Using eye tracking to explore design features in nuclear control room interfaces

Alexandra Fernandes, Sathiya Kumar Renganayagalu, & Maren H. Rø Eitrheim
IFE – Institute for Energy Technology and OECD Halden Reactor Project, Norway

Abstract

This study’s main goal was to analyse the impact of specific innovative design features in nuclear control room digital interfaces. A within-subject experimental approach was used, where the same participants responded to the same blocks of questions in two conditions: with innovative designs – including bar graphs, mini-trends, pie-charts, etc. – and a control condition where the same process information was presented through numerical information only. A simplified task was designed to collect the response time and accuracy through a tablet: the participants were presented with consecutive questions regarding the process status that required them to scan the process displays and report values of targeted components and decide on the accuracy of statements on current plant processes. Nine experienced operators participated and three wore eye tracking glasses. The current analysis focused on the questions that presented the larger differences between the control and the innovative conditions (both time and accuracy). The overall performance results reveal that the participants were more accurate in the innovative condition and showed equivalent response times in both conditions. The eye tracker enabled a further qualitative exploration of performance data, showing that dwell times and fixation counts tended to be lower in the innovative condition, and that average fixation duration were equivalent in both conditions.

Introduction

One of the current main challenges within the nuclear industry is to assess and compare safety, performance, and efficiency in analogue systems versus digital systems. Most of the currently operating nuclear power plants were designed and built between the 1960’s and 1980’s, implying that the main control rooms were designed with mostly analogue interfaces and manual controls. In the last few decades it has become increasingly difficult to maintain and replace these types of interfaces due to its obsolescence, lack of replacement parts, or unavailability of the vendors (Joe, Boring, & Persensky, 2012). Although analogue interfaces are still a central part of nuclear process control, many power plants have been involved in small or large-scale modernisation projects that introduce digital interfaces as an extention or replacement of anologue interfaces (Stubler, O’Hara, Higgins, & Kramer, 2004).

The impact of digitalisation on human performance and safe operations in the control room has become a central aspect within the nuclear industry (O’Hara, Stubler, & Kramer, 1997). At best, new interface features may enhance the ability to monitor and control the nuclear process, for example by providing memory aids and diagnostic support to the operators. At worst, the advanced graphics can confuse the operators and mislead their attention. There is a need to adjust existing methodologies and techniques of performance assessment to target these particular issues and to derive conclusions for the analogue versus digital debate (Hildebrandt & Fernandes, 2016). Within this context, the use of usability methods gains relevance, especially for techniques such as eye tracking that originate direct data that is independent of subjective reports and preferences.

Eye tracking is a set of techniques and methods used for recording and measuring eye movements. The term eye tracking or gaze tracking, as used here is the estimation of direction of the user’s gaze. In most cases, estimations of gaze direction imply identification of a target object. Eye movements can be interpreted as the result of constant interaction between cognitive and perceptual processes (Richardson & Johnson, 2008), and as such could enable a better understanding of search strategies and cognitive functions such as information processing, reasoning, and decision-making (Mele & Federici, 2012). Eye tracking methodologies have been considered promising for many years, but its use in applied contexts is still not the standard (e.g. Bojko, 2013). Many factors can be contributing to this, namely its cost, its ease of use, or the amount of training in data collection and analysis that is required when using the equipment. However, recent developments in the technology have made eye tracking less expensive and more robust and simple to use in laboratories as well as applied contexts especially in Human-Computer Interaction research (Jacob and Karn 2003). Currently, the most common eye tracking technique is video based eye tracking. Video based eye trackers are unobtrusive, easy to setup and collect data. There are two main eye movements that are measured by eye trackers: fixations and saccades. Fixations are minor eye movements around a point of interest. These minor eye movements are needed to keep points of interest in focus. Saccades are rapid eye movements changing the fovea to a new location of interest. There are many other metrics derived from these two eye movements. In the interface design context, number of fixations, fixation duration and its frequency are the measures for search and processing (Goldberg & Kotval, 1999) and provide usability data of interfaces. Related with these main measures are a set of metrics that can be derived from eye tracking data, namely dwell time and lookbacks that will be consired in the current study. Dwell time relates with the sum of the duration of all fixations and saccades within a predefined area of interest (in the current case the process displays picture). Lookbacks refer to the count of the number of intances where the participants’ gaze returned to a specific area of interest after leaving it.

The current paper describes a first attempt to use an eye tracking tool to further interpret and analyse performance data in a usability-based technique, where the search patterns, the ability to find different system elements, the correctness of the responses, and the response time is assessed objectively for different design versions of the same process components.
Method

Participants

Nine licensed nuclear power plant control room operators (three crews of three operators each) took part in the behavioural tasks in this study. From these, three operators volunteered to wear eye tracking glasses throughout the tasks. All the participants were male and had an average age of 43.7 years ($SD=11.1$) and 15.4 years of experience ($SD=10.5$) as control room operators.

Materials

Stimuli

The main stimuli in this study were process pictures taken from a control room interface developed at the authors’ institution. A set of 61 pictures of different displays, presenting different system status was selected. Figure 1 shows an example of a display picture presented in both the innovative and control conditions.

![Figure 1. Process pictures with innovative (top) and control (bottom) features](image-url)
Each of the pictures had two versions: 1) innovative, where new visualisation features for process data were introduced such as pie-charts to indicate flow after pumps, bar graphs to allow a comparison of values in sequential pumps, minitrends that showed the history of a specific parameter in the last 10 minutes, or pictorial displays that represented a combination of for example pressure and water level values; and 2) control, where the process data required to answer the questions was shown only in numerical format, i.e., digits representing the values of pressure, temperature, etc. The innovative visualisations are the results of long-term process of conceptualisation, design, development and implementation within the Halden Reactor Project and intended to support control room operators in a vast set of tasks, aiming at reducing workload, improving shared situation awareness, and minimizing secondary tasks, facilitating process monitoring (Svengren, Eitrheim, Fernandes, & Kaarstad, 2016). The control displays corresponded to a version of the innovative displays where all innovative features were removed and replaced by more conventional numerical representations of the process values.

Data collection App
The data collection was performed using a tablet app developed for this purpose. The participants were asked to answer a question and then swipe the screen so that they could continue with the following question. The participants were able to change their answers while in the question screen, but it was not possible to swipe back to previous questions. Figure 2 shows an example of one of the questions presented in the task. The participants were presented with a static image of a predetermined process status, so the answers for all questions were pre-set and were the same for all participants in both conditions (but the order of the questions within was randomised). All the questions corresponded to everyday tasks that the operators perform in the control room. Different questions targeted different features of the interface, and might be focused in only one feature (e.g. identifying the flow in a pump) or a combination of features (e.g. checking if a pump was open and confirming a value in another pump).

The participants were presented with different possibilities to answer each question: true/false; yes/no; multiple-choice between three or more alternatives; or typing specific parameter values in the tablet. An answer was mandatory in all questions and the participants had the option to answer Don’t know in a button presented in the lower right corner of the screen.

An experienced operator worked with the authors to define and select the scenarios represented in the pictures and to define the questions to be presented for each. These questions were intended to be representative of the types of tasks that the operators need to perform in their everyday work and focused on checking parameters (e.g. 712 PD1 has reduced flow) and identifying status of specific components (e.g. 314 VD3 is open), but also, in some instances, required more complex monitoring (e.g. The pressure in the reactor containment is increasing) and interpretation tasks (e.g. What 327 lines are pumping water into the RPV?). There were 37 different questions presented in this task – some of the questions were
presented only with a specific picture but others had more than one instance where different status in the same picture were presented.

![462 VD19 is in automatic mode](image)

Figure 2. Example of a question presented in the data collection app

**Eye tracking equipment**

SensoMotoric Instruments (SMI) develops the eye tracking glasses used in this study. It uses dark pupil tracking technology to track eye movements and takes 60 samples per seconds (60 Hz). The system consists of a wearable eye tracker (glasses) and a recording unit (modified Samsung note smartphone/portable computer). The SMI eye tracking system has automatic parallex compensation and provides binocular eye movement data in both real-time and recorded for later observations. The recorded eye movement data is stored in the recording unit and transferred to a computer through USB for later analysis. “BeGaze” is the software tool from SMI that is used in this study to analyse the eye movement data.

![Figure 3. SMI eye tracking glasses 2 with recording unit](image)

**Procedure**

The study set-up had one main independent variable – the type of interface - with two conditions – innovative and control. We analysed whether this variable would have an impact on the operators’ accuracy, response time, fixation counts, dwell time and average fixation duration.
This study was performed jointly with a full-scale simulator study. The participants had been working with the innovative interfaces for 3-4 days at the time of the study, including 6 hours of training. Data was collected simultaneously for each crew. The participants sat in independent stations and were instructed to perform the task individually using only the screen in front of them to see the stimuli and the tablet to answer the questions. In the beginning of the study the researchers asked one of the participants to volunteer to wear the eye tracking device during the task. After that, the eye tracker glasses were calibrated, some basic instructions regarding the eye tracker were given, and the participant was informed that if the device became uncomfortable it could be removed. The researchers then explained the task to the participants and conducted a training trial where all possible types of questions/answers were presented. Figure 4 shows the set-up for the experiment.

Each participant responded to 3 blocks of 40 questions each, separated by short breaks (approximately 5 minutes). In each block the participants were presented with innovative and control pictures of the interface in a pseudo-randomized way so that the same question/picture pairing would never be presented consecutively. Whenever the participant that was wearing the eye tracker glasses took a break between blocks, the eye tracker would be re-calibrated. The overall duration of each block was of 10-15 minutes. All the questions and blocks were randomised. The questions were synchronised with the main stimuli presentation so that everytime that the participant swiped the tablet to the following question a new, randomised display picture would show up in the main screen.

![Figure 4. Study set-up](image)

**Results and Discussion**

**Performance data**

The performance data was considered for questions and instances of questions that allowed a paired comparison between the interface conditions. The first phase of the analysis covered the accuracy and response time averages for the 9 participants in the microtasks. Figure 5 shows the overall accuracy (left) and response time (right)
obtained for the innovative and control conditions. The statistical analysis showed that the operators were on average better at answering the questions for the innovative condition, \( t(8) = 3.5, p = .007 \), and that the differences in the response time, showed a marginally significant result, \( t(8) = -1.9, p = .09 \), tending to benefit the innovative condition.

\[
\begin{align*}
\text{Average Accuracy} & \\
\text{Innovative} & = 0.89 \\
\text{Control} & = 0.85 \\
\text{Average Time [seconds]} & \\
\text{Innovative} & = 15.8 \\
\text{Control} & = 17.2
\end{align*}
\]

*Figure 5. Average results for accuracy (left) and response time (right)*

From this analysis it was possible to identify the questions that showed larger differences for accuracy and response time. The targeted questions for the eye-tracker analysis corresponded the questions were the accuracy scores showed a difference of at least 20% between conditions (6 questions); and where the time difference was equal or larger than 5 seconds (7 questions).

Three questions corresponded to these criteria: Q1: *The cooling function 316-322-721-712 is OK in sub (A/ B/ C/ D)*, Q8: *462 VD19 is in automatic mode (True/ False)*, and Q9: *712 PDI has reduced flow (True/False)*. For Q1 there were two separate instances of the questions (presentations of the same display pictures in different satus) – in the detailed analysis they will be differentiated as Q1a and Q1b. The participants responded to all occurrences of these questions, and only one occurrence had a “Don’t Know” answer – for the present analyses, this entry was removed from the data set. This fact illustrates the overall tendency of the participants to respond to all questions avoiding the use of the “Don’t know” button in most instances.

**Eye tracking data**

Considering that the study sample for the eye tracking data is very small, the analysis will focus on descriptive comparisons of the metrics. For this pilot study this information was thought to be able to generate insightful information and allow a more informed interpretation of the performance data. Three eye tracking metrics were chosen for comparing the performances on different conditions. *Fixation count, dwell time* and *fixation duration* since they are the measures for search and processing performance (e.g. Poole & Ball, 2006). We decided to use the three
measures, trying to explore the overall time the participants would spend looking at the process screens (dwell time); the number of different points they would focus on before answering each question (fixation count) and the time they would be attending to specific points within the process screen fixation time, and not just exploring it. We defined the main process screen had an overall area of interest for the analysis, since the participants had to search for the information in the overall screen for all questions.

Each of the three questions selected from the performance data was presented in both conditions. Q1 was presented twice (different status), amounting to a total of 4 questions answered by 3 participants wearing eye trackers. Figure 6 shows the average accuracy for the eye tracking participants in each condition, together with the averages of fixation duration, dwell time and fixation counts.

![Figure 6. Average accuracy (a), fixation duration (b), dwell time (c), and fixation count (d)](image)

In figure 6a it is possible to see that none of the participants in the eye tracking trials was able to correctly answer question Q1b in any condition. Other than that, the pattern for the other questions seems to correspond with the averaged performance data with the 9 participants in the sample: the innovative interface condition seems to enable better accuracy. The average fixation duration (figure 6b) looks similar for
both conditions. However, there seems to be a clear difference in dwell time (figure 6c) and fixation count (figure 6d) between the conditions.

Considering that the accuracy data is equivalent for question Q1b, it seems natural that this question is not differentiable in any of the eye tracking metrics. Three out of four questions had less dwell time (overall time spend staring at the process picture, including both saccades and fixations) and fixation count (number of times the participant was attending to a specific aspect in the process picture – gaze is stationary) in the innovative design. The control condition for question Q1b had a missing data point and it is likely that it influence the averages for question Q1b. A qualitative analysis into the individual performances of operators shows that 8 out of 11 questions had less fixation counts and dwell time in the innovative design.

Fewer fixation counts are expected to correspond to more efficient search patterns (Cooke, 2006). As such, it appears that the operators managed to better find the information in innovative than control displays. The results for Q9 might be particularly interesting, since the information was not available in the control condition and still all the participants responded to the questions, not using the “don’t know” button, meaning that all got a correct response in the innovative condition where the information was available, but all got a wrong response in the control condition. The overall fixation time did not capture this distinction, but the metrics in figure 5c and 5d are congruent with this, showing higher dwell times and fixation counts and also more variance for the control condition. This pattern of response needs further exploration in future studies since the participants seem to be more willing to provide a wrong answer than a “don’t know” answer. Nonetheless, this behavior might be explained by a series of factors, for instance the nature of the current tasks that is quite different from the usual way of work in a control room where there is no immediate time pressure and the operators are encouraged to take their time to analyse, interpret, and decide on a required action - “don’t know” is not an available option and might explain why most participants ignored the button in most trials.

Conclusions

The objective of this study was to explore the usefulness of eye tracking methodologies within the context of interface assessment for nuclear process control. The data on performance and eye tracking showed that eye metrics can contribute to the interpretation of the results and understanding of the performance patterns. The participants were able to wear the eye tracking comfortably throughout the whole study and reported that it was not obtrusive nor limited their task, which is a relevant feature of the equipment. The eye tracking glasses were easy to use for the participants and easy to set-up for the researchers: calibration can be performed in a few minutes and it can be corrected if needed during the data collection or afterwards, making it a robust tool in applied settings. The instructions to participants and the preparation were simple and the analysis was straightforward, providing objective data.

Nonetheless, this study has noticeable limitations. The eye tracking equipment was used as an add-on in a previously determined set-up where the participants used two
independent screens (main stimuli display and response tablet), which meant that often the participants moved their eyes and not their heads between screens, decreasing the discriminability of the eye movements and targets. Future studies should consider the optimal set-up for an eye tracking study and conduct a pilot test to assure sufficient quality of the eye tracking data collected. Also, eye tracking data were collected only from three out of nine participants and represented a sub-set of the whole database. The participants were also more familiar with the innovative displays since the training on the main simulator study focused on these types of displays, and probably led to higher needs of verification in the control condition, checking component labels and system numbers (the control displays were copies of the innovative displays where visualisation features were replaced by numerical values).

Regardless of the current limitations, the authors consider that eye tracking techniques might be particularly useful in well-defined contexts where access to the studied population is particularly restricted such as in process control studies within the nuclear industry. Here, the eye tracking data allowed deeper analysis of a sub-set of data to better understand how the participants used the different displays. Eye tracking is a valuable tool to obtain large amounts of objective data in relatively short periods on how the participants interact with a particular interface. This can be valuable when the target group is only available for a couple of hours in an interface study. Another significant advantage of eye tracking is that it provides both quantitative and qualitative data that can be analysed and recovered to interpret specific events, patterns, and individual results, contributing to a sorting of the outliers in the data set. Designing interface studies for eye tracking methods can enable the optimisation of both the amount and quality of the obtained data. In the current study, the use of eye-tracking in a simplified tasks allowed us to explore the search patterns of the operators while answering specific questions and looking for specific information in the displays – this corresponds to a unique opportunity to see how the innovative features can be advantageous or not. Moreover, even though the performance data regarding response times was not able to show a significant difference between interface conditions, we were able to notice that the control conditions tends to have a larger variability in dwell and fixation times.

One of the most acclaimed capacities of eye tracking data relates of course with its link to cognitive processes and its potential to establish an objective connection with concepts such as workload. The study of pupilometry as a way to assess workload is quite promising and has presented significant developments in the past years (e.g. Marshall, 2002). Pupilometry is a technique centred in the study of the variations in pupil diameter in relation with task difficulty or workload at any given moment. This seems to be a quite robust measure (e.g. Alnæs et al., 2014) and current research efforts are broadening its use beyond the laboratorial settings. There are two main advantages of this technique in comparison with the more traditional questionnaire-based workload measurements (like for instance NASA-TLX): 1) it allows online data collection, meaning that you can have live workload assessment while the participants are performing the task, contrary to the questionnaire approach where the participant has to recall and estimate his/her average workload throughout the task or at specific moments during the task and 2) it is a direct or
objective measure of workload, not relying on the participants’ interpretation of what a high/low workload is at any given moment. This is a topic that has particular interest within the nuclear context and it will be pursued in future dedicated eye tracking studies.

Acknowledgements

The authors would like to thank Håkan Svengren for his contribution in selecting the scenarios, questions, and displays presented in this study; and Aleksander Lygren Toppe and Hans Olav Randem for their work developing the tablet app used for the data collection.

References


