorkin, R. D. (1988). Why are people turning off our alarms? The Journal of the Acoustical Society of America, 84(3), 1107-1108.

Stanton, N.A., & Babar, C. (1995). Alarm-initiated activity. Ergonomics, 38(11), 2414-2431. http://doi.org/10.1080/00140139508925276

Stanton, N. A., & Edworthy, J. (1999). *Human factors in auditory warnings*. Ashgate.

Tatler B., Kirtley C., Macdonald R., Mitchell K., Savage S. (2014) The Active Eye: Perspectives on Eye Movement Research. In: Horsley M., Eliot M., Knight B., Reilly R. (eds) *Current Trends in Eye Tracking Research.* Springer, Cham. https://doi.org/10.1007/978-3-319-02868-2_1

Vicente, K. J. (1999). *Cognitive work analysis: Toward safe, productive, and healthy computer-based work*. CRC Press.

Vicente, K. J., Moray, N., Lee, J. D., Hurecon, J. R., Jones, B. G., Brock, R., & Djemil, T. (1996). Evaluation of a Rankine cycle display for nuclear power plant monitoring and diagnosis. *Human Factors*, *38*(3), 506-521.

Vicente, K. J., Roth, E. M., & Mumaw, R. J. (2001). How do operators monitor a complex, dynamic work domain? The impact of control room technology. *International Journal of Human-Computer Studies*, *54*(6), 831-856.

Wallis, K. (2006). A note on the calculation of entropy from histograms.

Wang, J., & Chen, T. (2014). An online method to remove chattering and repeating alarms based on alarm durations and intervals. *Computers & Chemical Engineering*, *67*, 43-52.

Wang, J., & Chen, T. (2016). Main causes of long-standing alarms and their removal by dynamic state-based alarm systems. *Journal of Loss Prevention in the Process Industries*, *43*, 106-119.

Watson, M., Sanderson, P., & John Russell, W. (2004). Tailoring reveals information requirements: the case of anaesthesia alarms. *Interacting with computers*, *16*(2), 271-293.

Wickens, C., & Colcombe, A. (2007). Dual-task performance consequences of imperfect alerting associated with a cockpit display of traffic information. *Human factors*, *49*(5), 839-850.

Wickens, C. D., Dixon, S. R., & Seppelt, B. (2005, September). Auditory preemption versus multiple resources: Who wins in interruption management. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *49*(3), 463-466. Los Angeles, CA: SAGE Publications.

Wickens, C. D., Rice, S., Keller, D., Hutchins, S., Hughes, J., & Clayton, K. (2009). False Alerts in Air Traffic Control Conflict Alerting System: Is There a “Cry Wolf” Effect? *Human Factors*, *51*(4), 446–462. https://doi.org/10.1177/0018720809344720

Woods, D. D. (1995). The alarm problem and directed attention in dynamic fault management. *Ergonomics*, *38*(11), 2371-2393.

Wu, X., & Li, Z. (2018). A review of alarm system design for advanced control rooms of nuclear power plants. *International Journal of Human–Computer Interaction*, *34*(6), 477-490.

Xiao, Y., & Seagull, F. J. (1999, September). An analysis of problems with auditory alarms: Defining the roles of alarms in process monitoring tasks. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *43*(3), 256-260. Los Angeles, CA: SAGE Publications.

Yu, Y., Zhu, D., Wang, J., & Zhao, Y. (2017). Abnormal data detection for multivariate alarm systems based on correlation directions. *Journal of Loss Prevention in the Process Industries*, *45*, 43-55.

Appendix A: Pseudo code for gaze-suppression methods

Global Variables:
AOI #= 1,2,3,4,5,6,7. Indicates what the component is and location of the AOI
Alarm= .wav file and yellow highlight of the triggered component
Parameter_value= the calculated amount at any given time for a component
High_alarm= parameter_value @ 135 F, 75%, or 40% depending on AOI #
Low _alarm= parameter_value @ 115 F, 25%, or 20% depending on AOI #
t=time

Proximity-based method

Local variables:
High_alarm_acknowledge_threshold = High_alarm - 10% of the range between the High_alarm and Low_alarm
Low_alarm_acknowledge_threshold = Low_alarm + 10% of the range between the High_alarm and Low_alarm

Rule: Suppress alarm when the user has two or more fixations on the designated AOI and the parameter value is at the suppression threshold value 4 seconds before the alarm occurs.

```
     IF fixation occurs on AOI [#] between t-1 and t-4 seconds
          AND parameter_value ≥ High_alarm_suppression_threshold between  t-1 and t-4
          seconds
          AND parameter_value ≥ High_alarm
               THEN Suppress alarm
   IF parameter_value < High_alarm OR t > 12 seconds
     Reset alarm

     IF fixation occurs on AOI [#] between  t-1 and t-4 seconds
          AND parameter_value ≤ low_alarm_suppression_threshold between  t-1 and t-4
     seconds
          AND parameter_value  ≤ low_alarm
               THEN Suppress alarm
   IF parameter_value > low_alarm OR t > 12 seconds
          Reset alarm
```

Prediction-based method

Local Variables:
Y=mx+b
Y= parameter_value of high_alarm (if positive rate of change) or low_alarm (if negative rate of change)
x= time_till_alarm
b=parameter_value when fixation occurs
window_start= t-4
window_end= t-1

window= 3 seconds
Slope (m) = [(parameter_value @ t-4) - (parameter_value @ t-3) /1 second]
Predicted_value 3 sec ahead = parameter_value @ window_start + (slope*(window+3)

Rule: Suppress the alarm when user has a fixation on the designated AOI and the parameter trendline suggests an alarm will be reached within the prediction time window

```
IF fixation occurs on AOI [#] between t-1 and t-4 seconds
    AND IF predicted_value $\geq$ high_alarm
            THEN Suppress alarm
        IF parameter_value $<$ high_alarm OR $t > 12$ seconds
            Reset alarm

IF fixation occurs on AOI [#] between t-1 and t-4 seconds
    AND IF predicted_value $\leq$ low_alarm
            THEN Suppress alarm
        IF parameter_value $>$ low_alarm OR $t > 12$ seconds
            Reset alarm
```

Entropy-based method

Local variables:
Histogram estimator of entropy $H(x) = -\sum f(p_k)\log (f(p_k)/f(w_k))$
Bin probability $(p_k)= d_k/N$
d= # of data points in a bin
N= total # of data points
Bin width(w)= 3 parameter_value
Bin number(k)= 1,2,3,4, [5,6,7,8,9,10] depending on the range between high and low alarm of component
H= amount of entropy
Build the histogram of the last 15 seconds

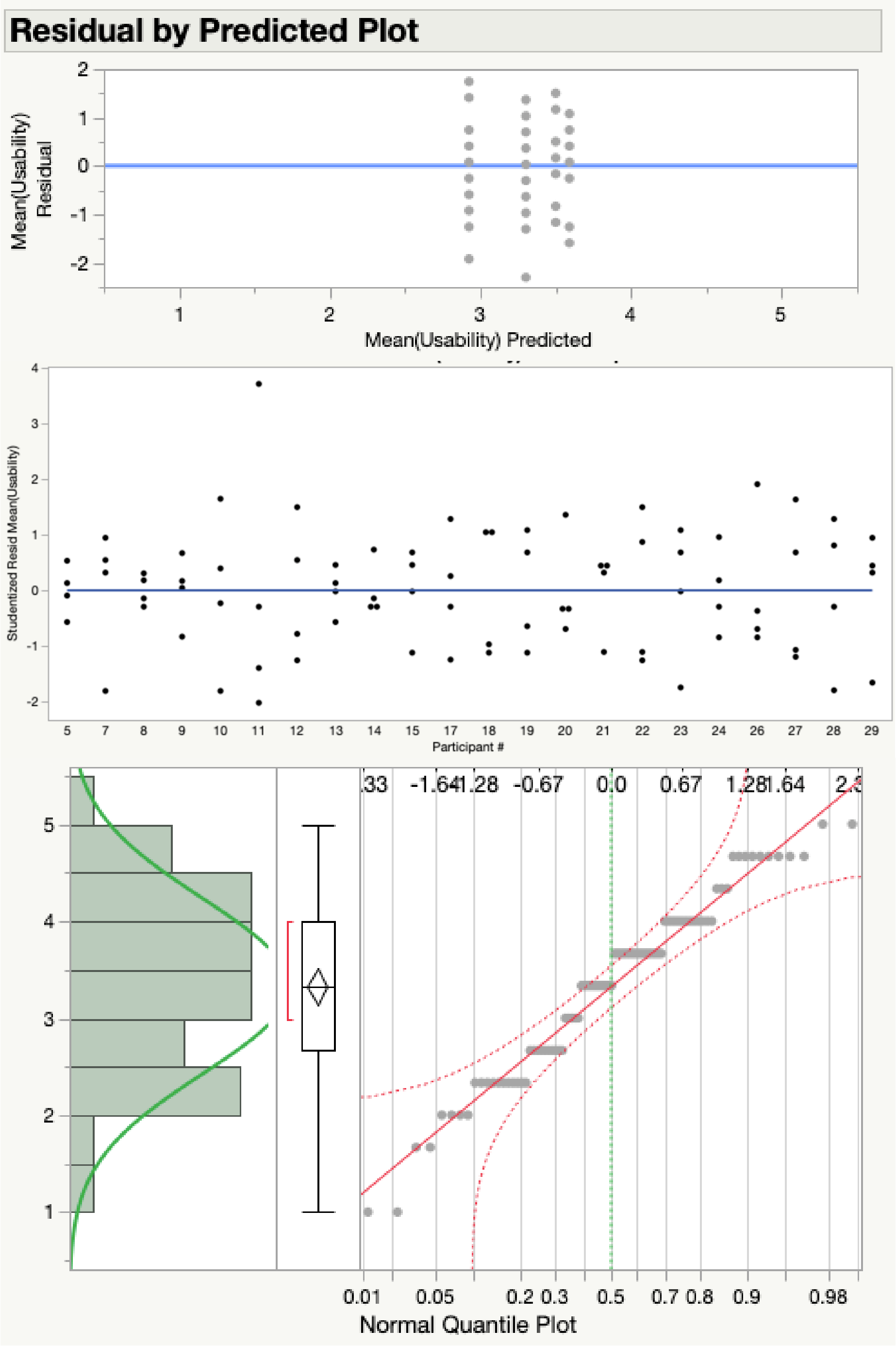
Rule: Suppress the alarm when the user has a fixation on the designated AOI and the parameter trendline has entropy greater than or equal to threshold.

```
IF fixation occurs on AOI [#] between t-1 and t-4 seconds
        AND IF $H \geq 1.20$
            THEN Suppress alarm
        IF $t > 12$ seconds
            Reset alarm
```

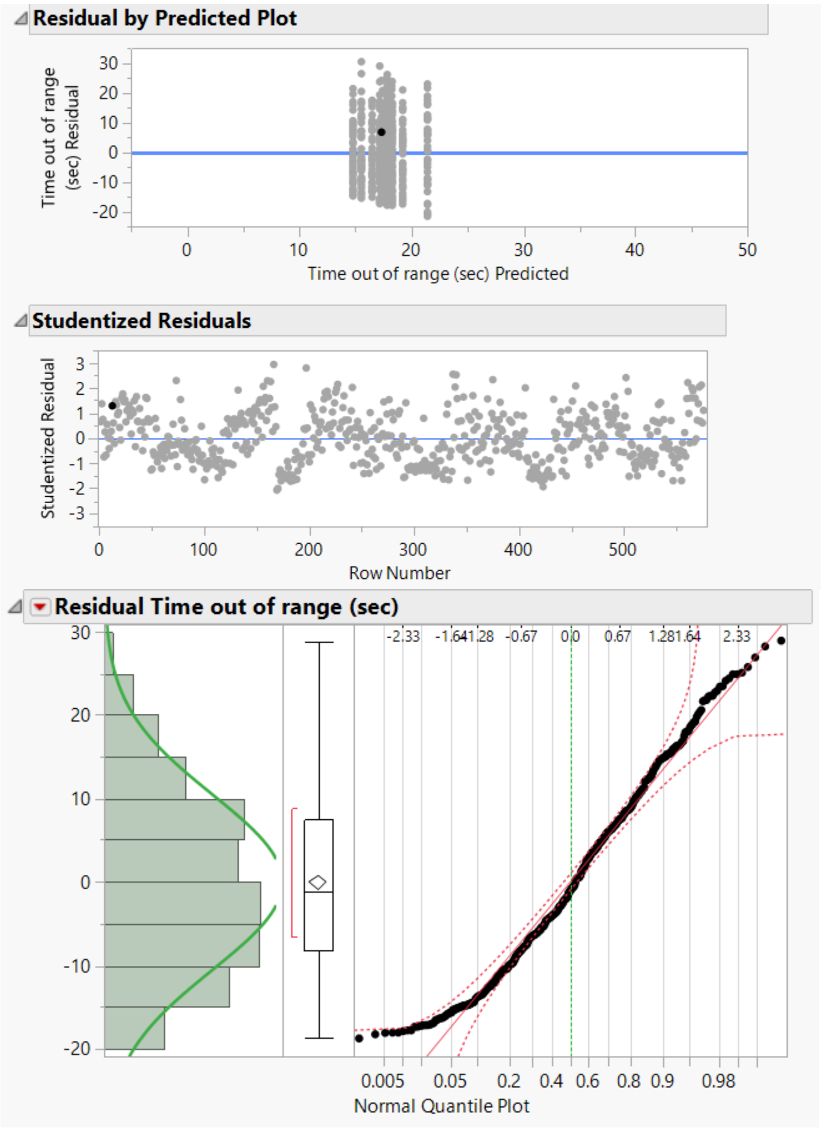
Appendix B: Normality Assumptions

Usability and Acknowledgement Method

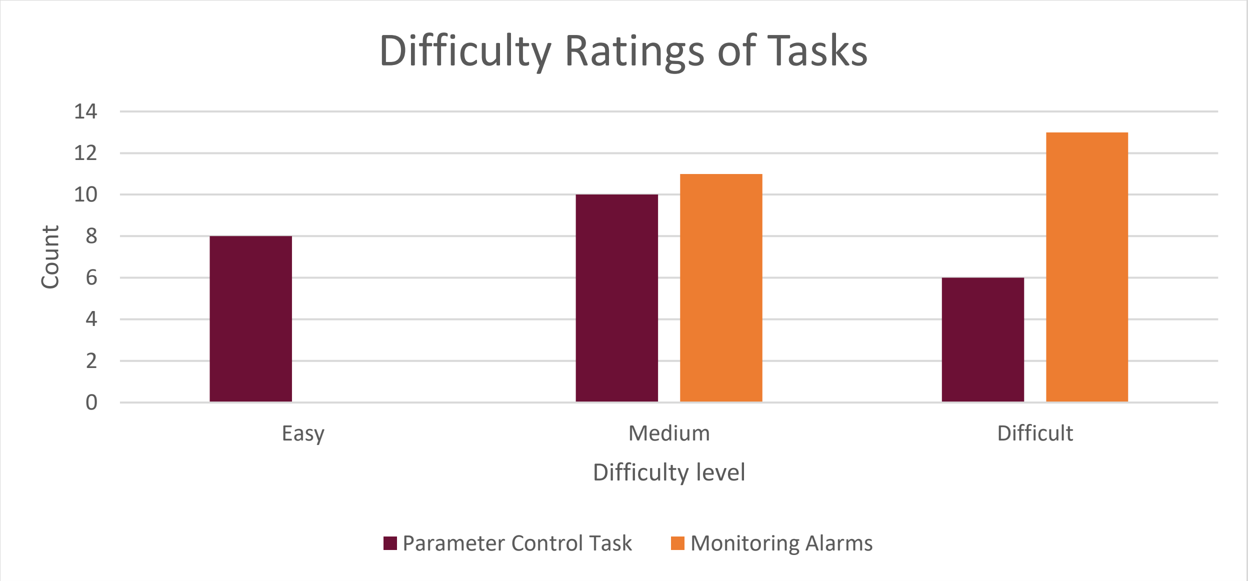
Shapiro-Wilk test W=.964 p = .0167



Control Task Performance with Scenario Type and Acknowledgement Method
Shapiro-Wilk test W=.9774, p < .0001



Appendix C: Debrief Results



Q: Were you able to perform the tasks to the best of your ability? Why or why not?

The alarms helped me to do the task better while I had more focus with the gaze-based acknowledgement.
Yes, I performed the task to best of my ability. Most of the alarms I was able to predict.
Yes, there was nothing inhibiting performance other than the nature of the tasks themselves.
Somewhat. It was difficult monitoring both tasks at the same time due to the rapid actions involved and unpredictability. Even though the primary task was more important, the existence of the second task took away from it. Due to that, shortcomings happened in both.
No, when the alarm continued to go off it was hard to reset my focus while it continued to chime. Also, I think the bottom left box didn't always highlight with the blue dotted box, and that was distracting because I kept trying to click it.
Yes, I could. It was a simple task. It was a bit difficult to look at two screens, but it was doable.
Yes, until the very end I got tired.
I could not perform them to the best of my ability because I had to focus on both at once. I could have focused on each task better if they were the only tasks at hand.
Yes, and I'm trying my best to balance between tasks as well. Meaning, maybe the primary task performance might be affected by my divided attention in secondary tasks
Yes. At the time of the study, I was feeling 100% healthy.
Yes, I had clear view and instructions.
I could not perform to the best of my ability as there were multiple tasks to perform at the same time. It got a bit overwhelming.
NO, I was not. I think I was facing trouble in monitoring both the screens at the same time
I was not able to perform the tasks to the best of my ability. There were too many things to look out for at once from the control task to six different graphs. As a result, I ended up missing a lot more than I got.

No. Because I think the visual representation of the 7 alarms in this screen size was not good enough to detect change easily. May be a bigger visual display will help to detect the alarms easily.
No, it was difficult having to do both tasks at once. I sacrificed the control task many times.
Yes, I think I tried my best to complete the monitoring task. I scanned 7 parameters and observed their trend to see if they would be out of the range. While doing so, I sometimes forget to monitor the secondary task. But I react to the secondary task as long as I am aware that it's out of range.
The sluggishness of the mouse may have adversely affected reaction time. Larger graphs for tracking metrics with narrow alarm bands would likely have increased my effectiveness.
I believe yes, I understood what I had to do and tried my best to perform the tasks asked.
Yes. I was paying attention.
Yes, I was able to watch the metric throughout and identify most times when an alarm would sound.
Yes, it was the best I could do coming into this task fresh.
Not really, because I was not very familiar with the task in the beginning, I got better later on.
Yes, I was able to perform the tasks to the best of my ability because there were no hindrances to performing the tasks. However, it was difficult to manage both tasks at the same time.

Q: What did you think about gaze-based acknowledgement?

Gaze-based provide needed focused.
It is a really good concept and saves reflex time for clicking a button in time critical scenario.
It was very effective in helping me to catch alarms that I might have not caught otherwise.
I am not sure if it's as effective. A worker performing two tasks might not notice due to no sounds and just the border. Maybe a red color, more signifying of danger can be used, along with a sound cue.
It was helpful to some degree, but some time it acknowledged that I was looking at a box when I was looking at the one above or below it.
It’s an interesting study. It will help in minimizing the reaction time if it works.
I thought that it detected my eye movement at first, but towards the last few experiments it did not detect my eye movement as much as it did early on.
It worked sometimes but other times I felt like I noticed an alarm, but the device did not pick up my eye recognition
Seems like an indicator of "you did a good job".
I think it is very helpful since the parameters oscillate and change their tendencies very quickly
Seemed to identify the alarms that I anticipated.
It was great in alerting me as and when the parameter thresholds are being exceeded.
I think that it’s an innovative idea but there is a scope for improvement. It does not always detect accurately.
I think it works about a bit more than half the time but there were times I felt it did not acknowledge an alarm I was looking at.
I think it will be really helpful for operators who works in such alarm monitoring condition.
It worked only when I felt like I had to had been looking/staring at that one parameter for a long time or only looking at it (intensely) for the duration of the trial. Other times, I failed

mostly due to my own lack of awareness of a quick dip in one of the parameters as I was either anticipating another to dip or focusing on the control task.
I think sometimes it works well as I was looking at a parameter with pending alarm. But sometimes it seems not working even though I noticed the parameter that was going to issue an alarm. Maybe it requires some time of fixation to trigger the acknowledgement. But the acknowledgement did makes it less startled when I anticipated the alarm.
It was pretty cool when it worked. I still don't understand how it determines if a gaze is "anticipating" or just scanning, but it's an interesting concept to pursue.
Overall, my experience with it was fine. Sometimes when I predicted an alarm it still went off but that was the only thing.
Interesting. I don't think I was exposed to this before today.
It was tedious but it kept me alert most times
I think it may have been easier to see when it was successful in predicting my notice of an alarm had the "success" color been different. I think I was mistaking the yellow of any sort as a failure in me predicting an alarm or properly clicking, which increased self-disappointment at being able to complete the task or not. Perhaps green for success, yellow for missed alarms?
It was effective in anticipating my acknowledged alarm.
I like the gaze-based alarm acknowledgement because it is less intrusive than the chime-based alarm.

Q: Any other thoughts you'd like to share.

The monitoring alarm UI some of the lines were pretty close to the borderline. So prediction became difficult for those component.
Neither task is extremely difficult alone, but both together is tough.
Alarm acknowledgement seems to be better but might not some tweaking in terms of the cues involved.
Anytime I'd let the control task slip it would always go off screen in the down direction, I think because the average downward slope was greater than the average upward slope. Also, I mentioned this to the facilitator but, both the mouse and the enter key are on the right half of the computer set up, so I felt like I would drift to the right while doing the task. My head and body did not stay where they were when the tracker was calibrating.
It was hard to tell sometimes if an alarm was about to go off, so I found myself guessing at times.
The line graphs and the value indicated in some parameters does not match which may have hindered my ability to detect the alarms.
It is incredibly difficult to perform both tasks even at less than perfect efficiency.
At times, I felt that I was looking at the parameter that did go off later as an alarm, but it was not picked up by the gaze-based system. Though this only happened twice. The other times where I was intentionally trying to only look at a certain parameter and it was the one that dipped/increased drastically it did catch it.
I believe that adding the audio when there is no gaze detected is very effective to attract the operator's attention to the alarmed parameter. Sometimes I think it was going to alarm and click it, but it did not issue an alarm. So, I moved to monitor other parameters, and then the previous parameter alarmed with audio warnings. It attracted my attention back to that alarm.

Perhaps instruct participants to focus on blue numbers beneath the control (1st) task instead of the line. Once I started doing that, I stopped trying to predict the line and was only reacting to the numbers on the last two rounds, which I found easier to deal with and pay attention to the 2nd task (alarms).