Høgskolen i Østfold

Master Thesis

Nuclear control operator simulator training: How eye tracking can be used to support training in a simulated environment

Jens-Patrick Langstrand

18.05.2016

Master in Applied Computer Science
The Faculty of Computer Sciences

Abstract

Working as a nuclear power plant (NPP) control operator is an important job, which requires costly and time consuming training in the use of safety critical systems. Since NPP control room simulators are used extensively for training it would be of great value if new technologies could be integrated to increase the efficiency of the operator, to better the learning outcome of the training, or to improve the safety of the operating environment. A technology that has shown promise in the field of training is eye tracking. Eye tracking opens up the possibility of registering where and how users focus their attention. This information can then be used to support trainees by giving them visual feedback, which is a starting point of this thesis. The thesis begins by providing a deeper understanding of the eye tracking technology by performing a review of literature focused on the use cases of eye tracking, best practices, and recent applications of the technology.

To explore how eye tracking can be used to support NPP operators a design study was performed. A prototype system integrating eye tracking and static information heavy display screens from the NPP simulator used at the OECD Halden Reactor project was implemented and evolved through pilot testing. A task based on monitoring process parameters, which is a part of NPP operators job, was created and used to evaluate the system. The goal of the task was to check numbers on an information heavy display screen and make sure that the numbers were inside specified safety ranges, if they were not they had to be marked.

The aim of this thesis is to explore how eye tracking can be used as a supportive technology during the number checking task. Using the location of the user's gaze the system is able to register which numbers the user has looked at, and provide visual feedback in the form of highlights. Three different supportive concepts were designed and implemented to assist the user with the task, "Highlight and Disappear", "Highlight Missed" and "Heat Map". The supportive concepts along with a baseline condition were tested with sixteen participants; seven employees at IFE and nine students from Østfold University College. Quantitative and qualitative data were collected which showed that two of the supportive concepts were able to support the participants, "highlight and disappear" increased the efficiency at which the task was completed, and "highlight missed" increased the participants' confidence during the task. The "heat map" concept was unable to directly support the participants during the task, but the generated heat map could still be used as a training feedback tool by highlighting areas that need more attention.

In conclusion eye tracking has been shown to be useful for supporting users during a number checking task. During the development and testing process several difficulties were encountered, these difficulties include the visualisation of the user's gaze point and gaze interaction problems, among other things. The requirements to solve or avoid the difficulties are described in the thesis. This thesis offer guidelines for how eye tracking can be used as a supportive technology for nuclear power plant control room simulation training and similar information heavy domains, and as such contributes to this field of research.

Keywords: Eye tracking, Simulation training, Nuclear power plant, Visual cues, Interface history

Acknowledgements

I would like to thank my thesis supervisor advisor Susanne Koch Stigberg, associate professor at the Faculty of Computer Sciences at Østfold University College and co-supervisor Michael Hildebrandt, senior human factors scientist at the Institute for Energy Technology, for their valuable feedback on this thesis. They were always available whenever I had a question about my research or writing. They consistently allowed this thesis to be my own work, but steered me in the right direction whenever they thought I needed it.

Finally, I must express my gratitude to my girlfriend for taking care of our baby and continually supporting me throughout the process of writing this thesis. This accomplishment would not have been possible without you.

Content

List of Tables

Table of Figures

1 Introduction

1.1 Motivation

This thesis is done in cooperation with the Institute for Energy Technology[1]; IFE is an international research foundation for energy and nuclear technology. IFE is also responsible for executing the OECD Halden Reactor Project[2] with the goal of generating key information in areas such as extending fuel utilization, degradation of core materials and Man-Machine systems. This thesis is focused on the latter, more specifically in the domain of human factors and man-machine interaction in NPP control rooms.

Working as a nuclear power plant (NPP) control operator is an important job, which requires costly and time-consuming training in the use of safety critical systems. Since the 1970's, NPP control room simulators have been used as an integral part of the training of control operators, aiding in the increase of operational and safety performance (IAEA, 2004, pp. 2 & 7). Over 1,000 hours yearly are spent running simulations in most NPP's, in some cases the time spent running simulations can be as much as 24 hours 7 days a week, more time is generally spent on the simulators if they are on site (IAEA, 2004, p. 4). The use of simulators results in a greater control over training, allowing dangerous situations to be reproduced safely so the trainee can learn how to overcome them and prepare for them ahead of time. In addition, tasks the trainee has difficulties with can be targeted and exercised with a simulator, increasing the effectiveness of the training.

Since NPP control room simulators are used extensively, it would be of great value if new technologies can be integrated to increase the efficiency of the operator, to better the learning outcome of the training, or to improve the safety of the operating environment. A technology that has shown promise in the field of training is eye tracking.

1.2 NPP control room simulators

The NPP control room simulators used for training range from basic principle simulators to full-scope simulators. A *basic principles simulator* can be used to teach general concepts, basic operation of complex systems, and the operation of a NPP. A *part-task simulator* contains detailed modelling of parts of NPP systems, and can be used to partially train for a job or task. Lastly, a *full-scope simulator* incorporates detailed modelling of the systems with the same operator interfaces as in the actual NPP control room. The NPP control room simulator used at IFE is the full-scope variant; see Figure 1.1 for an example display screen. For a more detailed explanation on the various types of NPP control room simulators refer to the International Atomic Energy Agency (IAEA, 2004, p. 2).

[1] IFE, 2015. About IFE. Retrieved from http://www.ife.no/en/about-ife

[2] HRP, 2015. The Halden Reactor Project. Retrieved from http://www.ife.no/en/ife/halden/hrp/the-halden-reactor-project

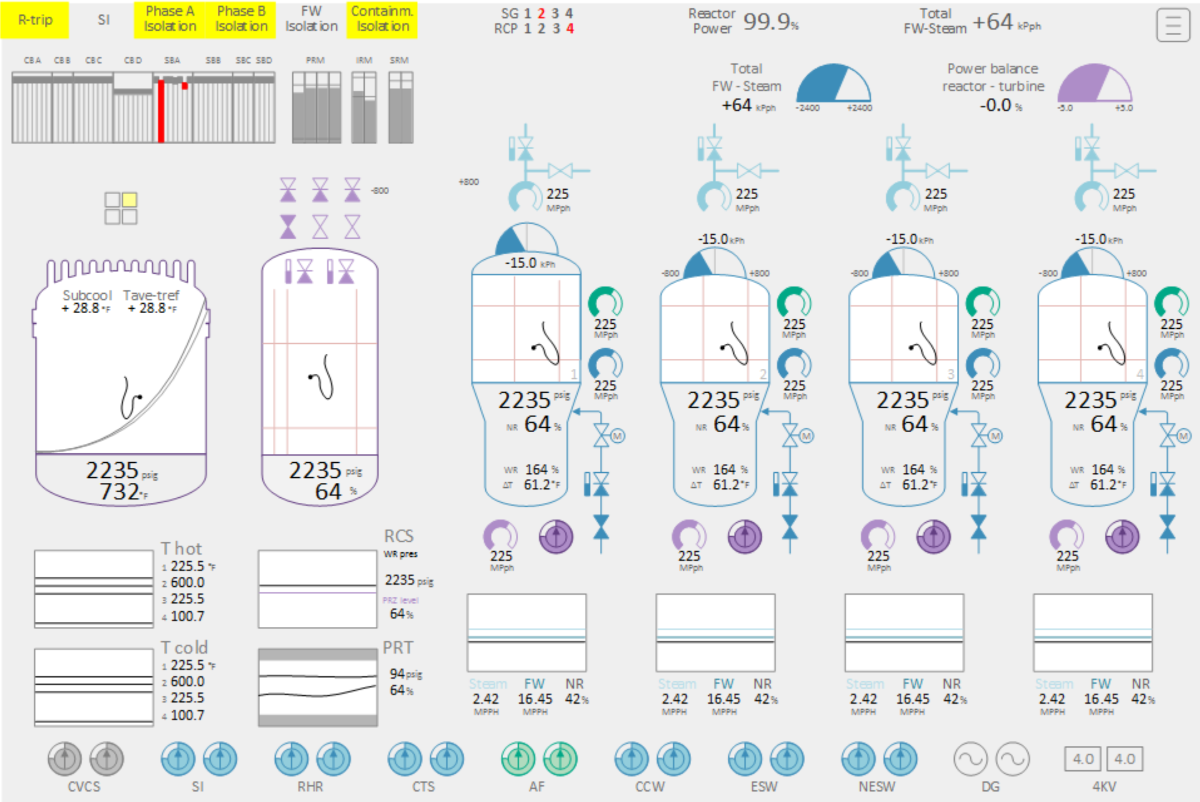


Figure 1.1 An example of an information heavy process display screen from IFE's NPP control room simulator

Training procedures are usually created for each control room operating- or supervisory position (IAEA, 2004, p. 11). In the early stages of training, the control room simulator is often used to familiarise the operator with the locations of important instruments and controls. Next, the simulator is used to demonstrate the operation of systems and components. Training exercises usually begin with demonstrated and coached exercises that involve reactor start-up and shutdown, while increasingly introducing malfunctions that are more complex to develop the skills and confidence of the trainees. A training scenario using the NPP simulator at IFE can consist of a list of tasks that the operator has to perform on a static information heavy display screen. The exercise is procedurally driven through the actions of the operator, an example task can be that the operator has to check the value of a steam pump and if it is above a certain value the operator has to open a valve, or if it is below the value the operator must close a valve.

Eye tracking support

Integrating the eye tracker with the NPP control room simulator makes it possible to know where the operator is looking. This information can be used to determine which components they have or have not looked at. In addition, the viewing order of the components can be registered, revealing the operator's scan pattern. Through analysis of the gathered data a better understanding of the trainees thought process can be achieved which can also improve the quality of the feedback given after a training session, as shown by Renganayagalu et al. (2014) and Sadasivan et al. (2005). Another possibility is to influence the simulation with auditory or visual cues based on the information gathered to assist the operator as shown by Booth et al. (2013).

The two approaches have different implications for changes in the simulated control room, using the data to give feedback after the exercise does not affect how the training is normally performed and is closer to a realistic work situation. The other approach uses the gathered data to guide the trainee during training, using for example generated cues based on the trainee's actions and gaze patterns. The prototype system used both forms of feedback to support the user. The goal was to find and test supportive concepts using the eye tracker that can be further developed and in the end integrated into the NPP simulator to support NPP control room operators.

In order to evaluate the concepts a simplified task based on one of the tasks NPP control operators perform, monitoring process parameters, was created. The task consists of checking numbers on an information heavy screen and responding to numbers that are outside of specified ranges by marking them.

1.3 Research questions and methodology

The study has two research questions:

> How can data gathered from eye tracking be used to support users with visual feedback during a number checking task?

And:

> What are the difficulties of using gaze based interaction with an information heavy display?

To learn more about eye tracking technology and its uses a literature review was performed. The focus areas were which domains eye tracking has been successfully applied to, what kind of research methodologies are used with eye tracking, and how eye tracking experiments are prepared. In addition, it was of interest to see if any best practices exist that should be adhered to in this study, the evaluation method of prototype systems using eye tracking, and lastly what eye tracking have commonly been used to by other researchers in the last few years. The information gathered in the literature review was used to assist in the creation of a prototype application and a design study.

The design study explored how eye tracking can be used to support users during a number checking task. A prototype system with three different ways to use eye tracking and highlighting as a supportive tool during a number checking task was implemented and tested. A static information heavy display screen from IFE's NPP simulator was used as the visual stimuli. The usage of an information heavy display screen could introduce complications due to the accuracy of the eye tracking system. The prototype system was tested to determine if a good accuracy could still be achieved.

In-house pilot tests were conducted to get feedback on the prototype system and to determine if further development was required before the final testing of the prototype system. See Figure 1.2 for an overview of the project structure.

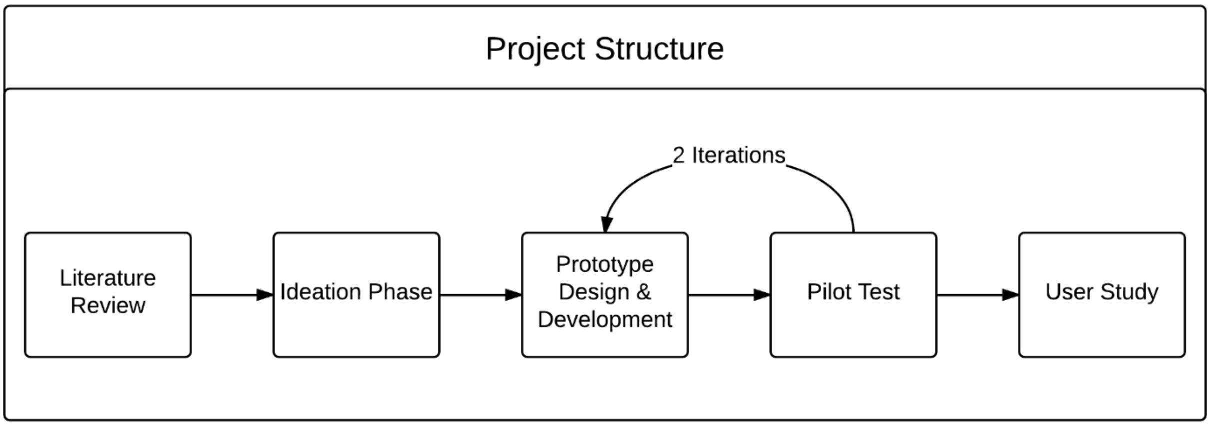


Figure 1.2 The structure of the project, showing the different stages of the design, development and testing of the prototype

The testing of the prototype system was performed with 16 participants, the number was chosen due to counterbalancing and the number of supportive concepts. The participants tried the different supportive concepts, and were given two questionnaires after the completion of each concept. The System Usability Scale (SUS) and the NASA Task Load Index (NASA TLX), both questionnaires were modified to better fit the prototype system. Since the prototype system is not compared to other systems the modifications of the questionnaires makes it easier to compare between the different supportive concepts. The SUS was used to evaluate how the participants found the supportive concepts. The NASA TLX was used to evaluate how difficult the participants found the tasks during the different supportive concepts.

After the test was complete, a semi-structured interview was given to learn what the participants thought about the different supportive concepts and their thoughts about eye tracking in general. The goal of the design study was to test different supportive concepts and see which concepts the participants preferred. In addition, it was desirable to see if the measured data showed that any of the concepts were able to support the participant by increasing their efficiency, learning or make them feel more confident during their task. If one or more of the concepts prove useful they will be further developed and possibly integrated into the actual NPP simulator to support real NPP operators.

To summarize, the purpose of this thesis was to explore the use of eye tracking as a supportive technology for users during a number checking task. The study was conducted using a static screenshot from the NPP control room simulator developed by the Institute for Energy Technology (IFE), for use with the Halden research reactor (HRP). At this stage in the study, different approaches to how eye tracking can be used to support the users was explored, implemented and tested. Evolutionary prototyping was used together with testing to improve the prototype in three iterations. The end goal was to use eye tracking to find ways to support the NPP operator during simulation training, to increase their efficiency, learning and confidence.

1.4 Thesis outline

The remainder of this thesis is structured as follows:

In chapter 2 our eye movements and how the eye tracking technology functions is explained. In addition, the chapter contains a review of literature on eye tracking technology, focused on the use cases, best practices and recent applications of the technology.

The design and implementation process of a prototype system integrated with the eye tracking technology is described in chapter 3. The process started with an ideation phase of how eye tracking can be used as a supportive tool. A number-checking task was created and concepts that support the user with highlights during the task were developed. The supportive concepts were tested in two pilot tests.

Chapter 4 describes the testing procedure, the participants and how the testing of the prototype system was set up. The study used repeated measures and the order the supportive concepts were given to the participants was counterbalanced to limit the suspected learning effects.

The results gathered from the logging of the prototype system, the Nasa Task Load Index and the System Usability Scale questionnaires, as well as the semi-structured interviews are presented in chapter 5.

In chapter 6 the quantitative and qualitative results are discussed and related to the research questions. The usefulness of the supportive concepts is investigated and the difficulties of using gaze interaction with an information heavy display screen are listed and discussed.

Chapter 7 concludes the thesis by answering the research questions through the findings of the study. In addition, the subject of the usability of the eye tracking technology with information heavy display screens is discussed. Lastly, the possible directions for future work on the prototype system and the supportive concepts are described.

2 Background

The way our eyes move and how the eye tracking technology functions was researched to better understand the requirements to create a prototype system that integrates the eye tracking technology. In addition, a literature review was performed to learn more about the use cases, best practices and recent applications of the technology.

2.1 Eye movement and eye tracking

Our eyes use four different types of eye movements, *saccades*, *stabilized fixation*, *smooth pursuit* and *vergence*. The most dominant eye movements are saccades and stabilized fixation. Saccades are fast movements which changes the direction of the eyes towards a new point of interest, while fixations are the periods of time where the gaze remains almost motionless between saccades (Land and Tatler, 2009, p. 13). Fixations allow the intake of information while saccades quickly direct the eyes towards a new point of interest, see Figure 2.1 for an illustration of saccades and fixations while reading. A third eye movement called smooth pursuit allows a small object to be tracked if it moves slowly enough, as it speeds up saccades will be used to support the pursuit, and if it is moving faster pursuit will be completely replaced by saccades (Land and Tatler, 2009, p. 23). Vergence is the last movement type, it alters the angle between the eyes so that they can converge along the path of the point of fixation (Land and Tatler, 2009, p. 24). The detection of these eye movements can be achieved with the help of eye tracking technology.

Figure 2.1 Example of tracked saccades and fixations while reading. The circles represent fixations and the lines represent saccades.

Several eye tracking technologies are available for consideration and the method to track the users' eyes varies with the technology. Four main categories can be used to describe the methodologies of eye tracking, Electro-OculoGraphy (EOG), Photo-OculoGraphy (POG) or Video-OculoGraphy (VOG), and video-based combined pupil and corneal reflection (Duchowski, 2013, p. 51). EOG measures the movement of the eyes by placing electrodes near the eyes. The measurements are relative to the head and are therefore generally not suited for measuring the gaze point of the user, unless combined with head tracking (Duchowski, 2013, p. 52). POG and VOG measures the distinguishable features of the eyes under rotation or translation, such as the shape of the pupil. The POG and VOG methods are not suited for point of regard measurements and many of them require the head to be fixed, using for example a chin rest (Duchowski, 2013, pp. 53 & 54).

EOG, POG and VOG are all suitable for measuring eye movements, but to measure the point of regard the head either has to be fixed or multiple ocular features has to be recognised (Duchowski, 2013, p. 54). This is required to separate head movements from eye movements. One way to achieve this without the need for head restraints is via video-based combined pupil and corneal reflection, which uses the ocular features corneal reflection and the pupil centre to measure the eye movements (Duchowski, 2013, p. 54). For the prototype system, video-based eye tracking with combined pupil and corneal reflection will be explored as it does not require the user's head to be restrained and is suitable for the acquisition of the point of regard.

Holmqvist et al. (2011, p. 51) gives an overview of the three types of video-based eye trackers, static eye tracker, head-mounted eye-tracker, and the head-mounted eye-tracker combined with head-tracking. The static eye trackers place the illumination and the eye cameras in front of the participant, or on their heads. They are dividable into two sub groupings, tower-mounted eye trackers that are close to the participant and restrain the head movements, and remote eye-trackers which can be located underneath a monitor in front of the participant.

The head-mounted eye tracker places the illumination and eye cameras on the participant's head mounted on a helmet or a pair of glasses. A scene camera is used to record what the participant sees. The third type of eye tracking combines position tracking with the head-mounted eye tracker, which enables increased accuracy due to the location of the head being known. See Figure 2.2 for example images of the different eye tracker types.

Figure 2.2 Images of three video-based eye-trackers produced by SMI (SensoMotoric Instruments).From left to right: Tower-mounted eye-tracker, remote eye-tracker attached to a monitor, head-mounted eye-tracker in the form of eye glasses.

The tower-mounted eye-tracker provides a high accuracy and precision as a result of the participant's head being restricted, and is suitable for use with a monitor (Holmqvist et al., 2011, p. 53). Head mounted eye tracking has the advantage of allowing the participant to be mobile (Holmqvist et al., 2011, p. 54), and is suitable when there are more than one screen to monitor or if the participant needs to walk to different consoles during tasks. The combined head-mounted and position-tracking system has a higher accuracy and makes it possible to automate the data analysis process due to the location of the participant's head being known (Holmqvist et al., 2011, p. 54). The remote eye tracker does not require the user to wear anything on the head and does not restrict the participant's head, but the data quality can be lower than with the other eye tracking types.

The remote eye-tracker will be used for this prototype system even though it can have a poorer data quality, the fact that the participant does not have to wear anything and that the tracker does not affect how the participant works outweighs this limitation. Remote eye trackers are suitable for user-interface studies (Cantoni and Porta, 2014), easy to operate and the participants tend to forget that the tracker is there (Duchowski, 2013, p. 53).

Infrared eye cameras are used to capture the participant's eyes in order to avoid natural light reflections. One or more infrared light sources are used to illuminate the participant's eyes. A good view of the pupil and the light reflecting from the cornea, which covers the outside of the eye, is important when utilizing video-based eye tracking (Holmqvist et al., 2011, p. 21). The images captured by the eye cameras are analysed to find the pupil and the corneal reflection caused by the infrared illumination (Holmqvist et al., 2011, p. 25). See Figure 2.3 for an example of what the output of the eye cameras can look like. When the location of the pupil and the reflection are known, it is possible to calculate where the participant is looking or how the eyes are moving.

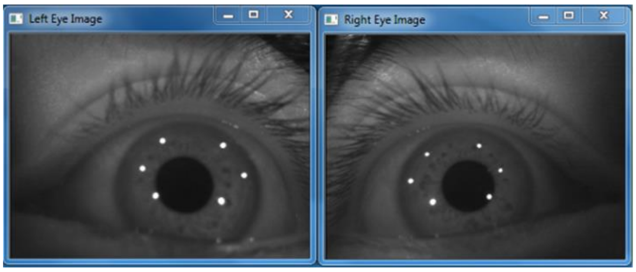


Figure 2.3 Images from the eye cameras of a head-mounted tracker

Limitations

Eye tracking comes with several difficulties, which can result in the loss of collected data due to lack of data quality. These difficulties include the calibration procedure and tracking issues, the quality and flexibility of the equipment, and the complexity of the experiment itself. Factors such as glasses, lenses, how wide the eyes open, the brightness of the iris, downward eyelashes or mascara on the eyelashes can also reduce the precision of the tracking and the quality of the data (Holmqvist et al., 2011, p. 141). Most of these problems can be mitigated by screening the participants leading to very good data, but the experiment then only represents a subset of the population. These are technical limitations that affect the performance and percentage of the population which can use the system, 2% to 5% loss of data from a non-pre-screened average population of Europeans can occur, but the number can get bigger depending on the setting, e.g. lab setting versus outdoor environments (Holmqvist et al., 2011, p. 141).

Even with these limitations, good results have been obtained using eye tracking technology. Renganayagalu et al. (2014) used eye tracking to improve the quality of feedback from maritime simulation training instructors by 43%. Sadasivan et al. (2005) showed the visual search strategy of an expert user to novices before training, which resulted in an increase in accuracy when inspecting an airplane cargo hold for faults. Eye tracking has also been used to give severely handicapped persons the ability to communicate through eye writing (Porta and Turina, 2008).

The purpose of this thesis however is to explore how eye tracking can be used to create supportive systems to assist NPP control room operators in the future. As such, these technical limitations will not be directly addressed in the thesis.

2.2 Literature Review

The literature review was conducted to gain a better understanding of the novel eye tracking technology and its use cases. The focus areas of the review were which domains eye tracking has been successfully applied to, and how eye tracking has been used as a supportive technology. Additionally, it was of interest to discover possible best practices when using eye tracking. Lastly, the recent applications of eye tracking and what has been accomplished was examined.

2.2.1 Literature search

To find relevant literature the conferences, databases and journals listed below were searched with the keywords: eye tracking, simulation, training, nuclear power plant, and control operator.

- The Eye Tracking Research & Applications (ETRA) conference
- The ACM digital library
- The Human Factors: Journal of the Human Factors and Ergonomics Society
- Research paper reference library maintained by Tobii

The papers were selected based on the following criteria: if they showed use cases of eye tracking, contained interface or interaction guidelines, examined ways to use gaze for interaction and as an input device, or showed eye tracking as a supportive technology.

2.2.2 Gaze-based interface guidelines and interaction parameters

When designing interfaces suitable for gaze-interaction there are some aspects that needs considering.

- The accuracy of the eye tracking system
- The eyes movements
- The definition of areas of interest (AOI)

The **accuracy** of the eye tracking system varies depending on hardware and is calculated by taking the difference between the true gaze position and the recorded gaze position (Holmqvist et al., 2011, p. 33). The minimum size of objects to be measured by gaze depends on this accuracy. Using the eyes as input cannot achieve the same level of accuracy as using a mouse cursor, since the eyes are always moving, even when fixating on a point (Cantoni and Porta, 2014). Different sizes for objects of interest are used in the different eye tracking systems, Porto and Turina (2008) used 190pixel wide squares for their hotspots, Putze et al. (2013) used 100pixel size objects based on the manufacturers accuracy specification.

Another important factor is the definition of how **eye movements** are to be categorised and defined in the terms of the gaze-interaction. Blaschek and Ertl (2014) defines saccades as rapid eye movements lasting approximately between 30ms to 80ms. Fixations as eye movements remaining on the same position within a radius of 20 to 50 pixels and lasting around 200 to 300ms. Statistical measures of fixation durations show that they last between 150ms to 600ms (Duchowski, 2013, p. 47). A sequence of fixations and saccades is called a scan path.

It is assumed that the three eye movement's, *fixation*, *smooth-pursuit* and *saccades* provide evidence of visual attention, however the possibility of involuntary movements cannot be discarded. Fixations naturally correspond to the desire to maintain gaze on an object of interest, similarly with smooth-pursuit. Saccades are considered manifestations of the desire to voluntarily change the focus of attention (Duchowski, 2013, p. 47).

Areas of interests (AOIs) are used to define areas of the interface that are to be observed. Using the AOIs it is possible to register when a person is looking at the defined areas. The AOIs can be parts of stimulus, such as the hotspots used by Porta and Turina (2008), or regions of interest in interfaces, for example a button. With AOIs defined it is possible to record the order AOIs are gazed at during a task and compare the order with other participant's results (Blascheck and Ertl, 2014).

With defined AOIs and the ability to register where the user is looking it is possible to create interfaces and games driven by gaze based interaction.

Gaze Based Interaction Examples

Gaze-based interaction is a new method for input that can be used to control user interfaces, the movement of game characters, or real world objects among others. For the prototype system it could be useful to control aspects of the system using gaze based interaction, gaze interactions could be used in the real NPP simulator and work environment.

A gaze-based interface for steering in virtual environments was created by utilizing a 2D overlay with activation regions (Stellmach and Dachselt, 2012). By fixating at the regions a sticky pointer appears which controls the direction and the velocity of a game character, a similar approach is used to control the movement of a character in a maze-game but without a sticky pointer (Krejtz et al., 2014). Another control scheme uses partial gaze-based steering, where the gaze controls the speed and the rotation of a drone, the other control aspects are handled via a keyboard, this shows that gaze-interaction can assist hands-busy operators when paired with other input devices (Hansen et al., 2014).

Gaze-based interaction can in addition to controlling game characters and real world objects, also be used to navigate and activate user interfaces. An application called GazeGalaxy controls a fisheye lens via gaze and a smartphone (Stellmach et al., 2011). The lens is used to navigate a large quantity of images (800) represented by thumbnails displayed on a Tobii monitor. By moving the lens over the thumbnails, they are enlarged, allowing the image to be viewed in full size.

While gaze-based interaction opens up new and exciting possibilities for interaction there are some limitations that needs to be considered before adopting gaze as a method for input. One of the most common problems with gaze-based interaction is the "*Midas Touch"* problem, which occurs when a user accidently triggers an action via gaze. This problem is especially prominent when gaze is used as the only method of input such as the maze-game (Krejtz et al., 2014), where the users felt that they couldn't visually scan the game scene without accidently activating a movement command.

A common way to overcome the "*Midas Touch*" problem is to use dwell-time (fixate on a location for a period of time) as a precondition to trigger an action, but it is time consuming (500ms-1000ms) (Cantoni and Porta, 2014). Porta and Turina (2008) use dwell-time to activate the hotspots for their eye writing application. However, using dwell-time increase the time required to trigger actions, resulting in dwell-time being a limiting factor as well. Stellmach and Dachselt (2012) use a sticky pointer to reduce the amount of dwell-time activations needed to control the character. The pointer itself is activated via dwell-time, but then stays in place allowing the user to visually scan the screen without accidentally triggering actions. The pointer is cleared by glancing at one of the defined stop zones.

Stellmach et al. (2011) implemented two control modes to overcome both the "*Midas Touch*" problem and the need for dwell-time activation. By pressing a button on a keyboard or touching a hotspot on the smartphone, the gaze-control is enabled giving users control over when the gaze is used as input. In the other mode, the users can scan the scene freely without worrying about accidently triggering actions. By combining gaze with additional input devices, the time-consuming dwell-time activation can be removed and the mental workload of the users can be reduced. This solution is also mentioned in the feedback from the users that tried the gaze controlled maze-game (Krejtz et al., 2014).

Most of these experiments were performed in a laboratory setting. Additionally, they consist of proof of concepts rather than field studies. Lastly, many of the studies are focused on the utilization of eye tracking in the gaming domain. Many cases show eye tracking functioning as an interaction device, if the interaction and the interface are designed to overcome the problem of "*Midas Touch*" and the time-consuming dwell-time activations.

From the literature search, it can be determined that there are no specific guidelines for the creation of gaze-based interfaces. Different sizes for AOIs are used throughout the reviewed papers and recommendations for positioning of gaze objects in regards to distance from each other were not found. Interaction parameters such as dwell-time are also different from paper to paper, indicating that such parameters might be application dependant. The numbers discovered in the papers offer some starting points for the development of the prototype but testing will be necessary to find suitable values for the prototype application.

2.2.3 Eye tracking as a supportive technology

How eye tracking has been used to support users and tasks are of interest as it relates to the purpose of the thesis. How others have used eye tracking as a supportive tool is investigated to see if any of the ideas are applicable to the prototype system.

Eye tracking has been used as a supportive technology successfully assisting users with tasks they could otherwise not have performed due to severe disabilities. Eye-S is a Microsoft Word add-on, which enables eye writing. It uses 9 defined hotspots, either hidden or visualized as a 2D overlay, to write letters by gazing at the hotspots in specified sequences (Porta and Turina, 2008). Bulling et al. (2009) use EOG eye tracking to recognize eye-gestures, the gestures are created through sequences of saccades in different directions and used to play a computer game.

Eye tracking can also be used to support activities that require high vigilance. An example activity is video analysis of surveillance footage, where the expert has to visually scan for dangerous events (objects or situations) and manually register them. Putze et al. (2013) combines eye tracking with EEG to automatically detect and tag dangerous events through synchronized data collection and automatic data analysis. The occurrence of an event is detected through automatic analysis of the EEG data, which causes the location of the first fixation after a saccade to be registered from the eye tracking data. Another study used eye tracking to discover eye- and head-based cues, particularly changes in blinking- and saccade-patterns, which may be indicative of individual vigilance levels during a repetitive baggage screening task (Langhals et al., 2013).

These examples show that eye tracking can be used to support people via gaze interaction, and to monitor the state of people working with repetitive tasks that requires a high level of vigilance. Another area where eye tracking has been used as a supportive technology is in simulation training.

2.2.4 Simulation training with eye tracking

Literature on the subject of simulation training supported by eye tracking was examined to gain an overview of what has been studied in this domain, and because it is directly related to the present study. There are two common approaches when using eye tracking to support simulation training, *feedback* and *feedforward*. A feedback simulation-training scenario can be a trainee performing a task in the simulator while an instructor observes and provides help and feedback when needed. When the task is complete, the instructor can give feedback based on what was observed. This type of feedback training is usual in the maritime domain which relies mostly on the instructor's verbal feedback (Renganayagalu et al., 2014).

Renganayagalu et al., (2014) propose a new training method in the maritime domain which incorporates eye tracking as a feedback tool. A live feed of the student's gaze video with the gaze point visualized is shown to the instructor, which allows the instructor to more closely follow the student and provide feedback that is more valuable. The results show that the instructors were able to give 43% more accurate performance assessments and discover bad practices such as over focusing on screens. Eye tracking during training allows the assessment of search patterns and can provide an effective way to assess and correct visual search skills.

With feedforward training, information is given prior to a task or during a task, for example showing an expert user's scan path to a novice before a task is performed (Sadasivan et al., 2005). Cantoni and Porta (2014) claims that novice inspection strategies tends to be more random due to the absence of memory, while expert strategies which tend to be more organised and come from experience. Sadasivan et al. (2005) attempt to use an expert aircraft inspector's scan patch as part of feedforward training, before the actual simulation task is started the trainees can see a simulated aircraft cargo hold with the expert's scan path overlaid on top. The idea is that the novices via the expert's scan path can adapt a systematic inspection instead of the usual random inspection. Results show that the novices that received feedforward training achieved a better accuracy, but at the cost of more time required in the simulator.

The literature search revealed some cases of eye tracking being used for training purposes, but not many. This indicates that while eye tracking is useful for both feedforward and feedback training there is room for more research in both areas of training, this thesis will focus on a combination of feedback and feedforward training. Booth et al. (2013) shows that visual cues helped the test subjects find the required objects faster and with less errors. Due to difficulties in finding relevant literature about visual cues that were useful for the prototype system, the decision to highlight the AOIs as visual cues was made. The prototype system differs from the other cases by utilizing eye tracking to actively influence the simulation, using the knowledge of where the user has and has not gazed to support the user by highlighting AOIs and removing highlights when necessary.

3 Prototype design and implementation

Three phases were visited before the prototype system was finished, an ideation phase, a development phase, and a pilot test phase. The ideation phase focused on how to use the eye tracking technology as a supportive tool. In the process a simplified number monitoring task based on a real NPP operator task were created. In the development phase a prototype system was made as a testing platform to use the eye tracker to support the user during the number monitoring task. During the pilot test phase two pilot tests with two employees from IFE were used to help evolve the prototype based on the participant's feedback, after each pilot test the prototype went back to the development phase. In this chapter the different stages of the prototype development will be explored and the effect the pilot studies had on the prototype's development are brought up for each section.

3.1 Ideation Phase

To begin with, ideation sessions were conducted with a senior human factors scientist at IFE. The scientist has knowledge about the eye tracking technology as well as how NPP operators train and work. The focus of the ideation sessions was to discover how to use eye tracking in a meaningful way. Four ideas were considered, a user interface study, a supportive training tool for novices, a NPP operator support tool, and a number monitoring support tool.

User interface study

One idea was to create a user interface study and test different parameters to learn how to create interfaces suitable for eye tracking. An abstract task and interface would be created where different interface design parameters could be tested. The parameters would be the sizes of the AOIs and the positioning of the AOIs. How close can one AOI be to another AOI without compromising the accuracy of the eye tracking system?

The screen would be divided into segments and populated with numbers, symbols or images of objects, and the user would be given a task such as "Look at all the odd numbers", or "Look at all the images that has objects starting with the letter C". The accuracy of the system would then be measured and used to create guidelines on what ranges the parameter values can be and how a user interface suitable for gaze data collection can be created. The benefits of this approach would be that it would produce a set of guidelines on how to create a gaze enabled user interface, and what the values of different design parameters such as the position and sizes of AOIs, and dwell-time activations should be.

Supportive training tool for novices

Another idea was to train novices to learn real NPP process display screens. Eye tracking would be used to support the novices by helping them find and remember nuclear components such as reactor coolant pumps and steam generators. An explorative phase would be used to let the novices learn about the different nuclear components. The components would be labelled with information and made gaze-intractable so that the novice can activate a component by looking at it. By activating the component, the information about the component would appear. The information would contain the components function and what its safe operation values are. However, this approach is not very valuable as actual NPP operators go through years of training before they even see a NPP simulator, and this idea does not provide a real experience that can lead to the novice understanding how NPPs operate.

This system could also be extended to teach simple scan patterns to the novices. By overlaying components with visual highlights that disappear and reappear at the next component the novices could learn suitable scan patterns for checking different parts of the display screens. The novices' knowledge about the display screens and scan patterns could then be tested and feedback could be given through the eye tracking system. The system could also assist the novices during the test, if the novice takes too long to find a component the system could highlight it. This approach would use eye tracking as an interaction tool as well as a supportive tool. The ability to overlay scan patterns and detect when the novice has looked at the components in the scan pattern could be a valuable tool during training.

NPP operator support tool

Another idea was to use the eye tracking system to assist the NPP control room operators during simulation training. Numbers would be associated with the nuclear components, the operator would then have to check the numbers and make sure that they are within specified ranges. If a number is outside a range, the operator has to open or close a valve. For example, if the pressure of reactor pump one is above the threshold the valve should be opened, if it is below the threshold it should be closed. If the number is within the specified range, do not change anything. The numbers would change over time and at different rates.

The operator would be assisted by the system through visual cues when a number is close to the specified range's threshold, when a number has not been looked at for some time, or if all the valves on a line are open. The cues would disappear when the operator has looked at them, resulting in alarm reduction. So in addition to supporting the operator the system would also reduce the number of active alarms by stopping the alarm when it has been registered that the operator has noticed that a problem is occurring in that area. In addition, the system can determine if the operator has already seen the problem and as a result not give a visual cue, resulting in further alarm reduction. The difference from having a non-eye tracking supported system would be that the system would not know if an operator has been made aware of a situation and as such all alarms would have to stay active until the situation has been handled. This can lead to the operator being overwhelmed with alarms and graphical cues that could make it difficult to focus on the task.

Number monitoring support tool

The last idea was to support the user during a task that resembles one of the tasks NPP control operator performs, to monitor numbers and make sure they are within specified ranges. The task is simplified so that any user is able to control the system. Static screens from IFE's NPP simulator are used as the systems interface, the screens are overlaid with numbers the user has to check. The NPP simulator screens are used to give the task a context, but it could just as easily have been made suitable for other domains, such as air traffic control.

The user is given several screens where the numbers change for each screen. The user's task is to check all the numbers and make sure that they are within specified ranges. If a number is outside the ranges, it should be marked. After all the numbers have been checked, the user presses a button to submit an answer. Here the user would be supported by the eye tracking system using visual feedback. Different ways to use the data gathered from the eye tracking system to provide visual feedback to the user can be tested. This approach has the benefits of allowing experimentation of different ways to use the eye tracker as a supportive tool, while not restricting the users of the system to NPP control operators. This makes it easier to perform testing and simplifies the development of the prototype as there is no need to replicate a proper NPP environment or connect the system directly to the NPP simulator.

Differences and similarities of the ideas

The user interface study would focus on the technical parameters of using eye tracking with an interface, it differs from the other ideas as it is not tied to any domains. The "supportive training tool for novices" and the "NPP operator support tool" ideas are both tied directly to the NPP domain, and as such would require access to the NPP simulator resulting in increased complexity of the development of the application. The "number monitoring support tool" did not need to be tied to any specific domain, but to give the idea a context it was decided to use the NPP domain. It has an advantage over the "user interface study" as it is tied to a domain and therefore becomes less abstract and more immediately applicable. The required functionality to integrate the eye tracking technology is the same for all of the ideas, as such regardless of the choice made the groundwork for the testing of other ideas in the future will be laid.

While there are benefits to every idea the "number monitoring support tool" idea was chosen, the simplified static task makes it easier to develop the prototype system. It is also easier to test the prototype as there is no need for actual NPP operators. In addition, the development of the prototype lays the groundwork for trying other ideas in the future, since the elements required to integrate the eye tracker are the same. The prototype system was created as a testing platform built around the number monitoring task. To be able to try different ways to use the eye tracker as a supportive system is useful to find supportive tools that can be further developed and integrated with the actual NPP simulator in the future.

3.2 Prototype architecture

The prototype system consists of two main parts, the eye tracking data transfer and the eye tracking application; see Figure 3.1 for an overview image of how the different parts communicate. The eye tracking data transfer consists of the SMI eye tracking server which streams the users gaze data, and a VRPN[3] client which connects to the SMI server. The gaze data is combined by adding the estimated gaze location of the users left and right eye and averaging the locations, this results in an estimated gaze point (X, Y coordinate) on the computer display screen. A VRPN server is used to stream the users combined gaze data to the eye tracking application part of the prototype system. A VRPN client connects to the server and passes the gaze data to the application where it is used.

[3] VRPN, 2015. Virtual Reality Peripheral Network. Retrieved from https://github.com/vrpn/vrpn/wiki

To reduce the amount of work needed to create a working testing platform a game engine was used. A game engine provides functions such as time synchronisation, mathematics libraries and graphics rendering tools which makes the development of the prototype system easier, as the focus can be on creating the task and the supportive tools themselves. JMonkeyEngine[4] was chosen due to previous experience with the game engine and because it is open source and released under the BSD license. The game engine is built on top of the Java programming language.

The system uses eye tracking to enable the collection of the user's gaze data. The system uses the SMI REDn Professional remote eye tracker attached to a 23" LG 1080p LED LCD monitor. The prototype is set to run at 1920x1080 screen resolution. The eye tracker is attached to the bottom part of the display case.

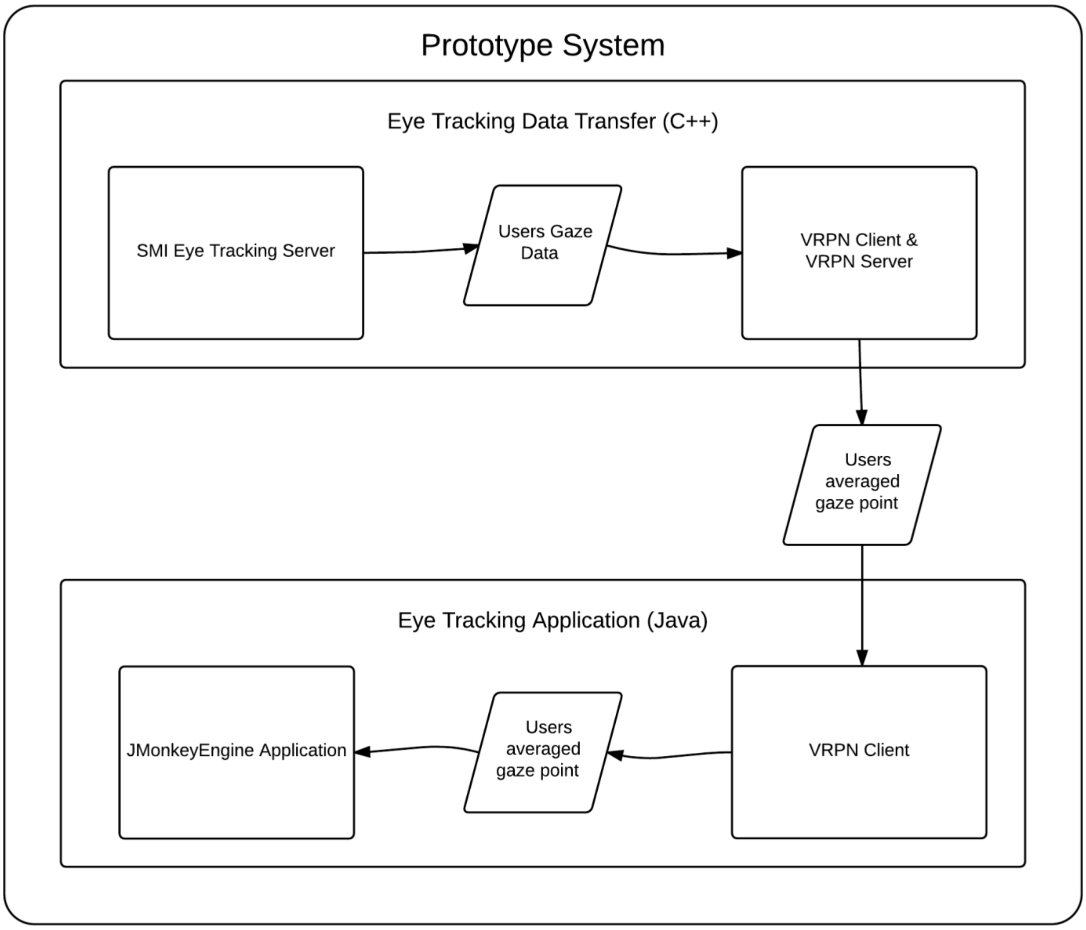


Figure 3.1 Prototype System Overview

The SMI red server communicates with the eye tracker, a C++ server and client is used to connect to the SMI red server and pass the gaze data on to a Java VRPN client. The Java VRPN client is required to use the data from the SMI red server with the prototype application built in JMonkeyEngine.

[4] JME, 2015. jMonkeyEngine 3.0. Retrieved from https://jmonkeyengine.org/

3.3 Prototype design

The prototype system was created as a testing platform where eye tracking can be used in different ways to assist the user. The prototype was developed through an iterative process, with three iterations. Based on the ideation phase, three low-fidelity prototype implementations were developed, and direct feedback on those prototypes was obtained from IFE staff through demoing and discussion. The purpose of this phase was to eliminate concepts that were not feasible, so that the user study could focus on viable studies. Once a set of viable concepts was identified, a pilot study was conducted with two participants from IFE. The pilot system's functionality was close to the system used in the user test. The purpose was to obtain more feedback, but also to test the functionality and reliability of the prototype system. Another function of the pilot test was to obtain a realistic estimate of the study duration, and test the debriefing format, including questionnaires. In the sections below, any modifications done to the prototype from the outcome of the pilot study is mentioned.

Test scenario setup

The prototype was designed to be flexible in terms of changing parameters such as dwell-time activation and the size of the gaze cursor, this allowed us to test different parameter values in real time without having to change the code. The prototype system was built around the idea of a monitoring task; the task partially resembles one of the tasks that NPP control operators perform. Two pilot tests were performed; the participants' feedback was used to further develop the prototype two times before the final test phase.

To give a context to the task the decision to use the NPP domain was made. A suitable image from IFE's NPP control room simulator was found and used as the background image for the task, see Figure 3.2. The image was altered to remove possible confusion from the users, first all the numbers in the image were removed in order to overlay numbers generated by the system. In the process the types of the numbers were removed as well, this was done to simplify the task. Next the SUMP and Water labels were removed from the containment component, this was done because the numbers and labels would be too close together to be able to distinguish them from each other with the eye tracking system. Lastly the numbers in the bottom right corner were moved slightly away from each other for the same reason as above.

The screen is overlaid with 32 numbers, the amount of numbers was selected based on the original amount of numbers in the image, and is therefore representative of a typical information heavy display screen. The users task is to inspect all the numbers, and make sure that they were within two specified number ranges. If a number is outside the specified ranges, it should be marked. The number values that are used for the screen were randomly generated to lie inside the ranges specified. Then a handful of the values were changed by hand to go outside the specified ranges. 8 of the displayed values are out of bounds for each concept. The numbers of values that are out of bounds are the same for each concept to ensure that the time taken to complete each design will be comparable. After all the numbers have been inspected and the user feels confident in the marking of the numbers, the user gives an answer by pressing a button in the top right corner of the screen.

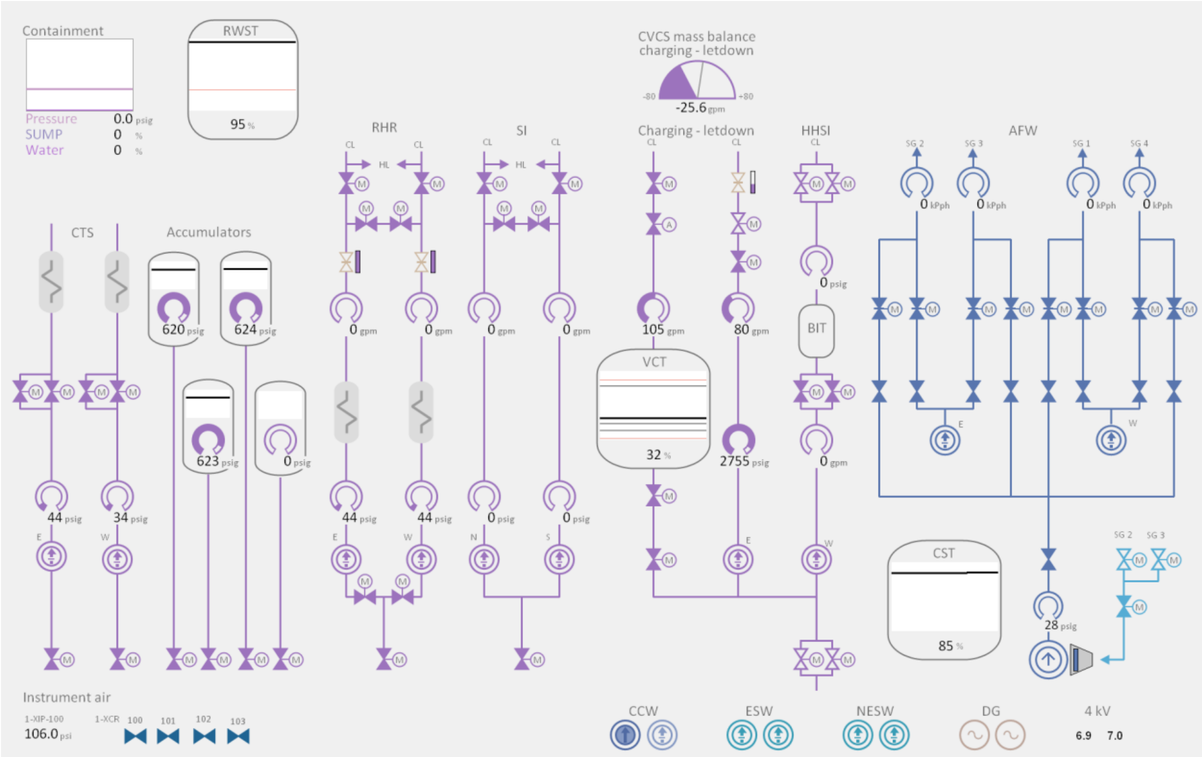


Figure 3.2 The image from IFE's NPP simulator used as the background image for the prototype system.

The numbers the user has to check are randomly generated to lie inside the specified ranges at the start of the application. To ensure that all the users are given the same numbers for each task the random generator is seeded with the same seed every time the application starts. Since the supporting concepts are given in different orders due to counterbalancing the numbers are generated in the same sequence of supporting concepts and tasks, this ensures that every task and concept is the same regardless of the order the concepts are given.

Pilot test results

The user's task was changed after the first pilot test. From the results it was found that it was too easy to be thorough during the task when you had to check all the numbers on the screen. The task was made slightly more complicated by making the numbers be a specific type. Four number types were used, %, psig, kPph and gpm. A real NPP process screen usually contains numbers in different units, so the change also made the task more realistic. 8 numbers of each type are distributed across the screen. The users task is now to check two of the number types and make sure the numbers are within the number types range. Each number type has a different range that is valid for the numbers of that type. In essence two of the number types are targets and the other two number types are distractors.

From the pilot tests, it was found that there is enough information to learn that it was necessary to let the participants have a warmup run before the actual experiment. The warmup consisted of three tasks in a baseline condition, where no support in terms of visual feedback was provided to the participant. The pilot tests were also used to test the written instructions that were prepared for the participants. The participants found the instructions to be clear and thorough. The participants were asked to describe the task and the understanding of the objective before the experiment was started. The participants were able to do this for each section of the instructions manual. The instructions as given to the participants can be seen in appendix A .

The eye tracking technology was used to simplify the task by creating concepts that support the user with highlights during the task.

3.3.1 The supportive concepts

Four different concepts to support the user with the number-checking task was thought of, always highlight, highlight and disappear, highlight missed, and heat map. The always highlight concept and the highlight and disappear concept are feedforward concepts; they provide assistance during the task. The highlight missed concept and the heat map concept are feedback concepts; they provide assistance after the task has been performed.

In "**always highlight**" all the numbers that the user has to check are highlighted, see Figure 3.3. The idea behind the concept is to reduce the amount of information the user has to look at so they can focus more on the task at hand. Initially the task was to check all the numbers, as such the concept highlighted every number on the screen.

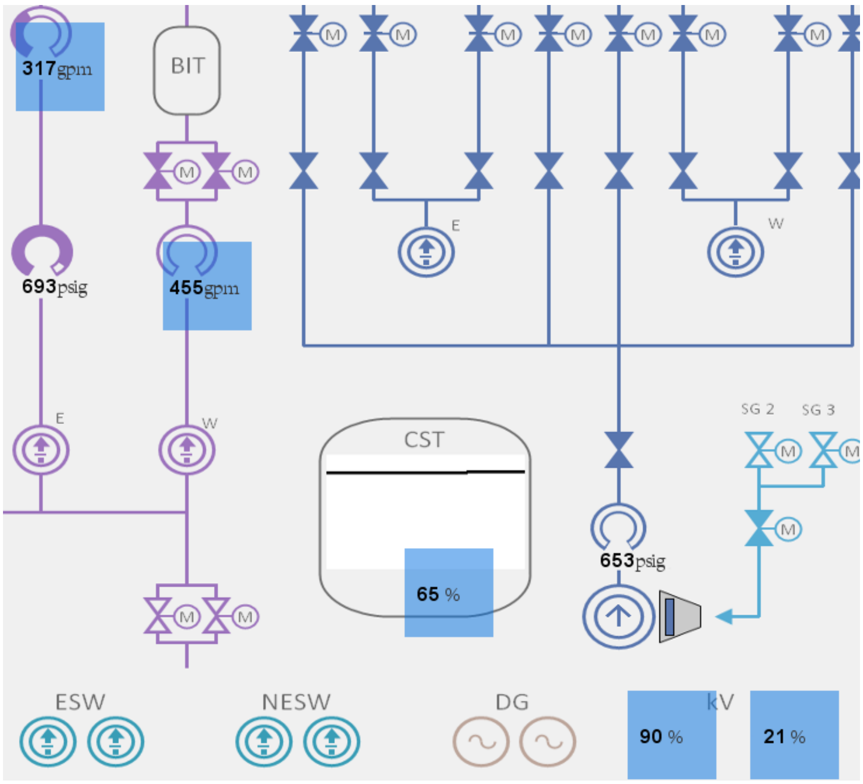


Figure 3.3 A part of the screen during the always highlight concept

The first participant in the pilot study felt the “highlight always” concept was a little annoying, as all the numbers had to be checked and it did not really help that all numbers were highlighted. In fact, it was kind of tiring for the participant’s eyes. The decision to remove the concept from the experiment after the first pilot test was made, the reasons were that it did not use the eye tracking information to provide support and that it was found to be annoying by the pilot test participant. In addition, the removal simplified the counterbalancing of the order the concepts were given to the participants, it also reduced the amount of testers required for the user study.

The “**highlight and disappear**” concept also highlights all the numbers the user needs to check, but in addition it also removes the highlights after the user has looked at the highlighted number, see Figure 3.4. When the blue highlight changes colour to red the system has registered that the number has been looked at, then it disappears after a little time. The goal of this concept is to help the user structure the scanning of the numbers so that they do not have to check numbers more than once.

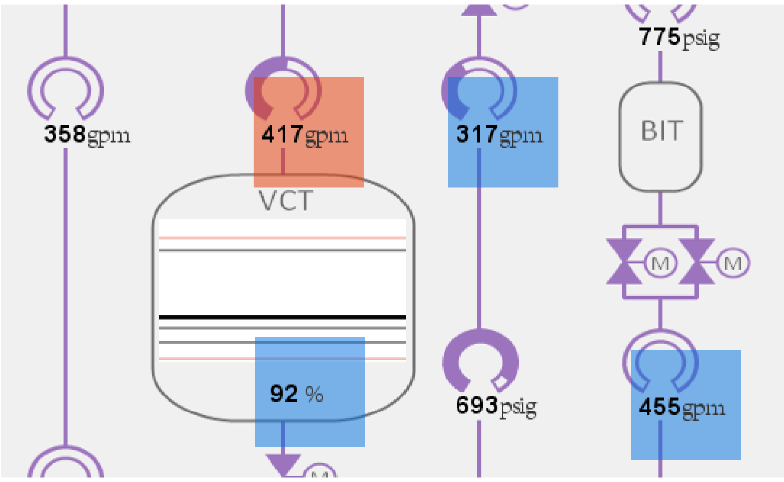


Figure 3.4 A part of the screen during the highlight and disappear concept

In the first pilot test the concept highlighted every number on the screen. The participant found the “highlight and disappear” concept annoying. The highlights disappeared while the participant was scanning the screen which resulted in the participant focusing on the highlights disappearing. In turn this made the participant’s gaze trigger more highlights to disappear causing a chain reaction. Afterwards it was difficult for the participant to remember where the participant was in the scanning process. This concept was the participant’s least favourite.

After the first pilot test the task was changed from checking all the numbers against ranges to checking two of four number types against their own ranges. The “highlight and disappear” concept is made more useful as it now highlights only the numbers the user needs to check, effectively reducing the information the user needs to process in order to complete the task. Even though the change would have made the “always highlight” concept more useful by providing information reduction as well, the decision to remove it from the study was still made, as it does not use the eye tracking information to provide support.

The second participant liked the “highlight and disappear” concept, it made the participant change the search strategy from one number type at a time to dealing with both in parallel. The participant felt more effective during the task. The participant felt the highlights disappeared a little too quickly. As such the time it takes for a highlight to disappear was increased from 1.5seconds to 2seconds.

In the “**highlight missed**” concept the numbers are not highlighted initially. The concept instead gives feedback after the user has inspected and marked out of bounds numbers and pressed the answer button. If the user forgot to inspect a number, it will be highlighted with a transparent yellow square which pulsates, see Figure 3.5. The user then has a chance to inspect the highlighted areas and mark numbers if necessary before giving the final answer. The idea behind this concept is to make sure the user inspects all the numbers.

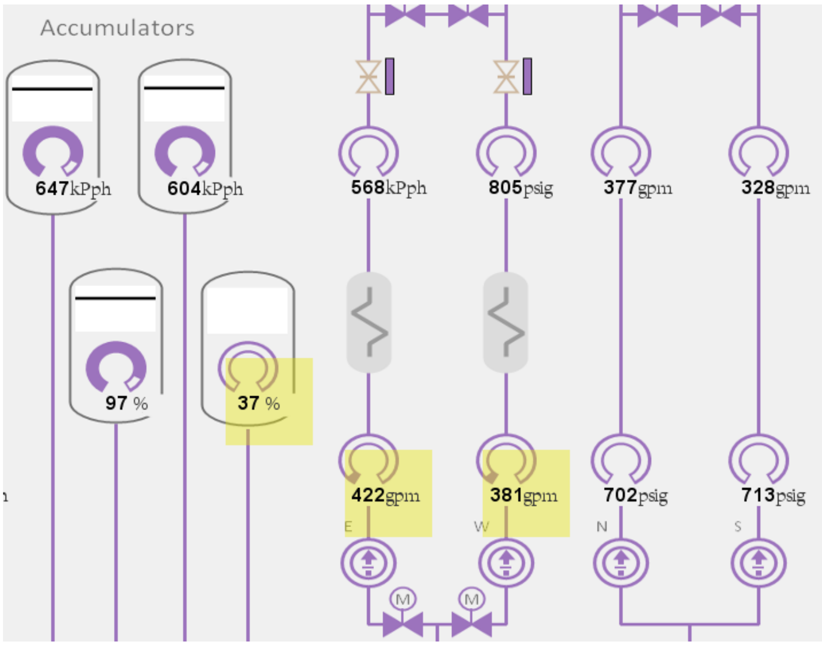


Figure 3.5 A part of the screen during the highlight missed concept

For the first participant two numbers were highlighted in two of three tasks. The participant felt like the numbers had been checked, although it might have been a quick glance. For the second participant no numbers were highlighted throughout the three tasks, the participant was very thorough and checked all the numbers. The participant liked the idea behind the concept and thought it would be useful.

The “**heat map**” concept also gives feedback after the user has inspected and marked the numbers. The concept overlays the numbers with a simplified heat map, see Figure 3.6. The heat map consists of a gradient between green and red. Green means that the number has been looked at for some time while red means that the number has not been looked at or looked at only for a short while. The idea behind this concept is to show the user how they spent their time during the task, and let them spend more time in the red areas.

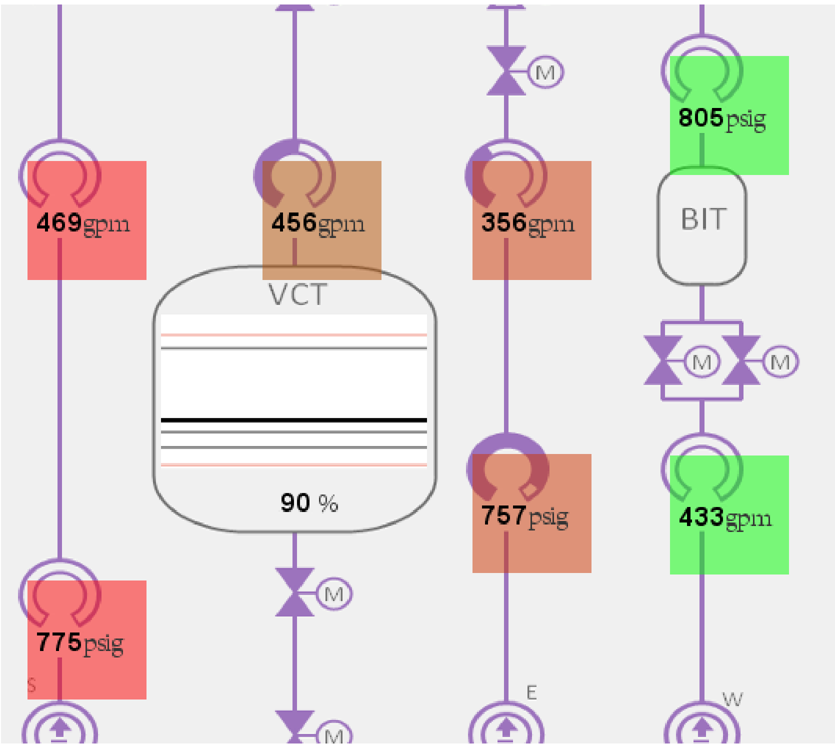


Figure 3.6 A part of the screen during the heat map concept

The first participant liked the “heat map” concept as it gave an overview of how the time was spent while checking the numbers. The participant felt like it would take longer to finish the task since there was so much information to process.

The second participant liked the “heat map” concept. The participant liked the colours and found the concept useful, but the participant did not use all the information the concept provided. If any numbers showed up as green they were ignored. The participant would have preferred the “highlight missed” concept since it provides less information and the information is more relevant.

In addition to the supportive concepts a **baseline** condition was made where the user does not receive any visual feedback from the system and has to complete the task to the best of ability, see Figure 3.7. This condition is used for comparison purposes to get more qualitative feedback as well as having a base condition to compare with the quantitative data. In order to determine if the user solved the task correctly the amount of wrong numbers the user successfully marked was used as a metric, see section 3.4.1 for details on how the users mark numbers.

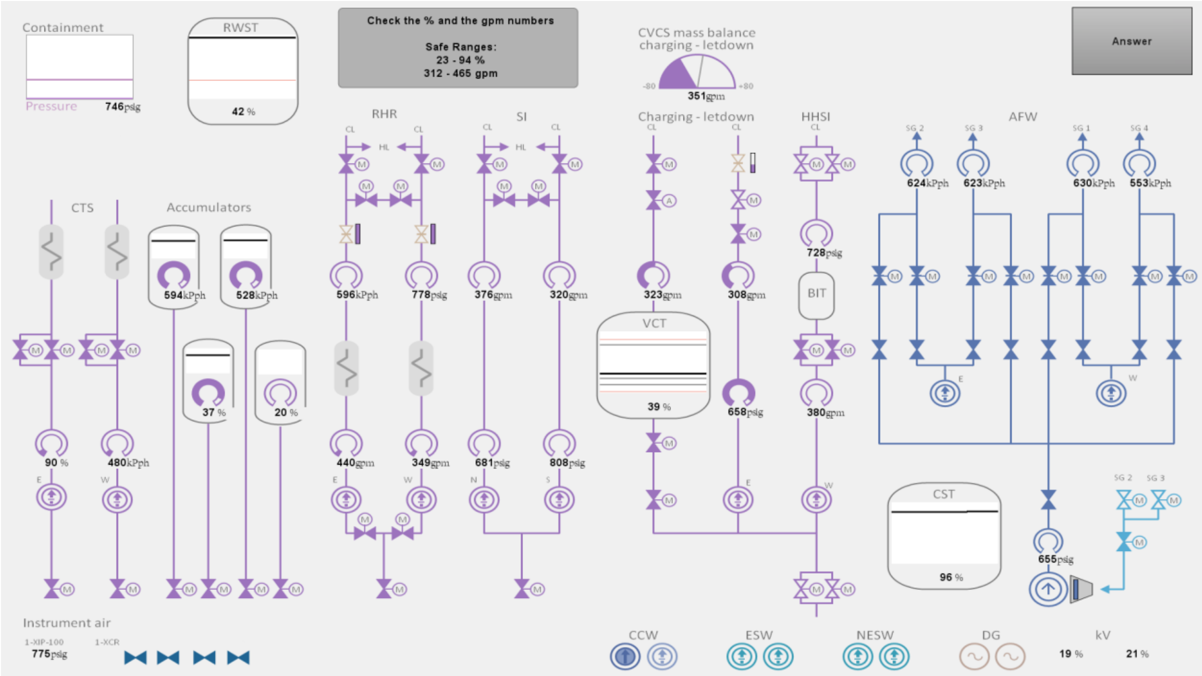


Figure 3.7 The screen as shown in the baseline condition

Table 1 shows the differences between the conditions. The task the user performs can be split into two phases, the task phase and the feedback phase. The task phase is the first time the user checks and marks the numbers on the screen before giving their answer, if the condition does not have a feedback phase the system continues to the next task. If the condition has a feedback phase the system goes into the feedback phase after the button is pressed instead. The feedback provides feedback based on the data gathered during the task phase, it allows the user to double check areas and mark numbers as desired before giving their final answer.

The baseline, “always highlight”, and “highlight and disappear” conditions only have the task phase and moves on to the next task after the user gives an answer. The “highlight missed” and the “heat map” concepts have the feedback phase as well, and provide feedback after the answer button is pressed.

Table 1 The differences between the conditions. Shows what support each condition provides during the task phase and the feedback phase of the task

Condition	Task Phase	Feedback Phase
Baseline	No support	-----
Always highlight	Relevant numbers highlighted	-----
Highlight and disappear	Relevant numbers highlighted, highlights disappear as the numbers are looked at	-----
Highlight missed	No support	Numbers that were not looked at are highlighted
Heat map	No support	The relevant numbers are overlaid with a simplified heat map

To enable the interface to provide the different types of feedback several steps were taken.

3.4 Gaze enabled interface

The first step taken was to define areas of interest (AOIs) that would be associated with the numbers, see Figure 3.8 to see what the defined AOIs look like in the application and their size compared to the number they encapsulate. Every number has an AOI that is responsible for handling the interaction between the user and the number.

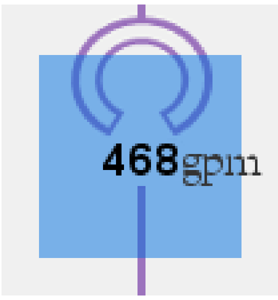


Figure 3.8 Area of interest defined to encapsulate the number

Because no clear direction was found on how big areas of interest should be, the prototype system initially defined the sizes of the AOIs to be big enough to encapsulate the desired area. The AOI size was set to be 38pixels in width and height. After some testing it was found that the size of the AOIs were too small to be able to accurately and repeatedly trigger the activation of the AOIs. The size of the AOIs were then extended by adding some padding around the numbers, the padding was increased by 6 pixels at a time, until the accuracy of the system was high enough to reliably and repeatedly register all the interaction of the Author. The final size of the AOIs were 76pixels when the accuracy was deemed well enough.

Since no clear guidelines for how close the AOIs can be together were discovered, little changes were made to how close the numbers were together except when the numbers were right next to each other. The shortest distance between two AOIs in the system is 24pixels, see Figure 3.9. This distance worked well when testing the system during development, but it could be an issue when a bigger population is used to test the system.

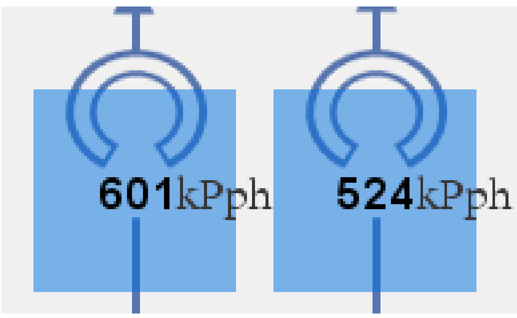


Figure 3.9 The closest distance between AOIs 24pixels.

With the AOIs defined, the next step was to use the position of the user's gaze to interact with the system. The user's point of regard is provided by the eye tracking system as two 2D coordinates on the screen, one for each eye. The position of the right and left eye coordinates were averaged and used as the gaze input for the system. If the user uses only one of the eyes the system uses the gaze point from that eye. When the user blinks, the location of the gaze stays the same as before the blink. This allows the user to interact with the system without worrying about blinking disrupting the flow. This is achieved by filtering out the values that are provided when the user is blinking, otherwise the gaze point would be set back to the top left corner of the screen as 0 is the default value when the eye tracker cannot detect the user's eyes.

The level of accuracy that can be achieved when using gaze as input is lower than the accuracy of a mouse cursor, the reason being that the eyes are unable to stay still when fixating on a point (Cantoni and Porta, 2014). To help make the system more accurate it was decided to change the location of the user's gaze from a point to a circle instead. The averaged gaze location is used as the centre of the circle. The area corresponding to the user's gaze was called the gaze cursor. To determine if the user is looking at a AOI a circle against square collision check was used. First it finds the AOIs closest point, then it checks if the distance from the centre of the gaze cursor is smaller than the gaze cursors radius. If the distance is smaller than the radius a collision has occurred meaning that the user is looking at the AOI.

The gaze cursor was created with a radius of 5 pixels initially, but the value was found to be too small to accurately interact with the system. The problem is that the calibration of the user's gaze is not perfect in all the sections of the screen. The estimated gaze position when the user is looking at the centre of the screen can be very good, but when the user is looking at the edges of the screen or between the edges and the middle the gaze position can be too far to the left, right, top or bottom of the actual location. The gaze cursors radius was incremented by 2.5pixels at a time until a level of accuracy that let the Author interact with the system without any accuracy issues was achieved. A satisfactory accuracy was achieved with a radius of 25pixels for the gaze cursor.

The increase in the gaze cursor size introduced a new source of error. Since the shortest distance between two AOIs is 24pixels and the gaze cursor is 50pixels wide, the gaze cursor can now overlap two AOIs at the same time. One solution was to move the numbers around to ensure that the gaze cursor never overlaps more than one AOI at a time, however the movement of all the numbers would result in an interface that is not representative of an information heavy display screen, therefore a different solution was found. To solve the issue another condition was added to the circle against square collision check. The system now stores the distance from the closest AOI to the centre of the gaze cursor, if more than one collision occurs the AOI that is closest to the gaze cursor is selected to be the active collision.

Using this approach to collision the AOIs can be quite close to each other while still retaining a high accuracy of the system. With the increased size of the gaze cursor and the improved collision check a very high accuracy was achieved when the Author tested the system. The systems accuracy was also tested with two pilot tests. The participant in the first pilot test wore glasses, so it was a good way to test if a good calibration could be achieved, even when the participant is wearing glasses. The participant was able to achieve a good calibration and was able to interact with the system with good responsiveness. There was only one problem area in the upper right corner of the screen where the participant sometimes struggled to activate a button, otherwise the number marking and button activations worked very well. Due to the good results it was decided to not screen participants based on their vision for the actual testing. The second pilot test participant was also able to achieve a good calibration and did not have any issues with the eye tracking accuracy during the test.

Initially the gaze cursor was made visible to the user, see Figure 3.10 to see the two types of gaze cursor visualisations that were used. However, it was found during testing that the gaze cursor had a detrimental effect on the user's task. Since the eye tracking system cannot achieve 100% accuracy the gaze cursor is never located exactly where the user is looking, therefore it is very easy to start trailing after the gaze cursor, as it moves when you look at it. In the end you forget what it was that you were supposed to do and just follow the gaze cursor around. Because of this effect it was decided to keep the gaze cursor hidden from the user and only use it to check the accuracy of the system.

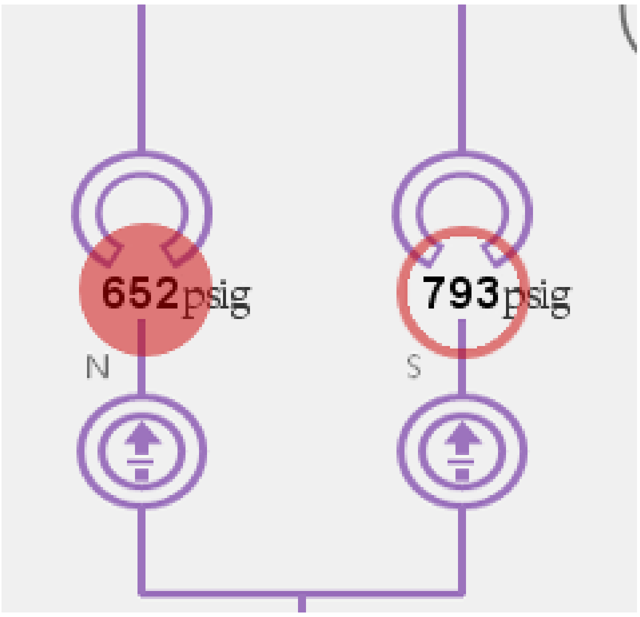


Figure 3.10 Example of the gaze cursors used. The gaze cursors in the image have a radius of 25pixels.

With the AOI and the gaze cursor in place the design of the interaction with the system could begin.

3.4.1 Interaction

The main interaction between the user and the system consists of looking at numbers, marking numbers that are out of range and pressing buttons. One part of the interaction is passive from the user's end, namely looking at the numbers. Here the system uses the defined AOIs assigned to the numbers and the location of the gaze cursor to determine if the user is looking at a number. To avoid the "Midas Touch" problem the prototype uses dwell-time activation before it registers a number as looked at. The value of the dwell-time is set to be between 150 and 600ms (Duchowski, 2013, p. 47) which is a statistical measure of fixation durations found in the literature review. In the end the dwell-time set the value in the lower end of the scale, 300ms, the value was chosen after testing the system. In addition, the prototype uses the dwell-time activations for logging purposes.

The active form of interaction between the user and the system is the marking of numbers and the activation of buttons. While the numbers are activated by dwell-time to decide if the user is looking at them, it was decided to use an approach that does not require the user to stare at the numbers in order to mark them. This could potentially result in problems with "Midas Touch" as well as feeling unnatural in the terms of how our eyes are normally used. Therefore, it was decided to combine the user's gaze with a keyboard interaction to remove the need for dwell-time activation as well as removing the risk of "Midas Touch". To mark a number, the user can look at it and press the "space bar". During in-house testing the interaction method felt quite natural and faster than having to mark numbers with the mouse cursor. To show the user that the number was successfully marked a red circle with a white cross inside it is placed on top of the marked number, see Figure 3.11. The icon was chosen as it is commonly related to errors. If the user incorrectly marks a number, it can be unmarked by repeating the marking process.

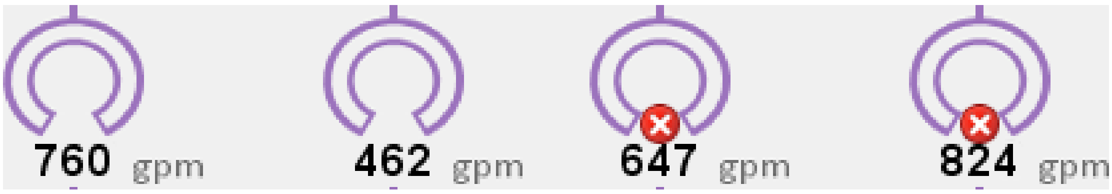


Figure 3.11 A red circle with a white cross indicates that the number has been marked.

To begin with the buttons were dwell-time activated, the user could look at the button and a visualization of the dwell-time activation was started. It was visualized using a red bar above the button that was filled with green as the dwell-time accumulated. It worked well and was fun to use, but from the first pilot test, it was found that it had the potential to trigger the "Midas Touch" effect. Right after the pilot test was started the participant made a quick scan of the first screen, when the participant scanned the gaze enabled button the dwell-time activation visualisation started. This in turn made the participant focus even more on the button, and in the end trigger the button's activation before the participant had time to read the instructions on the screen.

The accidental activation of gaze buttons happened twice in a row and the pilot test had to be started over, the participant was instructed to not look at the buttons before having read the information on the screen. This was never an issue when the system was tested in-house, but since it seemed to have a strong possibility of accidentally making the user activate buttons it was decided to change how the user interacts with the button. Two approaches to work around the problem was thought of, the first was to change it so that the button is activated in the same way that the numbers are marked, by looking at it and pressing the space bar. The second approach was to change it so that the button cannot be activated until the user has looked at the text for at least 2seconds to make sure that the user has actually read the instructions, and can't accidentally activate the button due to the visualization.

In the end it was decided to remove the dwell-time activation and visualisation, replacing it with the same interaction the user performs to mark numbers, looking at the button they want to activate and pressing the "space bar". While the other approach has the added benefit of making sure the user looked at the instructions it is a more complex solution to a problem that could easily be fixed using the interaction element that was already in place. In addition, using the same method of interaction for all parts of the system reduces the methods of interaction the user has to learn resulting in a system that is easier to use.

As a fall back solution every active interaction can also be performed using the mouse, this was decided in order to make sure the application is useable even if the eye tracking systems accuracy is low due to a bad calibration or due to the user having eyes that are difficult to track precisely.

From the pilot tests it was found that both participants really liked the gaze interaction, both felt that marking the numbers using gaze and the space bar was faster and easier than using the mouse cursor. One participant said it was good because there was no need to first locate the mouse cursor and then move it to the number to mark it, you could just look at it and mark it immediately. In this way the participant was able to focus on the number checking task. The other participant felt a need to stare at the numbers and the buttons when interacting, but the participant also thought the need to stare would disappear after using the system for a while.

3.5 Logging

A logging system was made to log the participants' performance during the experiment, the logging system logs the data to an excel file for easy analysis later. For each participant the system logs data such as their completion time and the amount of numbers that were out of bounds that was not marked. For each participant the system creates a new sheet in the excel workbook. A list of the data the system logs that were used for the analysis in the results section can be seen in Table 2.

For each task the system logs the data shown in Table 2, and at the completion of a concept the task data is aggregated to create the concepts data. The completed trial data is created by aggregating the data from the completed concepts data. For easy comparison between the concepts and the baseline condition a summary is created at the top of the log file.

Table 2 An overview of the data logged by the system during the experiment

Logged Parameter	**Parameter Description**
Completion time	The time it took the participant to complete a concept
Wrong values	The amount of out of bounds values the participant did not mark and the amount of numbers the participant incorrectly marked
Number activations	How many times the participant looked at numbers
Target number activations	How many times the participant looked at target numbers
Distractor number activations	How many times the participant looked at distractor numbers
Numbers viewed time	How long the participant viewed numbers
Target numbers viewed time	How long the participant viewed target numbers
Distractor numbers viewed time	How long the participant viewed distractor numbers
Safety ranges activations	How many times the participant looked at the safety ranges
Safety ranges viewed time	How long the participant viewed the safety ranges

In addition to the data in the table, the system logged the participant's, age, gender and the order the supportive concepts were given. It also logged the number of times the user marked a number using gaze or the mouse cursor. Similarly, the number of times the user activated buttons with gaze or the mouse cursor was logged. Lastly, the user's gaze sequence is logged after each task, the gaze sequence was logged in a separate file to avoid having cells in the excel file that cannot be automatically resized to fit its content. The gaze sequence consists of the numbers the user looked at and the safety ranges in the order they were looked at, an example of a short gaze sequence can be seen below.

SIupperRight->Instructions->Instructions->Instructions->CTSleft->CTSright->CTSleft->InstrumentAir->ACClowerRight->ACCupperRight->ACClowerRight->RHRupperLeft->RWST

4 Methodology

The study used mixed methods, qualitative data was gathered from semi-structured interviews, and quantitative data was gathered from questionnaires and the logging functionality of the prototype system. The study was conducted using repeated measures, the participants tried all the conditions during the experiment.

4.1 Questionnaires

It was decided to give two questionnaires were given to the participants after the completion of each supportive concept and the baseline condition. One was to be used to measure the participants' perceived performance and the other was to be used to measure the participants' perceived usability of the system.

Performance Metrics

To measure the participants perceived difficulty during the tasks with the support of the concepts and in the baseline condition it was decided to give a questionnaire which asked about performance metrics. Three questionnaires were considered, the After Scenario Questionnaire (ASQ; Lewis, 1990), the Subjective Mental Effort Questionnaire (SMEQ; Zijlstra and van Doorn, 1985), and the Nasa Task Load Index (Nasa TLX; Hart and Staveland, 1988).

The SMEQ asks one question which shows how difficult the participants found the tasks, more performance metrics were required to compare the supportive concepts so it was disregarded. The ASQ asks the participants three questions, how satisfied they were with the completion of the task, the amount of time it took to complete the task, and the support information available when completing the tasks. The ASQ did add an additional two performance metrics, however it was of interest to have more metrics available when comparing the concepts, as such the Nasa TLX questionnaire was selected. The Nasa TLX contains six questions which ask the participants about their mental demand, physical demand, temporal demand, performance, effort and frustration during the tasks, giving a larger amount of performance metrics to compare the supportive concepts.

System usability

To test the usability of the supportive concepts, two questionnaires were considered, the System Usability Scale (SUS; Brooke, 1996) and the Post-Study Usability Questionnaire (PSSUQ; Lewis, 1992). The PSSUQ measures the participants perceived user satisfaction of the system, it has four sub-scales which measures the system quality, information quality and the interface quality. In total the questionnaire has 16 questions. Since the participants are asked to fill in two questionnaires after each concept it was decided to use a questionnaire with less questions. Therefore, the decision to use the SUS questionnaire was made, the SUS contains 10 questions that measures the participants perceived usability of the system.

Modifications

The questionnaires were intended to be used to compare the supportive concepts against the baseline condition, and not to compare the prototype system to other eye tracking systems, as such they were modified to better fit the experiment. The third question of the Nasa TLX was modified from asking how high the "Temporal Demand" was, to how much "Eye Fatigue" was experienced instead. Since there was no time limit during the tasks the "Temporal Demand" question would not have given useful data, it was of interest however to know if some concepts caused more eye fatigue than other concepts. See appendix C.1 to see the modified Nasa TLX questionnaire as given to the participants.

The SUS was modified by the removal of the "I thought there was too much inconsistency in the system" statement as inconsistencies were not deemed a relevant part of the testing of the system prototype system. In addition, two of the statements were changed. The "I think I would need the support of a technical person to be able to use the system" statement was changed to "I thought this system was enjoyable to use", and the "I found carious features of this system were well integrated" statement was changed to "I thought the system was useful and supported me in my task". The changes were made to get feedback that was more appropriate for the supportive concepts. See appendix C.2 for the modified SUS questionnaire as given to the participants.

The total score of the questionnaires were not calculated and used to compare the concepts due to the way the questionnaires were modified. More care should have been taken during the modification process to ensure that the standard total score calculation was still applicable.

4.2 Semi-structured interview

After the participants complete the experiment a semi-structured interview was performed. The interviews questions were focused on the points listed below:

- The user experience
- The usefulness and usability of the supportive concepts
- Suggestions for improvements to the supportive concepts
- The perceived accuracy of the gaze interaction and the supportive concepts feedback
- What the participants thought of the gaze interaction
- Which other domains eye tracking technology could be useful

The interview was pre-coded using key words from the interview questions (See appendix B). The participants' statements were then grouped through similarities and differences. For the questions that were more open ended such as the likes and dislikes of the concepts, emergent codes were created from the key points the participants mentioned. An example of an emergent code that was used from the likes and dislikes of the "highlight and disappear" concept is "search strategy", many of the participants mentioned how the concept fit or did not fit their preferred search strategy. Another code from that concept was "activation time", the participants brought up that the activation time of the numbers were too short or too long, frequently during the interviews. In the following sections the participants' answers are presented.

4.3 User study

The prototype application was presented on a 23” LCD monitor running at a resolution of 1920x1080. An SMI remote tracking system (SMI RED-n Professional) was mounted below the screen and used to detect the participants’ eye movements. The eye tracker operated at a refresh rate of 60Hz. The data from the eye tracking system was processed by the prototype system as described in section 3.2.

4.3.1 Participants

Since good results were obtained with a participant that wore glasses during the pilot study it was decided to not pre-screen the participants for the user study. Sixteen unpaid volunteers were recruited from IFE staff and from students at Østfold university college (HiØ). Seven participants were from IFE; they were asked to participate in an eye tracking study via email. The remaining nine participants were recruited among the students in the Authors year and asked to participate in the master thesis study. The average age of the participants was 31 years. Ten of the participants were male and six were female. Of the IFE participants, one had a professional background in nuclear control room operations, and one had a professional background with the eye tracking technology. The other participants were researchers with a basic understanding of nuclear operations. None of the HiØ participants had experience with process control systems. One of the students had some experience working with eye tracking technology.

4.3.2 Testing protocol

The testing with the participants from IFE was performed in an office at IFE, and the testing with the HiØ participants was performed in a meeting room at HiØ. To begin with the participants were given written instructions (See appendix B) which outlined the background of the study, the testing protocol, and a description of the supportive concepts. After reading the instructions the participants were asked to sign a consent form.

The participants were seated in front of the monitor and instructed to find a comfortable position. The experimenter ensured that the participants were in a position where the eye tracking system could track their eye movements by utilising the head box provided by the SMI eye tracking software. Next the eye tracking system was calibrated to the participants, a nine-point calibration process was used to perform the calibration. The participants were instructed to focus on the red calibration dot and follow it around the screen. In cases where the calibration dropped points or the quality of the calibration was low, the calibration process was repeated until a satisfactory calibration was achieved. In the worst case the calibration process was performed five times before an adequate calibration level was obtained.

After the calibration process, the participants conducted three training runs in order to familiarize them with the task and the system. The training runs used the baseline condition where the participants received no visual support and had to complete the tasks to the best of their ability. Having completed the training run, the experiment begun.

The experiment consisted of four conditions, three conditions supported the participant using visual feedback. The last condition was the baseline where no support was given. Before each condition started the participants were shown an information screen which described how the condition worked. The participants then completed three tasks for each condition, between tasks they were shown a progress screen. This also allowed the participants to relax between tasks as time was only incremented during the tasks themselves. In total the participants completed twelve tasks during the experiment. See Figure 4.1 for an overview of the experiment flow.

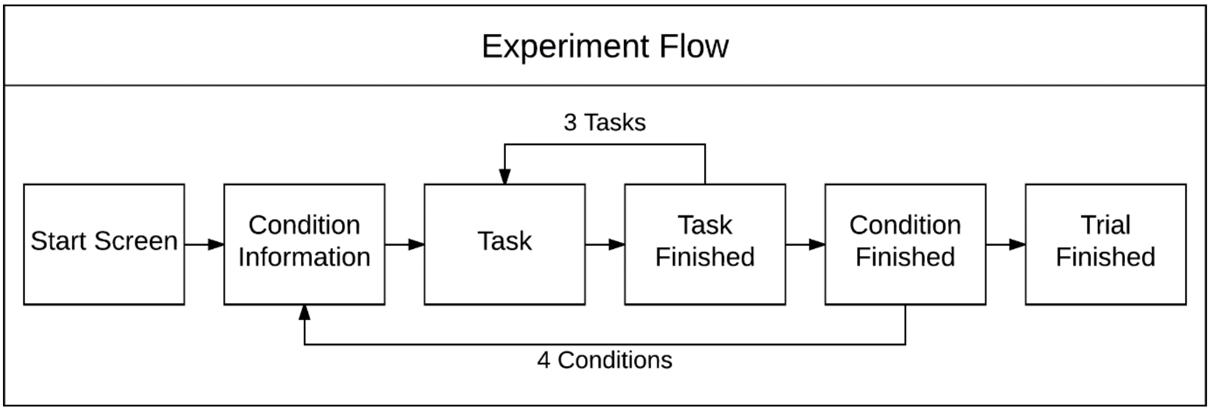


Figure 4.1 The experiment flow showing the different stages of the experiment.

There was no time limit for how much time the participants could spend on each task, when the participants were ready to answer they clicked the answer button in the top right corner of the screen. As explained in section 3.3.1 all the conditions have a task phase, but only the "highlight missed" and "heat map" concepts have a feedback phase as well. After the participants press the answer button it either goes to the task progress screen or the feedback phase if the concept has it. In the feedback phase the participants were shown the feedback created from the data gathered in the task phase. The participants could then re-check the displays and potentially change their answer.

After the participants completed each concept, they were presented with two questionnaires. One was the modified NASA TLX (see section 5.3.1), the other was the modified SUS (see section 5.3.2). Note that for the baseline, the SUS was omitted. After all 12 tasks were completed, a semi-structured interview was conducted (see section 5.1).

The order of presentation of concepts were counterbalanced using the Balanced Latin Square method (Campbell and Geller, 1980), as shown in Table 3. The counterbalancing was done to reduce an expected learning effect. The participants were equally grouped into four groups, one for each condition.

Table 3 The balanced Latin square used to decide the order of the concepts

Concepts\ Participants	Baseline	Highlight and Disappear	Highlight Missed	Heat Map
4	1	2	4	3
4	2	3	1	4
4	3	4	2	1
4	4	1	3	2

5 Results

Data was gathered from three sources, the prototype system logged the participants' performance during each concept, the participants were given two questionnaires after each concept, and lastly a semi-structured interview was conducted after the participants finished the experiment.

5.1 Interviews

In the semi-structured interviews, the focus was on the gaze interaction and the accuracy of the system, in the terms of the gaze interaction and the visual feedback. The search strategies of the participants were of interest to see how the strategies affected the usefulness of the concepts. In addition, the participants were asked what they liked and disliked about each concept, if they had suggestions for improvements and new ideas, and which concept they found the most and the least helpful. Lastly the participants were if they could think of other domains where eye tracking would be useful.

5.1.1 Gaze interaction

It was of interest to learn if the participants found the gaze interaction straining, and if they liked this method of interaction. In this case gaze interaction is defined as how the participants activated buttons and marked numbers by looking at them and pressing the “space bar”. The usefulness of gaze interaction was explored in order to determine if gaze interaction should be integrated unto the NPP simulator in the future.

Ten of the sixteen participants did not find the gaze interaction straining, four of these participants said that the gaze interaction felt natural and not straining at all. The six participants that found the gaze interaction straining did so for different reasons. One of the participants found it a little straining during the “highlight and disappear” concept. Another two of the participants found it a little straining in the beginning, but got used to it over time and it became less straining. Two of the other participants said it became straining due to accuracy issues, they had to shift the gaze to be able to mark numbers and press buttons. The last of the six participants said it was straining due to the participant wearing lenses, and that it would have been less straining if the participant wore glasses instead.

All sixteen participants liked the gaze interaction. Six of the participants felt it was quicker to interact using the eyes than the mouse. Three of the participants said they enjoyed the new concept of interaction, and one of the three was impressed at the accuracy of the system considering the close distance between some of the numbers. Two of the participants said it was more convenient than using the mouse, as it did not require you to first locate and move the mouse to the number in order to mark it, you could immediately mark the number when it was found to be out of bounds. Five of the participants said it was fun to interact in this way, and enjoyed marking the numbers and activating the buttons.

During the study it could be seen that some participants were able to grasp the new form of interaction faster than others. Some of the participants were able to use the gaze interaction without changing how they normally use their eyes. While other participants felt the need to stare at areas when interacting in this way. One participant had a tendency to lean in towards the screen when preparing to mark a number or activating a button, this caused problems for the eye tracking system as it was unable to track the participant's eyes.

The results show that all the participants enjoyed the gaze interaction, however six of the participants did find this form of interaction straining. Out of the six participants only one participant found the gaze interaction straining throughout the process, the other participants got used to it over time or found it straining due to unrelated reasons. This indicates that with a good calibration most users would enjoy this form of interaction and it would become less straining as they get used to the interaction.

5.1.2 Accuracy

The accuracy of the system was broken into two parts; the first part was the accuracy of the feedback provided to the participants by the supportive concepts. The second part was the accuracy of the gaze interaction, meaning the marking of numbers and activation of buttons using the eyes and the "space bar". It was of interest to learn how the participants' found the gaze interaction and if they felt that the visual feedback they were given by the system was correct.

The accuracy of the gaze interaction and the accuracy of the feedback provided by the supportive concepts were rated using the four possible values listed below.

1. Perfectly accurate
2. Problems once or twice
3. Several problems
4. Completely unreliable

Gaze Interaction accuracy

One participant marked the interaction accuracy as "perfectly accurate", the participant encountered no errors while marking numbers and could activate all the buttons on the screen without difficulties.

Fourteen participants said the interaction accuracy had "problems once or twice". Eleven of the fourteen participants experienced problems in the corners or along the edges of the screen. Three kinds of problems were experienced, some of the participants experienced more than one of the problems. Three participants had some troubles marking a number, five participants experienced that the number next the one they were trying to mark was marked instead, and four participants experienced difficulties when attempting to activate the answer button in the top right corner of the screen.

The remaining three participants experienced other kinds of difficulties. Two of the participants had a tendency to lean forward when marking numbers or activating buttons which made it difficult for the eye tracking system to track their eyes, when they leaned back again the system was able to register their interactions as normal again. The third participant had one problem where the participant moved the eyes before the “space bar” was fully pressed. The participant imagined that issues of that kind would be less of a problem once the user becomes more familiar with the new interaction method.

Only one participant marked the interaction accuracy as having “several problems”, the participant wore skinny glasses and experienced many problem areas in the corners and along the edges of the screen. The participant was unable to mark numbers in several areas of the screen and struggled to activate the answer button at the top right corner of the screen.

In general, the system worked perfectly for one participant and well for fourteen participants. The participants that marked “one or two problems” achieved a good calibration, but had one or two areas on the screen that were poorly calibrated. Most of those areas were in the corners or edges of the screen. This resulted in the marking of the number next to the one that was looked at, buttons that could not be activated, and in some cases numbers that could not be marked without shifting the gaze slightly or using the mouse. This could indicate that the positioning of important objects along the outer edges and corners of the screen should be avoided.

Concept Feedback Accuracy

Five participants rated the concept feedback accuracy as “perfectly accurate”, and did not experience any problems with the feedback provided by the system. Although the participants rated the concept feedback accuracy as “perfectly accurate”, four of the participants had some problem areas in terms of accuracy, therefore it is likely that the feedback information in those areas was entirely accurate.

Ten participants marked the concept feedback accuracy as having “problems once or twice”. Four main issues were experienced, some of the participants experienced more than of the issues during the experiment. Nine participants experienced highlighted numbers in the “highlight missed” concept that they felt they had looked at. Five participants felt that the system sometimes did not register that they had looked at numbers. Two participants thought some feedback in the corners of the screen could be wrong due to accuracy issues. One participant felt that the feedback from the “heat map” concept was too sensitive, there were many red coloured areas even though the participant had looked at the numbers in those areas.

The last participant rated the concept feedback accuracy as having “several problems”. The participant was unable to activate all the numbers in the “highlight and disappear” concept by looking at them. The participant also had several false highlights in the “highlight missed” concept, and the feedback in the corner and edges of the screen in the “heat map” concept was wrong.

The prototype system uses dwell-time activation to register a number as looked at, in some cases the system was unable to register that the participants had looked at a number. Three participants were able to determine that some numbers were in range before the dwell-time accumulated enough to activate by glancing at the numbers, resulting in the system not registering the numbers as looked at. One participant actually used the peripheral vision to check some numbers, in this case the system had no possible way of registering the numbers as looked at. It is difficult to set a dwell-time that works for all the users of the system, as such it is necessary to find a way to calibrate the dwell-time to fit the user of the system.

Calibration issues

Nine of the sixteen participants were in the group of people that eye tracking systems generally have problems tracking well, see the limitation section in chapter 2.1 to see the factors that can affect the precision of the eye tracking system.

Six of the participants wore glasses and had accuracy issues in some areas of the screen, usually in the corners and edges of the screen. Two of the six participants were unable to achieve a good calibration while wearing skinny glasses, and had to take them off in order to calibrate well with the eye tracking system. The other four participants wore bigger squared glasses and were able to achieve a good calibration with the eye tracking system. This could indicate that the shape and size of the glasses affect how well the eye tracking system can calibrate, and that bigger glasses should be preferred when working with eye tracking.

Two of the participants wore lenses, but the eye tracking system had no problem calibrating to their eyes. The last participant wore mascara which caused the eye tracking system to have some problems when the participant looked far down or up, due to the dark colour of the mascara the eye tracking system can have difficulties detecting the pupil at those angles as the eye lashes and the pupil blend together. The participant was still able to interact with the system and achieved good feedback from the system.

5.1.3 Search strategies

The participants were asked what their search strategy was when checking the numbers against the safety ranges. Six participants used a “one number at a time” search strategy. The participants checked one number at a time, first the number’s type was checked to determine if it was a target number, if it was then the safety range for that number type was checked and used to determine if the number was out of bounds or not. During the development of the prototype system this search strategy was expected to be the most natural and used strategy.

However, it was found that the other ten participants preferred to use a “one number type at a time” search strategy. The participants memorised the safety range of the first number type and then checked all the numbers of that type, next they memorised the second number type’s safety range and checked the numbers of that type.

The “highlight and disappear” concept did not work well with the “one number type at a time” search strategy, by the time the participants were finished with the first number type all the highlights were already gone. Meaning that the second number type had to be checked without any support. Six of the participants that used the one number type at a time eventually changed their search strategy to checking one number at a time during the “highlight and disappear” concept so that they could get support for both the number types. The remaining four participants did not change search strategy and continued to check the second number type without support.

5.1.4 Concept feedback

The participants were asked what they liked and disliked for each of the supportive concepts. They were also asked which concept they found the most helpful and which concept they found the least helpful. In addition, the participants were asked if they found the highlight and disappear concept confusing, since it was the only concept that actively affected the visual feedback during the task.

How the participants ranked the concepts can be seen in Table 4. The concept that was rated the most helpful was the “highlight missed” concept, seven of the sixteen participants found it the most helpful. The second highest rated concept, rated by five of the participants, was the “highlight and disappear” concept. Three of the participants could not decide which of the two concepts that were the most helpful. “Heat Map” was voted to be the least helpful concept by nine of the participants, followed by the “highlight and disappear” concept with five participant votes.

Table 4 Rating of the usefulness of the concepts from the interview with the participants. The + 3 are participants that could not decide between “highlight missed” and “highlight and disappear” when they rated the concepts.

Concept Ratings			
Most Helpful Concept		**Least helpful Concept**	
Concept	**Votes**	**Concept**	**Votes**
Highlight missed	7 + 3	Heat Map	9
Highlight and disappear	5 + 3	Highlight and Disappear	5
Heat Map	1	Highlight Missed	2

Highlight and Disappear

Ten of the sixteen participants found the “highlight and disappear” concept confusing. Seven of the ten participants thought the concept was confusing when the highlights disappeared, three of the seven participants had to check some of the numbers twice. Another two of the ten participants were distracted when the number next to the one they were looking at activated, it attracted the participants gaze and they forgot what they were doing. The last of the ten participants found it confusing that the highlights disappeared at different times. Out of the ten participants that found the concept confusing three of the participants said that they liked the idea behind the concept but not how it was executed.

Six of the sixteen participants found the task easier to complete with this concept’s support, four of the participants liked the search strategy the concept imposed as they felt they used less energy and time to complete the task. The other ten participants said the concept did not fit their preferred search strategy, while they were focusing on one number type the other numbers highlights disappeared as well. The task had to be repeated without support for the second number type. Three of the ten participants also felt that the concept made the task more stressful and tiring, they had to jump around more with their eyes from the number to the safety ranges. Two of the ten participants additionally said that they felt the concept made the task more mechanical and gave less freedom to which search strategy they could use.

Nine participants liked the way the concept reduced the information they had to search through and said that they were able to focus on the task without using energy to locate the numbers they had to check. Two participants liked that they got support during the task and another participant liked the concept better than the other concepts, while the other concepts give feedback it takes longer to process the feedback and finish the task.

Three problems with the prototype systems timing were encountered, some of the participants experienced more than one problem during the experiment. Five participants thought the numbers activated too quickly, and accidentally activated some numbers. One participant thought the activation time for the numbers was too high, and would have liked it if they activated sooner. Five participants found that the highlights disappeared too quickly. One of the five participants said that it would not be a bad thing to focus more to activate components, if you are going to check the numbers you need to do that regardless.

During the user study some of the participants were observed as confused during this concept, most of the participants understood how the concept worked by the end of task 3 and had changed their search strategy to a “one number at a time” strategy. The participants that already used a “one number at a time” search strategy were able to utilise the concepts support more successfully than the participants using the “one number type at a time” search strategy. One of the participants that found the concept confusing thought it was the best concept by the end of the third task.

Five of the sixteen participants rated this concept as the most helpful, in addition three participants were torn between this concept and the “highlight missed” concept. Five participants rated this concept the least helpful. The feedback shows that the concept has potential, but the timing problems with the number activation and the number disappearing has to be fixed.

Highlight Missed

Eight participants liked that the concept increased their confidence when giving an answer, since they knew where they had or had not looked. Three participants liked to be informed when a number had been forgotten and thought it was fun to see the forgotten areas. Five participants felt that the concept was useful as it gave concise and easy to understand feedback. One of the five participants thought the concept could be useful in a learning context, as it shows you areas that you forgot or did not know about and can help learn an interface quicker.

The participants mentioned three reasons for the concept not being helpful. Firstly, three participants said it was not helpful as it did not give support during the task. Secondly, two participants said it was not helpful as no numbers were highlighted. Thirdly, two participants felt less confident when no numbers were highlighted as they did not know if the system was working or if they had checked all the numbers.

Eight participants experienced false highlights on numbers they felt like they had checked. Two of the eight participants said that even though it highlighted some numbers falsely, it was not disruptive since they could quickly check the numbers before giving their answer.

Seven of the sixteen participants rated this the most useful concept, in addition there were three participants that could not choose between this concept and the “highlight and disappear” concept. Only two participants rated this concept the least helpful. The participants that did not find the concept helpful were very thorough during the task and received no highlights. The participants that found the concept the most helpful did so because of the feedback they received, it was useful and easy to understand.

Heat Map

Three participants liked that the concept always provided feedback, they also liked how the feedback was presented. One of the three participants liked the freedom of how to interpret the feedback, to begin with the participant assumed that green numbers were ok and only checked red and brown colours, in the next task the participant assumed that green and brown were correct and only checked the red colours.

Five participants felt it was interesting and useful to see how the time was spent during the task. Two of the five participants thought the concept could be useful in a learning context, one of the two participants said that the concept could be useful for NPP operators. Another two of the five participants liked that the values could be rechecked and that more time could be spent in the red areas where little time was spent. The last of the five participants felt assured that everything had been checked through the feedback.

The participants mentioned two reasons for the concept not being helpful during the tasks. Firstly, three of the participants did not feel like the concept helped during the task, one of the three felt confident in the answer even though areas appeared as red. Secondly, four participants said that there was too much information given by the feedback, two of the participants said there were too many colours in the heat map. One of the two participants said that a number was looked at for maximum 2seconds and that two colours would be enough to represent the data.

Four participants felt that numbers in red areas had to be checked again, but some numbers were easily determined to be in range and were still double checked as they appeared in red areas. One participant felt that the heat map was too sensitive, when the participant saw red numbers even though the number had been checked it felt discouraging, and It felt like the task was not done correctly. Three participants felt it took longer to finish the task in this concept compared to the other concepts.

Only one participant rated this concept as the most useful, while nine participants rated it as the least helpful concept. Most of the participants said it was the least helpful due to the amount of information they received and that much of the information was unnecessary in order to support them during the task. They also said that the “highlight missed” concept is a simplified version of the concept, which only shows the information relevant to the task. There were two participants that could see the concept being useful in a learning context where the user can see how time was spent and see how to spend it better in the future.

5.1.5 Improvements and new ideas

The participants were asked if they had suggestions for improvements or new ideas for how to use the eye tracking as a supportive tool during the number checking task. Four of the participants mentioned that they would like to have graphical visualisations to improve the feedback of the gaze interaction. Two of the four participants would have liked a graphical indicator of where they were looking. The other two participants would have liked a non-intrusive visualisation that shows when the system has registered that you are looking at something.

There were several suggestions for improvements of the “highlight and disappear”, two participants suggested ways to help the focus on one number type at a time. The first suggested to highlight one number type first, and after the numbers have been checked the other number type could be highlighted. The second participant suggested to colour code the numbers on their number types. Three participants suggested to not make the highlights disappear, but rather fade or change the colour so it could be double checked easily.

One of the participants said that with fine-tuning, one of the concepts would probably be useful for a real NPP operator task. Another participant suggested to control the opening and closing of reactor valves in the NPP simulator using gaze based interaction.

Four participants suggested to make changes to the interface itself. Two of the participants would like the way the number types are shown to be changed, one participant suggested the use of symbols to represent the different types, the other participant suggested to colour code the numbers so it is easier to see which type they belong to. Another of the four participants suggested to change how the safety ranges are presented so the eyes do not have to jump long distances every time to check a number. The last participant suggested to use circles instead of squares for the highlighting of the numbers.

Two participants suggested to make the concepts give support during the task instead of feedback after the task, they also had some ideas of how to achieve that. For the “heat map” concept the suggestion was to make a live heat map, where the numbers change colour as they are looked at. The colour of the numbers would be independent of each other so the colours of other numbers are not changed while a number is looked at. The testing revealed that change of colour and the appearance and disappearance of shapes easily draws the attention of the user, therefore it is better to keep the change to where the user is already looking.

To make the “highlight missed” concept give support during the task it was suggested that the numbers are highlighted as they are looked at instead of highlighting numbers that were not looked after the task has been performed. Another participant suggested a mix of the “highlight missed” and the “heat map” concept, instead of highlighting only the numbers that were missed it could highlight the numbers that were looked at for only a short time, this would remove the redundant information from the heat map and allow the user to focus on the important information for the task.

5.1.6 Other domains where eye tracking can be useful

The participants were asked if they could think of other domains where eye tracking could be useful. Eight participants said that gaming is a natural domain for gaze-interaction. Three of the eight participants mentioned the controlling of weapons, using the eyes to aim. Another four participants suggested to use gaze interaction to control different things, the game character, the steering of a car, the activation of skills, or controlling where the camera is facing. The last participant suggested horror-games, the players gaze could be used to determine when to show or hide horror elements.

Six participants thought eye tracking could be used to support the driver while driving a vehicle. One suggestion was to remind the driver to look out of the window if the driver spent too much time looking at the phone or fiddling with the radio. Another suggestion was to remind the driver to check blind spots before changing lane, if the blind spots were not checked. One participant mentioned that if you are driving long trips it is easy to lose focus over time, the loss of focus could be registered using eye tracking and used to help the driver regain focus. Another suggestion was to automatically reduce the speed of the car if the driver fails to notice that an obstacle or that a dangerous situation is occurring.

One participant thought eye tracking could be useful in any safety-critical domain where monitoring is required. The participant said that humans are not so good at staying in the loop, so eye tracking can detect when the human has disengaged and help reengage the human in the loop. Four participants said eye tracking could be useful to support training in the maritime or the air traffic control domain. One participant had two ideas that was not mentioned by any of the other participants, to use eye tracking in police investigations to make sure that all areas of the crime scene have been covered, and to use eye tracking to support the training of hospital surgeons.

Ten participants mentioned cases where eye tracking could be used for interface interaction. Four of the participants suggested to replace the mouse cursor as an input device with eye tracking combined with the keyboard. Another two participants thought gaze interaction could be useful to support severely handicapped people. Two of the participants said that gaze interaction could enable new kinds of interfaces. Lastly, two participants said it could be interesting to see eye tracking used with virtual reality.

From the participants' responses it can be seen that the eye tracking technology has potential in various domains and can be used for many different purposes, both for entertainment, interaction, training in safety-critical systems, or as support while operating vehicles.

5.2 Logged data

During the experiment the prototype system logged data for each participant, see Table 2 in section 3.5 for an overview of the logged parameters. The data was averaged on all the participants, and then for the participants split into two groups based on their search strategy, see 5.1.3 for an explanation of the search strategies. In the following charts the orange bars represent the standard deviation.

5.2.1 Completion time

Looking at the average completion time for all the participants (See Figure 5.1-A) it can be seen that both the "highlight missed" and the "heat map" concepts have a higher completion time than the baseline condition. This is as expected considering the participants first solve the task as if it was the baseline condition before receiving feedback, then interpret the feedback and make changes accordingly before giving the final answer. The "heat map" concept has a higher completion time than the "highlight missed" concept, since it gives the participant more information to process than the "highlight missed" concept. The "highlight and disappear" concept has a slightly faster completion time than the baseline concept. The goal of the concept was to reduce the amount of information the participants have to look at in and to help structure the scanning of the numbers, so that the completion time is lower than that of the baseline condition is to be expected.

With the completion time data split on the participants search strategies (See Figure 5.1-B) the average results follow the same trend as when the data was for all the participants. However, the results shot that the participants that used a one number at a time search strategy completed all the conditions faster than the participants that used a one number type at a time strategy. The standard deviation is also lower for the one number type at a time search strategy, this could indicate that the one number at a time search strategy is more suited for these kind of tasks.

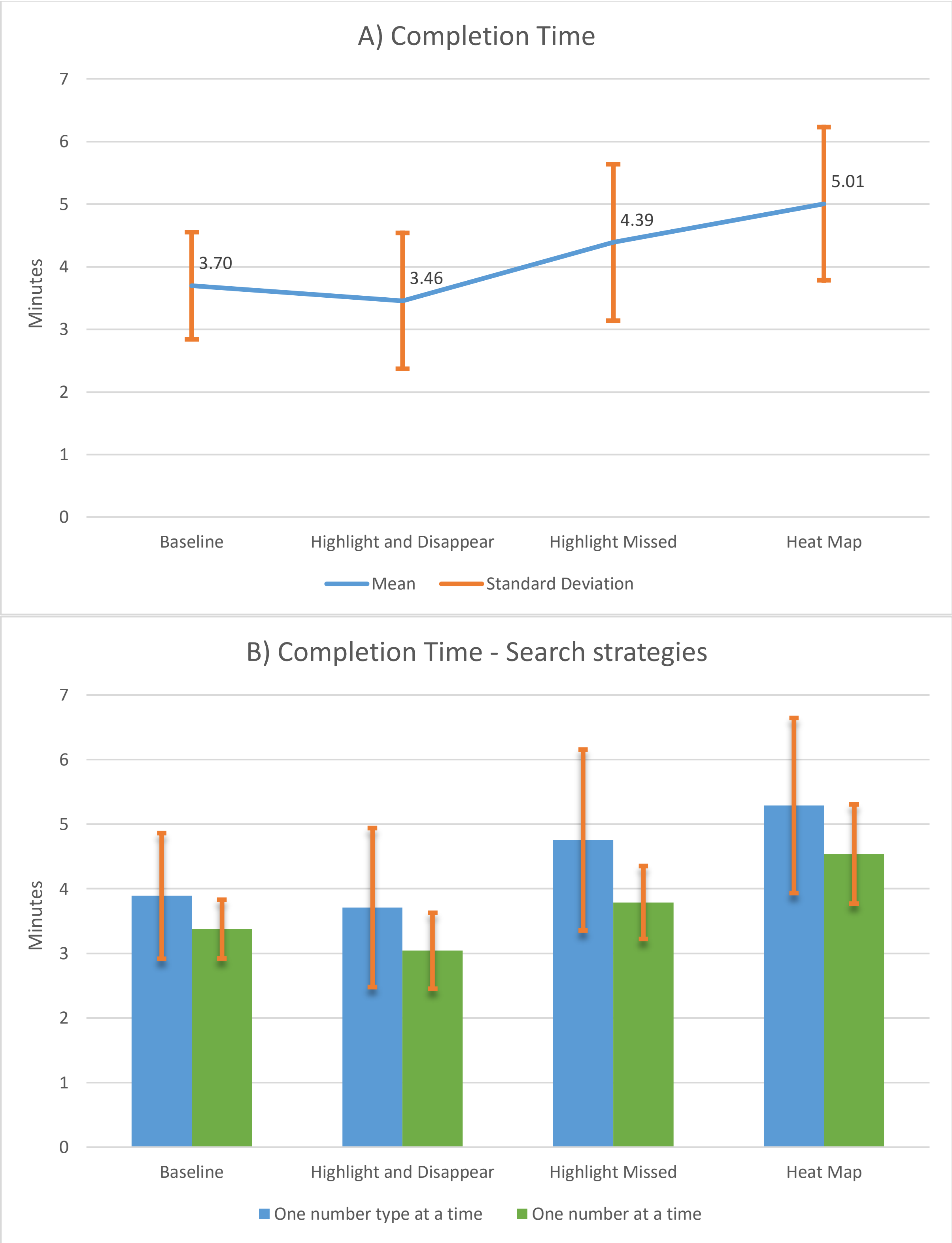


Figure 5.1 The average completion time of the participants for each condition

5.2.2 Missed wrong numbers

Another interesting source of data is the number of missed wrong numbers. A wrong number is a number that was out of bounds that the participant did not mark, or a number that was falsely marked. On average for all the participants (See Figure 5.2-A) the “highlight missed” and the “heat map” concept had less wrong numbers than the baseline condition. The “heat map” concept had the least errors with 0.5 less errors that the baseline condition. The “highlight and disappear” concept had a 0.22 higher rate of wrong numbers than the baseline condition. With the data split on the participants search strategy (See Figure 5.2-B) it can be seen that the participants that used a one number at a time strategy on average had less wrong numbers than the other participants, they also had a lower standard deviation. The trend is the same as with all the participants’ data.

The reason for the “heat map” concept having less errors than the other concepts could be that the concept always gives feedback, and gives the participants a chance to check the numbers again. This comes with a price though as the completion time is on average more than a minute higher than the baseline condition. However, the average number of errors for all the conditions are very low, in addition the standard deviation is very high, therefore the results can be seen as insignificant.

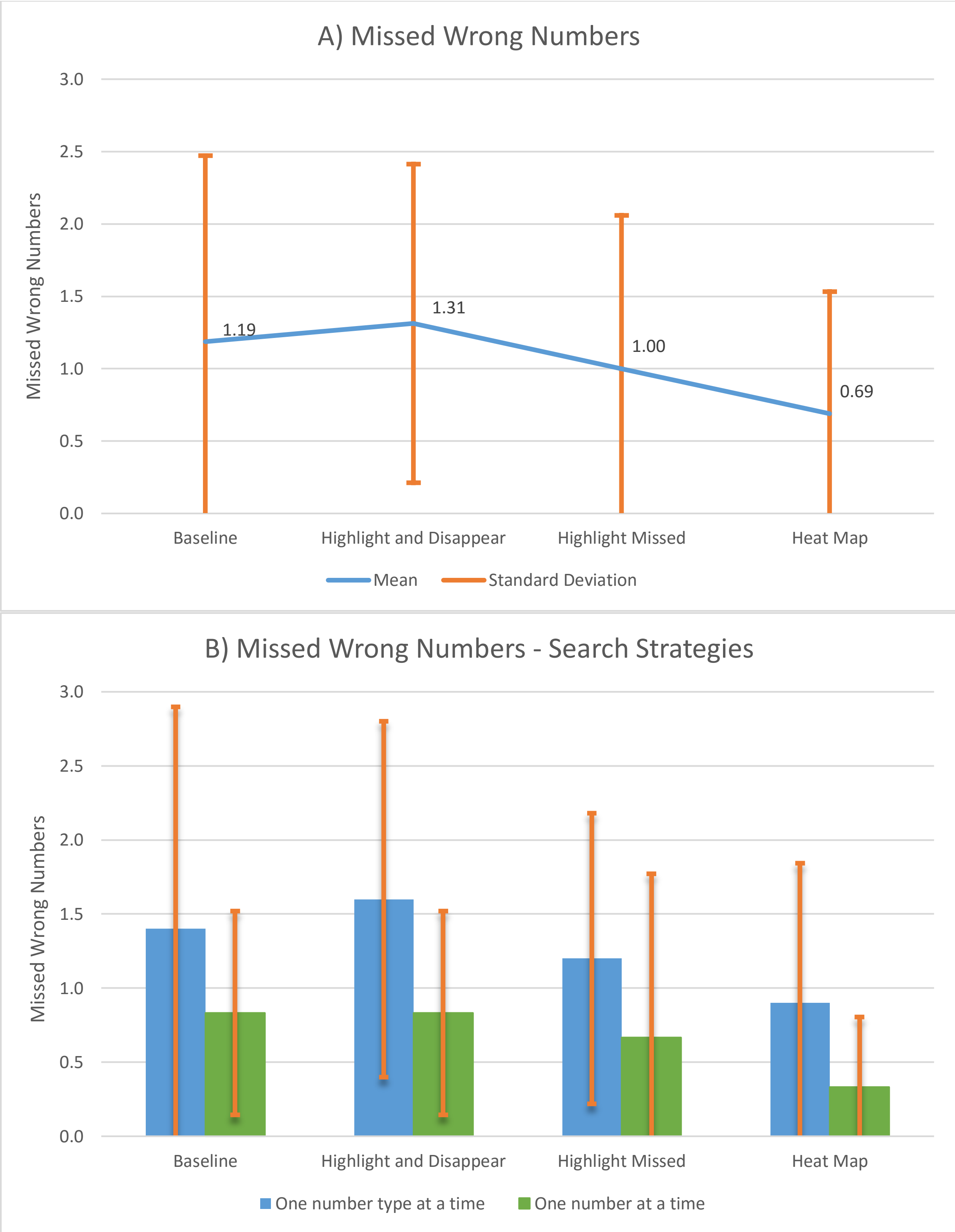


Figure 5.2 The average number of out of bounds values the participants missed for each condition

5.2.3 Total number activations

The number of times and how long the participants looked at the numbers is also interesting. On average all the participants (See Figure 5.3-A) looked at the most numbers in the "highlight missed" and the "heat map" concept. There are slightly more numbers looked at in the "highlight missed" concept than in the baseline condition, this is as expected considering that only a few numbers would be highlighted during the concept. The "heat map" concept has 59 more number activations on average than the baseline condition, the concept offers more feedback than the others, so it is as expected that there are more number activations than in the other conditions. The "highlight and disappear" concept has 29 less number activations than the baseline condition on average, since the concept highlights the numbers that needs to be checked it is as foreseen that there are less activations. The participants that used a one number at a time strategy looked at less numbers in average compared to the participants that checked one number type at a time (See Figure 5.3-B). The standard deviation is also lower for the participants that checked one number at a time. The trend of the concepts is the same as with all the participants' data. The trend of the numbers activations matches the trend of the completion time numbers (See Figure 5.1).

In addition to the number of times the numbers were looked at the system also logged how long the numbers were looked at. On average the participants looked at the numbers longer in the "heat map" and the "highlight missed" concepts, and slightly less in the "highlight and disappear" concept (See Figure 5.4-A). The numbers are similar when the data is split by search strategies (See Figure 5.4-B), however here the baseline condition and the "highlight and disappear" concept has almost the same values. Again the standard deviation is lower for the participants using the one number at a time search strategy. The trend lines of the time the numbers were looked at matches the trend lines of the number of number activations.

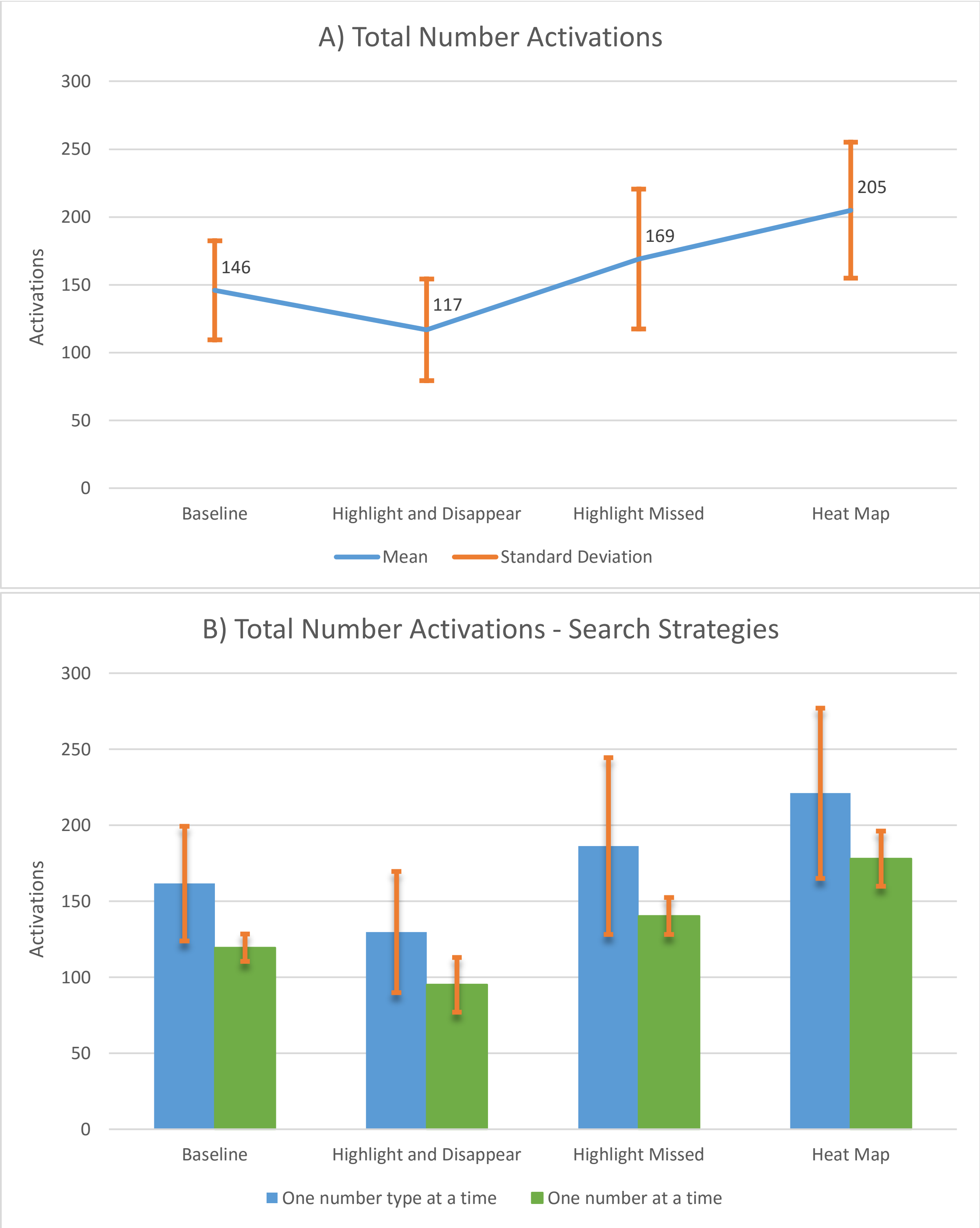


Figure 5.3 The average number of times the numbers were looked at for each condition

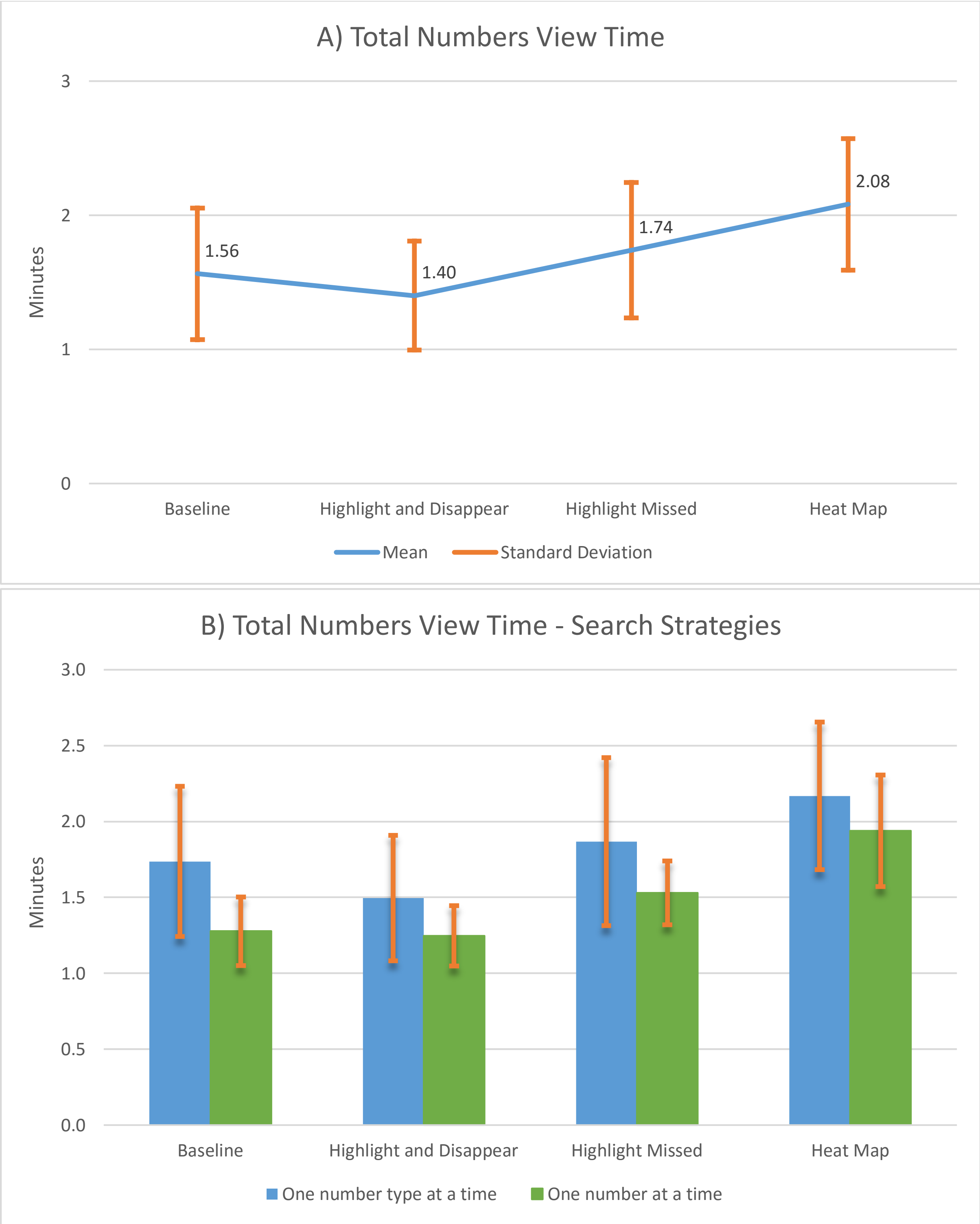


Figure 5.4 The average time the numbers were looked at for each condition

5.2.4 Target and distractor number activations and view time

The numbers are also split into target and distractor numbers. The target numbers are numbers that should be checked during a task, while the distractor numbers are numbers that do not need to be checked. For all the participants on average (See Figure 5.5-A) there were more target activations than in the baseline condition. The “heat map” concept had 54 more number activations on average than the baseline condition. With the data split on the participants search strategy (See Figure 5.5-B) the data looks similar, however the participants with a one number at a time search strategy has less activations and a lower standard deviation.

The distractor numbers were on average activated less than the target numbers in all the conditions (See Figure 5.6-A). In the “heat map” and the “highlight missed” concepts the distractor numbers were activated slightly more than in the baseline condition. The “highlight and disappear” concept resulted in a decrease (Mean = 16.75, SD = 15) in the amount of distractor number activations compared to the baseline condition (Mean = 50.13, SD = 14). The decrease was statistically significant, $t(-10) = 3.09E-08$, $p < .05$, two-tailed. Since the “highlight and disappear” concept shows the participants which numbers they have to check it is as expected that the number of distractor activations are less than in the other conditions. In addition, the participants using the one number at a time search strategy activated less distractor numbers and had a lower standard deviation than the participants using the one number type at a time strategy (See Figure 5.6-B).

The target numbers were viewed longer in the concepts than in the baseline condition on average (See Figure 5.7-A). Since the concepts highlight the target numbers during or after the task it is not surprising that they are viewed longer than in the baseline condition. The target numbers are viewed the longest in the “heat map” concept, since the concept provides the most feedback the results are as expected. With the data split on the two search strategies (See Figure 5.7-B) the participants using a one number at a time strategy has a slightly shorter view time and a lower standard deviation than the other participants.

What is more interesting is how long the distractor numbers were viewed (See Figure 5.8-A). The baseline condition, the “highlight missed” and the “heat map” concepts have very similar view times. The “highlight and disappear” however, resulted in a reduction (Mean = 0.12, SD = 0.12) of the time distractor numbers were viewed compared to the baseline condition (Mean = 0.44, SD = 0.17). The reduction was statistically significant, $t(9) = 2.01E-07$, $p < .05$, two-tailed.

The baseline and the two feedback concepts have similar distractor view times, that is as expected due to the task being the same during the first part of the concept. During the feedback part the target numbers are highlighted, as such it as foreseen that the distractor numbers are not looked at for long after the highlights appear. Again, with the data split by search strategies (See Figure 5.8-B) it can be seen that the participants checking one number at a time has lower values than the other participants. Examining the distractor and target number data it would appear that the “highlight and disappear” concept is better at focusing the participants’ energy on the actual task than the other concepts.

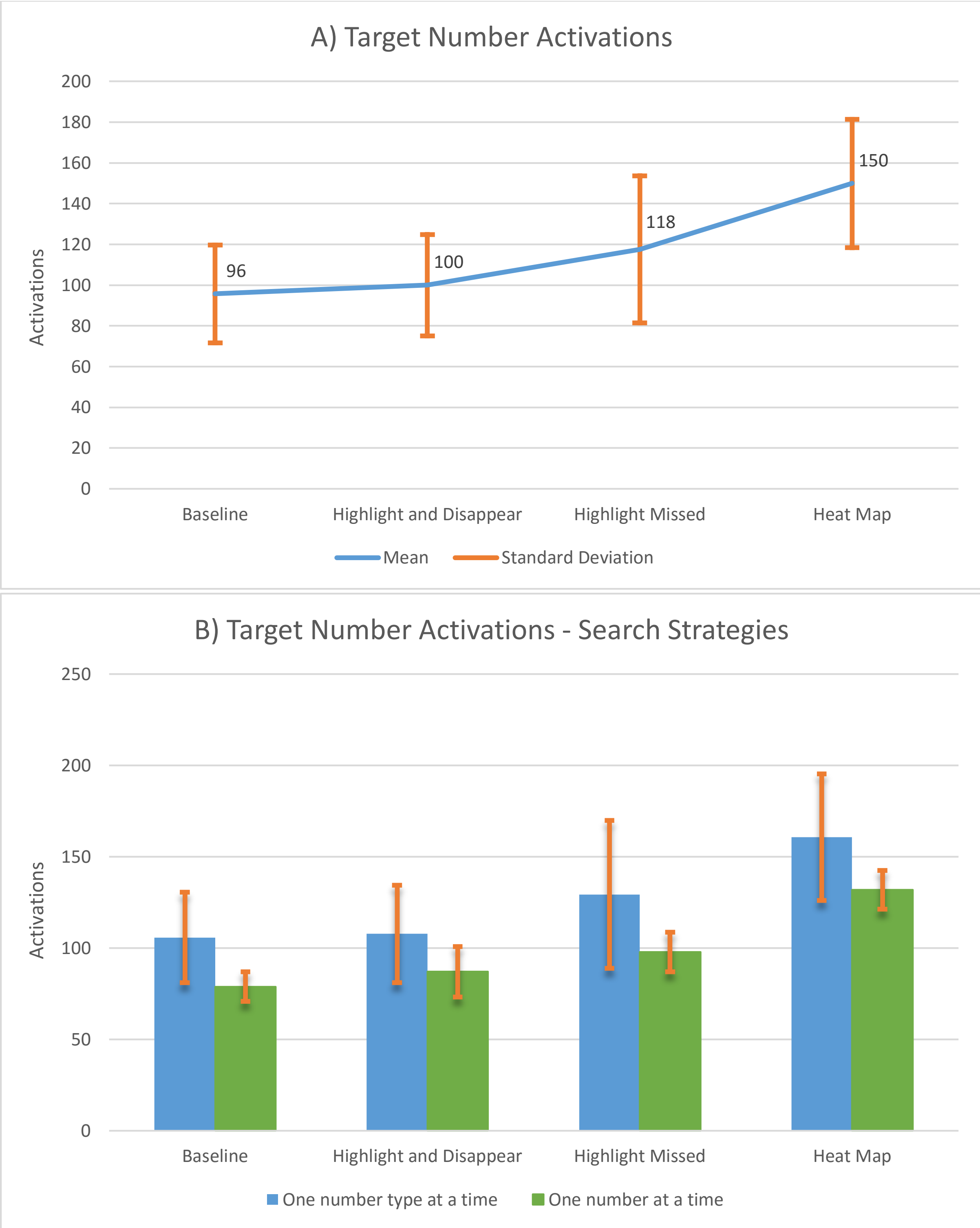


Figure 5.5 The average number of time target numbers were looked at for each condition

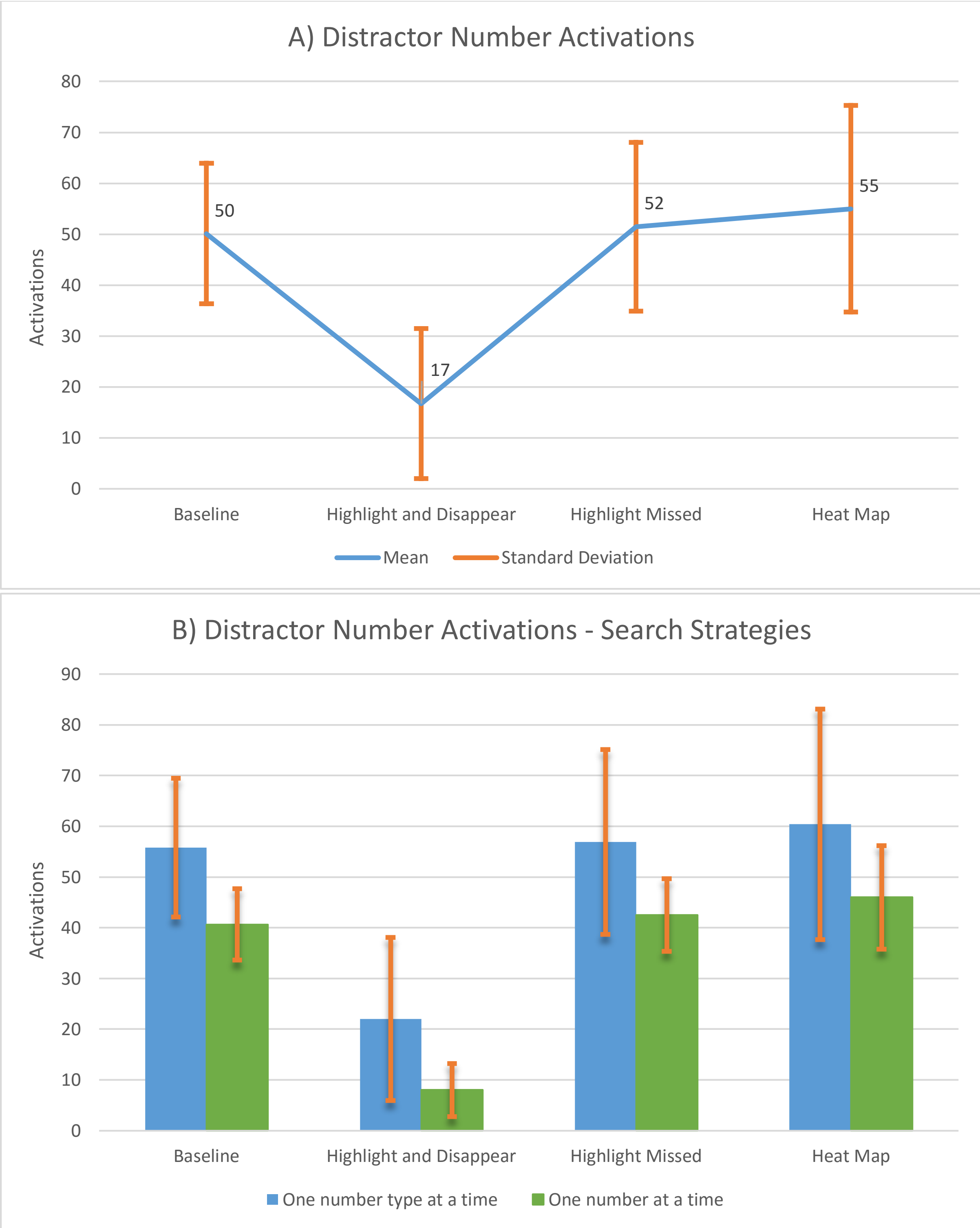


Figure 5.6 The number of times distractor numbers were looked at for each condition

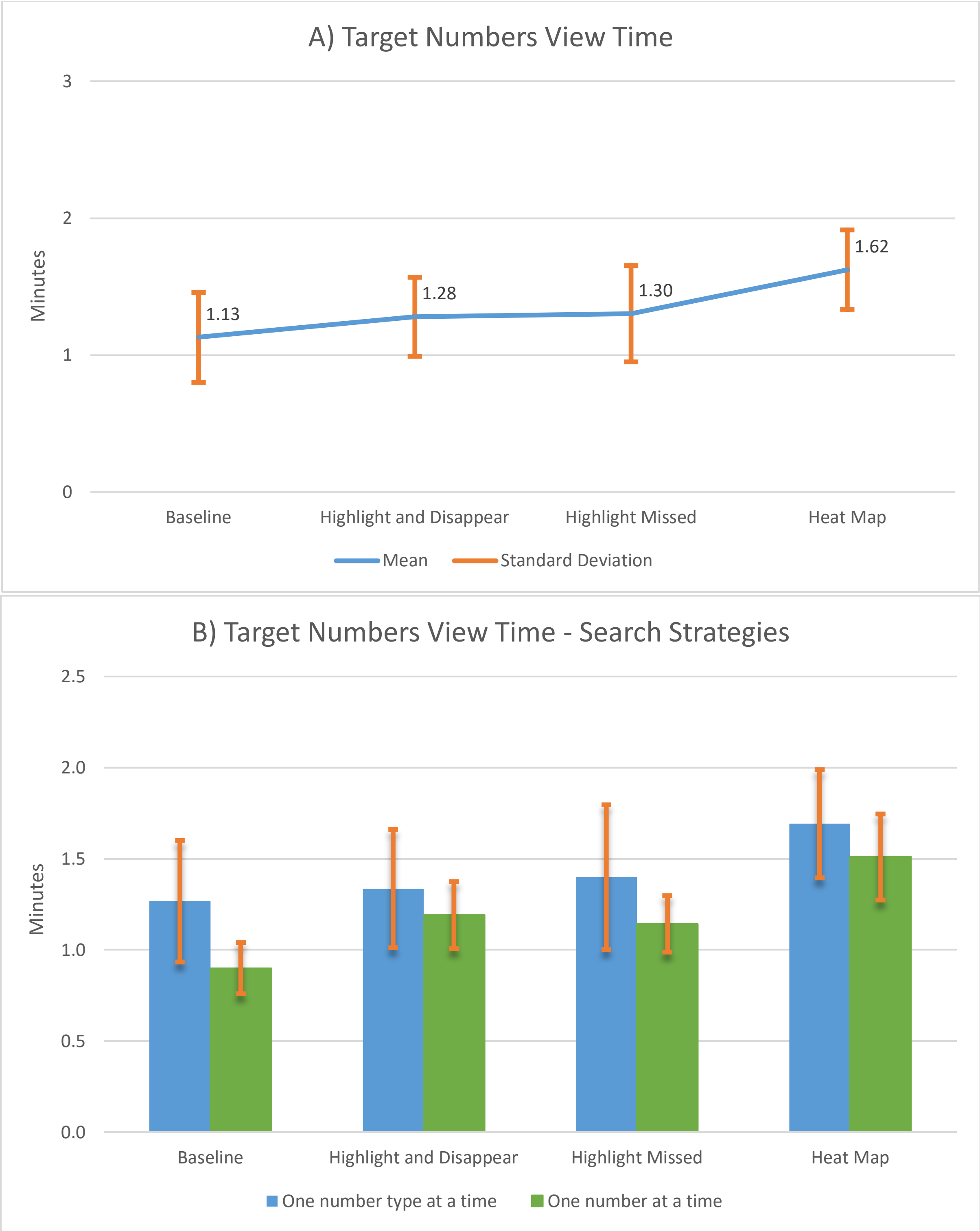


Figure 5.7 The average time target numbers were looked at for each condition

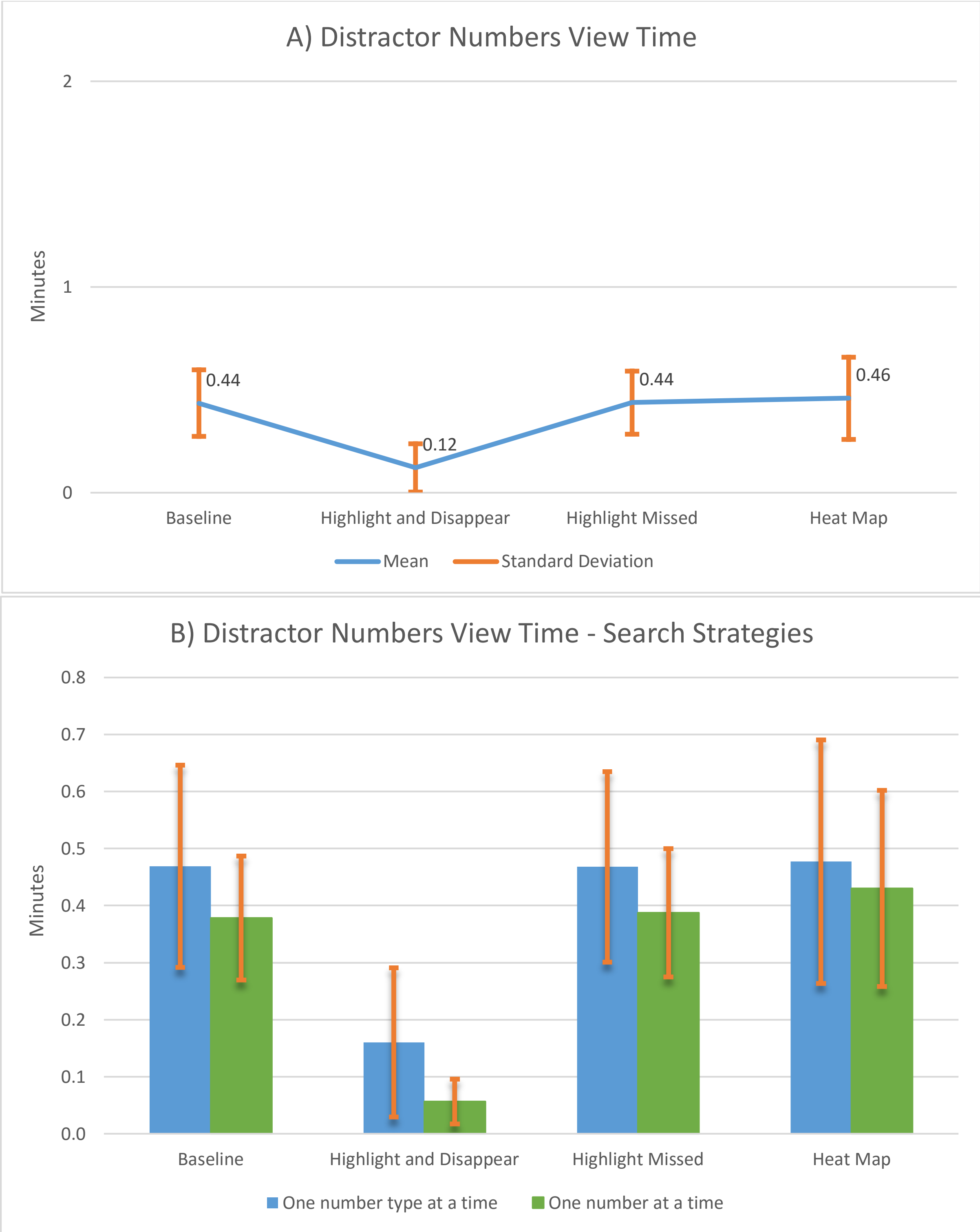


Figure 5.8 The average time distractor numbers were looked at for each condition

5.2.5 Instructions

The last data the system logged was the number of times the participants looked at the instructions and how long they looked at the instructions. On average the participants looked at the instructions between 26 and 33 times during each condition (See Figure 5.9-A). With the data split on the participants search strategies (See Figure 5.9-B) it can be seen that the participants using a one number type at a time search strategy looked at the safety ranges less than the other participants. The participants that used that strategy attempted to memorise one range before checking the numbers, as such the results are as expected. Compared to the baseline the participants checked the safety ranges more often in the other concepts, this is probably because the highlights do not distinguish the number types from each other so the participants had to check the safety ranges more often. The participants that used a one number at a time search strategy looked at the safety ranges about the same amount for all the conditions except the “highlight and disappear” concept, where the number of times the safety ranges were looked at was reduced.

On average all the participants looked at the safety ranges for about the same amount of time during all the conditions, with about a 3 second difference at the most (See Figure 5.10-A). With the data split on the participants search strategy a difference between the two groups can be seen (See Figure 5.10-B). During the “highlight missed” and the “highlight and disappear” concepts the two groups have similar numbers, during the baseline and the “heat map” condition however the participants with a one number at a time search strategy spend more time looking at the safety ranges compared to the other participants.

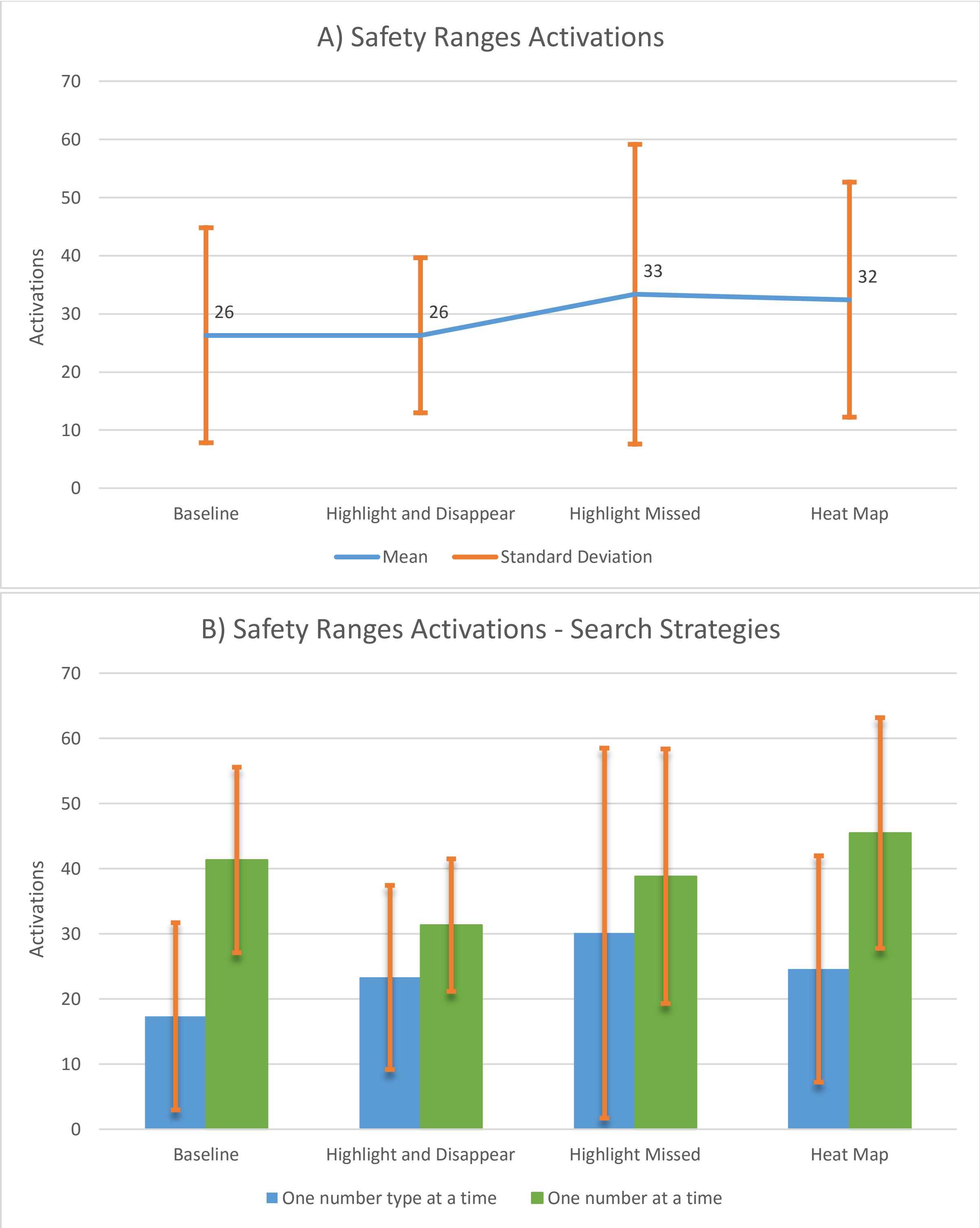
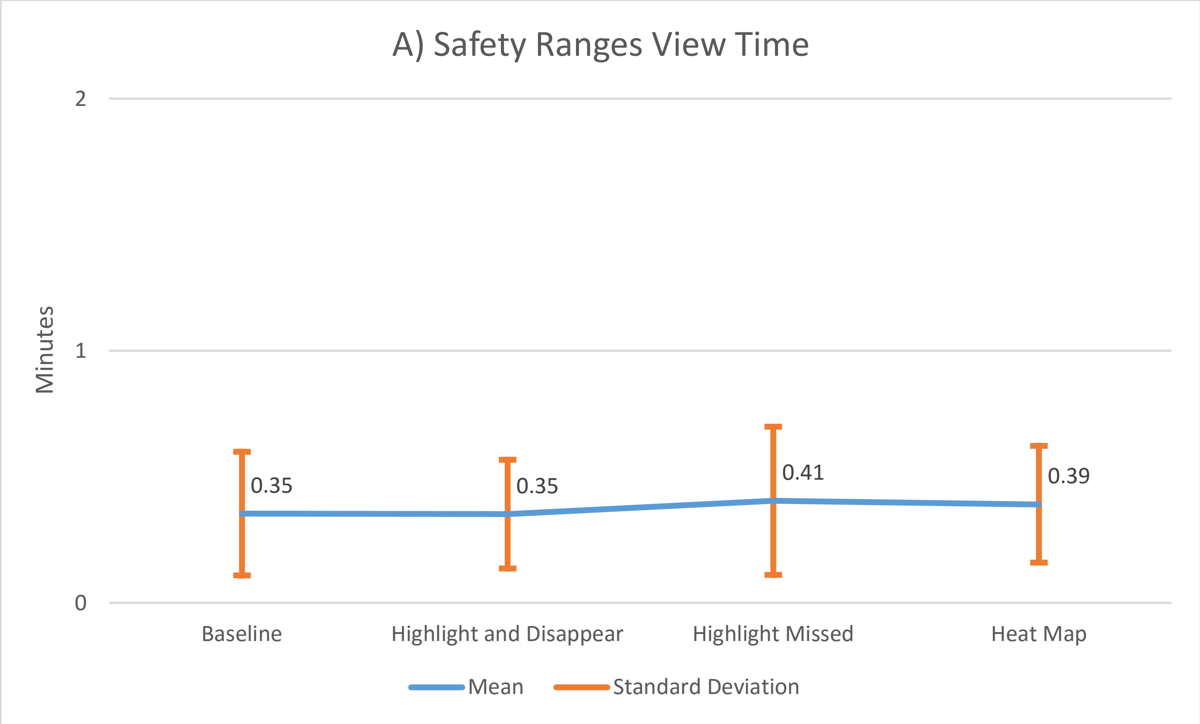


Figure 5.9 The average number of times the participants looked at the safety ranges for each condition



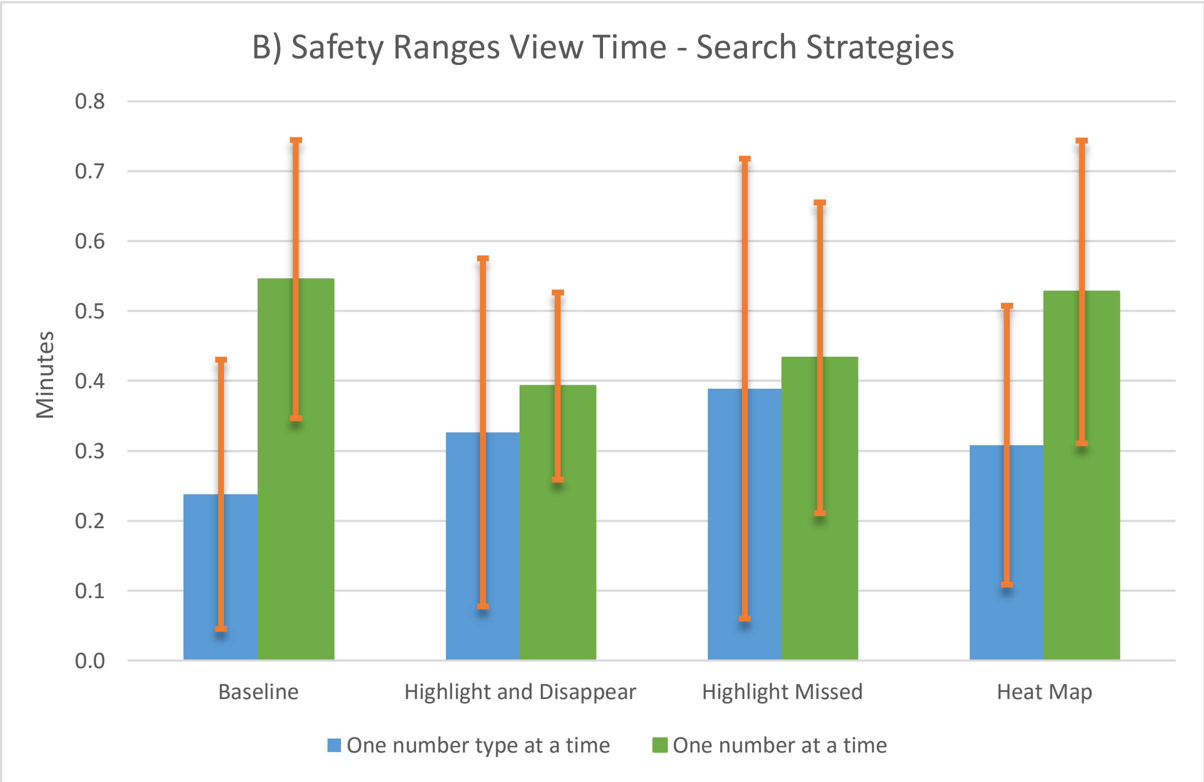


Figure 5.10 The average time the participants looked at the safety ranges for each condition

5.3 Questionnaires

After each concept the participants were handed two questionnaires, a modified Nasa task load index (See appendix C.1) and a modified system usability scale (See appendix C.2). Since the prototype system is not compared to other eye tracking systems but rather the supportive concepts are compared to each other, some of the questions were modified to better fit the experiment, see section 4.1 to see how the questionnaires were modified.

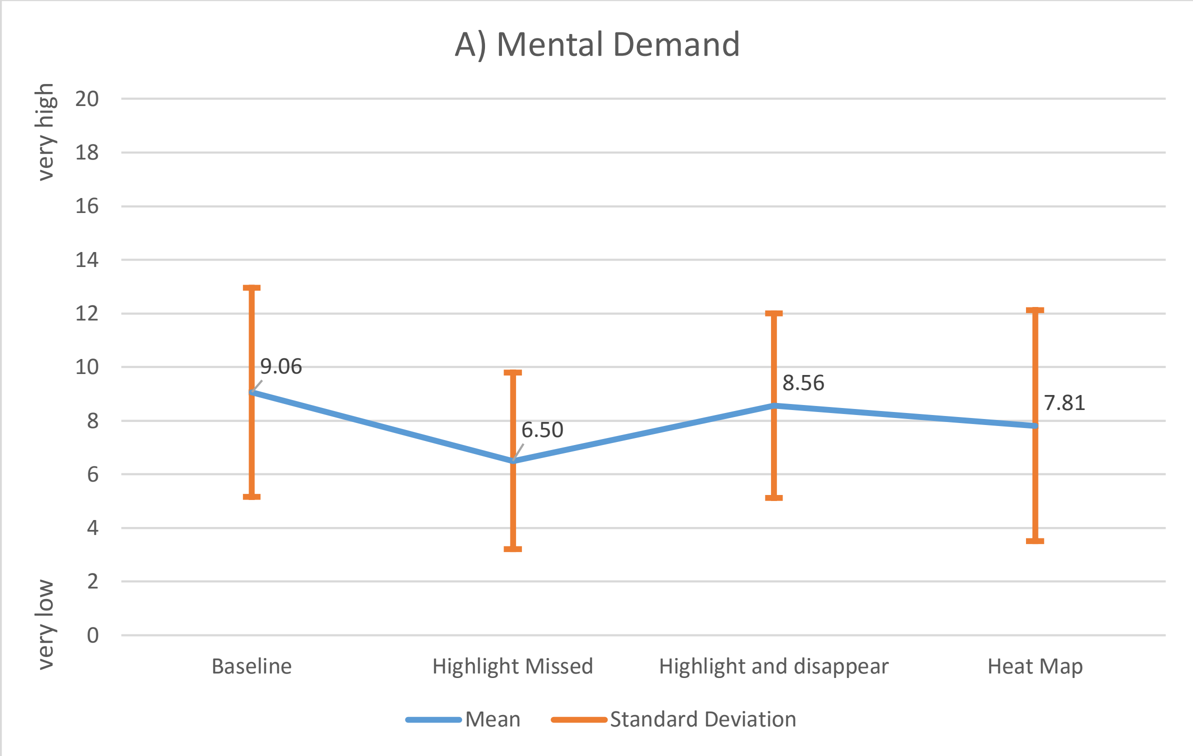
5.3.1 Nasa TLX

The modified Nasa TLX questionnaire contains six questions to determine the mental demand, physical demand, eye fatigue, performance, effort, and the frustration of the participant after each condition of the experiment. The results were used to measure how the participants perceived the tasks during the baseline condition and how they found the tasks during the different supportive concepts. Nasa TLX operates with a scale between 1 and 20, where 1 is very low and 20 is very high. On the fourth question "How successful were you in accomplishing what you were asked to do?", a rating of 1 means perfect while 20 means failure.

Mental Demand

The participants were asked how mentally demanding they found the tasks for each supportive concept, see Figure 5.11-A. On average the participants rated that the concepts were less mentally demanding than the baseline condition where they received no support. The standard deviation between the different concepts are quite similar and lie between 3.5 and 4, only the "highlight missed" concept shows an indication of being less mentally demanding compared to the other conditions.

The results of the mental demand question was split on the participants search strategy (See Figure 5.11-B) to see if there was a difference in the "highlight and disappear" concept. It seems like the participants that used a "one number at a time" search strategy found the concept slightly less mentally demanding than the participants who used a "one number type at a time" search strategy.



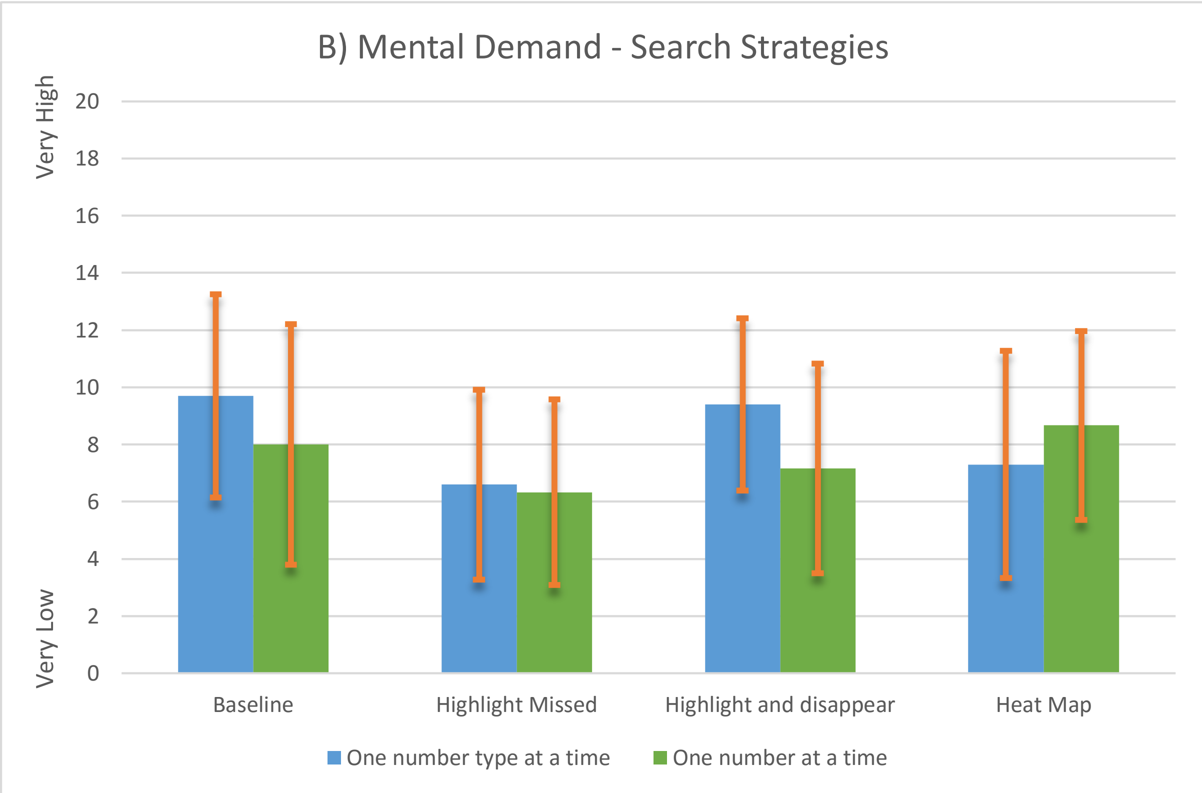


Figure 5.11 Nasa TLX question 1: How mentally demanding the participants found the tasks during the different conditions.

Physical Demand
The participants were also asked how physically demanding they found the tasks for each concept, see Figure 5.12. It can be seen that the values are quite similar. Since the participants are sitting in front of a screen with the task of checking numbers it was as expected that the answers to this question were very low. Again it can be seen that the concepts have a slightly better average rating than the baseline condition.

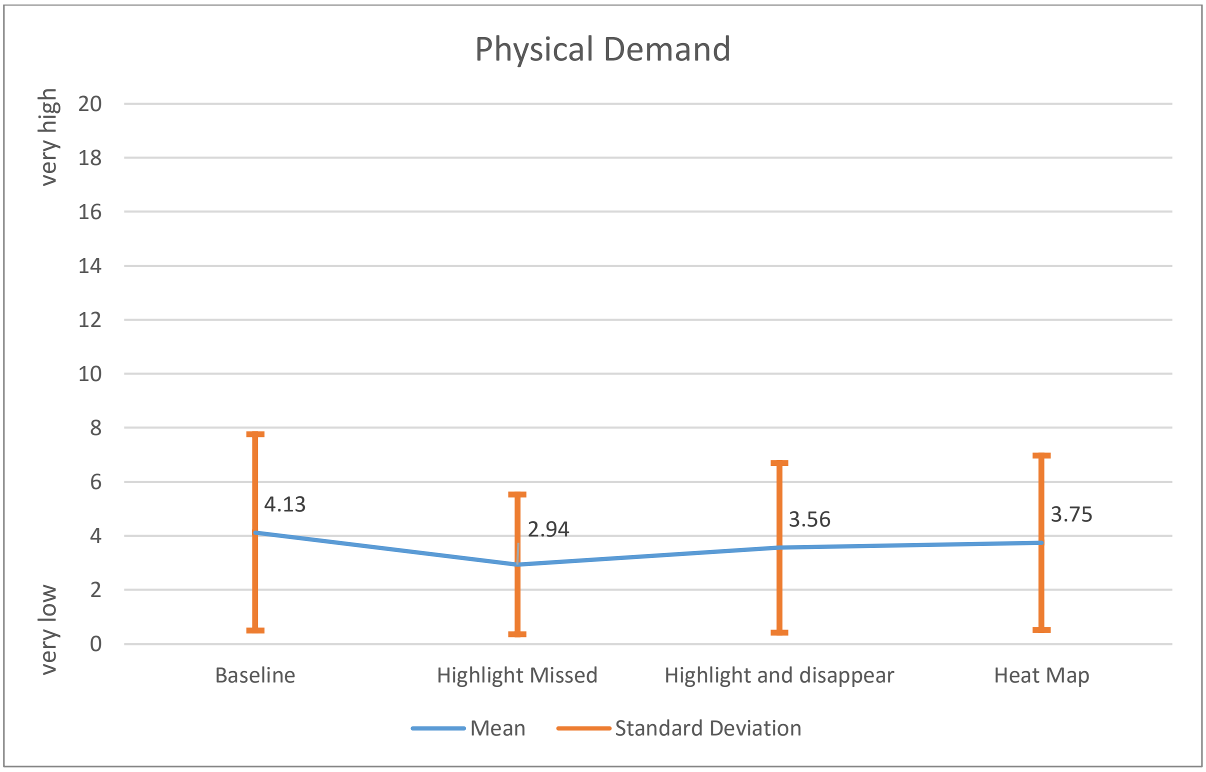


Figure 5.12 Nasa TLX question 2: How physically demanding the participants found the tasks during the different conditions.

Eye Fatigue
The participants were asked how fatigued their eyes were after completing all the tasks of each concept, see Figure 5.13. On average the participants found the supportive concepts less fatiguing for their eyes, the "heat map" concept was rated the least tiring. That is unexpected considering that the concept has the most visual feedback. The variance for the conditions are large, but they are also similar which could indicate that the supportive concepts did not make the tasks more fatiguing for the eyes.

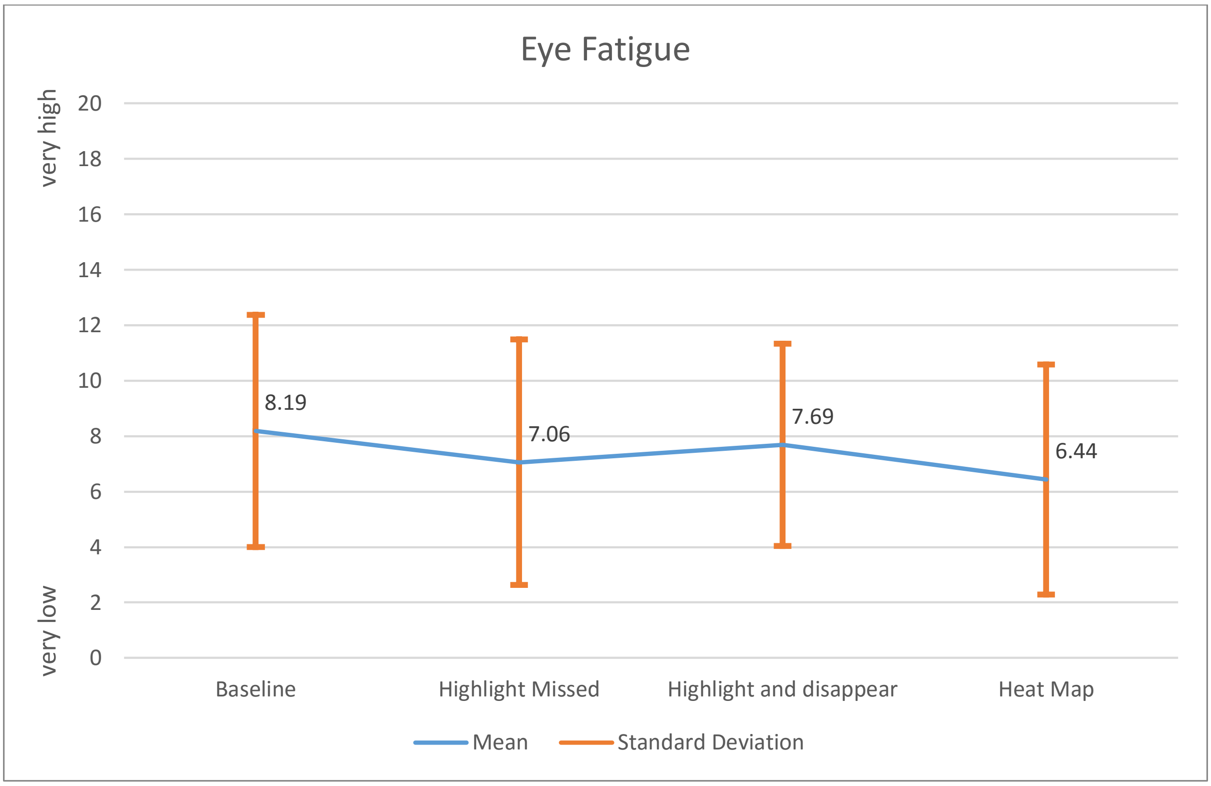


Figure 5.13 Nasa TLX question 3: How fatigued the participants eyes were after each condition.

Performance

The participants were asked how good their performance was for each concept, see Figure 5.14-A. Performance means how successfully the participant was at completing the task. Based on the average results it can be seen that the participants felt like they achieved a worse performance during the "highlight and disappear" concept. The other two concepts are rated slightly better than the baseline condition. Only the "highlight missed" concept has a low enough variation to indicate that the concept performed better than the baseline condition. A possible reason for the "highlight and disappear" concept ranking lower than the baseline is that it was confusing for many of the participants due to the visual feedback changing during the task, and how the concept is interacted with.

The search strategy of the participants had a big impact on the rating for the "highlight and disappear" concept, see Figure 5.14-B. Looking at the participants that used a one number type at the time search strategy, it can be seen that the average performance value for the concept is worse than for the baseline condition. However, looking at the participants that used a one number at a time search strategy the concept is rated better than the baseline condition. This indicates that this concept in particular was experienced differently depending on the participants search strategy.

Figure 5.14 Nasa TLX Question 4: How well the participants rated their performance on a scale from 1-20, where 1 is perfect and 20 is failure.

Effort

The participants were also asked about the level of effort they used in order to achieve that level of performance, see Figure 5.15. From the chart it can be seen that all the concepts on average reduced the level of effort the participants had to put in to the task compared to the baseline. The standard deviation is high, leaving the results unclear. There is however an indication that the "highlight missed" and the "heat map" concepts required the least effort during the tasks.

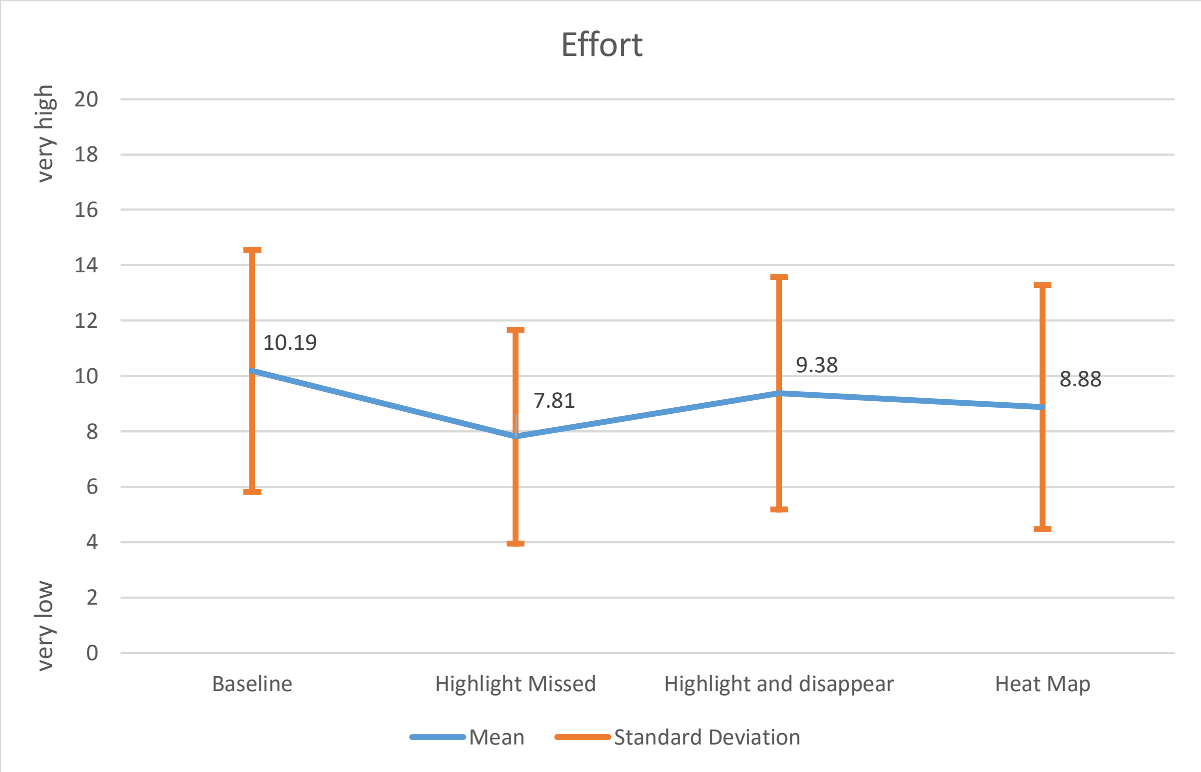


Figure 5.15 Nasa TLX question 5: How much effort the participants had to put in to achieve their level of performance.

Frustration

Lastly the participants were asked about their frustration level during the tasks in the different concepts, see Figure 5.16. On average the frustration levels for the "highlight and disappear" concept was slightly higher than for the baseline condition. The least frustrating concept was the "highlight missed" concept followed closely by the "heat map" concept. The standard deviation is high for all the conditions except for the "heat map" concept, however only the "highlight missed" concept has a variation that goes below the baseline condition which could indicate that the concept is the least frustrating to use. The "highlight and disappear" concept was probably frustrating due to the way the highlights are activated and disappear, it is difficult to find a dwell-time activation threshold that suits every person.

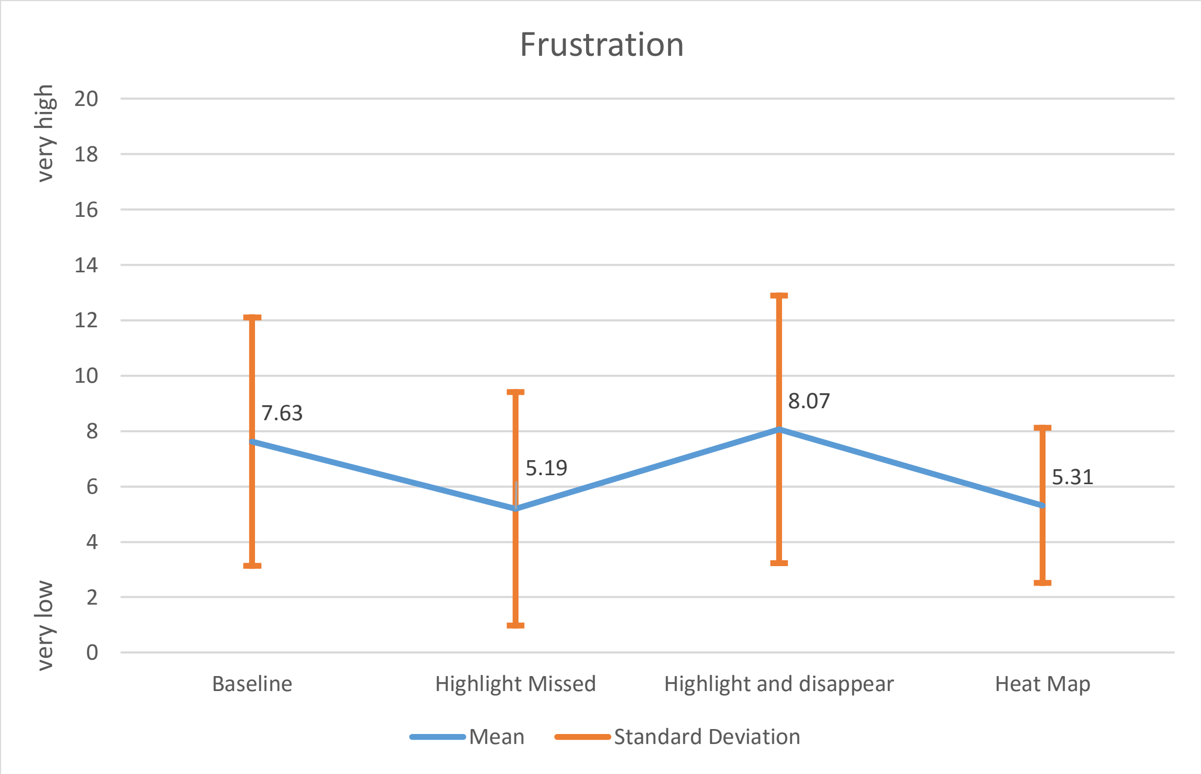


Figure 5.16 Nasa TLX question 5: How frustrating the participants found the tasks during each condition.

5.3.2 System Usability Scale

The modified system usability scale was used to see how the participants rate the supportive concepts in terms of complexity, usefulness, and likability. The questionnaire consists of nine statements that are answered with a number between 1 and 5, where 1 means “strongly disagree” and 5 means “strongly agree”. The statements contain the word system; the participants were instructed that system corresponded to the concept they had just finished during the experiment. The results were used to measure how the participants felt about the different supportive concepts. Note that the participants did not receive the system usability scale questionnaire after the baseline concept.

I think I would like to use this system frequently

The first statement was if the participant would like to use this system frequently, see Figure 5.17. The highest rated concept on average was the “highlight missed” concept, while the other two concepts were quite similarly rated. The “highlight missed” concept also had the lowest standard deviation indicating that the participants liked to use this concept the most.

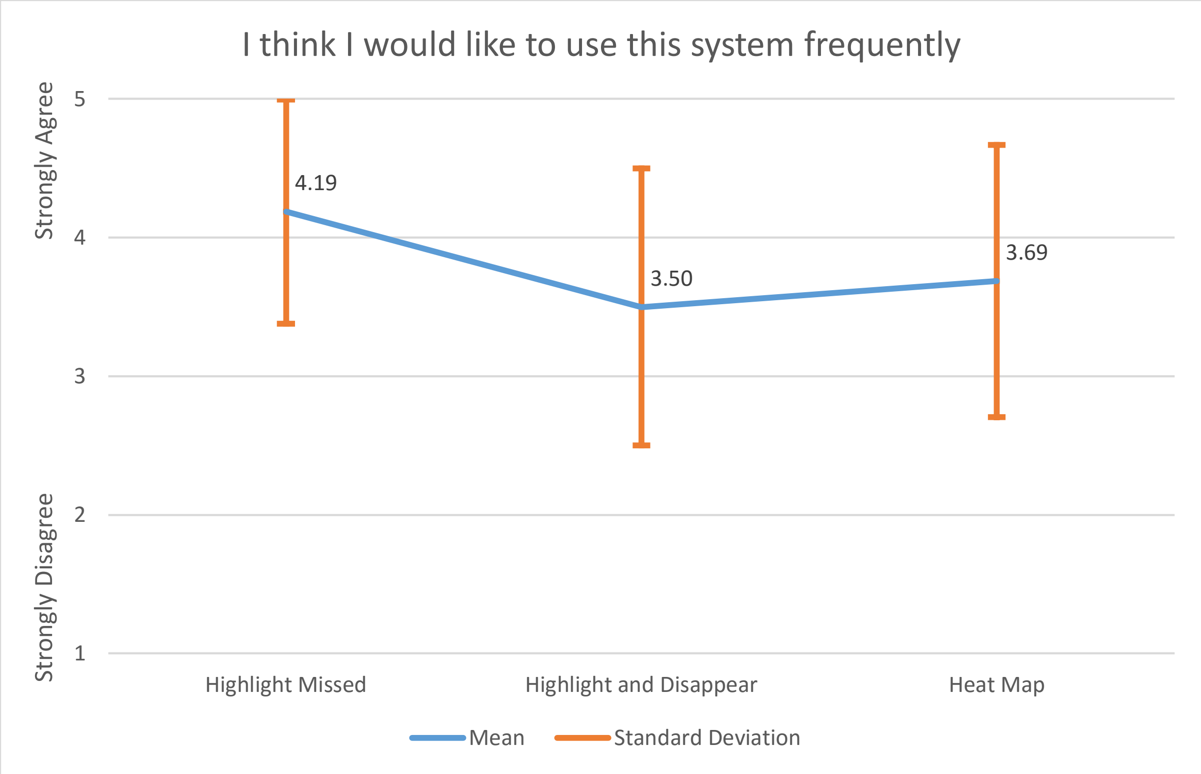


Figure 5.17 SUS question 1: How frequently the participants felt like they would use the supportive concepts.

I found the system unnecessarily complex

The second statement was “I found the system unnecessarily complex”, see Figure 5.18. None of the concepts were considered very complex. The average value for the concepts are similar, but the “highlight missed” concept has the lowest value and the lowest standard deviation, indicating that the participants found the concept the least complex to use. In terms of the concepts feedback the “highlight missed” concept provides simpler and less feedback than the other concepts which could explain the good rating.

I thought the system was easy to use

The next statement was “I thought the system was easy to use”, see Figure 5.19. The “highlight missed” concept was rated best on average; it also has the lowest standard deviation. The other two concepts were rated similarly, however their standard deviation is so large that it is unclear if they are less or more complex than the “highlight missed” concept. Since “highlight missed” has the least feedback and simple to interpret feedback it is as expected that the concept was rated easiest to use.

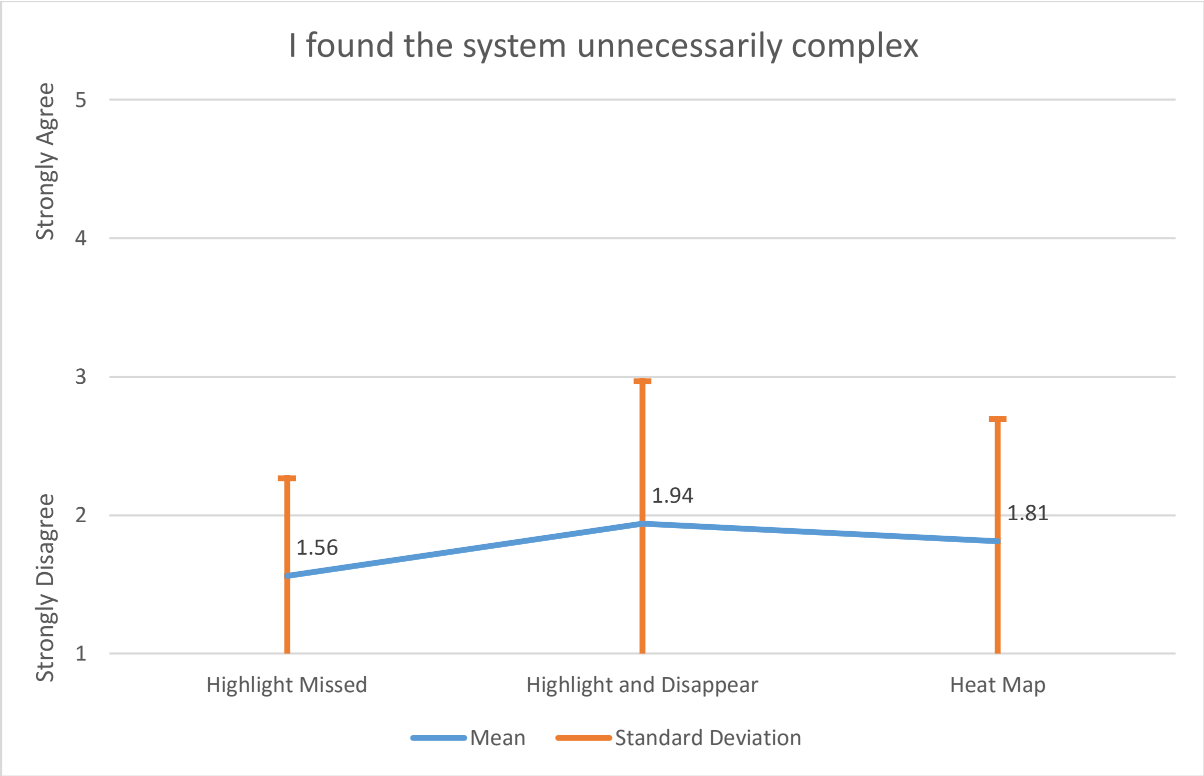


Figure 5.18 SUS question 2: How complex the participants found the supportive concepts to be.

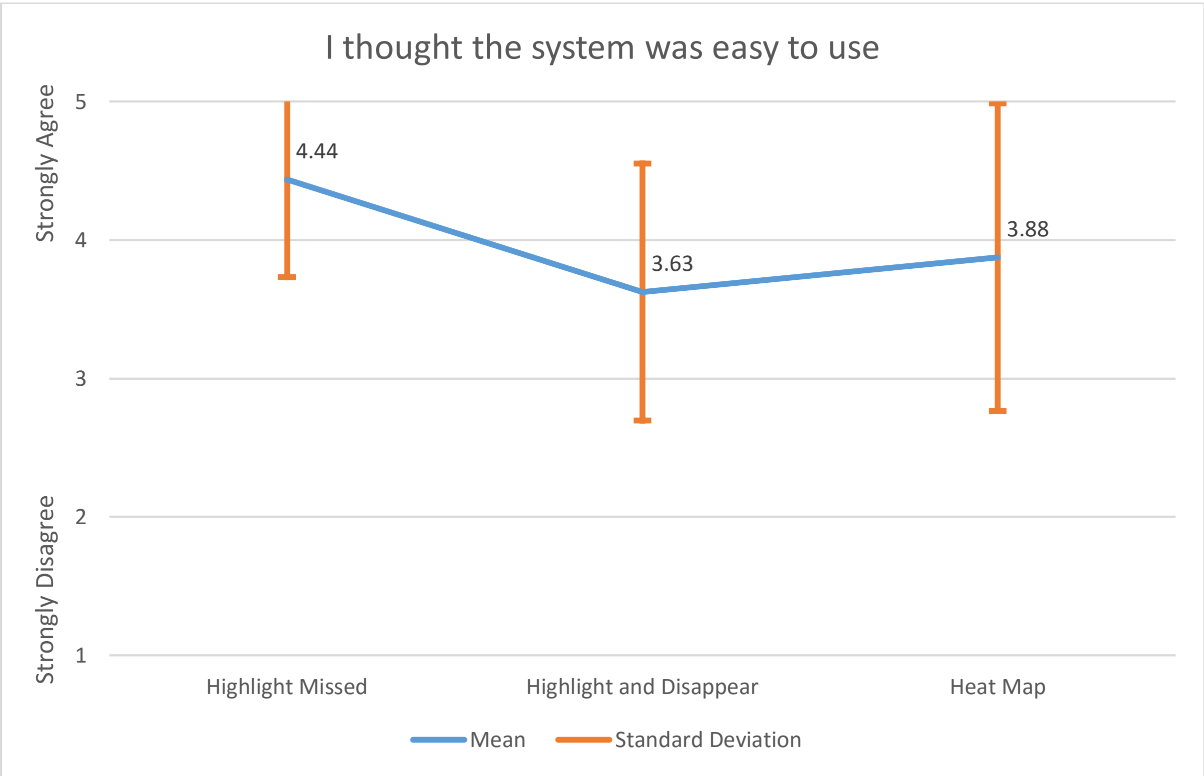


Figure 5.19 SUS question 3: How easy to use the participants thought the supportive concepts were to use.

I thought the system was enjoyable to use

The fourth statement was "I thought the system was enjoyable to use", see Figure 5.20. The "heat map" concept was rated the most enjoyable to use on average, followed closely by the "highlight missed" concept. The "heat map" concept offers feedback that is varied and colourful which might be why it is rated slightly higher than "highlight missed". Even though the "highlight and disappear" concept was last it was still fairly close to the other two concepts. The high variance of the results make it unclear which concept was the most enjoyable to use.

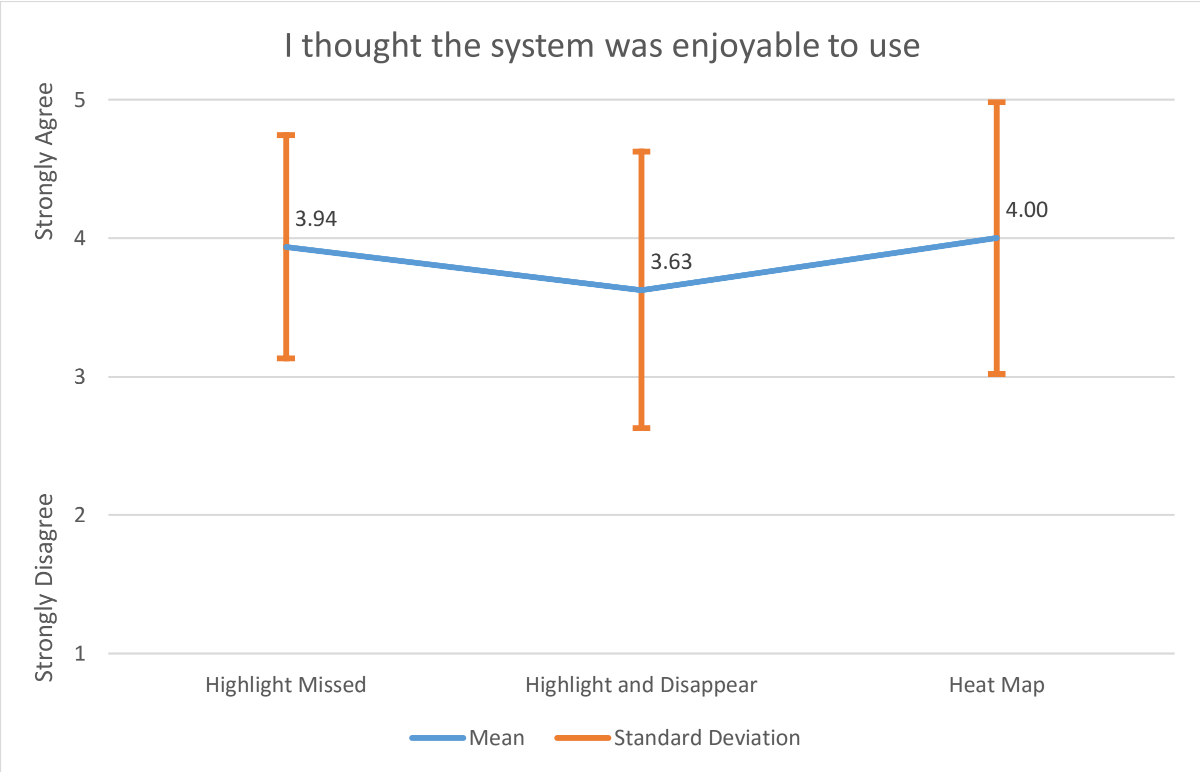


Figure 5.20 SUS question 4: How enjoyable to use the participants found the different supportive concepts to be.

I thought the system was useful and supported me in my task

The next statement was "I thought the system was useful and supported me in my task", see Figure 5.21. For this question all three concepts were rated fairly equally on average. The "heat map" concept has a very small lead over "highlight missed", but the standard deviation for "highlight missed" is smaller. The variation is big for all the concepts making it unclear which concept was the most useful and supportive, however since all the concepts rated highly on average it could indicate that they are all useful for a monitoring task of this nature.

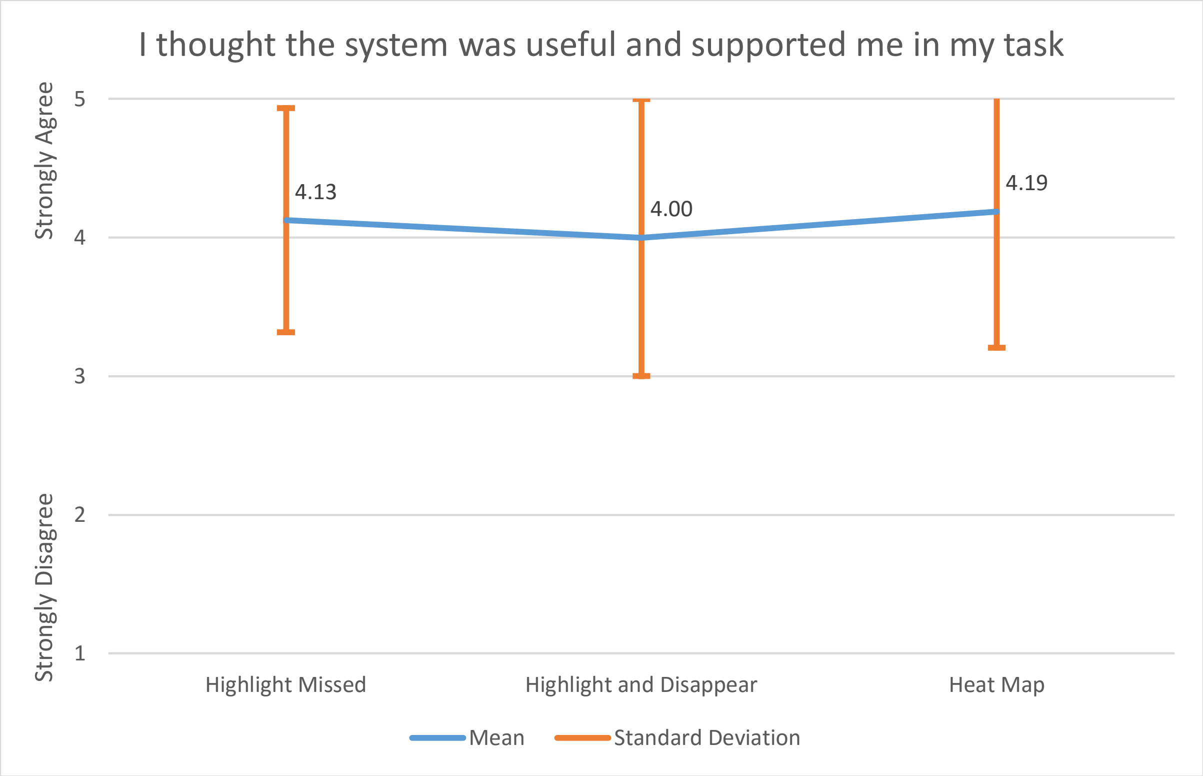


Figure 5.21 SUS question 5: How useful and supportive the participants found the different concepts to be during the task.

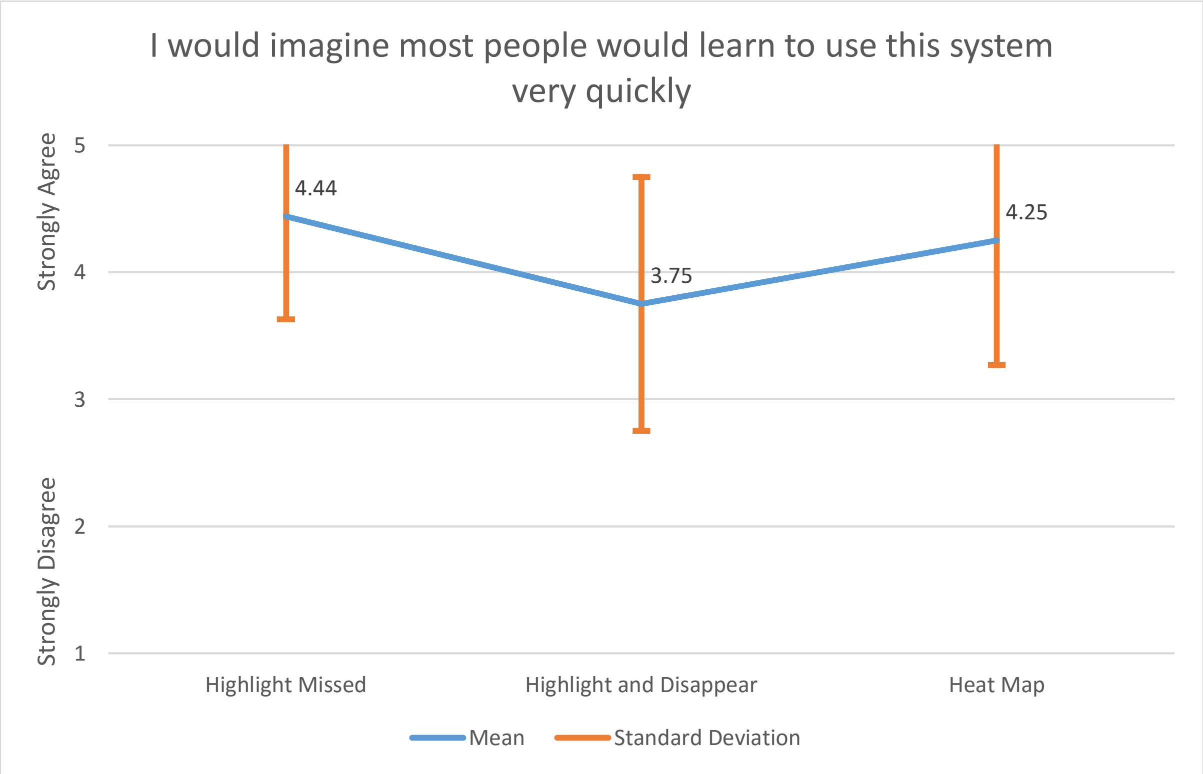


Figure 5.22 SUS question 6: How quickly the participants thought other people would learn to use the different concepts.

I would imagine most people would learn to use this system very quickly

The next statement was “I would imagine most people would learn to use this system very quickly” see Figure 5.22. The “highlight missed” concept is rated the highest in addition it has the lowest standard deviation, indicating that the concept was the easiest to learn how to use. The result is as expected considering the participants have to learn the least to use this concept. The “highlight and disappear” concept was rated lower than the other concepts, it is as foreseen considering that the participants actively interact with the concept. In the other concepts the feedback is static, making them easier to learn.

I found the system very cumbersome to use

The next statement was “I found the system very cumbersome to use”, see Figure 5.23. Again the “highlight missed” concept is rated best with the smallest standard deviation, indicating that more participants agreed that this concept was the least cumbersome to use. The “highlight and disappear” concept is rated worst, it is still rated between two and three meaning that it is not very cumbersome to use either. Considering the “highlight and disappear” concept changes the visual support during the task and that the participants had to learn more to use the concept it is as expected that the concept was rated most cumbersome to use.

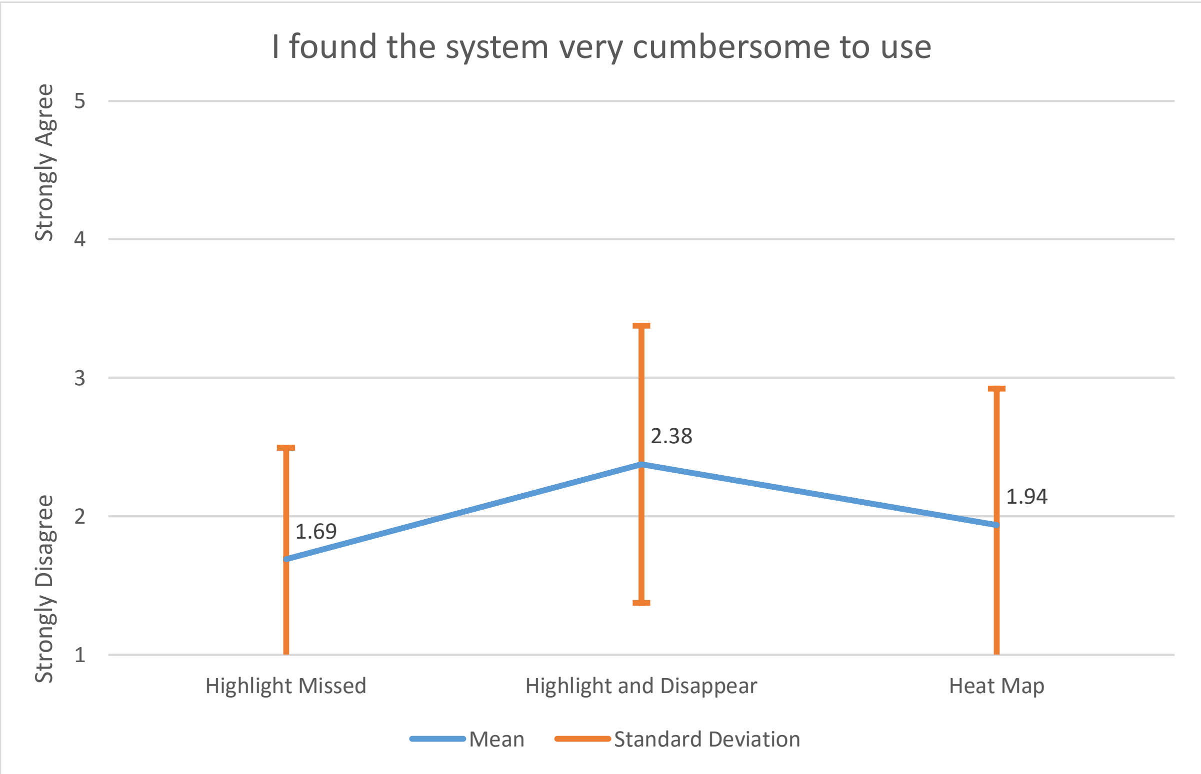


Figure 5.23 SUS question 7: How cumbersome to use the participants found the concepts to be.

I felt very confident using the system

Statement eight was “I felt very confident using the system”, see Figure 5.24. The “highlight missed” and the “heat map” concepts are rated the same on average, while the “highlight and disappear” concept is slightly lower. During the experiment it could see that some of the participants found the concept confusing, which could explain why it is rated lower than the other concepts. The large variations make the results unclear and make it difficult to determine which concept made the participants feel the most confident.

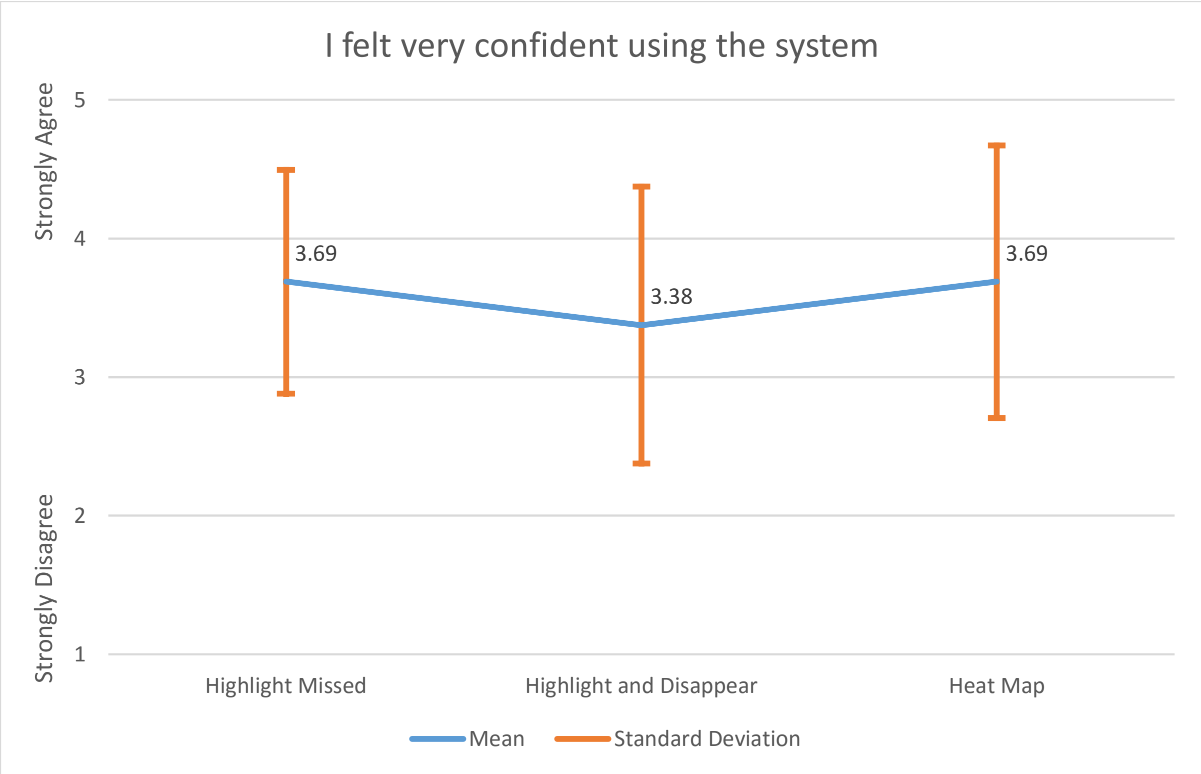


Figure 5.24 SUS question 8: How confident the participants felt using the different concepts.

I needed to learn a lot of things before I could get going with this system

The last statement was “I needed to learn a lot of things in order to get going with this system”, see Figure 5.25. The “highlight and disappear” concept was rated highest in terms of having to learn more, which is as expected since the concept is the only one that gives support actively during the tasks. The second highest rated concept was the “heat map” concept, this concept provides the most feedback and the feedback consists of different colours, as such it is as expected that the participants felt like there is more to learn compared to the “highlight missed” concepts. The “highlight missed” concept has the lowest average rating as well as the lowest standard deviation, indicating that the concept was generally agreed to be the easiest to learn to use. The other two concepts have a high variance making the ranking between the concepts unclear.

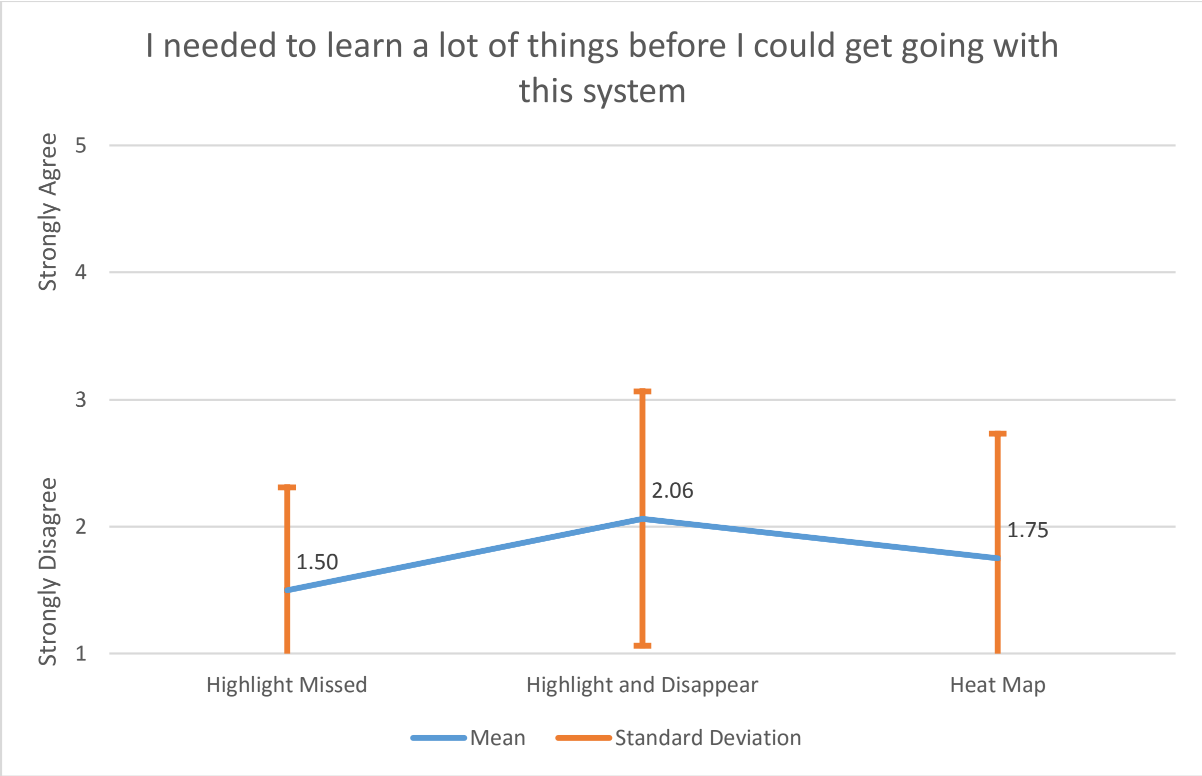


Figure 5.25 SUS question 9: How much the participants felt they had to learn before they could use the different concepts.

6 Discussion

6.1 Interview

6.1.1 Gaze interaction

In general participants liked the gaze interaction, they found marking numbers and activating buttons using their eyes to be easy and effective. Six of the participants did find the gaze interaction straining, but some of these six got used to it over time. Of the six participants that found the gaze interaction straining, only the participants that wore lenses found it straining throughout the process, this could indicate that the use of lenses makes gaze interaction more straining. Our results show that gaze interaction can be used as input in an interface as an alternative to using the mouse cursor when combined with the keyboard. In the case of monitoring and responding to events that occur, it seems that gaze interaction is faster as it allows the operator to keep focus on the task and respond to events without first having to locate the mouse cursor, and then move the mouse cursor which is slower than simply focusing with the eyes and pressing a button. However, a study that compares the two interaction methods should be used to determine if there is a difference and how big it is.

Accuracy

Accuracy problems seem to be quite common, only one of the participants did not experience any problems with the gaze interaction and the concepts feedback. The other participants had accuracy issues in one or two areas of the screen that were poorly calibrated. This led to the fact that most of the participants had difficulties marking numbers in these areas. In addition, some of the participants had problems activating the button in the top right corner of the screen used to proceed to the next phase of the experiment. This also affected the visual feedback accuracy in those areas of the screen. The calibration problem is a limitation of the eye tracking technology, but it could be avoided by improving the interface design. Most of the problem areas were in the corners or edges of the screen, by avoiding those areas when positioning elements some of the accuracy issues could possibly be negated.

Search strategy

As mentioned in 5.1.3, the participants used two different search strategies. A limitation of the system was that this possibility was not accounted for in the design of the concepts. In the "highlight and disappear" concept the "one number type" at a time search strategy had an effect on the systems performance, which caused confusion. While the participants checked one number type the highlights of the other number types disappeared as well, leaving the participants without support during the checking of the second number type. One possible solution to this problem is to colour code the highlights so the number types have their own colour, allowing the user to focus on one number type at a time without triggering the other number type.

Summary

Gaze interaction is fun and easy to use, it shows that gaze interaction can be used intuitively by most people without any significant training, especially if combined with the keyboard. The participants also felt that they were able to interact with the system quicker through the use of gaze interaction. As mentioned previously some participants found the gaze interaction straining, but they also got used to it over time. The main problem with the gaze interaction was that the calibration of the eye tracking system had to be very good in order to be able to use the system without problems, otherwise there would be areas on the screen where number marking or button activation did not work. In information heavy display screens components and numbers are usually close to each other, making it difficult to distinguish which number the user is looking at. This shows that the eye tracking systems accuracy is not yet ready to be used with information heavy display screens.

6.1.2 Supportive concepts

When evaluating the concepts, the participants focused on the time they spent during the concepts, the usability of the concepts, and the feedback provided by the concepts. In addition, the participants' search strategy and confidence when giving their answer was brought up. When asked to rate the concepts, the participants said that the "highlight missed" concept was the most helpful concept, with the "highlight and disappear" concept close behind. The "heat map" concept was rated as the least helpful concept by most of the participants.

Highlight and disappear

In the "highlight and disappear" concept the participants liked how the highlights reduced the amount of information they had to go through and that it supported them during the task. In addition, some of the participants felt like they finished the task more effectively during this concept. They disliked the timings for the activation of the numbers and the way the highlights disappeared. The concept did not fit the search strategy used by ten of the participants, who checked one number type at a time, resulting in some confusion. Most of the participants that used a "one number at a time" search strategy found this concept more useful than the other participants.

Our results indicate that the information reduction from the "highlight and disappear" concept can be useful in other areas as well, any task that requires the scanning of partial information on a screen could benefit from incorporating information reduction. The main limitations of the concept were that it did not fit the search strategy of several participants, as well as being experienced as confusing by many of the participants. Some of the confusion came from the activation of the highlights being too sensitive, with a way to calibrate the dwell-time activation threshold to the user of the system some of the confusion could be removed. A possible solution to the search strategy problem, colour coding, was described in the search strategies section of 6.1.1.

Highlight missed

The participants thought that the "highlight missed" concept gave concise and easy to understand feedback. Half the participants liked that the concept gave a sense of confidence that everything had been checked before they gave their answer. Some of the participants felt less confident when the concept did not highlight any numbers, they did not know if the system was working or if it had failed. Many of the participants experienced that the system highlighted numbers that they had inspected, two of the participants said that it was not disruptive and they could quickly check the numbers and continue with their task.

That the participants felt less confident when giving their answer if no numbers were highlighted is a weakness of the “highlight missed” concept. If the user is questioning if the system is working, they are spending energy on something that is not related to their task. This problem could occur regardless of which domain or what task the user has to perform. A possible solution would be to give the user a message “You did not miss any numbers, good job”, to reassure them that the system is working. Highlighting important information that has been forgotten should be useful in many domains, especially when the task consists of monitoring.

Heat Map

The participants liked that the “heat map” concept always gave feedback. They found it useful to see how time was spent during the task, and that they could recheck numbers that had not been looked at for long. Some of the participants thought the concept gave too much feedback, they had to spend time rechecking numbers in red areas even though they were already confident in their answer. Many of the participants felt that this concepts tasks took the longest, this correlates to the amount of time used to complete the concept logged by the system.

To get an overview of what was looked at and for how long it was looked at should be useful in any training situation. In addition, it could be useful to make sure that all the important components the user has to check are looked at long enough. For a task as simple as the one the participants performed the heap map probably contained too much information, since the participants only spent seconds looking at numbers a gradient of colours between red and green might be too many colours. Perhaps it would have been better to use two or three colours instead.

Summary

The negative feedback from the “highlight missed” and the “highlight and disappear” concepts are mostly related to limitations of the prototype system itself. The participants found the “highlight missed” concept the most useful during their task. They were more confident when answering the tasks during this concept. The “highlight and disappear” concept was rated the second most helpful, the participants felt more effective during their task, and liked the information reduction the concept provided. This concept had some difficulties in terms of activation times, the value set for the system did not work well for every participant. It also had problems with the way the highlights disappeared. With further development and testing this concept could become the most useful concept, as it provides support during the task. Both concepts should be considered for further development.

The “heat map” concept was found to be the least helpful concept. The participants liked the information they received, but thought it was too much information. As mentioned above, the “heat map” concept might be too complex for a simple task like the number checking task, however it could be more useful in complex monitoring tasks, or dynamic tasks where the information changes over time. As such it could still be useful to further develop the “heat map” concept as well. Lastly, it would be interesting to adapt some of the ideas the participants had to improve the concepts and make the “highlight missed” and “heat map” concepts provide support during the task as well.

6.2 Quantitative

6.2.1 Logged data

See Table 2 in section 5.2 for an overview of the data that was logged by the system. The only parameters that showed significant differences in the results were the number of distractor activations (3.09E-08, $p < .05$) and distractor number view time (2.01E-07, $p < .05$). Some of the other parameters show indicative differences for one or two of the concepts when compared to the baseline condition. The other results are affected by a high amount of variance, making them non-significant.

Heat Map

The participants missed fewer out of bounds numbers in the "heat map" concept than in the other conditions. Due to the low error rate for all the conditions and because the standard deviations are very high, the result were deemed insignificant. The reason for the high variance and standard deviation could be that the task the participants had to perform was too simple. The participants spent significantly more time on finishing the tasks in this concept, this is most likely due to the amount of feedback the participants receive and have to process before giving their final answer. This also corresponds to what the participants said in the interview, where they felt it took longer to complete this concept. From the logged data it can be seen that this concept increased the completion time without reducing the number of errors significantly, this indicates that the concept might not be suitable for supporting the user during the task. It could still be suitable in a training situation however, as it gives an overview of how the user spends their time during the task.

Highlight and disappear

From the perspective of efficiency, the "highlight and disappear" concept performed better than the other conditions. The participants completed this concepts tasks slightly faster than during the baseline condition, this matches the findings of Booth et al. (2013). The amount of numbers that were looked at and how long they were looked at is lower for this concept. In addition, the number of distractor numbers that were looked at and how long they were looked at are significantly lower compared to the other conditions. This correlates to what most of the participants said during the interview, that they felt more effective during this concepts tasks. The concept did slightly worse than the baseline when comparing the amount of wrong numbers, the standard deviation however is so big that the amount of wrong numbers is insignificant. The large standard deviation could mean that the task was too simple, or that there were not enough testers to get a clear result. Based on the logged data the "highlight and disappear" concept is the most effective at supporting the user during the task, the results are as expected since this concept is the only one that gaze support from the beginning of the task. In addition, the "highlight and disappear" concept also limited the amount of numbers the participants had to check. Even though the completion time is only slightly faster than for the baseline condition, the time is spent better as the focus is mostly on target numbers and the safety ranges.

Highlight Missed

The "Highlight missed" concept has a higher completion time than the baseline condition, but due to a high standard deviation more testing is required to get clearer results. Again the amount of wrong numbers is slightly lower than the baseline condition, but due to the high variance the values are insignificant. Further testing with more testers and a more difficult task is necessary to achieve clearer results. Both the amount of number activations, as well as the number view time is slightly higher than the baseline condition, though they are not significantly higher. The target number activations and view time are slightly higher than in the baseline condition, as expected because the concept highlights missed target numbers. The concept also has the highest amount of safety ranges activations and view time, though once again the standard deviation is high making the results unreliable. The higher values are as expected considering the participants get the chance to recheck highlighted numbers and would need to check the safety ranges again. From the logged data alone the concept does not appear to be very useful in supporting the participants with their task.

Search strategy

The logged data was split on the participants search strategies. The participants using the “one number at a time” strategy had a slightly lower completion time, a lower amount of wrong numbers, due to relatively large standard deviations the results can only be seen as indicative. Further investigation is required to determine if the “one number at a time” search strategy is in fact better than the other. Lastly they had slightly less number activations, both in total and split on target and distractor numbers. They had significantly more safety ranges activations and viewed the safety ranges longer, even so they completed their task faster. The standard deviation for the participants using the “one number at a time” search strategy was significantly lower in almost all the logged data except for the safety ranges activations and view time. The results reveal that the “one number at a time” search strategy could be more suitable for this kind of tasks than the “one number type at a time” search strategy. It could be useful to investigate the difference between the search strategies performance more extensively in the future.

Limitations

The validity of the gathered data cannot be guaranteed as there are inconsistencies due to the dwell-time threshold, the results can still be treated as indicative. The value of the dwell-time activations was set to a value that worked well during the testing of the system during development, however every person processes information at a different rate. If a participant is able to process the data quicker than the dwell-time threshold the system would be unable to register that the participant looked at a number, this is a limitation of the system itself. By improving how the dwell-time threshold is set so that it adapts to the current user, the validity of the data can be improved, not to mention this being a very useful feature in an actual eye tracking system, no matter the application. In addition, there were accuracy issues for most of the participants, resulting in some poorly calibrated areas which affected the systems registering of number activations and view time. These issues arise from the limitations of the eye tracking technology.

Most of the results had a high amount of variance, with a bigger test population the results should become less sensitive to the outlier values. Since the average amount of errors the participants made were so low for each condition it could indicate that the task was too easy. If the task was more difficult participants would possibly make more errors, which would make it easier to detect differences between concepts.

6.2.2 Questionnaires

The participants' answered the questions in the modified Nasa TLX questionnaire with a high amount of variance, the mean value of the answers are also close to each other for every condition. The supportive concepts scored better than the baseline in most of the questions, except for the "performance" and "frustration" questions, where the "highlight and disappear" concept scored slightly worse than the baseline. Due to the high amount of variance it is difficult to say anything conclusive about the results.

The modified system usability scale questionnaire answers were similar in values and had a high amount of variance. In general, only the highest rated concepts have a difference that is bigger than the variance, the lower ranked concepts are usually similar in value and have a high amount of variance. It is not possible to determine which concept is the best in most of the questions, except for the "I think I would like to use this system frequently" and the "I needed to learn a lot of things before I could get going with this system" questions which both showed that the "highlight missed" concept was better.

It could be that the task the participants had to perform was too simple to create any big differences between the concepts, or that the high amount of variance is due to the number of participants being too small. Another possibility is that the selected questionnaires were not suitable in this context.

Highlight and Disappear

In the modified Nasa TLX the "highlight and disappear" concept was rated the worst in five of the six questions. In the "performance" and "frustration" questions the concept was rated worse than the baseline condition, the results are understandable as this concept is the most intrusive of the supportive concepts. However, the variance is high that the results are not conclusive. With the results split on the participants search strategy, the participants using a "one number at a time" search strategy rated the "performance" of the concept better than the baseline, the "frustration" level was still worse than in the baseline. The "mental demand" of the participants also varied with their search strategy, the participants using a "one number at a time" search strategy had a lower mental demand than the participants using the other search strategy. This indicates a limitation of the concept when the user has a different search strategy than a one number at a time strategy. The other questions did not show any big differences when split on the participants search strategy.

In the modified system usability scale the "highlight and disappear" concept scored worst in all nine questions. This indicates that the participants thought the concept was the most difficult to use and that it did not support them well during the task. The participants were undecided about how helpful the concept was in the interview, where five participants plus three participants, that could not decide between the "highlight missed" and the "highlight and disappear" concepts, rated it most helpful and five rated it least helpful. This implies that the concept has potential if the negative feedback such as the activation time and the way the highlights disappear can be corrected. The variance of the answers is big however, and in many of the questions the concept rates close to the second best rated concept.

Highlight Missed
The “highlight missed” concept was rated the best in five of the six questions in the modified Nasa TLX and seven of the nine questions in the modified system usability scale. The participants found the concept the least mentally demanding and they achieved the best performance with the least amount of effort. The participants found the concept the easiest to learn how to use and rated it to be the concept they would use frequently. They also found the concept to be the least complex, cumbersome and frustrating to use. The participants felt confident when giving their answer during this concepts tasks. The concept scored well in both the questionnaires and the scores correlate well with what the participants said during the interview, where they ranked this concept to be the most helpful. The variance of the participants’ answers is lower compared to the other concepts, which could indicate that more participants agree that this concept is the most beneficial.

Heat Map
The participants rated the “heat map” concept to be the least fatiguing for the eyes, which is unexpected considering the concept gives the most visual feedback. The concept was also rated as the most enjoyable to use and the most useful and supportive during the tasks. In the rest of the questions the concept was rated slightly better than “highlight and disappear”. The participants also felt more confident in their answers during this concept. Despite scoring better than the “highlight and disappear” concept in all but one of the questions, most of the participants still rated this concept as the least helpful in the interview. This could indicate that the participants liked to use the concept, but that it was not helpful during their tasks.

Limitations
The high variance and the similar mean values of the participants’ answers make it difficult to analyse the data, most of the results are not clear enough to draw conclusions from. A bigger population sample could reduce the high variance of the results. A more difficult or realistic task could create a bigger difference between the different conditions so that clearer results can be obtained. In addition, the rating scale of the system and usability scale is only between 1-5, meaning that the distance between answers from participants that are unsure about their rating becomes quite large. A bigger scale would shorten the distance.

6.3 Research question 1: How can data gathered from eye tracking be used to support users with visual feedback during a number checking task?

The first research question has resulted in the creation of three supportive concepts, “highlight and disappear”, “highlight missed” and “heat map”. The “highlight and disappear” concept provides support during the task (feedforward) while the two concepts “heat map” and “highlight missed” provide support after the task (feedback). The concepts show three different ways in which eye tracking can be used to support a user during a number checking task through the use of visual cues.

The concepts supported the participants in different ways, the “highlight and disappear” concept was able to support the participants by increasing the efficiency at which they completed their tasks. “Highlight missed” supported the participants by making them more confident in their answers since they could be sure every number had been checked. Lastly, the “heat map” concept supported the participants by giving detailed information of how they spent their time looking at the numbers, the participants could then recheck areas that were not looked at for long. A comparison of the different conditions and their rankings can be seen below, in Table 5.

Table 5 A comparison table that shows the conditions average completion time, average number of errors and the participants ranking of the conditions. The rankings were created by counting the highest and the lowest scoring concepts.

Condition	Completion time	Number of errors	Interview ranking	Nasa TLX ranking	SUS ranking
Baseline	3min 40sec	1.19	-	-	-
Highlight & disappear	3min 27sec	1.31	2	3	3
Highlight missed	4min 23sec	1.00	1	1	1
Heat map	5min 6sec	0.69	3	2	2

Highlight and Disappear

The concept that was most successful at supporting the user with their task was the “highlight and disappear” concept. The concept reduced the time it took for the participants to complete their tasks, and it reduced the time spent viewing distractor numbers by a significant amount. However, it was also the most difficult concept to use, as it was confusing for many of the participants. Ten of the participants used a “one number type at a time” search strategy which the concept did not support very well. While the participants checked one number type the highlights of the other number type also disappeared, leaving the participants without support while checking the second number type. Another confusing aspect of the concept was the activation time of the highlights and the time it took for the highlights to disappear, which could result in the participants losing track of what they were doing. It can be induced that these are the reasons for the concepts low ranking in both the questionnaires. Even with these issues five of the participants and three participants that could not decide between this concept and the “highlight missed” concept rated it as the most helpful during the interviews.

Highlight Missed

“Highlight missed” was rated as the favourite concept by the participants in the interview as well as in both the questionnaires. From the logged data however the concept did not support the participants in any significant ways. The participants found the concept easy to use and understand, in addition many of the participants felt more confident when giving their answers during this concept as they could be sure that all the numbers were checked. Some of the participants however felt less confident when giving their answers if none of the numbers were highlighted, they did not know if the system was working or not. By providing the users of the system with a message or a visual cue that the system has registered where they are looking would remove this limitation of the concept. Helping the users of the system complete their task with confidence is a valuable supportive tool, both in training and in real work.

Heat Map

The “heat map” concept was rated as the least helpful concept by nine of the sixteen participants during the interviews. Even though the concept provided detailed information about how the participants spent their time during the tasks, they did not feel that the information was very helpful in supporting their task. The participants said that some of the numbers that were not looked at for long were easy to check and that felt they had to recheck the red areas even though they were confident in their answer. They also said that they felt like they spent more time on the concepts tasks compared to the other conditions, this also matches the logged data where the participants' completion time is significantly higher than for the other conditions. The participants' answers and the logged data indicates that the concept was unable to support the participants' during their task. The concept could still be useful in a training situation, the user of the system could check how they spent their time and if they should make changes to their work flow for the next time. Or the heat map could be checked by experts who could provide suggestions for improvements to the user.

Concept Improvements

By fixing the timing issues causing confusion in the "highlight and disappear" concept, and making the concept able to support different search strategies, the concept could become the most useful concept, able to provide support during training or in work situations. The "highlight missed" concept can be improved by adding visual indications of when the system has registered that the user is looking at something, this would make sure that the user knows if the system is working or not during the task. The concept would then be able to support the user by making them more confident during their task. The "heat map" concept turned out to not be very supportive during the tasks, it could still be useful in training situations however, as it provides an overview of how time was spent and how the user can make changes to increase the efficiency of their work flow. The heat map can also be used by the trainee's supervisor to provide suggestions for improvements to the trainee.

Limitation

After the first pilot test the "highlight always" concept was removed due to bad feedback from the test participant and since the concept did not use the eye tracking information to provide feedback. In hindsight it would have been better to retain the concept during the second pilot test and the user study itself, as it would have provided another metric to compare the concepts that use eye tracking against a concept that provides similar support without the need of an eye tracker.

Summary

In summary three supportive concepts using use the eye tracker were created and tested, two of the concepts "highlight and disappear" and "highlight missed" were able to support the participants during their tasks by increasing the efficiency of the participant or by increasing the confidence of the participants' answer. The last concept "heat map" was unable to provide useful support during the task, but it could still be useful in training situations. All three concepts have potential and should be further developed to remove some design limitations and increase their usefulness.

6.4 Research question 2: What are the difficulties of using gaze based interaction with an information heavy display?

Throughout the development and testing of the prototype several difficulties were encountered, they are listed below.

- The calibration of the eye tracking system
- The accuracy of the prototype application
- The size and positioning of the AOIs
- The size and visualisation of the gaze cursor
- Which gaze based interaction methods to use
- The dwell-time threshold
- The “Midas Touch” problem
- Different search strategies

The calibration of the eye tracking system is a known difficulty, and factors such as glasses, lenses, mascara and how wide the eyes open can affect the calibration of the eye tracking system (Holmqvist et al., 2011, p. 141). The accuracy of the prototype application was affected by the accuracy of the eye tracking system, that is to be expected since it relies on the location of the user’s gaze. The gaze interaction methods used in the prototype application were discovered in the literature review, no new interaction methods were used.

It is to be expected that people use different search strategies, it was a limitation in the supportive concept design that resulted in the search strategies becoming a problem. The “Midas Touch” problem is well known and several examples of it was found during the literature review, even though the prototype application used interaction methods which attempt to avoid the problem it still occurred due to dwell-time activations.

Different values for the AOI sizes were found in the literature review and the prototype application used a value that was smaller than the related papers. The positioning of AOIs were not specifically mentioned in relation to each other in the reviewed literature, the prototype system uses a collision detection algorithm that allows for close positioning of AOIs, see 3.4. In the reviewed literature it was not clear how the systems used the gaze point of the user, if they used the point provided by the eye tracking system or if they used a circle as the prototype system does.

Calibration of the eye tracking system

There were difficulties when calibrating the eye tracking system, especially when the participants wore glasses as expected from the literature review, see the limitations section in 2.1. The participants that wore bigger squared glasses seemed to be able to calibrate better than the participants wearing smaller skinny glasses. In some cases, the participants wearing skinny glasses had to remove them in order to achieve a proper calibration, leading to a more straining experience. Almost all of the participants, including the ones that wore glasses, experienced some poorly calibrated areas of the screen, especially in the corners and edges of the screen, even though the calibration results were quite good. It could be that the optical resolution of the prototype system was too low (number level) compared to the accuracy of the eye tracking technology, and that the problem areas would be less noticeable with a higher optical resolution (e.g. component level).

The accuracy of the prototype application

The calibration level had implications for the accuracy of the prototype system. For almost all the participants there were one or two areas of the screen that were poorly calibrated. This resulted in participants having difficulties marking numbers and activating buttons in those areas. In addition, the data gathered by the prototype system is affected by this as the system cannot accurately register when the participants are looking at numbers in those areas, resulting in data that cannot be completely trusted and inaccurate feedback being provided to the participants, such as false highlights in the “highlight missed” concept. The use of gaze interaction helped reveal accuracy issues that otherwise could have been unnoticed.

Area of interests

As mentioned in section 3.4 Gaze enabled interface, the sizes of the AOIs were changed from 38pixels to 76pixels through testing by the Author. An increase of 6 pixels was repeated until the accuracy of the system was good enough to register all the interactions of the Author. The final size of the AOIs is smaller than the sizes used in the related literature, Cantoni and Porta (2014) used 190pixel wide squares for their hotspots and Putze et al. (2013) used 100pixel size AOIs. The system was still able to register all the interactions when used by the Author and one of the participants. Another participant had only one problem marking a number, and the second time it worked. This shows that a small AOI can be used if the calibration of the eye tracking system is good.

The positioning of the numbers and their AOIs were not altered beyond the changes described in 3.3 Prototype Design, where some numbers were removed and two numbers were moved slightly apart from each other. One thing to consider when positioning the AOIs is the possibility of the gaze cursor overlapping more than one AOI, if a close proximity between AOIs is required the distance collision check solution used by the prototype application as described in section 3.4 can be adapted. The closest distance between two AOIs in the prototype was 24pixels. Some of the participants experienced that the system was unable to distinguish between the numbers AOIs at this distance. When the system was used by the author the issue was not experienced.

Gaze cursor

The sizes of the eye interaction point was not clear from the literature review, so a small radius (5pixels) was selected as the starting point for the gaze cursor used in the prototype system. The radius of the gaze cursor was increased by 2.5pixels until it became 25pixels, at which point the accuracy of the system became reliable for the Author.

As mentioned in section 3.4 the gaze cursor was visualised to show the user of the system where the gaze cursor was located. However, an unexpected finding occurred during the Authors testing of the system. Due to the movement of the gaze cursor it draws the attention of the eyes, and the imperfect calibration of the eye tracking system causes an offset from the point the gaze cursor is visualised and the point the user is actually looking at. When the user is looking at the gaze cursor it moves in the direction of the offset, making the eyes follow it around the screen.

Gaze interaction methods

The literature review uncovered different methods to interact using gaze as input (See section 2.2.2 gaze based interaction examples). One method was to look at hotspots to activate them, immediately or by dwell-time activation. Another method was to activating a sticky pointer by fixating at an area, the sticky pointer activates the area until it is cleared. Lastly, the combination of gaze input with other input devices such as the keyboard can be used to activate by pressing a button or to toggle gaze input on and off.

Which interaction method that is appropriate to use depends on the needs of the application. If the application is simple and gaze is the only intended input it could be suitable to activate immediately after looking at an object, or use a dwell-time activation in case the “Midas Touch” effect is harmful for the application. If the activation has to be quick but controlled, it is useful to combine it with another input-device, or to use a sticky pointer. If gaze is not the primary input but is used occasionally it would be suitable to activate and deactivate the gaze input via another input device. The prototype application shows that the interaction methods can be combined to accomplish different goals.

The prototype system uses two gaze interaction methods, the first was dwell-time activation to determine when a number was looked at for the purpose of feedback and logging. Secondly, gaze combined with the keyboard was used to mark numbers as out of bounds by looking at them and pressing the “space bar”. Initially dwell-time activation was used to activate the buttons by looking at them, but the visualisation of the dwell-time activation caused the “Midas Touch” effect to occur. The interaction method to activate the buttons was changed to match the marking of the numbers to avoid the “Midas Touch” effect in the testing of the prototype. The Author would recommend the combination of gaze and another input device interaction method when the goal is to trigger activations, as the participants found this interaction method easy and fun to use.

Midas Touch

As mentioned above the “Midas Touch” effect was triggered when the buttons were activated with dwell-time, in addition the effect was triggered in the “highlight and disappear” concept. Some of the participants accidentally activated highlights when exploring the screen, similar to the accidental activations in the maze-game made by Krejtz et al. (2014), where the users accidentally activated movement commands while scanning the screen. Even though interaction methods that counter the “Midas Touch” effect were used for the active interaction the effect still occurred for the passive interaction of the system, looking at numbers. This occurrence of the “Midas Touch” effect was due to the dwell-time activation.

Dwell-Time
The dwell-time threshold was through testing of the system, a value that worked well for the Author was selected (300milliseconds). In addition, the time before the highlights disappeared in the "highlight and disappear" concept was set to 1.5seconds which matched the Authors information processing speed. This time allowed the Author to check the number against the ranges before it disappeared. After the second pilot test the value was changed to 2seconds based on the feedback from the participant.

During the testing of the prototype system the participants gave conflicting feedback, the activation of the highlights (dwell-time activation) in the "highlight and disappear" concept was too quick for some participants, and it was too slow for another participant. This shows that every person has different information processing speeds, making it difficult to set a value that works for everyone. If the value is set so low enough to work for the fastest people it would cause accidental activations for people with slower processing speeds. If it is set high enough to work for the people with slower processing speeds it would force the faster people to focus longer than necessary in order for the system to register their interactions. As mentioned earlier a way to calibrate the threshold to the user of the system would be very helpful, to increase the usability of the system, to make the data more valid and not to mention that it would be a useful feature for any dwell-time based eye tracking application.

Search strategy
The interviews with the participants unexpectedly revealed that they used two search strategies. As mentioned before, the "highlight and disappear" concept did not work well with one of the strategies. When designing support and interaction systems using eye tracking it is important to think about the different search strategies the users can employ, and the effects the different search strategies would have on the system. To make sure the system functions at all times a particular search strategy can be enforced, or the system can be made in a way that handles the different search strategies.

Summary
In summary several difficulties were encountered during the testing and development of the prototype application, some of the difficulties can be solved, such as the calibration of the dwell-time threshold so It matches the information processing speed of the user. This would also help reduce the "Midas Touch" effect as the timings can be tailored to the user of the system. In addition, the search strategy of the participants can be investigated using pilot testing, the system can be made to fit all search strategies, the most popular ones or enforce the use of one strategy.

Other difficulties such as the calibration of the eye tracking system require the eye tracking technology itself to improve before it can be solved. The improvement of the eye tracking technology would also reduce the accuracy issues of the prototype system, and make the visualisation of a gaze cursor more helpful, as it would remove the "cat and mouse" effect it produces with an imperfect calibration. Lastly it would increase the flexibility of the AOIs, as they can be defined and positioned with less limitations due to the accuracy of the eye tracking system.

7 Conclusion

The study explored how the novel eye tracking technology could be used to support nuclear power plant (NPP) control operators during their simulation training. The simulation training is used extensively, as such it is of great value to find ways to support the control operator during training in order to increase the efficiency of the training, better the learning process of the training or increase the operator's confidence during tasks.

A prototype application with flexible functionality for interface features and detailed human performance measurement functions was created as a testing platform around a simplified number checking task based on one of the tasks NPP operators perform, namely number monitoring. Many design decisions were made along the way, the methods of gaze interaction used, the size and positioning of the area of interests (AOI), and the value of the dwell-time threshold. General ideas of how to use the data from the eye tracking system to support the users during the number checking task were developed into concrete supportive concepts. The prototype application and the supportive concepts were developed using an iterative process with three iterations and two pilot tests before the application was ready for user testing. A user interface testing process was created, which included an automatic counterbalancing of the order the supportive concepts were given. A user study was performed with sixteen participants recruited from a diverse group of test subjects. Quantitative and qualitative data was gathered from the study, analysed and reported.

The study also found and highlighted difficulties of using gaze based interaction with an information heavy display screen throughout the development and testing of the prototype application. The difficulties were listed and discussed, some of the difficulties were solved during the study, while other difficulties require more development or the improvement of the eye tracking technology itself in order to be solved.

The study sought to answer two questions:

1. *How can data gathered from eye tracking be used to support users with visual feedback during a number checking task?*

To address this research question, three supportive concepts were created and tested, "highlight and disappear", "highlight missed" and "heat map". The concepts use data from the eye tracking system to support the user with visual cues in the form of highlights. All three concepts have in common that they aim to support the user by showing which numbers have and have not been checked, so that the users can more easily recognise when a number was forgotten on an information heavy display screen. The gathered data showed that two of the concepts were able to directly support the participants, "highlight and disappear" supported the participants in terms of *efficiency,* and "highlight missed" increased the *confidence* of the participants. The last concept "heat map" was unable to directly support the participants with their task.

The tasks of the "**highlight and disappear**" concept were completed slightly faster than in the baseline condition, in addition the participants were able to focus on the target numbers, spending significantly less time looking at distractor numbers compared to the other conditions. However, there were problems with the concept. Many of the participants found it confusing to use, and the participants' different information processing speeds meant that many of the participants' found the dwell-time activation too fast or too slow. The increase in efficiency make it valuable to further develop this concept.

The "**highlight missed**" concept increased the confidence of the participants during their tasks, they felt more confident since they knew that they had checked all the relevant numbers by the time they gave the final answer. The concept produced false highlights for many of the participants due to accuracy issues, and because the dwell-time threshold did not match the processing speed of the participant. The concept increased the completion time of the tasks by a small amount, but if it makes the user of the system feel more confident in their performance it is arguably a negative that is outweighed by the positive.

The "**heat map**" concept was rated as the least helpful concept by nine of the sixteen participants. The participants did not feel that the overview of how they spent their time checking the numbers was helpful in supporting them during the tasks. The participants' answers from the interview and the questionnaires in addition to the logged data indicates that the concept was unable to support the participants' in their task, but that it did increase the completion time significantly instead. The concept does provide an overview of how the user spends their time during a task however, and as such could be useful in a training context where the information can be used to improve the work flow in the future.

While there are many ways in which eye tracking technology can be applied to support users with information-heavy tasks, three concepts were defined and explored in this thesis. Two of the concepts, "highlight and disappear" and "highlight missed" were shown to be effective, and the third concept "heat map" has potential as a training feedback tool. These concepts show that eye tracking can be used to support users during a number checking task, and potentially similar tasks. They also form the ground work for future studies that wish to employ these specific concepts. There were difficulties encountered in the process however, which leads to the second research question.

2. *What are the difficulties of using gaze based interaction with an information heavy display?*

The calibration of the eye tracking system: Only one of the participants was able to achieve a calibration with the eye tracker that was good enough to not experience any difficulties with the gaze interaction. The other participants experienced difficulties with the gaze interaction due to calibration issues. In order for the eye tracking technology to become more viable the calibration process needs to be improved so that people who require visual aids (e.g. glasses) can achieve a good calibration as well. Based on this finding, current eye tracking technology cannot be recommended for daily use in a training context.

The accuracy of the prototype application: The participants experienced false feedback from the prototype system due to two difficulties, the first difficulty is the same as the one mentioned previously, the calibration of the eye tracking system. Many of the participants experienced poor calibrations in the corners and edges of the screen, one way to limit these issues could be to avoid placing small AOIs in these areas of the screen. The second difficulty was caused by some of the participants' having a higher information processing speed than the dwell-time activation threshold. In both cases the system was unable to register numbers as looked at and increase the time they were viewed. The dwell-time activation difficulty can be solved with further development, one approach could be to have the user of the system focus on several pieces of relevant information and press a button after each piece of information has been processed. The average processing speed could be used as the dwell-time threshold.

Area of interests: A close distance (22pixels) between some of the AOIs caused activation difficulties for some of the participants, where the number next to the one they were looking at was activated instead of the one they were actually looking at. This problem would be solved with an improved calibration process of the eye tracking system. Another solution would be to move the AOIs further apart, but since the purpose of the study was to use an information heavy display screen this is not really a solution. With a good eye tracking calibration, the Author and one of the participants were able to interact with the prototype application without any difficulties.

Gaze cursor: Due to imperfect calibrations with the eye tracking system the visualisation of the gaze cursor created a "cat and mouse" effect where the eyes would follow the gaze cursor. The gaze cursor is not located at the place where the user is looking, but has an offset from that position which changes depending on the calibration. The eyes are drawn to moving objects such as the gaze cursor, and attempting to look at the gaze cursor shifts its position towards the offset. In order to visualise the gaze cursor without creating this effect a near perfect calibration with the eye tracking system is necessary. An alternative to visualising the gaze cursor could be to visualise when the system has registered an object as looked at by fading in and out a representative icon close to the object.

Gaze interaction methods: The prototype system uses two forms for gaze interaction methods, firstly a dwell-time activation is used to determine when the user is looking at a number, secondly a combined gaze and keyboard is used to mark number and activate buttons. The user looks at a number or button and presses the "space bar" on the keyboard to trigger an activation. The combined gaze and keyboard input was found to work well, and was enjoyed by the participants who found it easy to use. However, the participants found the dwell-time activation to be confusing when it was used to activate the highlights in the "highlight and disappear" concept. That problem could be solved by finding a way to calibrate the dwell-time threshold to the user of the system. In addition, it was found that activating buttons with dwell-time and visualising the activation progress caused the "Midas Touch" effect to occur. The user was exploring the screen, and when the gaze was directed at the button the visualisation started, which in turn made it more compelling to watch the button. Care should be taken when visualising the dwell-time activation so that it does not trigger accidental activations.

Midas touch: In addition to occurring when a dwell-time activation was used for the buttons, the "Midas Touch" effect occurred during the "highlight and disappear" concept. Some of the participants accidentally activated highlights while they were exploring the screen. The "Midas Touch" occurrence in the "highlight and disappear" concept can be remedied by finding a way to make the dwell-time threshold correspond to the user's information processing speed.

Dwell-time: The dwell-time activation threshold caused two difficulties during the user study. Firstly, it caused the "Midas Touch" effect to occur in the "highlight and disappear" concept. Secondly, it affected the validity of the supportive concepts feedback. The dwell-time activation threshold was set to 300ms, a value which worked well for the Author, however during the user testing it became apparent that the value did not work for all the users of the system. Most of the participants found the dwell-time threshold to be too low, and accidentally triggered highlights to activate in the "highlight and disappear" concept. One participant found the value to be too high, and had to focus longer than necessary to trigger activations. If the activation threshold is set so low that it suits the fastest users of the system, it can cause accidental activations for the slower users. If the value is set high enough to avoid accidental activations from the slower users, it causes the faster users to have to focus more than necessary in order to activate something. By finding a way to calibrate the dwell-time threshold to the user's information processing speed the "Midas Touch" effect can be negated and the time required to activate something be tailored to the user of the system.

Search strategy: A difficulty with the "highlight and disappear" concept occurred due to the unexpected use of different search strategies by the participants. The user testing showed that the concept was not able to fully support the search strategy where the users memorised a number types safety ranges and then checked the numbers of that type, before moving on to the second number type. However, after having checked the first number type all the highlights had disappeared, leaving the second number type to be checked without any assistance from the concept. This is a limitation of the concepts design which occurred because the possibility of the user using other search strategies was not thought about.

Implications

Several difficulties were uncovered during the creation and testing of the prototype application, some of them can be solved with further development, while others require the eye tracking technology to improve in order to be solved. The question then becomes, is the eye tracking technology ready for usage with information heavy interfaces? First of all, the technology is unable to achieve good calibrations with participants that wear glasses to correct their vision. That alone is a limitation which makes it realistically impossible to deploy a system that can be used by everyone. However, with the assumption that the technology is able to calibrate with every user, the accuracy achieved was still not high enough for most of the participants to be able to use gaze interaction without difficulties. Based on these limitations and the experiences from the user study the immediate answer would be that eye tracking is not a technology which is ready for deployment.

The technology can still be useful however, depending on the optical resolution required from an application. The prototype system had a fairly low resolution and focused on the numbers on the information heavy display screen, this proved to be too low of a resolution for the eye tracker to accurately perform with all the participants. It did however work flawlessly for the Author and one of the participants, showing that it has the potential to work at low resolutions if the calibration process is improved. If the optical resolution was higher, and focused on components on the interface, such as the containment tank and the accumulators, the accuracy of the system would most likely be high enough to accurately register these components as looked at. Even with the calibration achieved during the user study. For many control room tasks, the accuracy requirements are much lower. For instance, to assess if an operator checks the alarm list regularly, the area of interest would be the size of a full screen or perhaps half a screen. It might be on this less detailed level that eye tracking technology can be productively applied in the short term.

From the Authors experience the eye tracking technology seem to be suitable for both training and real operation. In addition, there are other domains than nuclear control rooms where the technology has an application, as shown by the literature review (Section 2.2), where the Maritime domain, Aircraft safety hold inspection and the assistance of severely disabled people has been explored. In the interview the participants were also asked if they could think of other domains where eye tracking could be useful, driving and gaming where the two most mentioned domains.

The supportive concepts should be applicable in any kind of monitoring tasks, for example in the air traffic control domain. However, developers of gaze-enabled interfaces should keep the difficulties of using gaze interaction in mind and attempt to solve or avoid them by using appropriate gaze interaction methods.

7.1 Future Work

For future work the concepts should be further developed based on the feedback received from the user study. In addition, ideas for new concepts were found through the interviews with the participants. The improved and new concepts should be tested again with a more realistic and possibly dynamic task. The test should be performed with more participants than the sixteen which participated in this user study, in order to gather more clear data and be able to draw some conclusions about the usefulness of the different concepts. The improved concepts should then be integrated into the actual NPP simulator and tested by supporting NPP operators during an actual task that they perform.

The user study was conducted with a remote eye tracking system. Such a system can only track a single monitor. To make the concepts more relevant for different domains such as air traffic control training, the system should be adapted to mobile eye tracking glasses that can be worn by the participants. Such a system would enable tracking across a complex information environment with multiple monitors and other information sources. In such an application, accuracy requirements may be significantly lower, perhaps down to screen-level accuracy instead of number-level accuracy.

In addition, it would be of great interest to find a way to calibrate the dwell-time activation threshold to the user’s information processing speed. It would make the system more user friendly, reduce the chance of the “Midas Touch” effect occurring, help validate the gathered data from the system, and it would be useful for any application that use the eye tracking technology. In addition, it would be useful to find a subtle way to indicate to the user that the system has registered that they are looking at an object, this would improve the user's confidence that the system is working.

8 Bibliography

Blascheck, T., Ertl, T., 2014. Towards analyzing eye tracking data for evaluating interactive visualization systems. ACM Press, pp. 70–77. doi:10.1145/2669557.2669569

Booth, T., Sridharan, S., McNamara, A., Grimm, C., Bailey, R., 2013. Guiding attention in controlled real-world environments. ACM Press, p. 75. doi:10.1145/2492494.2492508

Brooke, J., 1996. SUS - A quick and dirty usability scale. Usability evaluation in industry 189, 4–7.

Bulling, A., Roggen, D., Tröster, G., 2009. Wearable EOG goggles: eye-based interaction in everyday environments. ACM Press, p. 3259. doi:10.1145/1520340.1520468

Campbell, G., Geller, S., 1980. Balanced Latin Squares.

Cantoni, V., Porta, M., 2014. Eye tracking as a computer input and interaction method. ACM Press, pp. 1–12. doi:10.1145/2659532.2659592

Duchowski, A., 2013. Eye Tracking Methodology: Theory and Practice, 2nd ed. Springer Science & Business Media.

Hansen, J.P., Alapetite, A., MacKenzie, I.S., Møllenbach, E., 2014. The use of gaze to control drones. ACM Press, pp. 27–34. doi:10.1145/2578153.2578156

Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX. Human Mental Workload.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Weijer, J. van de, 2011. Eye Tracking: A comprehensive guide to methods and measures. OUP Oxford.

IAEA, 2004. Use of Control Rom Simulators for Training of Nuclear Power Plant Personnel. IAEA, Vienna.

Krejtz, K., Biele, C., Chrzastowski, D., Kopacz, A., Niedzielska, A., Toczyski, P., Duchowski, A., 2014. Gaze-controlled gaming: immersive and difficult but not cognitively overloading. ACM Press, pp. 1123–1129. doi:10.1145/2638728.2641690

Land, M., Tatler, B., 2009. Looking and Acting: Vision and Eye Movements in Natural Behaviour. OUP Oxford.

Langhals, B.T., Burgoon, J.K., Nunamaker, J.F., 2013. Using Eye-Based Psychophysiological Cues to Enhance Screener Vigilance. Journal of Cognitive Engineering and Decision Making 7, 83–95. doi:10.1177/1555343412446308

Lewis, J.R., 1992. Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 36, 1259–1260.

Lewis, J.R., 1990. Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the ASQ. ACM SIGCHI Bulletin 23, 78–81. doi:10.1145/122672.122692

Porta, M., Turina, M., 2008. Eye-S: a full-screen input modality for pure eye-based communication. ACM Press, p. 27. doi:10.1145/1344471.1344477

Putze, F., Hild, J., Kärgel, R., Herff, C., Redmann, A., Beyerer, J., Schultz, T., 2013. Locating user attention using eye tracking and EEG for spatio-temporal event selection. ACM Press, p. 129. doi:10.1145/2449396.2449415

Renganayagalu, S. kumar, Komandur, S., Rylander, R., 2014. Maritime simulator training: Eye-trackers to improve training experience, in: International Conference on Applied Human Factors and Ergonomics. Kraków.

Sadasivan, S., Greenstein, J.S., Gramopadhye, A.K., Duchowski, A.T., 2005. Use of eye movements as feedforward training for a synthetic aircraft inspection task. ACM Press, p. 141. doi:10.1145/1054972.1054993

Stellmach, S., Dachselt, R., 2012. Designing gaze-based user interfaces for steering in virtual environments. ACM Press, p. 131. doi:10.1145/2168556.2168577

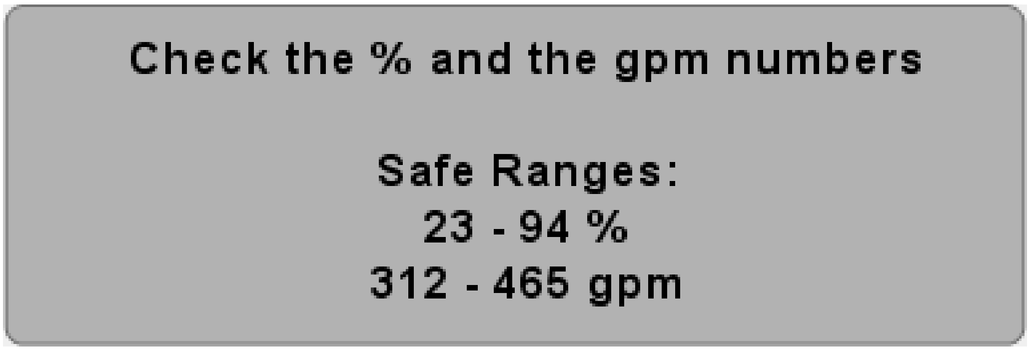
Stellmach, S., Stober, S., Nürnberger, A., Dachselt, R., 2011. Designing gaze-supported multimodal interactions for the exploration of large image collections. ACM Press, pp. 1–8. doi:10.1145/1983302.1983303

Zijlstra, F., van Doorn, L., 1985. The construction of a scale to measure subjective effort. Technical Report.

A Participant Instructions

Gaze-based Support of Nuclear Power Plant Control Operators

Imagine that you are a nuclear process control operator working with an information heavy display screen. **Your task is to monitor the numbers on the screen and make sure that their values are within the specified safety ranges. If a value is outside the specified ranges, it should be marked.** The safe number ranges will be displayed at the top of the screen at all times.



The job requires speed and accuracy to ensure safe operation of the nuclear process; therefore we want to support you with other technologies to make the job less demanding. A technology that has shown promise is **eye tracking**.

Eye tracking enables us to know where a person is looking. We have linked eye tracking with the system so the computer knows which numbers you have or have not looked at. Using this information we want to support you during your task by providing visual cues and visual feedback. We have prepared three concepts to aid in the process. These concepts are: Highlight and Disappear, Highlight Missed Areas, and Display Heat Map. The concepts are described and illustrated on the pages following how the experiment works.

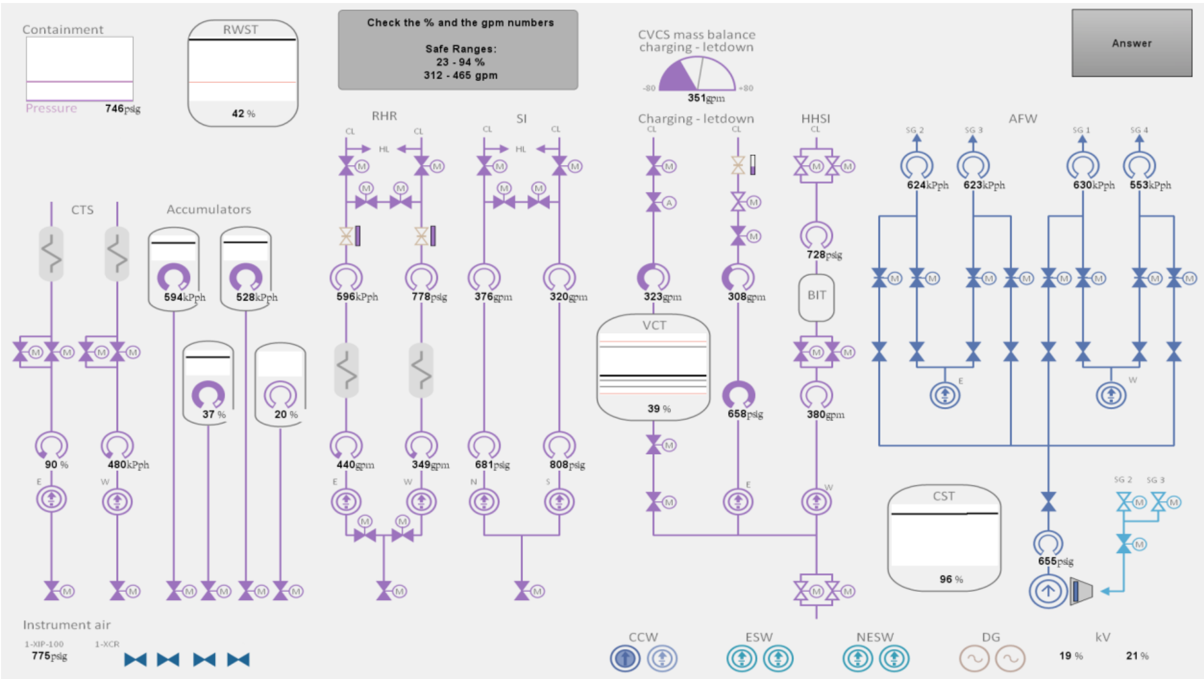
How the experiment works

Calibration

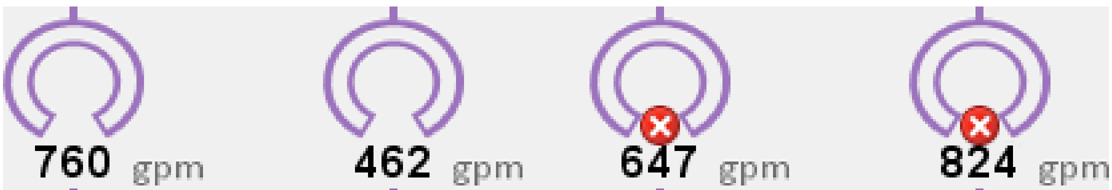
First it is necessary to calibrate the eye tracker to your eyes. This is a quick process where you have to sit in front of the display while trying to keep your head as still as possible. A grey screen will appear and on the screen a sequence of 9 red dots surrounded by a white circle will be displayed. Focus your eyes on the red dots as they appear, the surrounding white circles will shrink towards the red dot to help you focus. Depending on the results of the calibration this step might have to be repeated. After the calibration has been completed we can proceed with the experiment.

The Experiment

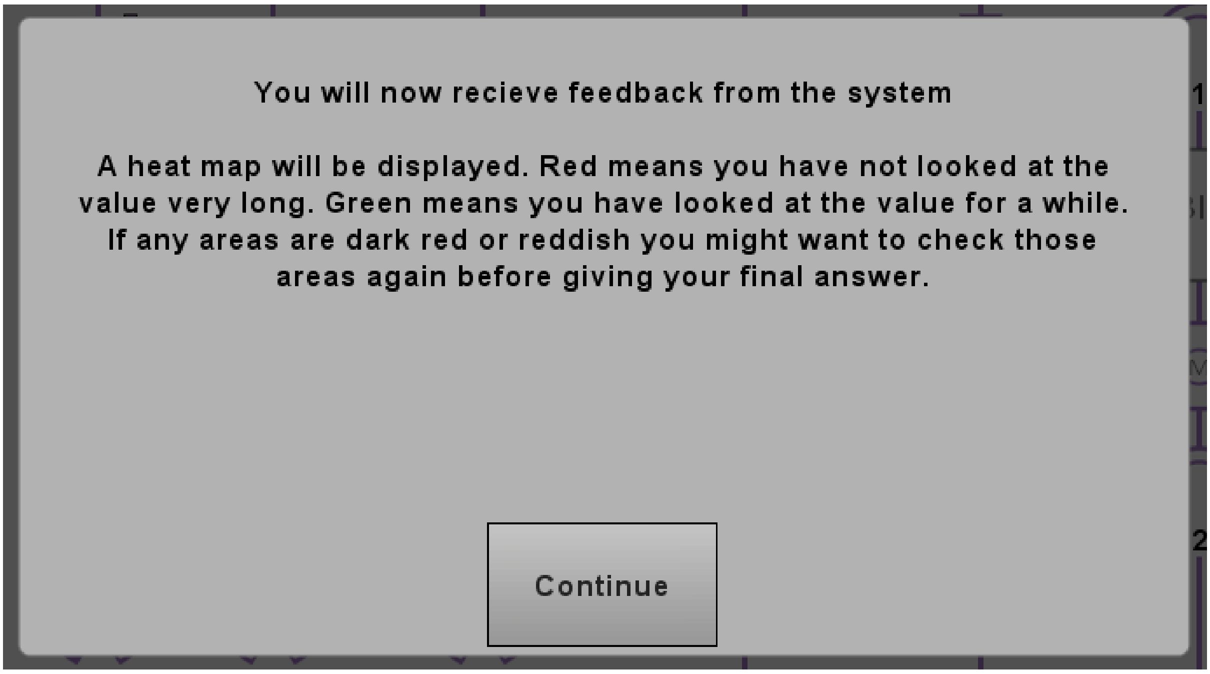
First we will explain each concept to you. Then you will perform 3 tasks for each concept. Each task consists of a nuclear power plant information screen with many symbols and numbers. The symbols are irrelevant for your tasks, and can be ignored. There are four different types of numbers, %, psig, gpm and kPph, for each task you will have to check two of these number types. The safety ranges are different for each number type, so make sure to check the number types you need to monitor at the top of the screen for each task.



Each information screen will have different numbers which you need to check. If a number is outside the safe ranges it should be marked. A number can be marked by clicking it with the mouse, or by looking at it and pressing the "space bar". You can tell that a number has been marked by the red circle with a white cross which appears above a marked number. If you accidentaly marked a number you can unmark it by marking it again. Buttons can be pressed by looking at it and pressing space.



After you have finished checking the screen and marking the numbers, you can give your answer by pressing or looking at the button labelled “Answer”. If the concept provides feedback a popup will appear with more information. After pressing the “Continue” button you will be brought back to the screen where you can see the feedback and make changes accordingly.



After all the tasks within one concept have been completed the next concept will start. This continues until all four concepts and the baseline has been completed. A table showing the amount of concepts and tasks you need to complete can be seen below. **The order of the concepts can vary**.

Table: Concepts and tasks overview

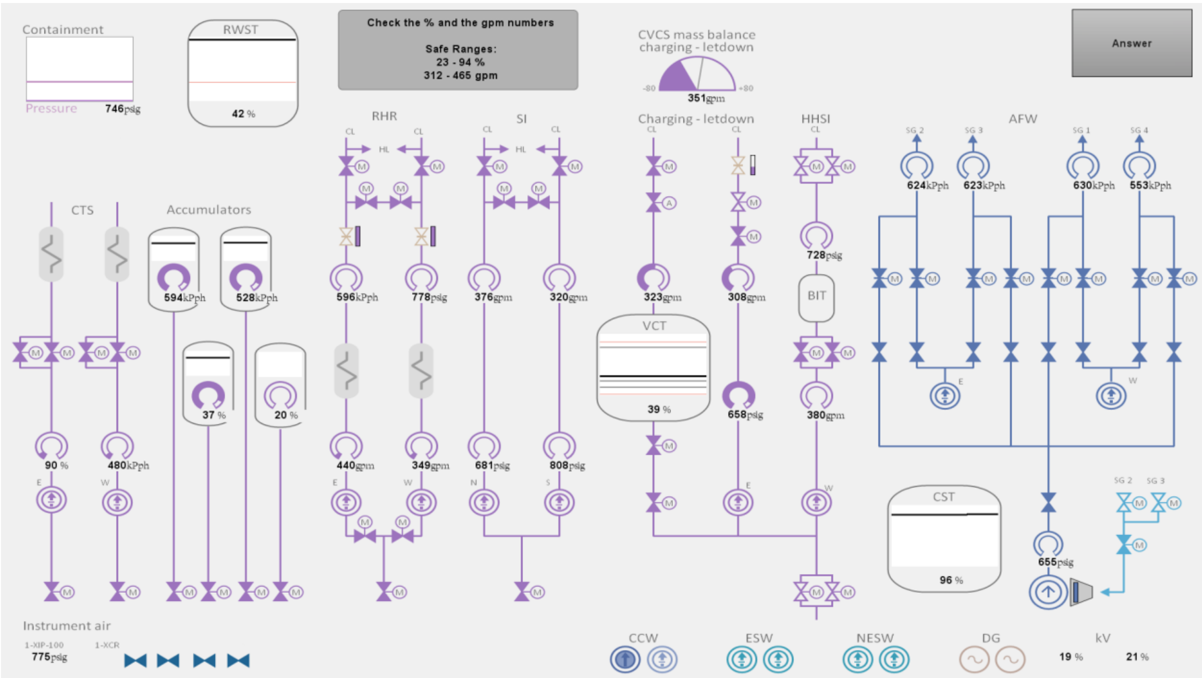
	Task 1	Task 2	Task 3
Baseline			
Highlight and Disappear			
Highlight Missed Areas			
Heat Map			

Task reminder

Your task is to inspect all the numbers on the display and make sure that they are within the specified ranges. The ranges can be viewed at the top of the screen. Mark any values that <u>ARE NOT</u> within the safety ranges by clicking on it, or by looking at it and pressing the space bar.

Baseline

In the baseline the eye tracker is not used to provide any support. You will have to complete the 3 tasks to the best of your ability. An example screen of what you will see is provided below.

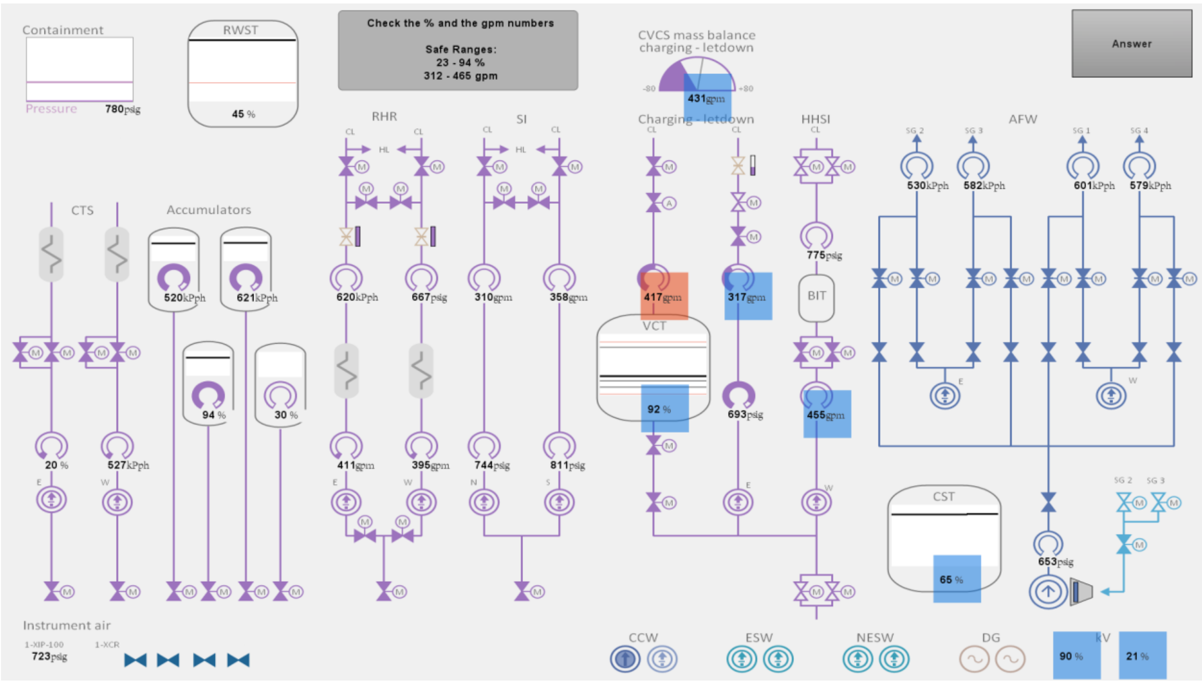


Task reminder

Your task is to inspect all the numbers on the display and make sure that they are within the specified ranges. The ranges can be viewed at the top of the screen. Mark any values that ARE NOT within the safety ranges by clicking on it, or by looking at it and pressing the space bar.

Concept 1: Highlight and Disappear

The "Highlight and Disappear" concept highlights the relevant numbers you need to check with blue squares. This concept uses the eye tracker to register when you have looked at a value, changing the highlight to red before it disappears. This concept is meant to help you structure your scanning of the values by highlighting the values that you have yet to check while removing the highlight from the values you have checked.

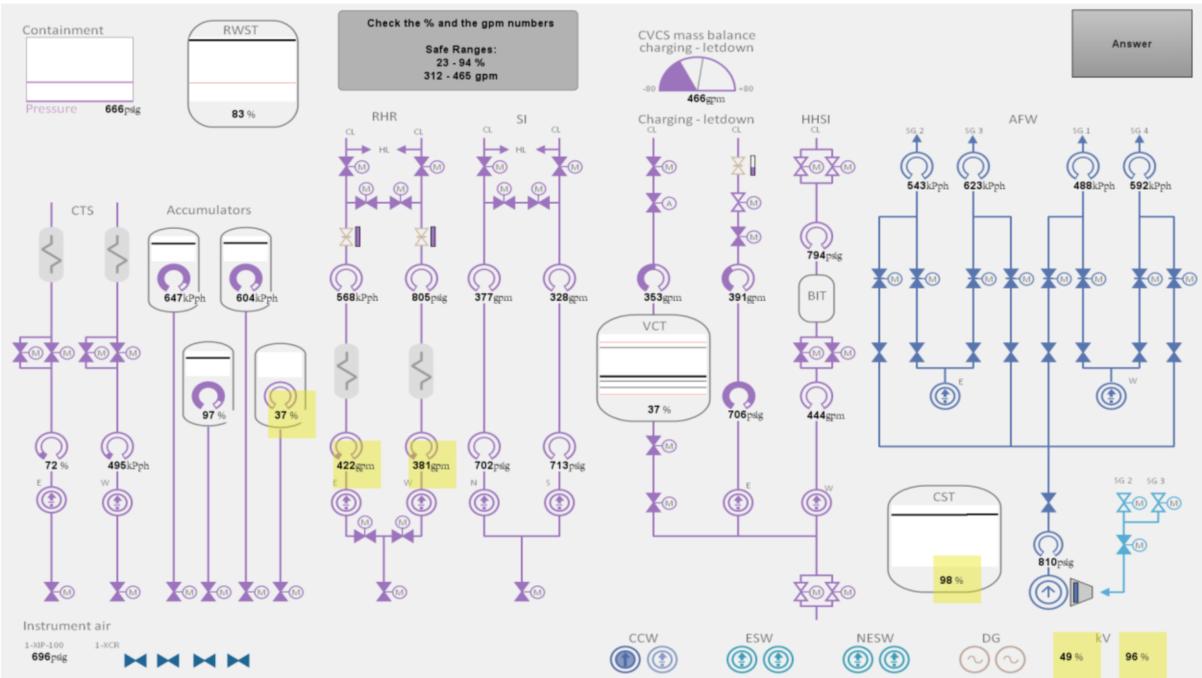


Task reminder

Your task is to inspect all the numbers on the display and make sure that they are within the specified ranges. The ranges can be viewed at the top of the screen. Mark any values that <u>ARE NOT</u> within the safety ranges by clicking on it, or by looking at it and pressing the space bar.

Concept 2: Highlighted Missed Areas

The “Highlight Missed Areas” concept does not highlight the values you need to check initially. The concept uses eye tracking to know which numbers **you have not looked** at, and will highlight them with blinking yellow squares, after you press the “Answer” button. If a number is highlighted you should check the area before giving your final answer. This concept is meant to ensure that you have checked all the values before giving your final answer.



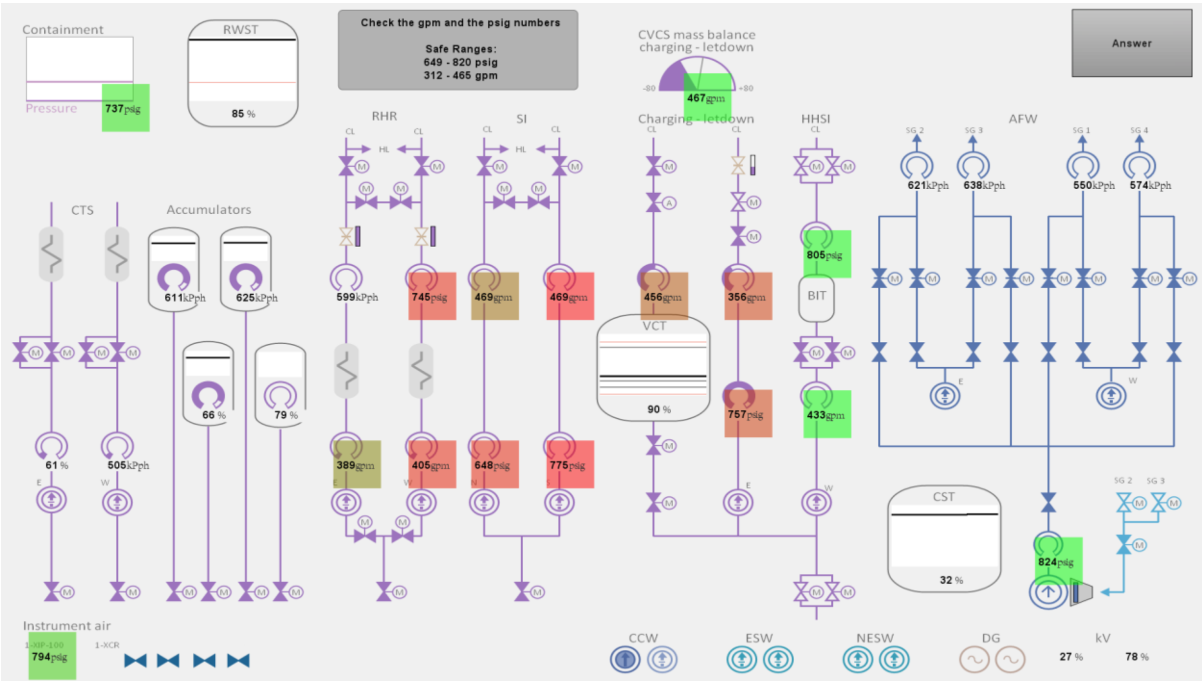
Task reminder

Your task is to inspect all the numbers on the display and make sure that they are within the specified ranges. The ranges can be viewed at the top of the screen. Mark any values that <u>ARE NOT</u> within the safety ranges by clicking on it, or by looking at it and pressing the space bar.

Concept 3: Display Heat Map

In the "Display Heat Map" concept the values are not highlighted initially. While you perform your task of checking numbers the system registers the amount of time you spend looking at components. After you press the "Answer" button a heat map will be displayed on top of the relevant numbers. The colours of the heat map will be a gradient between red and green. Red means that little or no time was spent viewing the value, and green means you looked at the value a fair amount of time.

If any areas are dark red or reddish you might want to check those areas again before giving your final answer. This concept gives you an overview of how you spent your time checking the values, and shows you areas that might not have been checked thoroughly enough.



In Summary

Your task is to monitor the numbers on the screen and make sure that their values are within the specified safety ranges. If a number's value is outside the specified safety ranges it should be marked by "left clicking" it with the mouse or by looking at it and pressing the "space bar". Any buttons can be pressed using the mouse or by looking at them and pressing the "space bar".

In the baseline condition you will not receive any support during the tasks and will have to complete them to the best of your ability.

In the concept "Highlight and Disappear", you will get visual support **while** you perform your task of checking the numbers.

In the concepts "Heat Map" and "Highlight Missed Areas" you will receive feedback **after** you have checked the numbers and pressed the "answer" button the first time.

In total you will have to complete 4 concepts and 12 tasks.

Table: Concepts and tasks overview

	Task 1	Task 2	Task 3
Baseline			
Highlight and Disappear			
Highlight Missed Areas			
Heat Map			

B Semi-structured interview

Participant Number:

Did you find the gaze interaction straining?
Comments:

Did you like or dislike the gaze interaction?
Comments:

How accurate was the eye tracking solution? Think about ***marking numbers*** *and* ***clicking buttons with your eyes****.*
___ Perfectly accurate
___ Problems once or twice
___ Several problems
___ Completely unreliable

Comments:

How accurate was the eye tracking solution? Think about ***the concepts: Highlight Missed, Highlight & Disappear, and Heat Map****.*
___ Perfectly accurate
___ Problems once or twice
___ Several problems
___ Completely unreliable

Comments:

Did you find the ***Highlight & Disappear concept*** *confusing*?
Comments:

What was your search strategy?

Which concept did you find the most helpful?
Which concept did you find the least helpful?

If you were to rank the concepts, how would you rank them?

*Give your opinion about each of the concepts. What did you **like/dislike**?*
Highlight and Disappear

Highlight Missed

Heat Map

Do you have suggestions for improvements or new ideas?

Which other domains could you see eye tracking being useful? E.g. driving, gaming (How?)

C Participant Questionnaires

C.1 Modified Nasa-TLX

Participant Number:

In the questions below, when you are asked about “the system”, think specifically about the concept “Baseline / Highlight Missed / Highlight & Disappear / Heat Map”.

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Eye Fatigue — How tiring was the task for your eyes?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed were you?

Very Low — Very High

C.2 Modified System Usability Scale

Participant Number:

In the questions below, when you are asked about “the system”, think specifically about the concept “Highlight Missed / Highlight & Disappear / Heat Map”.

	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently	1	2	3	4	5
2. I found the system unnecessarily complex	1	2	3	4	5
3. I thought the system was easy to use	1	2	3	4	5
4. I thought the system was enjoyable to use.	1	2	3	4	5
5. I thought the system was useful and supported me in my task	1	2	3	4	5
6. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
7. I found the system very cumbersome to use	1	2	3	4	5
8. I felt very confident using the system	1	2	3	4	5
9. I needed to learn a lot of things before I could get going with this system	1	2	3	4	5