INL/EXT-15-37311
Revision 0

Light Water Reactor Sustainability Program

Measuring Human Performance in Simulated Nuclear Power Plant Control Rooms Using Eye Tracking

November 2015

U.S. Department of Energy

Office of Nuclear Energy

INL/EXT-15-37311
Revision 0

Measuring Human Performance in Simulated Nuclear Power Plant Control Rooms Using Eye Tracking

Casey Kovesdi
Brandon Rice
Gordon Bower
Zach Spielman
Rachael Hill
Katya LeBlanc

November 2015

Prepared for the
U.S. Department of Energy
Office of Nuclear Energy

ABSTRACT

Control room modernization will be an important part of life extension for the existing light water reactor fleet. As part of modernization efforts, personnel will need to gain a full understanding of how control room technologies affect performance of human operators. Recent advances in technology enable the use of eye tracking equipment to continuously measure an operator's eye movement, which correlates with a variety of human performance constructs such as situation awareness and workload. This report describes eye tracking metrics in the context of how they will be used in nuclear power plant control room simulator studies.

CONTENTS

FIGURES

TABLES

ACRONYMS

AHP	Analytical Hierarchy Process
AOI	Area of Interest
APCPS	Average Percent Change of Pupil Size
BeGaze™	Behavioral and Gaze Analysis
ETG	Eye Tracking Glasses
FIR	Fixation to Importance Ratio
HSSL	Human Systems Simulation Laboratory
ICA	Index of Cognitive Activity
INL	Idaho National Laboratory
LWR	Light Water Reactor Sustainability
LWRS	Light Water Reactor Sustainability Program
NASA TLX	National Aeronautics and Space Administration Task Load Index
NPP	Nuclear Power Plant
OPAS	Operator Performance Assessment System
PCPS	Percent Change of Pupil Size
PV	Peak Velocity
R&D	Research and Development
RO	Reactor Operator
SA	Situation Awareness
SACRI	Situation Awareness Control Room Inventory
SAE	Selective Attention Effectiveness
SAGAT	Situation Awareness Global Assessment
SMI	SensoMotoric Instruments
SRO	Senior Reactor Operator
SVM	Support Vector Machines
U.S.	United States
USB	Universal Serial Bus

Measuring Human Performance in Simulated Nuclear Power Plant Control Rooms Using Eye Tracking

1. Introduction

This research is a part of the United States (U.S.) Department of Energy-sponsored Light Water Reactor (LWR) Sustainability (LWRS) Program conducted at Idaho National Laboratory (INL). The LWRS program is performed in close collaboration with industry research and development (R&D) programs, and provides the technical foundations for licensing and managing the long-term, safe, and economical operation of current nuclear power plants (NPPs). One of the primary missions of the LWRS Program is to help the U.S. nuclear industry adopt new technologies and engineering solutions that facilitate the continued safe operation of the NPPs and extension of the current operating licenses.

One important part of extending the life of existing light water reactors is control room modernization. Previous research under the LWRS program has identified ways to support control room upgrades and the transition to hybrid control rooms (e.g., Boring et al. 2014; Boring and Joe 2014; Hugo et al. 2013; Ulrich et al. 2014). Research has also investigated how new technologies can be effectively integrated into hybrid control rooms to enhance operator performance (LeBlanc, Boring, Joe, Hallbert, & Thomas, 2014; LeBlanc, Powers, Joe, Spielman, Rice, & Fitzgerald, 2015). This research has facilitated early stages of the transition to hybrid control rooms, but both lines of research will soon rely on a thorough understanding of the effect that upgraded control room technologies have on human performance.

Understanding how technology affects human performance in NPPs is complex, and often requires a large suite of human performance measures to capture the nuance in how the technology affects different aspects of human performance. Several efforts have defined suites of performance measures to capture aspects such as plant performance, operator performance, situation awareness (SA), and workload (Endsley, 1995b; Ha & Seong, 2009 & 2014; Hogg et al 1995; Skraaning, 2004; Taylor, 1990). Most human performance measurement suites capture a variety of measures including objective and subjective measures. One promising method of capturing objective data with relation to human attention is eye tracking. State-of-the-art eye tracking technology enables researchers to continuously measure several aspects of human performance including attention, SA, workload, and fatigue in a minimally intrusive manner.

This report describes eye tracking technology, the general metrics used in eye tracking research, and the metrics that will be used to measure human performance in full-scale simulator research supporting LWRS research on control room modernization.

2. Description of eye tracking technologies

Eye tracking is a technique used by researchers to pinpoint, record, and live stream with high accuracy the direction in which one's eyes are pointing. Along with single points of reference, eye tracking technology is also able to map gaze points, and generate heat maps to define what seems to draw a participant's attention. When a participant looks at an object multiple times, the heat map would illustrate not only how many times the object had been looked at, but also for how long. Primarily used in laboratories and academia, eye tracking has branched out into various other areas, including aviation, automotive, and NPP control room research.

There are two main categories of eye tracking technologies: remote eye trackers and wearable eye trackers. Both technologies use infrared emitting technology to detect eye direction and movement. The infrared light is reflected off of the cornea, which acts as a mirror, to create a glint. Eye direction can then be accurately determined by tracking each glint with a camera (Hansen, et al., 2010).

Wearable eye trackers are used when researchers want to avoid restricting participant movement and when a scenario requires free movement around a room. To accomplish this, eye tracking glasses are worn by the participant and paired with a data recording device that is convenient enough to be carried by the user. This is an efficient way to collect eye tracking data in an environment that is not confined to a small area (e.g., a grocery store or control room). For this research, wearable technology is ideal because the operators of a NPP control room are rarely stationary. This is especially true during procedures that require specific readings of individual dials that are located throughout the control room.

The advantage of wearable eye tracking is that it allows the participant to move freely and interact with his environment naturally. Disadvantages of the wearable technology begin to surface when it comes to post-processing the eye tracking data. Due to the fact that participants can move freely in three dimensional space, it is difficult to automatically map where the participant is looking to the environment. Typically, this places the responsibility of somehow mapping the fixation point(s) to the environment manually on the researchers. For experimental scenarios that could last up to two hours, this becomes time consuming for the researchers. However, once the fixations have been mapped, the analysis is straightforward.

Other physical issues are present with the wearable technology. Prescription eye glasses may interfere with the eye tracking. While the trackers work with most glasses, the accuracy may not be as precise. Also, placing gear on the face of a participant may be a distraction or may result in inadvertent movement of the trackers' position on the nose. Such issues could add to post-experiment processing time for frequent recalibration needs for some participants. Some models of wearable eye trackers, including SensoMotoric's, restrict the peripheral vision of the participant and may therefore force changes in the participant's behavior.

Remote eye tracking is done in a stationary environment, such as at a desk or kiosk. A device with infrared emitting diodes is placed below the participant's eyes and underneath the area of study. One example of research using remote eye tracking is website user interface design. While the participant's location has to be static, the area of interest can be dynamic (e.g., video, cycling images, etc.). Within a limited range, the remote eye tracker can directly identify where a participant is looking without requiring time-consuming manual post-processing of fixation points.

Because remote systems are meant to be in stationary locations, it is not ideal for experiments that require participants to be mobile. It is also not ideal for experiments that require a great deal of time as participants can end up slouching or not sitting within the area in which the trackers can read eye position. To accommodate both contexts (i.e., stationary and mobile scenarios) INL researchers selected two eye tracking technologies to conduct research on human performance in NPP simulator studies, the SensoMotoric Instruments (SMI) eye tracking glasses (ETG), and the SMI remote eye tracking system.

2.1 Eye tracking glasses

The ETG consists of three cameras: two cameras for eye tracking, and one outward-facing scene camera, which records up to 1280x960p resolution. The ETG must be calibrated to each new set of eyes using 0 (automatic), one, or three point modes. Once calibrated, the tracking accuracy of eye position is 0.5° between 40cm (15.75 inches) to infinite distances. To capture eye position, there are 12 infrared emitting diodes with six around each eye.

The wearable trackers can be paired with a laptop via Bluetooth or an ad-hoc wireless network. This allows the researcher to get a live view of what the user is seeing. While there is an option for full control over the experiment from the wirelessly attached device, the most logical use is for observation due to time constraints and efficiency. Since most experiments will involve more than one simultaneous participant, performing the calibration and setting options on the trackers themselves makes the most sense. SMI's glasses do not have built-in motion sensors to directly record information about head position or orientation. Connected to the ETG is a Samsung Galaxy S4 running SMI proprietary software, which is based off of the vendor's own version of the Android operating system. This software allows for input of participant information, individual experiment annotations, calibration, live feedback observation and validation, and audio recording. Its small form factor and light weight allows for easy placement into a participant's pocket or waist pack. The data from the experiment is stored on a micro-SD card, which can then be removed and placed inside of a computer for transfer and further analysis.

2.2 Analysis tools

Once transferred to the computer, SMI Behavioral and Gaze Analysis (BeGaze™) software is used for data review. Post processing analysis allows for generating graphs, heat maps, annotations, area of interest (AOI) definitions, and post-experiment recalibration if needed. Individual data points can be exported as well, including items like pupil diameter or head direction in graph form.
The visual representation of the eye movement can be displayed in several forms. Scan path, for example, allows the researcher to see the movement path of the eyes. Focal points can also be represented as individual dots that fade in and out as desired.

AOIs can be dynamically assigned within BeGaze software with a variety of options and mapping techniques. Having variety allows the researchers to collect data for a wide range of scenarios, as well as apply post-experiment data points for other research.

2.3 Remote eye tracker

The SMI REDn Professional Eye Tracker is a touchless eye tracking system, so the participant does not need to wear any gear. The device itself is placed beneath the participant and visual scene of interest to track eye movement and head position using built-in infrared emitting diodes. It is light weight (75 g), universal serial bus (USB) 3.0 powered, and secured to a surface area using adhesive pads. It can do automatic calibration, but offers two, five, and nine points if needed.

The remote eye tracker uses the same software as the ETG, so the capabilities are similar. One advantage to being a stationary device is that heat mapping can be done autonomously within the software. This saves a lot of time for the researcher and makes for faster data turnaround. Other useful features include automatic or guided setups, automatic calibration, and plug-and-play capabilities that provide faster preparation times.

3. Eye tracking metrics

Eye tracking metrics typically used in human factors research can be generalized to one of three distinct categories: fixations, saccades, and pupillometry. A brief description of each group is listed below, followed by a more in-depth discussion of specific metrics commonly used.

3.1 Fixations

Fixations are pauses in which the fovea rests on a particular region of space. Fixations involve additional processing of the raw eye-movement data to ensure that minute eye movements attributed to non-significant factors (e.g., drift, tremors, flicks) are filtered from higher-level analyses involving visual processing behavior (Salvucci & Goldberg, 2000). The eye-mind hypothesis suggests that the region fixated on can be traced to where one's attention is being directed. As such, a fixation is commonly regarded as an indication of visual information processing. Likewise, the time spent fixating can indicate various characteristics of one's cognitive ability and density of information for a given area of interest (Jacob & Karn, 2003).

Fixation Frequency is the number of fixations observed for a given task (Kovesdi, Barton, & Rice, 2012). All things equal, the greater number of fixations required to complete a task, the more inefficient the scanning strategy is, ultimately indicating interface design deficiencies when used to compare the effect of task efficiency on interface design (Goldberg & Kotval, 1999; Kovesdi, Barton, & Rice, 2012).

Fixations per Area of Interest is similar to fixation frequency but constrained by specific spatial regions of the environment (Poole, Ball, & Phillips, 2005). This metric is helpful when interested in understanding what specific region of an environment (e.g., control room board or interface) is being attended to the most. In this case, regions with higher fixations would infer that operators are attending to those regions more.

Fixation Duration is expressed as the length of a fixation, and is often used as an aggregate measure (i.e., average) when comparing differences in visual search (Kovesdi, Barton, & Rice, 2012). Fixation duration is sensitive to the level of effort required for extracting information from foveal vision (Marquart, Cabrall, & de Winter, 2015). For example, measuring fixation duration may help compare cognitive effort by tasks, or areas of interest (i.e., within an interface) in a given task, depending on the research objective. *Dwell* is commonly defined as the time a contiguous series of one or more fixations remains within an AOI (Jacob & Karn, 2003). Dwells are usually reported as a sum of fixation durations within a given AOI, and is suggested to be representative of the amount of attention spent within a particular AOI (Jacob & Karn, 2003; Poole & Ball, 2006). For instance, longer dwell durations may indicate more attention was devoted within one AOI over other AOIs.

Time to First Fixation is a measure of time from the onset of an event (e.g., experimental trial) to the first fixation on an AOI. That is, time to first fixation measures how long it takes someone to first look at a desired AOI. This metric can be used to assess how salient a particular set of features are for gaining one's attention (Byrne et al., 1999).

3.2 Saccades

Saccades are ballistic eye movements occurring between fixations when a particular area of interest is brought into foveal vision (Kovesdi, Barton, & Rice, 2012). During saccades, no information processing takes place; however, the extent of coverage (i.e., amplitude) and rate (i.e., velocity) has been used as measures for assessing the information complexity of a visual stimulus (Kovesdi, Barton, & Rice, 2012; Jacob & Karn, 2003). To note, combining fixations and saccades have provided researchers useful ways to assess what particular visual regions receive the most attention within a visual environment such as a user interface (Goldberg & Kotval, 1999).

Saccade Frequency is the number of saccades observed for a given task (Goldberg & Kotval, 1999; Jacob & Karn, 2003). Analogous to fixation frequency, in which more fixations will denote more saccades, saccade frequency can also be used to compare differences in required degree of searching, in which more saccades indicate more searching. Such a metric may serve useful when comparing search efficiency between two interface designs or tasks.

Saccade Amplitude or *Length* is the visual extent of an eye movement. Amplitude is commonly measured in visual degrees, and is typically averaged when comparing differences in visual search (Van Orden et al., 2001). Saccade amplitude is used as an indication across several constructions. As such, it is important to understand potential confounds this metric poses depending on what is being measured. For instance, saccade amplitude has been known as an indicator of mental workload in which shorter amplitudes have denoted greater mental workload (Van Orden et al., 2001). Moreover, saccade amplitudes can indicate object salience when a particular AOI yields greater attention (Goldberg & Kotval, 1999; Jacob & Karn, 2003).

Saccade Velocity is the rate of spatial displacement of an eye movement within the environment over time (e.g., degrees per second). While some research has used average saccade velocity to show relations of cognitive process with visual search skill across development (e.g., Kovesdi & Barton, 2013), the most commonly used metric is peak velocity (PV). PV is defined as max velocity for a given saccade (Di Stasi, Catena, Macknik, & Martinez-Conde, 2013). PV is known to correlate positively with saccade amplitude and has been suggested to be an indicator of mental workload and fatigue where lower PV correlates to greater workload and fatigue (Di Stasi, Antolí, & Cañas, 2011; Di Stasi et al., 2014).

Regressive Saccades refer to a leftward saccadic movement during reading. Opposite of progressive or rightward saccades, regressive saccades have been known to indicate the level of difficulty reading text in which more regressions infer greater difficulty reading (Sibert, Gokturk, & Lavine, 2000).

3.3 Pupillometry and blinks

Pupillometry covers measurement of pupil diameter, and has been indicative of various psychological constructs, including mental workload, arousal, and fatigue. Additionally, pupil diameter is sensitive to changes in ambient light levels, which has been a concern for confounds when using this measure in practical domains that have little control over ambient light.

Pupil Diameter is a well-known physiological measure of cognitive workload (Beatty, 1977). There are various metrics derived from pupil diameter that have been used to observe workload in various domains such as driving and hypermedia text (e.g., Igbal, Adamczyk, Zheng, & Bailey, 2005; Di Stasi, Antolí, Gea, & Cañas, 2011). Representative measures include average pupil diameter, percent change of pupil size (PCPS), average percent change of pupil size (APCPS), and the index of cognitive activity (ICA). Average pupil diameter is the mean diameter of the pupil (i.e., general addressed in millimeters) across a task. Due to individual variability in an individual’s pupil size, a within-subjects design is recommended to account for individual differences in pupil size (Xu, Wang, Chen, & Choi, 2011). Generally, larger diameter suggests larger workload. The PCPS is the difference between the pupil size measured at a certain point to a baseline pupil size divided by the baseline. This measure can be averaged (i.e., APCPS) across time. PCPS values are known to correlate positively with task difficulty (Palinko et al., 2010). Finally, the ICA is a sophisticated measure, which measures abrupt discontinuities in the signal of pupil diameter via wavelet decomposition. During effortful information processing, the pupil responds with rapid reflexive actions separate from reflexive changes to light. The ICA can be compared by counting the number of abrupt changes for each task. More difficult tasks would yield a greater ICA (Marshall, 2002).

Blinks have been used as a way of measuring workload. Common metrics include blink counts, blink rates, and blink durations. Counts represent the total number of blinks for a task. An increase in counts has been shown to correlate negatively to workload. Further, blink rates is the frequency of blinks per time on task. Blink rates have shown to also correlate negatively to workload (Nourbakhsh, Wang, & Chen, 2013). Finally, blink duration is the temporal latency of a blink, which is negatively correlated with workload and positively correlated with fatigue (Marquart, Cabrall, & de Winter, 2015; Schleicher, Galley, Briest, & Galley, 2008).

3.4 Aggregate metrics

Aggregate metrics are measures that combine multiple eye-tracking metrics to infer specific properties of the human-machine interaction most often in relation to attention allocation across the visual scene.
Scan Path Duration relates to the total time searching for a target. As such, scan path duration is the sum

of time from the first fixation to the final fixation when searching for a specific target. This metric accounts for the time spent fixating, as well as making saccades. Greater duration is suggested to attribute to increased information-processing complexity of a scene (Goldberg & Kotval, 1999).

Scan Path Length relates to the total saccade length required to find a target during visual search. That is, scan path length is the sum of saccades (i.e., in visual degrees or in pixels) for finding a target. Greater scan path lengths infer less efficient search behavior. Although this metric doesn't distinguish between visual search and information processing times, this metric can be used to compare different interface attributes as part of evaluating an optimal visual search (Goldberg & Kotval, 1999).

Scan Transition Matrices involve a temporal component to assessing scan paths. That is, transition matrices involve the evaluation of the transitions from one AOI to another. This evaluation takes the form of comparing counts of unique transitions from one AOI to another, usually presented as a percentage where the frequency of unique transitions is divided by the total number of opportunities. As shown in Figure 1, the search pattern on the left is less efficient where there is a transition density of 0.3125 compared to the efficient search of 0.25 on the right. Worth noting, evaluations of AOI can be context dependent or independent. In the dependent case, AOIs would be defined based on specific interface regions such as an indicator within a display. These AOIs may or may not be of the same geospatial size. The independent case would divide the entire visual area captured into a grid as part of comparing search (Goldberg & Kotval, 1999).

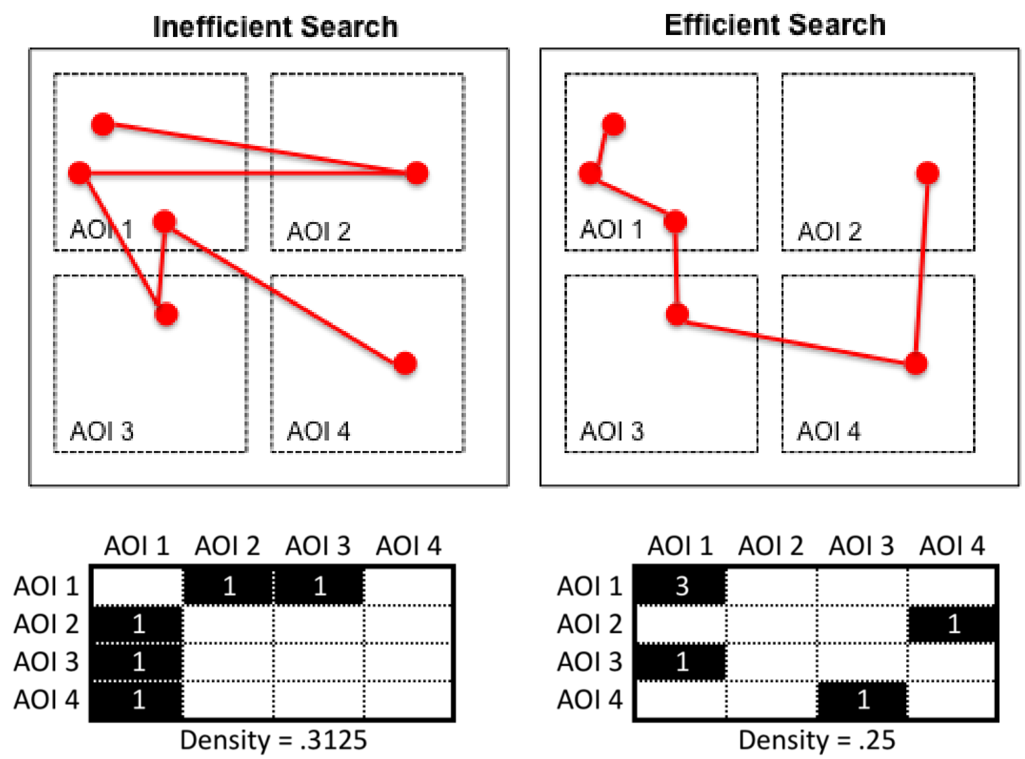

Figure 1. Scan transition matrices of two illustrated search patterns

Heat Maps are graphical representations used to compare the level of attention allocated to different areas in a defined visual space. Combining fixation frequency and duration, heat maps show a tiered coloring scheme to illustrate overall fixation density as a way of indicating how much attention an area received. Often heat maps use cool colors to show areas given little attention and hot colors as areas receiving more attention. The resultant pattern looks as though there are hot spots indicating focal points thus lending the name "heat maps."

Fixation to importance ratio (FIR) relates the time a participant spends attending to an AOI to its importance. The measure is indicative of how effectively the participant is utilizing their time and attentional resources to complete the task at hand. When analyzing the fixation to importance ratio researchers should consider using number of fixations along with average gaze duration. For instance, Ha and Seong (2007) calculated FIR as an aggregate of fixation frequency (FIR^N) and fixation duration (FIR^D) in a NPP simulation (i.e., $FIR^N + FIR^D / 2$). Analyzing average gaze duration alone for FIR may confound importance with a difficulty extracting data (Fitts, Jones, Miltion, 1950).

Selective attention effectiveness (SAE) aggregates all FIR's into a single value representing the participant's overall ability to focus on the necessary visual elements while ignoring the unnecessary elements to complete the task. SAE values closer to zero denote greater overall attentional resource effectiveness (Ha & Seong, 2007). The participant's ability to do this effectively contributes to their efficiency and effectiveness at completing the task. The inability to discriminate between the irrelevant and the relevant visual stimuli may lead to errors or slow down the diagnostic process.

Both FIR and SAE are inferential measures of perception and diagnosis. Perception has often been evaluated using questionnaire formats such as Situation Awareness Global Assessment Technique (SAGAT) and the adapted form for NPP observation Situation Awareness Control Room Inventory (SACRI). These forms are sensitive to either correct or incorrect determinations; however with eye tracking, we can determine whether the participant failed to see the information or saw but misinterpreted the information.

Table 1 summarizes the metrics described in this section and their relationship to constructs of human performance.

Table 1.The relationships between eye tracking metrics and the constructs they correlate with.

Construct	Measure	Correlation to Construct (+/-)
Scan/ Search Efficiency	Fixation Frequency	(-)
	Saccade Frequency	(-)
	Saccade Amplitude	(-) During Visual Search with Spatially Known Target
	Scan Path Duration	(-)
	Scan Path Length	(-)
	Scan Transition Matrix Density	(-)
Importance Inferences	Fixations per Area of Interest	(+)
	Dwell Duration	(+)
Workload	Fixation Duration	(+)
	Saccade Amplitude	(-) During Visual Search with Spatially Unknown Target
	Peak Saccade Velocity	(-)
	Regressive Saccades	(+)
	Pupil Diameter	(+)
	PCPS/ APCPS	(+)
	ICA	(+)
	Blink Counts	(-)
	Blink Rate	(-)
	Blink Duration	(-)
Situation Awareness	FIR	(-)
	SAE	(-)

3.5 Discussion of issues and constraints

While eye tracking offers great promise in human performance research, there are several considerations when planning for an eye tracking study to ensure valid and reliable data. Some issues are dependent to the eye tracking technology at hand, while others are technology independent. The following sections discuss the technology-independent followed by technology-dependent considerations.

3.5.1 Technology-independent considerations

3.5.1.1 Accuracy

Prior to running an eye tracking study, an important step is ensuring the eye tracker provides an accurate mapping of one's sight to the visual scene. Calibration is the process that familiarizes the technology with the individual's eyes to achieve the closest matching between person's point of fixation and the visual scene. The first important constraint to be aware of is the degree of accuracy offered by the eye tracking system. The degree of accuracy will determine the limits of how small an area can be defined as an AOI to be used in eye tracking analysis. For instance, assuming an eye tracker has an accuracy of 0.5 visual degrees (i.e., 30 minutes of arc) and the user's viewing distance is 24 inches, the system could track one's foveal vision within an area around 0.21 inches. All things equal, if the viewing distance is increased to 10 feet the eye tracking system could track an area around 1.05 inches. These differences could change the scope of the research questions such as examining an AOI that represents text on a screen to defining an AOI as a general region that contains said text (Eye Tracker System, Version 3.01).

Of particular importance in the context of NPP simulator studies is the fact that the viewing distance cannot be easily controlled if the realism of the scenario is to be preserved. During a control room study operators are on the move and depending on where information is located an operator may, at one point, be within two feet of a display and at another be 10 feet from the same display to collect information. The space differences could change the scope of the research questions such as examining an AOI that represents text on a screen to defining an AOI as a general region containing this text (Eye Tracker System, Version 3.01). Certain concessions may be required to accommodate the fluctuating resolution of the eye tracker such as using an entire control panel as an AOI versus a single indicator. Another consideration may be analyzing finer grained visual elements using a remote eye tracker during an isolated condition versus a full control room fault scenario.

3.5.1.2 Individual characteristics

Another important calibration related issue is how an individual's eyes and facial characteristics can influence adequate tracking. Users with bifocals or trifocals may not calibrate appropriately. Thick-rimmed glasses and use of make-up (e.g., Mascara) can cause similar problems. Participants who have long eyelashes or bangs can obstruct visibility of the eye tracking camera to the pupil. With some eye tracking systems, the color of one's iris (i.e., dark eyes) can be problematic for calibration. Finally, those with eye conditions such as glaucoma, cataracts, permanently dilated pupils, or an offset fovea can cause problems with accuracy (Eye Tracker System, Version 3.01; Pernice & Nielsen, 2009). The root cause of this problem is typically related to light never reaching the fovea or reflecting poorly due to lenses or other obstructions causing the system to misrecognize the pupil or corneal reflection.
Fortunately, eye-tracking technology has developed such that few conditions exist that cannot be accommodated by calibration. However, when screening participants a few considerations are still required. Requesting participants to wear corrective contact lenses in place of glasses if available on the day of testing can work around thick rims. Unfortunately for NPP Control Room Simulator studies the participants are difficult enough to find that an operator not compatible with ETG must forego eye tracking in that particular situation instead of selecting a participant solely for their compatibility with eye tracking.

3.5.1.3 Environment

Environmental considerations are a third major contributor to inaccuracies during calibration. Ambient light levels can lead to problems with the calibration process, ultimately causing poor data. Avoiding sunlight, as well as incandescent and halogen bulbs can improve calibration and data collection. Having a controlled room with adequate light levels from fluorescent bulbs is ideal. Other considerations related to lighting regards ensuring that participants avoid wearing brimmed hats or other clothing that can alter the light levels near the eyes (Eye Tracker System, Version 3.01; Pernice & Nielsen, 2009). Here is a case in which working in a NPP simulator context is advantageous. Having access to a laboratory with controlled lighting and environmental conditions can accommodate any lighting sensitivities the eye tracking gear may have with relative ease. No environment is perfectly uniform. Calibration issues aside, pupillometry and blink rate reflect a participant's physiological response to changing lighting conditions after head and body movements, as well as the effect of workload and stress.

3.5.2 Technology-dependent considerations

3.5.2.1 Wearable systems

Evaluating new technologies in the context of realistic NPP control room operations using eye tracking requires the mobility wearable eye tracking systems allow. The user may move freely about an environment without, or with minimal, spatial constraint. However great freedom of movement is accompanied by longer data processing for a couple reasons. First researchers must consider the processing time to spatially map all the data. Because the user is now mobile, the collection environment is dynamic, which requires manually mapping of fixation points to three dimensional space. Typical methods for mapping these eye movements to either take the form of reviewing a video and matching various AOIs frame-by-frame, or mapping fixations manually to a reference image of the scene concurrently to reviewing a participant's video. In either case, data processing can be labor and time intensive. However processing costs can be reduced by focusing on the areas that relate to the question under investigation and filtering out regions or time spans that do not provide meaningful information.

Second, the eye cameras are usually confined within the wearable system's glasses frame. Glances where the pupil is positioned at extreme angles exceeding the bounds of the glasses can result in missing data. The issue can be particularly prevalent in control rooms where screens or information located in the extremities of the room such as near the ceiling may lead the user to peer beyond the equipment's trackable region.

3.5.2.2 Remote systems

Remote systems provide non-intrusive eye tracking requiring no physical contact with the eye tracking equipment. Rather, a typical remote eye tracker is positioned below the viewing scene (i.e., usually a computer monitor). With remote systems, the experimenter gives up freedom of movement to have more efficient data processing. The remote system can map the dimensions of the monitor to form the boundaries of the visual scene. Since the boundaries of the scene are assumed not to move, coordinates of AOIs are readily available. However, a potential issue with remote systems is that it restricts the movement of the participant during data collection, an unrealistic control room constraint. Potential restrictions caused from the remote system may alter the way a user interacts, ultimately jeopardizing external validity. If the eyes fall outside (e.g., from moving one's head) the workable region of the remote system, data will be lost. Similarly, adequate eye height from the camera is important to ensure adequate sampling. Another issue with remote systems is ensuring any objects that are within view of the eye tracking cameras do not distract from recording the pupil and corneal reflection. For instance other people looking within view of the cameras, or other shiny objects are problematic when tracking one's pupil and corneal reflection (Eye Tracker System, Version 3.01).

4. Measures planned for HSSL studies

The research conducted under the LWRS program in support of control room upgrades has been described in Le Blanc et al. (2015). The research is investigating the impact several proposed control room technologies have on human performance in NPP control rooms. The proposed benefits of each of the technologies selected for the first phase of research are summarized in Table 2. To accurately assess whether these technologies provide the proposed benefits, the researchers will compare performance using the candidate technologies to baseline performance without the technologies (LeBlanc, Boring, Joe, Hallbert, & Thomas, 2014). Researchers will use a set of performance measures described in LeBlanc, Joe, Rice, Ulrich, & Boring (2015) in addition to eye tracking measures to investigate the benefits of these technologies for human performance.

Performance will be assessed using realistic full-scale NPP simulation scenarios. The scenarios used in these studies are modeled after the scenarios used to train NPP operators in their full scale simulator. A typical scenario will require a crew of operators to respond to a combination of instrument and sensor failures along with injected faults (such as faulted equipment or leaks). A crew of operators will use existing plant knowledge along with procedures (normal, abnormal, and emergency procedures) to monitor, diagnose, perform corrective actions, and perform recovery actions. These scenarios are complex and require the operating crews to utilize information contained on the analog display panels, annunciator panels, digital display systems, and written procedures. Scenarios can last from 30 minutes to several hours depending on the complexity of the scenario and the goals of the study (LeBlanc, Joe, Rice, Ulrich, & Boring, 2015). Collecting meaningful eye tracking data in this context requires careful consideration of how the eye tracking metrics correlate to meaningful constructs of human performance.

The majority of eye tracking methods use objective observation of a participant's attention to infer something about the participants understanding of the visual environment. In an NPP control room, the visual environment is complex, and operators are constantly scanning the environment to monitor important parameters and determine whether the plant is running normally. During an emergency, operators will be reading emergency operating procedures, using the analog control panel to verify that safety systems are functions properly, and utilizing the procedural guidance to diagnose the problem. The operators will move around the entire control during this time (LeBlanc, Powers, Joe, Spielman, Rice, & Fitzgerald, 2015). During each scenario, the researchers will select periods of time in which each the eye tracking metrics make sense given the context of the scenario and the way the operator interacts with his visual environment.

In order to evaluate the proposed benefits of the control room technologies summarized in Table 2, the researchers have selected eye tracking metrics to evaluate SA, workload, and monitoring and detection of important indicators. The quantitative metrics discussed and summarized previously will be included to evaluate the benefits of candidate control room technologies. Further, for full-scale simulation, the wearable eye tracking systems will be used for each operator to monitor eye movements during selected simulation events of interest. These selected metrics collected from the eye tracking systems will be treated as dependent variables, and will be used to compare performance with and without the technology upgrades.

Table 2 provides an overview of the relationship between technologies, benefits, and eye tracking measures. A description of this relationship is provided below.

Table 2. Relationship between technologies, proposed benefits and eye tracking metrics

Technology	**Benefit**	**Selected Eye Tracking Measures**
Overview Displays	Reduced workload Enhanced SA Enhanced detection of off-normal conditions Enhanced crew coordination	**Measures of workload** Fixation Duration Pupil Diameter PCPS/ APCPS Blinks (i.e., counts, rate, duration) **Measures of SA** FIR & SAE Fixations per AOI Fixation Rate Saccade Frequency Dwell Duration **Measures of detection** Fixation Frequency Time to First Fixation
Advanced Alarm Systems	Reduced workload Enhanced diagnosis Increased efficiency	**Measures of workload** See workload measures above **Measures of diagnosis** Time to First Fixation **Measures of efficiency** Fixation & Saccade Frequency Scan Path Length & Duration Scan Transition Matrix Density
Computer-Based Procedures	Enhanced performance Reduced errors Enhanced efficiency	**Measures of performance & errors** Scan Transition Matrix Density for monitoring tasks **Measures of efficiency** See efficiency measures above

Below is detailed description of these measures as they relate to the technology and benefits.

4.1 Measures of workload

Reduced workload is a claimed benefit of many advanced control room technologies. Overview displays are designed to provide at-a-glance information and deliver larger amounts of diagnostic information by borrowing principles of ecological interface design. It is expected that operators will be able to extract the same amount of information at a quick glance from overview displays as they would from searching, extracting and remembering multiple values from multiple indicators as many do currently. Eye tracking can provide detailed insight into the effort operators muster to extract important information from the overview displays or from current configurations and compare the two. Using specific eye tracking metrics we can then extrapolate workload levels afforded by the different technologies can then be extracted.

Another control room technology that may reduce operator workload is an advanced alarm system by aiding the user in monitoring and decision making through alerting, prioritizing, and providing decision support of important plant issues.

In corroboration with subjective workload measures (e.g., NASA-TLX), workload can be evaluated continuously through mobile eye tracking by collecting fixation duration, pupil diameter, PCPS/ APCPS, and blink characteristics during each scenario. Specifically, instances in which the scenario triggers increased workload can be identified from these measures, and cross-referenced to other data collection methods such as simulator logs and computer-assisted operator performance assessment system (OPAS) to understand current plant state, the audio-visual environment at a given time and the specific visual elements operators were attending to that likely increased workload. The following indicators would yield such information:

- Increase in fixation duration,
- Increase in pupil diameter,
- Increase in PCPS/ APCPS,
- Decrease in blink counts,
- Decrease in blink rate, and
- Decrease in blink duration.

4.2 Measures of situation awareness

Overview displays support global SA by providing visibility to important information in one place at all times, as opposed to only a subset. Having this “big picture” view of critical plant states supports prioritization of goals and enabling projection of future events (Endsley, 2011). Further, advanced alarm systems may support SA by alerting and readily providing the most critical information for addressing changes in plant state.

To investigate possible SA benefits, mobile eye tracking will be one way of evaluating the perceptual attributes of SA (i.e., level 1) through tracking the specific display elements that were fixated on. From the eye mind hypothesis, AOIs with greater fixations and dwells are inferred to have had greater attention allocated towards them. As such, use of mobile eye tracking will be able to examine display elements that were attended to most, as well as what important display elements were missed during each scenario. Per scenario, important elements of the display relative to the primary goal will be defined as an AOI, as a post data collection activity. AOIs will be considered at both the component and indicator level. All AOIs will account for potential inaccuracies through having a 0.5-degree edge around each of them. Eye tracking measures will include FIR, SAE, fixations per AOI, fixation rate, saccade frequency, and dwell duration. FIR and SAE measures will require additional importance weighting of each AOI relative to each task, which will be quantified using the Analytical Hierarchy Process (AHP). FIR and SAE will be calculated in accordance to Ha and Seong (2014), where FIR will be a composite measure of both ratio of fixation frequency per AOI to AOI importance and ratio of fixation duration per AOI to AOI importance. SAE will comprise an overall value of FIRs divided by total AOIs. FIRs and SAEs of lower values will suggest greater SA, through indication of greater attentional resources towards more critical display elements. Like FIR and SAE, greater fixations, fixation rate, saccade frequency, and dwell duration per AOI are expected to indicate greater SA. Worth noting, the latter measures do not distinguish level of importance like FIR and SAE. Rather, fixation frequency, fixation rate, saccade frequency, and dwell duration infer SA through dichotomizing whether an AOI was attended to or not.

4.3 Assessing operator performance of monitoring tasks

Existing methods for assessing crew performance during a scenario typically involve an observer recording the successful completion of a subtask, or relying on the simulator software logging the time at which some key action was taken. One situation where these methods are inadequate is when the procedure directs the operators to monitor some condition on an ongoing basis, and only take action if an abnormal condition is found during one of those routine scans. Such instructions are displayed on special fold-out pages of the Senior Reactor Operator's (SRO) procedure manual, so that the operator can carry out the scan repeatedly while other steps in the procedure are ongoing. The operators do not typically have to touch the panel while scanning or announce verbally every time they conduct a scan unless they detect a condition requiring a response. An eye tracker, however, can be used to detect the times at which the operator conducts the scan specified in a fold-out procedure. A direct comparison of how frequently the scan was accomplished vs. the scan rate specified in the procedure is now possible, rather than merely assessing the effectiveness of the crew's scan based on whether the crew successfully detected an abnormal condition if one arose.

If a complete mapping of all fixations in a scenario has been made, this analysis can be done by defining appropriate AOIs on the panel for the elements that the procedure says to include in the scan. Absent such a complete mapping, a count of the scans can be made quickly by having a reviewer watch the recorded video on fast-forward and making a note of each time that the operator scanned the appropriate area.
To evaluate detection, time to first fixation on important indicators such as alarms will be measured. Researchers and subject matter experts will identify points in the scenario in which critical information is displayed on the alarm panels or indicators.

4.4 Use of advanced displays

To investigate how reactor crews utilize the new overview displays, the operating procedures do not specify when the crews should look at the new displays or how they should use them. Typically the procedure directs a reactor operator (RO) to determine if some plant parameter such as a temperature or flow rate is in or out of a range and react appropriately to his findings. In cases where the relevant information appears both on the new and old displays, eye tracking footage can be reviewed to determine whether the operator looks at the overview display, the old gauge, or crosschecks them both, before reporting his observation to the senior operator. If a scenario includes several opportunities of this type, a quantitative measure of the degree to which the operators are choosing to use the overview display rather than the old detail displays can be created.

More generally, one can review an operator's gaze patterns over the course of a scenario, and record when they choose to look at the overview displays. The simplest such measure is totaling how much time each operator spends looking at the new displays. This can be done by defining the entire overview display as one AOI.

A more complex analysis aimed at assessing *how* the crews use the displays rather than simply *if and when* they look at them, might define several small AOIs on the overview display, and several large AOIs elsewhere in the room, each one corresponding to the portion of the larger control panel that is being summarized by a particular element of the overview display. A simple comparison of fixation counts or dwell times on the elements of the overview display provides a crude measure of how much each element is being used; these counts could easily be combined with expert judgment about the components of the display to create fixation-to-importance-ratio measures. A transition matrix, recording which portions of the control room operators view immediately before or after looking at the overview display, will provide insight as to whether the operators are glancing at the overview display *instead of* scanning a region of the main panel, or glancing at the overview display *in preparation for* a more detailed scan of the main panel, in reaction to what they see on the overview display.

5. Conclusions

It is clear eye tracking technology has evolved to facilitate its use in a complex environment such as a NPP Control Room wherein multiple operators fill varied roles. While not every issue has yet been addressed there are still collectable measures that would clarify how operators allocate their attention when perceiving and diagnosing faults or transients in a NPP setting.

Some measures afford more qualitative interpretations of the data. Heat maps generate a general picture of the visual elements that receive attention while simultaneously providing a hierarchical scale of which elements received the most attention. Such information allows for quick comparison between technologies of how different elements are interacted with. Depending on supporting performance metrics conclusions about the visual elements importance or perceptive difficulty may be made.

To understand how display type can affect operator performance, any number of the quantitative metrics described above could prove useful to interpreting operator attention allocation patterns. This is particularly true for those metrics pertaining to scan/search efficiency, importance inferences, and workload measures.

The measures that may provide insight into SA include the aggregate measures; FIR and SAE. FIR is intended to identify the ratio of fixations the operator made on visual elements critical to the task at hand versus those that were not indicating a higher level of system awareness. SAE is a broad score combining the FIR of all visual elements to ascribe an operator with a measured level of, as the name implies, Selective Attention Effectiveness (Ha & Seong, 2014).

Further, when planning an eye tracking study, a thorough understanding of the constraints of the eye tracking technology at hand is important to ensure that an acceptable level of accuracy is achieved. To summarize, such considerations include defining AOIs that are within the accuracy limits of the eye tracking system, accounting for individual differences, having a grasp of required data processing resources for certain measures, and controlling for environmental constraints such as ambient light levels.

A benefit of current eye tracking gear is in the a posteriori processing flexibility. When tracking a participant’s gaze, the eye tracking software records all the data necessary to calculate all the above metrics. Having the data at hand allows researchers to adjust which metrics to include without compromising the data in any way. Applicability and time or resources are the only constraints to utilizing the full range of metrics every time.

A final consideration involves use of appropriate statistical techniques for evaluating constructs of operator attention allocation. For instance, univariate approaches like linear regression or analysis of variance may not be appropriate due to nonlinear characteristics of the data. Future investigation should consider more methods that account for nonlinearities of eye movement data in the NPP domain. For example, research in other domains (e.g., driver distraction) have shown that Support Vector Machines (SVM) or Naïve Bayesian classifier can accurately classify eye movements to constructs like workload and distraction (Liang, Reyes, & Lee, 2007; Nourbakhsh, Wang, & Chen, 2013).

6. References

Beatty, J. (1977). Pupillometric measurement of cognitive workload. California University Los Angeles Department of Psychology, (No. TR-22).

Byrne et al. (1999). Eye tracking the visual search of click-down menus. *In Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 402-409). ACM. doi:10.1145/302979.303118

Di Stasi, L. L., Antolí, A., & Cañas, J. J. (2011). Main sequence: an index for detecting mental workload variation in complex tasks. *Applied ergonomics,42*(6), 807-813. doi:10.1016/j.apergo.2011.01.003

Di Stasi, L. L., Antolí, A., Gea, M., & Cañas, J. J. (2011). A neuroergonomic approach to evaluating mental workload in hypermedia interactions. *International Journal of Industrial Ergonomics, 41*(3), 298-304. doi: 10.1016/j.ergon.2011.02.008

Di Stasi, L. L., Catena, A., Canas, J. J., Macknik, S. L., & Martinez-Conde, S. (2013). Saccadic velocity as an arousal index in naturalistic tasks. *Neuroscience & Biobehavioral Reviews, 37(*5), 968-975. doi: 10.1016/j.neubiorev.2013.03.011

Di Stasi, L. L., McCamy, M. B., Macknik, S. L., Mankin, J. A., Hooft, N., Catena, A., & Martinez-Conde, S. (2014). Saccadic Eye movement metrics reflect surgical residents' fatigue. *Annals of surgery, 259*(4), 824-829. doi: 10.1097/SLA.0000000000000260

Endsley, M.R. (1995b). Measurement of situation awareness in dynamic systems. Human Factors, 37, 65-84. doi: 10.1518/001872095779049499

Endsley, M. R. (2011). Designing for situation awareness: An approach to user-centered design. *CRC Press.*

Eye Tracker System (Version 3.01) [Computer software]. Bedford, MA: Applied Science Laboratories

Fitts, P.M., Jones, R.E., & Milton, J.L. (1950). Eye movements of aircraft pilots during instrument landing approaches. *Aeronautical Engineering Review, 9*(2), 24-29.

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics, 24(*6), 631-645. doi:10.1016/S0169-8141(98)00068-7

Ha, J. S., & Seong, P. H. (2009). HUPESS: Human performance evaluation support system. *Reliability and Risk Issues in Large Scale Safety-critical Digital Control Systems,* 197-229. doi:10.1007/978-1-84800-384-2

Ha, J., & Seong, P. (2014). Experimental investigation between attentional-resource effectiveness and perception and diagnosis in nuclear power plants. *Nuclear Engineering and Design, 373*(1), 758-772. doi:10.1016/j.anucene.2015.01.021

Hansen D. W., & Ji, Q. (2010). In the eye of the beholder: A survey of models for eyes and gaze. *PAMI, 32*(3), 478–500. doi:10.1109/TPAMI.2009.30

Iqbal, S. T., Adamczyk, P. D., Zheng, X. S., & Bailey, B. P. (2005). Towards an index of opportunity: understanding changes in mental workload during task execution. *In Proceedings of the SIGCHI conference on Human factors in computing systems,* (pp. 311-320). ACM.

Jacob, R. J., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind, 2*(3), 4.

Kovesdi, C., Barton, B. K., & Rice, L. (2012). Visual efficiency-detection index: A new composite measure of visual search. *Journal of Eye-Tracking, Visual Cognition and Emotion, 2*. ISSN: 1647-7677.

Kovesdi, C. R., & Barton, B. K. (2013). The role of non-verbal working memory in pedestrian visual search. *Transportation research part F: traffic psychology and behaviour,* 19, 31-39.

LeBlanc, K., Boring, R., Joe, J., Hallbet, B., & Thomas, K. (2014). A Research Framework for Demonstrating Benefits of Advanced Control Room Technologies, *INL/EXT-14*-33901 Revision1

LeBlanc, K., Joe, J., Rice, B., Ulrich, T., & Boring, R. (2015). Benefits of Advanced Control Room Technologies: Phase One Upgrades to the HSSL and Performance Measures, *INL/EXT-15*-3538 Revision 0.

LeBlanc,K., Powers, D., Joe, J., Spielman, Z., Rice, B., & Fitzgerald, K. (2015). A Pilot Study Investigating the Effects of Advanced Nuclear Power Plant Control Room Technologies: Methods and Qualitative Results, *INL/EXT-15*-36432 Revision 0.

Liang, Y., Reyes, M. L., & Lee, J. D. (2007). Real-time detection of driver cognitive distraction using support vector machines. *Intelligent Transportation Systems, IEEE Transactions on*, *8*(2), 340-350. doi:10.1109/TITS.2007.895298

Marshall, S. P. (2002). The index of cognitive activity: Measuring cognitive workload. *In Human factors and power plants, 2002. Proceedings of the 2002 IEEE 7th conference on* (pp. 7-5). IEEE. doi:10.1109/HFPP.2002.1042860

Marquart, G., Cabrall, C., & de Winter, J. (2015). Review of eye-related measures of drivers' mental workload. *Procedia Manufacturing, 3*, 2854-2861. doi:10.1016/j.promfg.2015.07.783

Nourbakhsh, N., Wang, Y., & Chen, F. (2013). GSR and blink features for cognitive load classification. *In Human-Computer Interaction–INTERACT* (pp. 159-166). Springer Berlin Heidelberg.

Palinko, O., Kun, A. L., Shyrokov, A., & Heeman, P. (2010). Estimating cognitive load using remote eye tracking in a driving simulator. *In Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (pp. 141-144). ACM. doi: 10.1145/1743666.1743701

Pernice, K., & Nielsen, J. (2009). How to conduct eye tracking studies. *Nielsen Norman Group*.

Poole, A., Ball, L. J., & Phillips, P. (2005). In search of salience: A response-time and eye-movement analysis of bookmark recognition. *In People and Computers XVIII—Design for Life*, (pp. 363-378). Springer London. doi: 10.1007/1-84628-062-1_23

Poole, A., & Ball, L. J. (2006). Eye tracking in HCI and usability research. *Encyclopedia of human computer interaction, 1*, 211-219. doi: 10.4018/978-1-59140-562-7.ch034

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

Schleicher, R., Galley, N., Briest, S., & Galley, L. (2008). Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired?. *Ergonomics,51*(7), 982-1010. doi:10.1080/00140130701817062

Skraaning, G. (2004). Experimental Control Versus Realism: Methodological Solutions for Simulator Studies in Complex Operating Environments, HWR-361, *Halden Reactor Project*, Halden, Norway.

Sibert, J. L., Gokturk, M., & Lavine, R. A. (2000). The reading assistant: eye gaze triggered auditory prompting for reading remediation. *In Proceedings of the 13th annual ACM symposium on User interface software and technology*, (pp. 101-107). ACM. doi:10.1145/354401.354418

Taylor, R. M. (1990). Situational Awareness Rating Technique(SART): The development of a tool for aircrew systems design. *AGARD, Situational Awareness in Aerospace Operations* 17 p(SEE N90-28972 23-53).

Van Orden, K. F., Limbert, W., Makeig, S., & Jung, T. P. (2001). Eye activity correlates of workload during a visuospatial memory task. Human Factors: *The Journal of the Human Factors and Ergonomics Society, 43*(1), 111-121. doi: 10.1518/001872001775992570

Xu, J., Wang, Y., Chen, F., & Choi, E. (2011). Pupillary response based cognitive workload measurement under luminance changes. *In Human-Computer Interaction–INTERACT 2011*, (pp. 178-185). Springer Berlin Heidelberg. doi:10.1007/978-3-642-23771-3_14