Bachelor's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Engineering
Department of Ocean Operations and Civil Engineering

Gustav Frost, Hanne Kirkerød, Martin Sæter

Maritime Behavioral Analysis Under Passive Task-Related Fatigue

Bachelor's thesis in Nautical Science
Supervisor: Baiheng Wu
June 2022

NTNU
Kunnskap for en bedre verden

Gustav Frost, Hanne Kirkerød, Martin Sæter

Maritime Behavioral Analysis Under Passive Task-Related Fatigue

Bachelor’s thesis in Nautical Science
Supervisor: Baiheng Wu
June 2022

Norwegian University of Science and Technology
Faculty of Engineering
Department of Ocean Operations and Civil Engineering

NTNU
Kunnskap for en bedre verden

Abstract

Fatigue is seen as a prevailing health related problem for seafarers, as well as a contributing factor to accidents. This thesis explores fatigue in the maritime industry and task-related fatigue in navigation. Our hypothesis was that navigators would perform worse due to fatigue caused by several hours of work. This thesis analyzes correlation in data sets from two separate simulations of identical complexity in the Strait of Dover. To compare navigational performance in navigators, the data sets included quantitative data from eye tracker recordings and the simulator, and qualitative data from questionnaires. The participants were graduating students at NTNU and had similar levels of navigational expertise. They were asked to perform one experiment at 07:00 and another at 16:00. On average, participants slept 6.1 hours the night before the experiment, and had been awake for 69 minutes before the first simulation. Three out of six participants felt more tired in the afternoon, two did not feel any change in tiredness, and one participant felt less tired in the afternoon. Most of the subjects felt uncertain because of the long time since they had been in the Strait of Dover. The average level of fatigue experienced by our subjects in this study was not enough to affect their performance as navigators. Uncertainty due to time absent from certain navigational tasks and individual differences were more important factors when it came to performance. Further research should be devoted to both active task-related fatigue for seafarers, as well as the effect length of absence from navigational tasks has on navigational performance.

Acknowledgments

We would like to sincerely thank Baiheng Wu for being our supervisor and helping with implementing the eye tracker, Tron Resnes for testing and feedback on the simulations, as well as feedback during the writing process, Terje Ole Slinning for data extraction from the simulator, and all the subjects for their participation.

May 25, Ålesund
Martin Sæter, Hanne Kirkerød, and Gustav Frost

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Our topic

We looked at how fatigue affects navigators. Our research question was the following: *How does fatigue affect the performance and risk-taking of nautical navigators?* Our hypothesis was that the navigators' performance would be negatively affected.

1.2 Why we chose our topic?

We decided early on that sleep was a topic we found very interesting, especially for nautical officers who live and work on a ship with a varied schedule. There is a lot of responsibility resting on the navigators' shoulders, and mistakes can have severe consequences. Getting enough sleep can be challenging at sea, as conditions for proper sleep might not be ideal. In the maritime industry there seems to be a culture of under-reporting hours worked to stay within regulations (Jepsen, Zhao, and Leeuwen, 2015, Allen, Wadsworth, and Smith, 2020). This practice, along with a lack of consideration of the effects of fatigue, can create dangerous situations where officers push themselves far beyond what is responsible and safe. It seemed very important to gain more knowledge in this field for our own sake as future navigators. We wanted to see the effect of fatigue on the risk-taking and navigational skills of navigators, as we have all experienced being fatigued after a long day of work.

1.3 Approaching our topic

When we had to choose how we wanted to approach our bachelor thesis, we decided to not only read the research of others, but conduct our own experiment as well. We were curious and wanted first-hand experience, as well as our own data to analyze and compare to existing research.

We decided to run experiments in a simulator, with test subjects from the graduating class of Nautical Science at the Norwegian University of Science and Technology (NTNU). They would come in early in the morning to run a scenario, and another one in the afternoon. We had them fill out a survey after each run to gather data, and utilized eye tracking in the simulator.

For the simulation we decided to do a crossing of the Strait of Dover. In a crossing of a traffic separation scheme (TSS) a navigator uses tools and techniques to find a specific point to execute a safe crossing-maneuver (Zhao, Li, and Zhang, 2020). Crossing the Strait of Dover simulates a high-risk operation. Focus and alertness is important to navigate in compliance with the Convention on the International Regulations for Preventing Collisions at Sea (COLREGs) of the International Maritime Organization (IMO), especially in heavily trafficked areas (Irving, 1982). The Strait of Dover is frequently used in navigational training at NTNU.

1.4 Structure of our thesis

Our thesis is divided into the following chapters: introduction, background, methods, results and discussion, and lastly conclusion. Background contains research and results from other experiments and articles about fatigue and sleep. Here we wished to broaden our understanding of these topics and look at what was done before us. The methods chapter describes how we performed our experiment and the specific methods and equipment that was used. Results and discussion includes analysis and discussion of the results from the experiment.

Chapter 2

Background

2.1 Research on fatigue

2.1.1 Definition

In colloquial terms fatigue is commonly treated as interchangeable with tired, but we needed to define the term properly. The term fatigue is defined by the American Heritage Dictionary as “physical or mental weariness resulting from effort or activity” (*American Heritage Dictionary Entry: fatigue* n.d.). For it’s guidelines on fatigue IMO defines fatigue as "A state of physical and/or mental impairment resulting from factors such as inadequate sleep, extended wakefulness, work/rest requirements out of sync with circadian rhythms and physical, mental or emotional exertion that can impair alertness and the ability to safely operate a ship or perform safety-related duties." (*IMO - Guidelines on Fatigue* 2019). In our case we were looking at mental fatigue, defined by Van Cutsem et al. as “a psychobiological state caused by prolonged periods of demanding cognitive activity” (Van Cutsem et al., 2017). It can be divided further into task and sleep-related fatigue, and task-related fatigue can be either passive or active. These two states are contrasting, where passive fatigue is caused by a lack of mental stimulation typically seen in monotonous work not requiring much mental effort, and active fatigue is caused by high mental effort during cognitively difficult tasks. Sleep-related fatigue is caused by reduced sleep and poor quality of sleep, as well as a disruption of the circadian rhythm (Martinez-Marquez et al., 2021).

2.1.2 Seafarers' fatigue

There are many factors that affect fatigue in humans. It can also be hard to separate them seeing as many are interconnected. Workload, rest periods, and stressors in our environment are some key elements that can affect someone's level of fatigue (Grech, 2016, Allen, Wadsworth, and Smith, 2008).

It is generally believed that sleep is important for all humans (*What is Sleep & Why is It Important for Health? - ASA* n.d.). Sleep does not always come easy to everyone, even with all the right conditions to facilitate sleep. While on a ship there are many factors that affect sleep quality for the seafarers. The motion of the ship affected by the weather, the sound of engines, crew and alarms, being on the ship for the duration of your rotation, and being woken up should the need arise (Allen, Wadsworth, and Smith, 2008). In studies, seafarers themselves have reported a lack of sleep and inconsistent sleep times as a huge contributor to fatigue, which could mean that they are not given a proper chance to get adequate sleep (Grech, 2016).

Exactly how long someone must sleep for it to be adequate can be individual. Some research suggests that we need approximately 7 hours of sleep (Watson et al., 2015). The exact number of hours needed varies depending on individual factors and can range from anywhere between 5 to 10 hours (*What is Sleep & Why is It Important for Health? - ASA* n.d.). Despite individual differences one study found that the seafarers only slept for 5 hours on average (Oldenburg and Jensen, 2019).

Much of shipping is done around-the-clock, facilitating a need for shift-work to keep everything going. This means that some crew work through the night and sleep during the day, which is when the circadian rhythm rather encourages activity. This can cause the sleep they get to be short and of bad quality (Lützhöft et al., 2010), and impair their performance throughout their shift (Ferrara and De Gennaro, 2001).

This continuous cycle of work, and the competitive nature of the industry could be demanding more of the crew than can be sustained in the long run. Restricting crew-size to the bare minimum for financial gain puts a large amount of pressure and strain on the employees remaining (Grech, 2016). From a questionnaire given to 6461 seafarers the average hours worked per week was found to be between 68 and 69 hours. However a quarter of the participants reported working 84 hours a week, and 5% worked over 90 hours a week (Jensen et al., 2006). A study done on seafarers on a variety of ships found that working hours had a big impact on fatigue (Smith et al., 2003). It also found that exposure to a sum of variables, like noise, work hours, and shift type, seemed to have a greater adverse effect on health and fatigue when combined.

Among seafarers there are individual differences depending on which occupational group they belong to, or which maritime industry they work in. There are also big differences in ship types, flag of registration, voyage lengths, and type of work carried out. A study done with 323 sailors found that nautical officers had longer working hours on average, and the shortest sleeping time, compared to engine room personnel and deck ratings. Experiencing the most frequent sleep interruptions, 67% of the nautical officers reported having sleep deficiency (Oldenburg and Jensen, 2019). Another study (Smith et al., 2003) found differences based on which part of the industry the seafarers belonged to, comparing ferries and tankers in the short sea shipping industry, as well as comparing these to their previous research, in which they looked at the offshore oil industry. Ship type was an important indicator of fatigue, with higher levels of fatigue on the ferries. In general, the participants from the short-sea industry reported more negative moods and scored worse on performance tasks than those from the offshore oil industry. Short-sea ships also had a higher level of fatigue, which could suggest a correlation between the length of tours and level of fatigue.

2.1.3 Challenges when measuring fatigue

All the studies mentioned so far had their own method for how to measure fatigue, tiredness, and sleep. They looked at slightly different factors, in different parts of the industry, and used different baselines for comparison. There was a mix of subjective and objective methods of measurement, where many of them utilized some form of survey or questionnaire as well as some form of equipment to monitor the subjects. Tests were frequently used to check the status of the participants after a certain period of sleep deprivation. There were vigilance tests based on sound, reaction time tests, and addition tests, to mention a few.

Surveys and questionnaires had no set standard in their wording and set-up, using a variety of scales to rate the subjects' own perceived fatigue. These differences made it hard to compare data across the industry, as the scales used were not interchangeable with one another. For example, the Stanford Sleepiness Scale (SSS) measures the momentary value of tiredness by giving the subject seven options to choose from, while the Epworth Sleepiness Scale gives a collection of scenarios and four options to rate the likelihood of nodding off (Phillips, 2014, Gillberg and Åkerstedt, 1994). These subjective measures might not show the complexity of the accumulation of sleep restrictions, possibly not being sensitive enough to catch the longer-term effect as well as some monitoring equipment can (Ferrara and De Gennaro, 2001).

The different experiments were varied in their approach to partial sleep deprivation. Sleep duration varied from 0 to 7.5 hours of sleep, and the duration of these experiments could be from 1 to 2 nights up to weeks of reduced sleep. It is also important to remember the individual differences between the subjects, such as sleep need, which would result in two participants subjected to the same conditions producing different results. Establishing a baseline for each subject, such as their ideal sleep duration could help here, but it was not common practice. Sleep curtailment could for example be easier to handle for habitual short sleepers, than those more used to sleeping for long (Ferrara and De Gennaro, 2001).

2.1.4 Effect and consequence of fatigue

There have been multiple studies trying to figure out the effect sleep restriction has on humans. Houtman et al. found an association between seafarers' fatigue and a decrease in vigilance, alertness, and perception (Lützhöft et al., 2010). In another study the subjects were deprived of sleep to varying degrees and tasked with tests to measure the effect it had on them. Two days a week, over several weeks, they were allowed to sleep for different amounts of hours (7.5, 5, 3, 2, 1, 0 h of sleep). Their vigilance and mental speed were tested, and the study found a significant drop in vigilance and a steady decline of mental speed when allowed 3 or fewer hours of sleep. While the mental speed went down steadily, the percent of errors peaked at 1 hour of sleep. All results worsened on day two of sleep restriction, showing the cumulative effect partial sleep deprivation can have on both vigilance and mental speed (Wilkinson, Edwards, and Haines, 1966). Wilkinson et al. stated that "the reduction of sleep by about half on a single night can produce a significant fall in working efficiency." Both of these studies showed a decrease in vigilance, and yet another study (Ferrara and De Gennaro, 2001) also reported a drop in accuracy and speed of response in a vigilance task, after restricting the subjects' sleep to 5 hours. The mood of the participants had also gotten worse, compared to their usual 8 hours of sleep.

A similar study (Gillberg and Åkerstedt, 1994), which allowed its participants 4 hours of sleep, also found a negative effect on the participants' subjective feeling of being well-rested, worse performance of reaction time, as well as shorter sleep latency. This combination of subjective and objective measures was explored further using the SSS and Multiple Sleep Latency Test (MSLT) in a study restricting sleep to 5 hours a day for a week. Immediately, the subjective scale showed an increase in sleepiness during daytime, but the sleepiness measured objectively only increased after the second day of partial sleep deprivation. Later in the week, subjective measures leveled off after the fourth day, while sleepiness measured with MSLT kept increasing throughout the days (Ferrara and De Gennaro, 2001).

As mentioned above in the last study, our own perception of our level of fatigue does not always correlate to the more objective measures. Regardless of the studies above showing evidence of negative effects from fatigue, there is a culture of negligence within the maritime industry. Despite rules and regulations placed by IMO, working through fatigue is by some seen as professional. Whether it is ignorance of or willful subjection to the effects of fatigue, seafarers do not seem to take them seriously. The sentiment that dealing with fatigue is part of the job is prevalent throughout the industry, and there have been studies finding discrepancies between the recorded hours and actual hours of work and rest on ships (Grech, 2016, Jepsen, Zhao, and Leeuwen, 2015). One such study found that official working time records did not correlate with the documentation done by onboard examiners (Oldenburg and Jensen, 2019), while another study found that 40% of the participants admitted to under-reporting their work hours to various degrees. Out of these 40%, 11.7% reported frequently or always under-recording their work hours (Allen, Wadsworth, and Smith, 2020). This study also suggested that the measured 40% of under-reporters could be higher, considering that some seafarers might be reluctant to admit to this breach of regulations. A general view that the regulations in place are not practically applicable to the necessary work done in shipping can cause this lack of compliance to the legislation (Phillips, Nævestad, and Bjørnskau, 2015).

The lack of practical legislation and enforcement, as well as the pressure of efficiency and cost-cutting from the industry, leave seafarers with few choices (Allen, Wadsworth, and Smith, 2007). It cannot solely be seen as a personal problem on the part of the seafarers, when the industry as a whole perpetuates this behavior of regulation infringement. In the maritime industry, the disregard of fatigue can have severe consequences, causing accidents which can lead to injury, fatalities, damage to valuable cargo, as well as environmental disasters (Lützhöft et al., 2010, Grech, 2016).

2.2 Research on eye-tracking

Eye movements' importance to the individual's perception of and attention to the visual world is implicitly acknowledged (Hansen and Ji, 2010a). Eye movement recordings can give researchers a lot of useful information about behavior and attention. Initial studies in the maritime domain using eye tracker mainly focused on safety aspects, bridge design, and training programs (Trondheim, Odd, and Hareide, 2019, Lützhöft and Dukic, 2007). With better technology and improved means of analysis, the amount of eye tracker research is increasing in high-risk industries such as aviation, maritime, and construction (Martinez-Marquez et al., 2021).

Analyzing behavioral logic and attention using a wearable eye tracker is now a standard solution. Skvarekova and Skultety looked at the number of saccades per minute, dwell time, and number of fixations to find objective differences between experienced and inexperienced pilots (Skvarekova and Skultety, 2019). Using eye tracking to establish the relationship between workload and communication, eye tracker data can help to identify abnormal behavior at an early stage of navigational training (Streilein et al., 2020). Eye tracking is also used in assessing electronic navigation competency in maritime training (Atik and Arslan, 2019). An analysis of the collected data in (Muczyński et al., 2013) found that basic eye tracking characteristics, namely, number of fixations, and frequency of fixations, saccades, and blinks, can be used as an indicator of mental workload.

Eye movement as a measure of fatigue is well established. Fatigue was identified as the eighth most studied application of eye tracking, in a 2021 review (Martinez-Marquez et al., 2021). According to (Gupta et al., 2019) in a submarine environment, the only identified feasible measure of different fatigue factors was eye tracking. An evaluation on eye metrics as a detector of fatigue, using approximate entropy (ApEn) and total eye closure duration, found that analysis of eye metrics can indicate the onset of fatigue in advance of significant changes in operator performance (McKinley, 2011).

Performance specifications of commercial eye tracking devices are usually provided by their manufactures or distributors (Martinez-Marquez et al., 2021). However, there are studies reporting variation in the eye trackers' accuracy, compared to those provided by the performance specifications (Zhang and Hornof, 2011, Komogortsev and Khan, 2008, Hansen and Ji, 2010b). Implementing filters responsible for how fixation data are calculated improves overall accuracy, but a test comparing the fixation filter data with an observer's subjective impression is recommended (Tobii, 2019).

As well as with performance, there are challenges to conducting research using eye tracker technology. Some include gaze location error, fixation definition, and defining metrics (Goldberg and Helfman, 2010). Understanding the limitations is important to conducting valid research.

A validity study of maritime usability with eye tracking data, (Hareide and Ostnes, 2018) reported uncertainties in eye tracker data when participation was low, and recommended for similar studies to support the quantitative measurements from eye tracker with qualitative data.

Chapter 3

Methods

3.1 Experiment Design

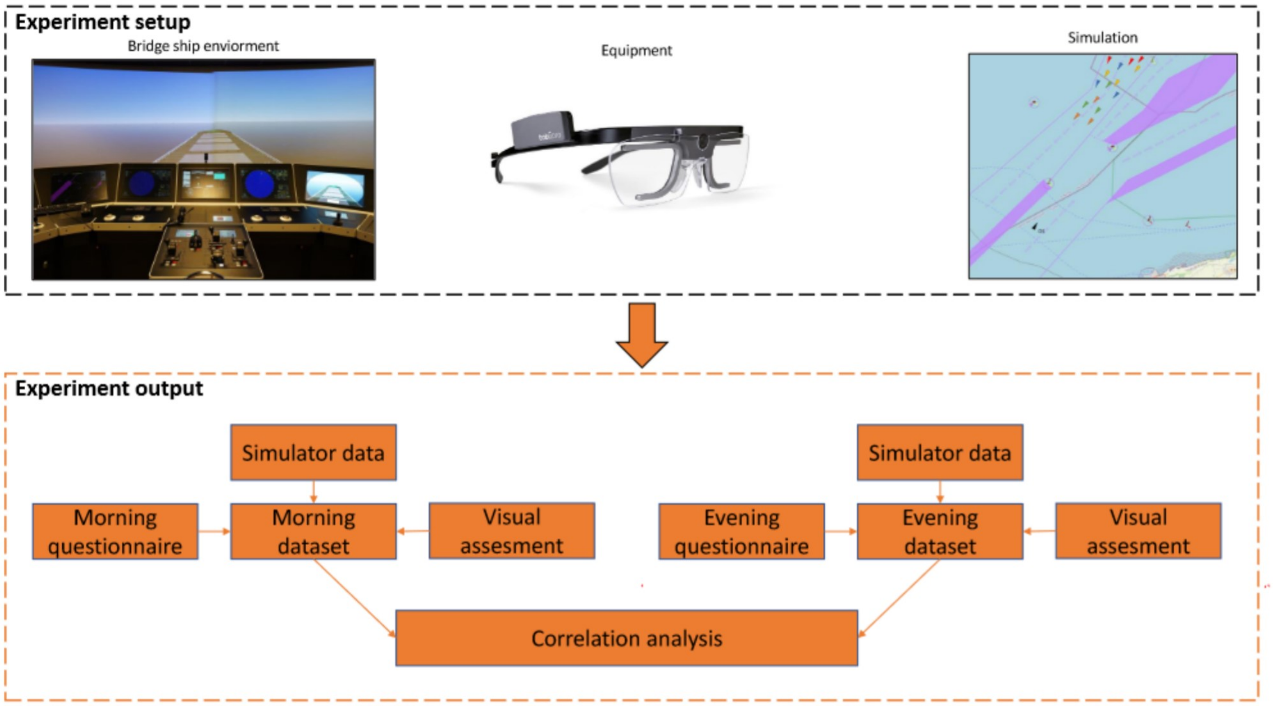


Figure 3.1: Experiment workflow. (Picture of Tobiipro glasses 2 is creditied to the official Tobii website).

Environment

The workflow of the experiments was as illustrated in figure 3.1. Experiments were conducted in the Kongsberg bridge simulator (k-sim), designed for training and certification of maritime personnel. The simulator system complies with STCW 78 as amended, and DNVGL-ST-0033 Maritime Simulator Systems standard, and achieves an appropriate level of physical and behavioral realism. (*DNVGL-ST-0033.pdf* 2020).

The bridge ship simulator has the ability to toggle various functions for training purposes. For this experiment all basic functions to aid in navigation were included in the simulation. This included two radars, with Automatic Radar Plotting Aid (ARPA) and Automatic Identification System (AIS) ability. One Electronic Chart Display and Information System (ECDIS) with GPS. One conning station, for the navigator to get feedback on parameters like speed, wind, course, and rate of turn. The outside view was projected on a wall, and there was a separate screen for binoculars with a zoom function.

Equipment

A pair of eye-tracker glasses powered by Tobii Pro, with a sampling rate of 100 Hz, was used. The glasses direct invisible near-infrared light into navigators' eyes and measure the reflection of this light using high-definition cameras. Recordings of this reflection can determine where and when navigators are directing their attention inside the ship-bridge simulator.

Software

The simulations were designed, implemented, and run in K-Sim Spirit 2.6.10.107.

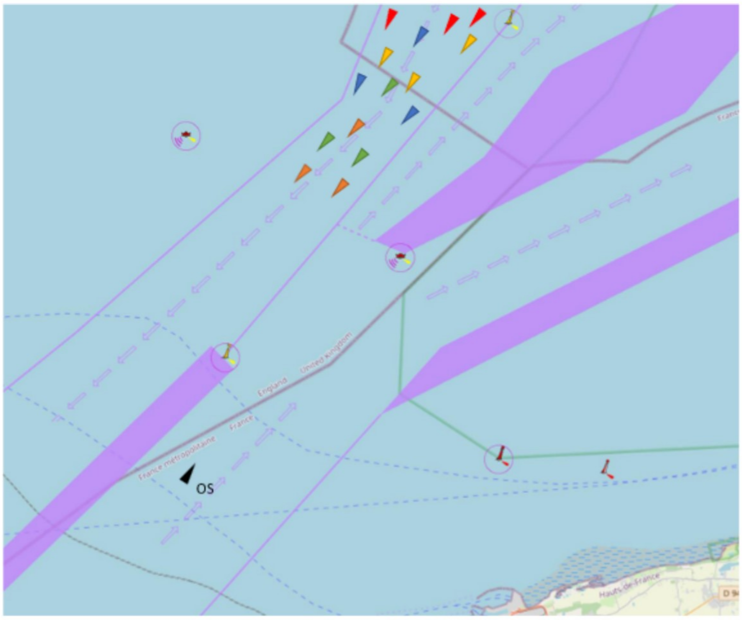
Tobii Pro Lab v. 1.181.37604 was used for processing raw eye tracking data, and also for manually delineating areas of interest (AOI). All fixations recorded by the wearable eye tracker were filtered with Tobii I-VT (attention) filter (Olsen, 2012).

All the quantitative analysis was done in LibreOffice Calc 7.3.2.2.

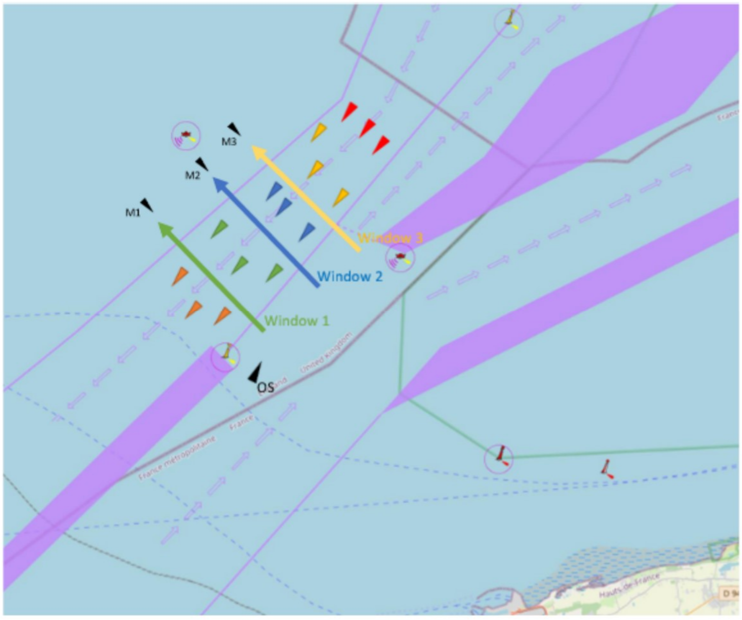
Simulation

Two nearly identical simulations were created. One morning simulation and one evening simulation. Traffic in the simulations was set up to reflect a realistic scenario in the Strait of Dover. In addition to ships sailing along the TSS, three small motorboats were also added. The boats were placed near likely places of crossing, and crossed the strait in the opposite direction of the subjects. They were used to reflect unforeseen circumstances, and test the navigators' visual attention.

The simulation was designed with three safe windows for crossing, with the first window being the most difficult one to detect and time, and the last one being the easiest (figure 3.2).



(a) Initial traffic



(b) Safe crossings

Figure 3.2: Illustration of traffic problem solving in Dover.

Weather and sea conditions were kept the same for the morning and afternoon simulations. Both wind and currents were removed due to problems with the physics of the simulation during testing. There were small differences in traffic to minimize recallability. Light conditions were the same for both voyages, with the time of day in the simulation set to 12:00.

Data Collection

For this study, quantitative data was collected from two sources: the eye tracker, and the simulator, and qualitative data was collected from questionnaires.

The eye tracker recorded eye movements inside the ship bridge simulator. These recordings were used to assess visual attention, and produce quantitative visual attention distributions.

Data collected from the simulator included voyage duration, position, distance to other ships, and rudder usage. When looking at the performance of the participants, the main focus was on whether they were too close to other ships or not, how many ships they were too close to, how close they got, and whether they got too close in front of or behind other ships. Other performance factors were voyage duration, rudder usage, and whether they saw any of the motorboats coming towards them.

Two questionnaires were created, one for each voyage, that asked about the participants' age, sex, experience navigating, and caffeine and nicotine usage. In addition, the questionnaire had a series of statements, that they were asked to rate their level of agreement with, having the following options: strongly disagree, disagree, neutral, agree, and strongly agree.

For the purpose of data collection and consent, this study was related to the bigger project *Remote Control Center for Autonomous Ship Support*, which was approved by the Norwegian centre for research data (NSD).

Participants

Six participants were chosen randomly from graduating students doing a bachelor's degree in Nautical Science at NTNU. The six participants had the same background in ship bridge training, navigation, and traffic separation schemes. It had been two years since any of the participants had been in a simulation of the Strait of Dover. One of the participants was female, and the other five were male. The participants had an age range of 22-24. All the participants gave consent for their anonymized data to be used for this study.

Stimulants

As the study would look at fatigue after a somewhat normal day, it was important to prevent the participants from artificially reducing their level of fatigue, or the effects thereof.

Originally, preventing the participants from both caffeine consumption and nicotine usage prior to the experiments was considered. It was decided that the participants had to refrain from caffeine usage on the day of the experiment, because the stimulating effects of caffeine would keep them from getting tired (Snel and Lorist, 2011). Restricting nicotine usage was decided against, due to difficulties with compliance, and the varied effects of both nicotine usage and withdrawal (“Nicotine and health” 2014). It was deemed better to keep the nicotine users at their normal baseline.

3.2 Implementation

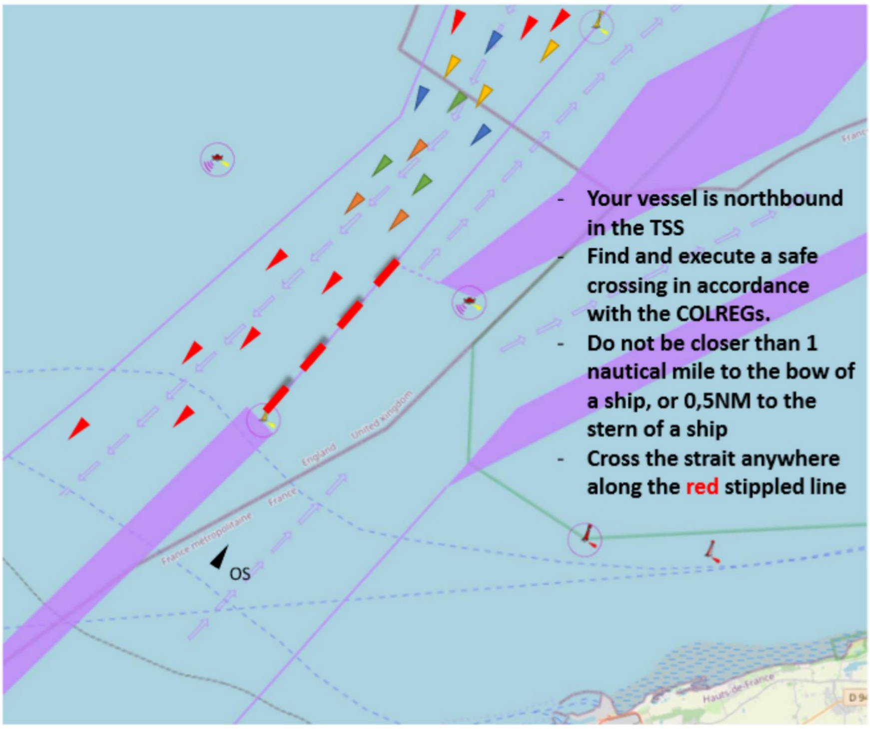
The six participants were each given two separate time slots in a day. For convenience, the first simulation started at approximately 07:00, and the second at approximately 16:00 as shown in table 3.1. The nine-hour window between the experiments was chosen to reflect a regular working day. Due to differences in schedules, as well as unforeseen circumstances, the shortest window between the start of the morning and afternoon voyage was 8.7 hours, and the longest was 12.1 hours, with 9.5 hours being the average.

Participant	1	2	3	4	5	6
Start morning	07:07	07:28	07:08	07:11	07:08	07:13
Start afternoon	16:16	16:10	16:16	16:04	19:16	16:11

Table 3.1: Start times for participants.

A verbal introduction to the experiment was given to the participants a day or two before the simulation, where they were told where they would sail, and also instructed not to consume caffeine on the day of the experiment.

Verbal guidelines were also given to the participants inside the bridge ship simulator before each simulation. They were told where to cross the strait, and not to be closer than 1 nautical mile (NM) to the bow of a ship, or 0.5 NM to the stern of a ship (figure 3.3).



- Your vessel is northbound in the TSS
- Find and execute a safe crossing in accordance with the COLREGs.
- Do not be closer than 1 nautical mile to the bow of a ship, or 0,5NM to the stern of a ship
- Cross the strait anywhere along the **red** stippled line

Figure 3.3: Verbal guidelines given inside the ship bridge simulator.

Participants were fitted with eye-tracking glasses, that were then calibrated. This was done to assure that the participants' fixation matched the visual representation on the computer. Once the eye tracker was calibrated, three minutes were given to the participants to set up navigational instruments and plan the voyage, after which simulation started.

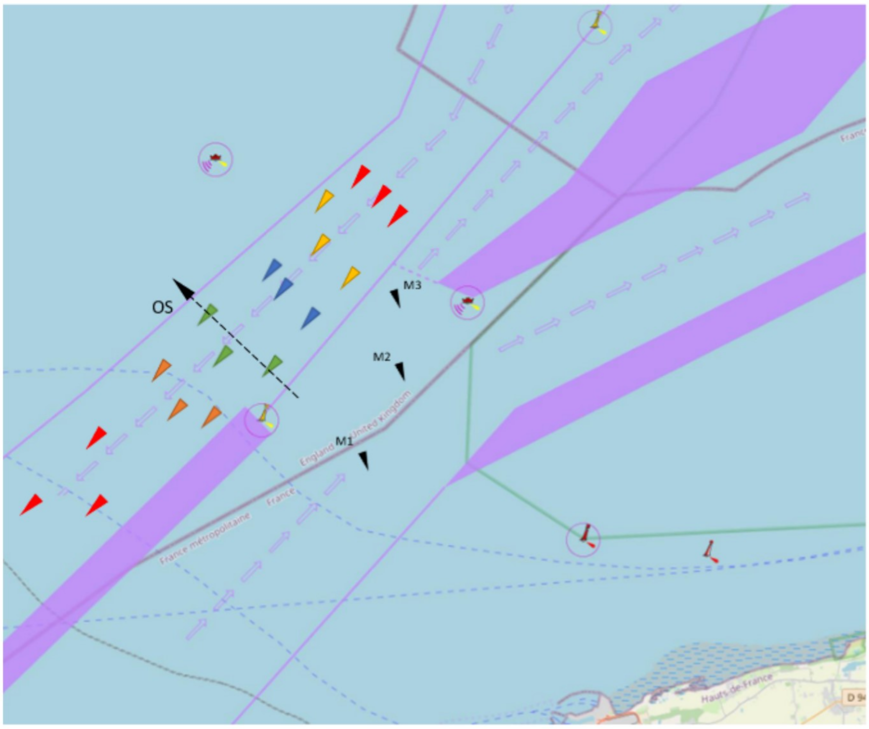


Figure 3.4: Simulations ends outside the TSS. Window 1 was chosen in this example.

Simulation ended when participants cleared all traffic and were outside of the TSS, illustrated in figure 3.4. Afterwards participants filled out the questionnaire for the voyage (see appendix A).

The light levels and temperature were kept constant for both voyages for all participants.

3.3 Analysis

Eye tracking

The eye tracker data included metrics of saccades, fixations, glances, and visits (Tobii, 2021). Tobii glasses use fixations, glances, and visits as metrics for when the eye is focused. Fixations are periods where the eye is focused. Saccades are rapid eye movements. Meanwhile, glances and visits look at time spent in an AOI. Glances are defined as the time between the start of the saccade leading into the AOI until the end of the last fixation on the AOI. The exit saccade is not included. Visits also do not include the entry saccade (ISO, 2020). Visits per minute, i.e. how many times per minute the participant changed what AOI they were looking at, was used as a general measure of visual activity.

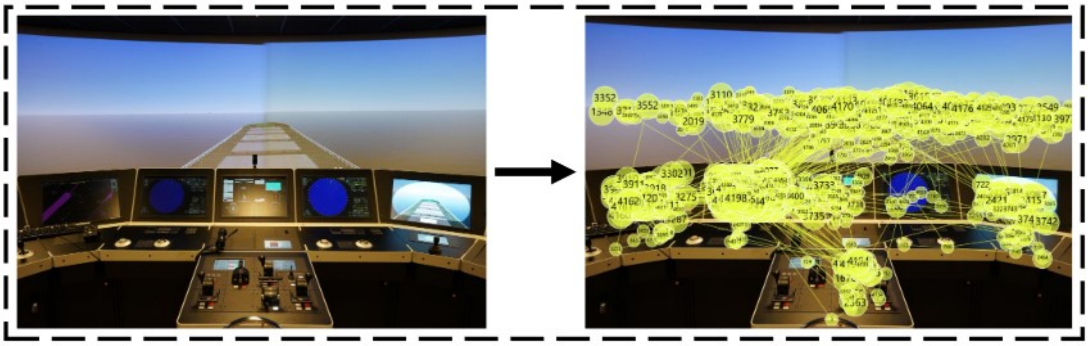


Figure 3.5: Manual mapping of fixations.

Eye-tracker recordings of the participants were attached to a snapshot, i.e. a still image of the ship bridge environment, in which the participants were located. Every fixation was then mapped from the recording into the snapshot, for further analysis as shown in figure 3.5.

Visual attention distribution can be presented in heatmaps as seen in figure 3.6 or gaze plots in figure 3.7. Heatmaps show static visual attention distribution. Red indicates the highest number of fixations or the longest time fixating there, and green the least. Gazeplots show the location (circle), order (number), and time (size of circle) spent looking at locations on the snapshot.

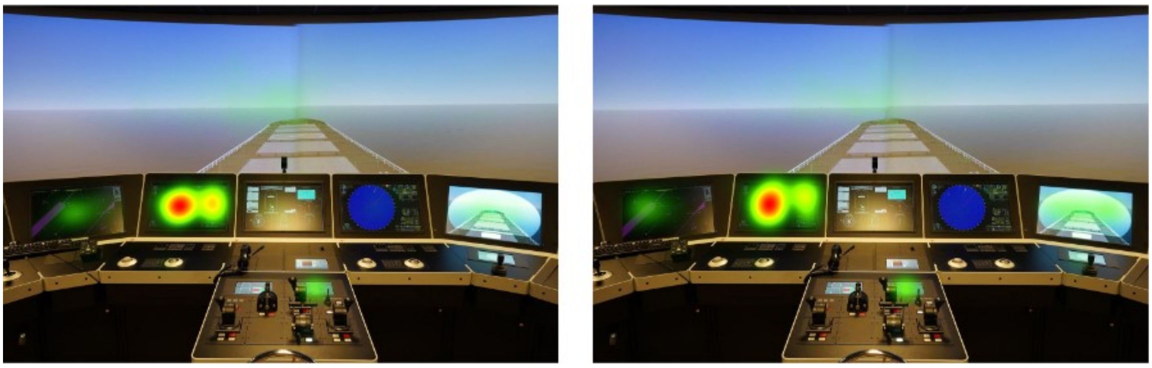


Figure 3.6: Heatmap from random participant, morning and afternoon.

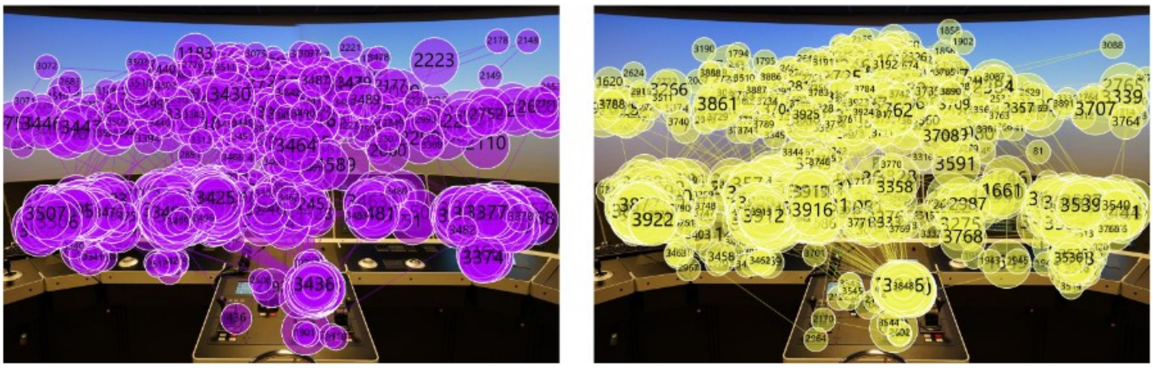


Figure 3.7: Gaze plot from the same participant as in figure 3.6, morning and afternoon.

The snapshot was appointed six AOI. In this configuration, every AOI selected was a screen in the simulator: ECDIS, radars, conning station, binoculars, and the outside view (figure 3.8).

Each recording was split into predefined events.

1. Start of simulation
2. Crossing. The point in time in which navigators make a distinguished maneuver to cross.
3. Clear of traffic. The point in time in which navigators have cleared all target ship vectors.
4. End of simulation

Figure 3.8: AOI applied to snapshot.

From these events, three times of interest (TOI) were applied: start of simulation to crossing, crossing to clear of traffic, and clear of traffic to end of simulation.

Correlation analysis

There are many different methods for statistical and correlation analysis. One of the most commonly used methods for finding the correlation between two variables is the Pearson correlation coefficient (Berman, 2016), denoted by r_{xy}. It measures linear correlation, with a score ranging from -1 to 1, with -1 being a perfect negative correlation, and 1 being a perfect positive correlation.

The Pearson correlation coefficient is calculated by the following formula:

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2} \tag{3.1}$$

Where n is the number of data points.

The Pearson correlation coefficient should not be used on ordinal data directly. Instead, both data sets being compared should have their values replaced with numerical ranks from lowest to highest or highest to lowest. If there are several data points in a set with the same value, the average rank of those values is given to all of those data points. The calculations for obtaining the Pearson correlation coefficient should then be performed on the ranks, instead of the values of the ordinal data (Zar, 2005). Doing this results in obtaining Spearman's rank correlation coefficient, denoted by r_s.

A weakness of the Pearson correlation coefficient is that it is highly sensitive to outliers, in which case it is preferable to use Spearman's correlation (Rousseau, Egghe, and Guns, 2018).

Another weakness of Pearson, that Spearman's does not suffer from, is that it only works with linear data (Hauke and Kossowski, 2011).

While Spearman's correlation has some advantages over the Pearson correlation, care should be taken not to overinterpret the significance of the strength of the correlation of data when only looking at Spearman's coefficient, as the values of r_{xy} and r_s can differ vastly for the same data sets, in both magnitude and significance (Hauke and Kossowski, 2011).

When comparing data sets in our study that included ordinal variables, only Spearman's correlation was used, however, when comparing two sets of non-ordinal data, both the Pearson correlation and Spearman's correlation were used. In addition, regression analysis was performed on data sets of particular interest, that showed a correlation by looking at either the r_{xy} or r_s scores, to obtain a p-value for an indication of the significance of the correlation.

There is no clear consensus in the literature about the interpretation of Pearson and Spearman's correlation coefficients, and what r-value indicates a strong correlation is greatly dependent on the field of study. The three most commonly used interpretations are from Dancey and Reidy (psychology), Quinipac University (politics), and Chan YH (medicine) (Akoglu, 2018).

Nettleton suggests that an r-value of 0.7 shows a considerable correlation (Nettleton, 2014).

Another suggestion is to use the following statistically justified rule of thumb for a rough satisfactory threshold for the r-value (Krehbiel, 2004): $\frac{2}{\sqrt{n}}$

Where n is the number of data points.

Using the rule of thumb for the r-value gave a value of 0.82 for this study. Such a high threshold is justified by the small sample size, however, it could potentially result in missing interesting or valuable correlations. As this study used a mixture of objective quantitative data, and qualitative data from questionnaires, a cut-off threshold of 0.5 was chosen, as it lies between the thresholds for moderate strength correlations of the most commonly used interpretations for psychology (0.4), a largely qualitative field, and medicine (0.6), a largely quantitative field (Akoglu, 2018).

Correlation does not necessarily indicate causation, and thus it is important to find the significance of a correlation. The p-value, a highly popular measure of significance, is an index of the discrepancy between the data and the null hypothesis, having a value from 0 to 1 (Tanha, Mohammadi, and Janani, 2017). Both the p-value and statistical significance have complex definitions, which leads to oversimplification and thus misinterpretations (Greenland and Poole, 2011). The statistical significance of a result should not be confused with the importance of a result, as the p-value only indicates the former (Tanha, Mohammadi, and Janani, 2017). Neither is the p-value a good measure of evidence for a model or hypothesis by itself (Wasserstein and Lazar, 2016).

Fisher, the creator of the p-value, advocated $p < 0.05$ (5%) as a threshold, but not an absolute rule, for rejecting the null hypothesis. He also argued strongly that researchers have to interpret the p-value themselves (Sterne and Smith, 2001). Different studies interpreting the p-value differently is especially important because a high enough sample size or measurement precision can lead to a small p-value, no matter how small the effect is, and vice versa (Wasserstein and Lazar, 2016). A study with a very small sample size should thus not assume the insignificance of a finding because of a greater p-value than what convention dictates would indicate significance.

Regression analysis was performed on data sets of particular interest, that had showed a strong enough correlation by having an r_{xy} or r_s value > 0.5, to obtain a p-value for the correlation in order to gauge its significance.

For categorical data with two categories, it is suitable and advantageous to give numerical values of "0" and "1" for the purpose of quantitative analysis (Newcombe, 1992). Categorical data was thus converted to numerical data, with "0" representing "no" or "false," and "1" representing "yes" or "true."

Ordinal data was converted to a scale of 1 to 5, with 5 representing "strongly agree."

Chapter 4

Results and Discussion

4.1 Questionnaires

Participant	**Morning**						**Afternoon**					
	1	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**
Sleep quality	4	4	3	4	4	4	-	-	-	-	-	-
Usually sleep quality	4	2	4	4	4	3	-	-	-	-	-	-
Tiredness	2	4	4	3	2	3	4	4	4	1	4	4
Neg. effect of no caffeine	1	4	2	1	2	5	3	4	2	1	2	5
Neg. effect of recording	2	3	2	2	1	4	2	3	2	2	1	4
According to plan	4	3	5	4	5	4	5	4	5	3	5	4
Treat the simulation like reality	4	3	4	4	4	3	4	3	4	4	4	3
Usually stressed in simulator	3	3	2	3	2	4	-	-	-	-	-	-
Uncertain due to long time since Dover	5	4	2	4	2	4	-	-	-	-	-	-
Felt more prepared second run	-	-	-	-	-	-	5	4	2	4	3	4

Table 4.1: Ordinal data from the questionnaires.

Ordinal data from the morning and afternoon questionnaires is shown in table 4.1. Other data from the questionnaires will be presented in their relevant sections. Age data, besides the age range, will not be presented to protect the participants from de-anonymization. Only one participant used nicotine, and so no correlation could be discerned.

Data from caffeine usage was difficult to quantify and analyze due to the range of answers and uncertainty expressed by the participants (appendix B). Furthermore, there can be severalfold differences in the caffeine concentration of coffee, depending on the choice of species, brewing time, water temperature, pressure, degree of roast, grinding degree, water type, and the ratio of coffee to water (Olechno et al., 2021). This made it impossible to know how much caffeine was actually consumed by the participants on the day prior to the experiment, thus making any speculation about the effects of such consumption meaningless.

4.2 Analysis

Tiredness

Participant	1	2	3	4	5	6
Slept (h)	6.5	7	5	6	5	7

Table 4.2: Hours slept.

The participants slept on average 6.1 hours (range 5-7 hours) the night before the experiment (table 4.2). The average sleep quality for the previous night was 3.8 (range 3-4), and the participants rated their usual sleep quality at an average of 3.5 (range 2-4). Tiredness during the experiments had a greater range: 1-4, with an average tiredness of 3.0 for the morning voyage, and 3.5 for the afternoon. There was, however, no correlation found between sleep duration or sleep quality and tiredness.

Morning voyage	1	2	3	4	5	6
Awake (min)	82	98	58	61	53	63

Afternoon voyage	1	2	3	4	5	6
Awake (min)	631	620	606	594	781	601

Table 4.3: Number of minutes since waking up.

On average, the participants had been awake for 69 minutes (range 53-98 minutes) when starting the morning voyage, and 639 minutes (range 594-781 minutes) at the start of the afternoon voyage (table 4.3). A moderate correlation was found between duration of having been awake and tiredness (r_s = 0.54, p = 0.072). The correlation is likely significant because of the small sample size, despite the convention of a threshold of $p < 0.05$, due to the reasons mentioned in the methods chapter.

There was a moderate correlation ($r_s = 0.56$, $p = 0.058$) between participants reporting being negatively affected by not consuming caffeine and their level of tiredness. This lends credence to our decision to deprive the subjects of caffeine prior to the experiment, so that they do not prevent themselves from getting tired due to exogenous means.

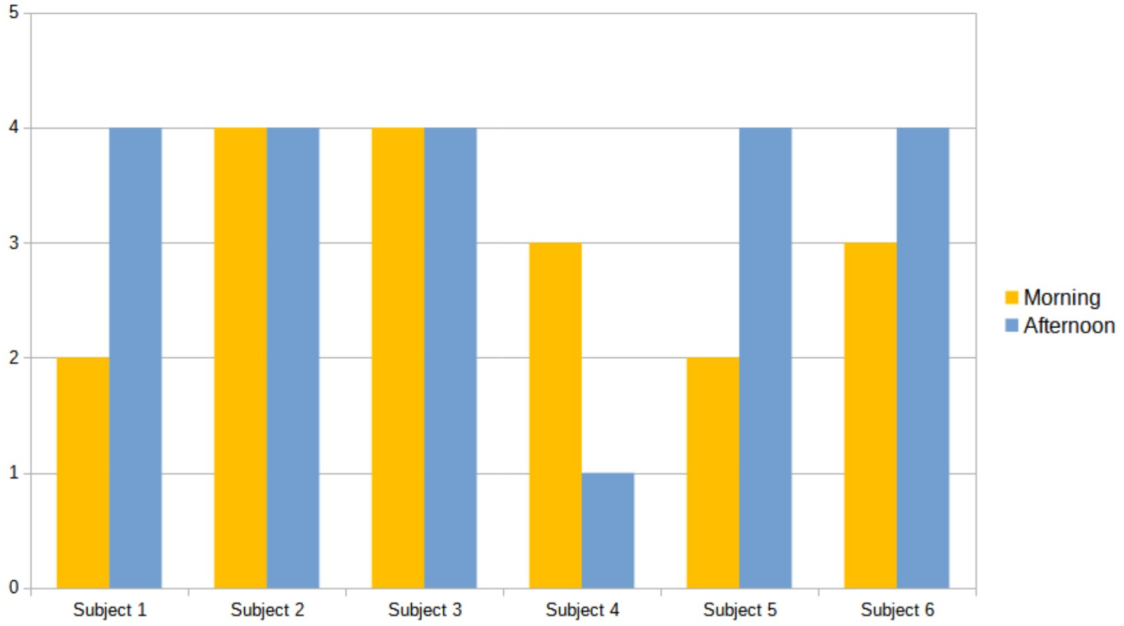


Figure 4.1: Tiredness in the morning and afternoon.

Greater levels of tiredness were expected in the afternoon, however, it seems like only being awake for 9 hours, with a significant amount of that time spent in passive lectures, and having plenty of free time, was not enough to cause significant levels of fatigue. As can be seen in figure 4.1, three of the participants felt more tired in the afternoon, but two of them did not feel any change in tiredness, and one participant even felt less tired in the afternoon, showing how differently individuals can be affected by the same things in the same environment.

Performance

Morning voyage	1	2	3	4	5	6
Too close?	Yes	Yes	No	Yes	No	No

Afternoon voyage	1	2	3	4	5	6
Too close?	Yes	Yes	No	Yes	No	Yes

Table 4.4: Showing whether a participant was too close to another ship or not.

Half of the morning voyages, and two thirds of the afternoon ones, had the participant being too close to other ships (table 4.4). On average, the tiredness from morning to afternoon increased from 3.0 to 3.5, indicating that there might be a connection between being tired and coming too close to other ships. However, when doing correlation analysis, contrary to expectation, no correlation was found (r_{xy} = -0.13, r_s = -0.08).

The only meaningful correlation found for whether a participant was too close to any other ships, was to the reported uncertainty of the participant, due to not having been in the Strait of Dover for a long time. This was both a very strong (r_{xy} = 0.83, r_s = 0.80) and a very significant ($p < 0.002$) correlation. This was also the only important correlation that came close to the correlation coefficient threshold given by Krehbiel's rule of thumb (0.82), which takes into account the small sample size. We received feedback from several participants that they felt uncertain both because it had been a long time since they were in the Strait of Dover, but also because it had been a long time since they used the trial function of the radar, which they all used in this experiment.

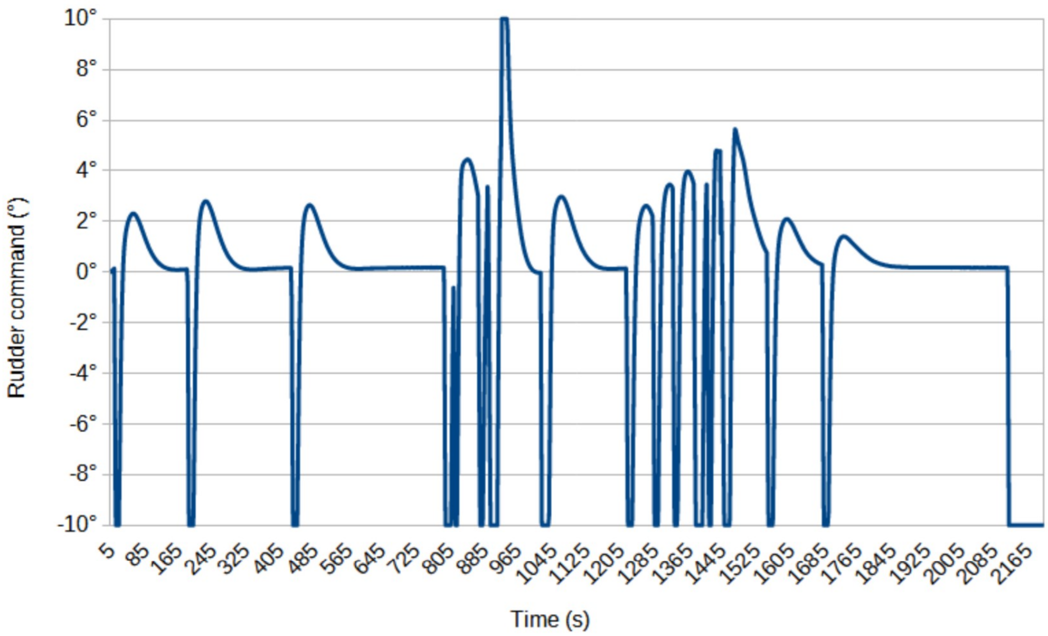
Participant	1	2	3	4	5	6
Uncertainty	5	4	2	4	2	4
More prepared 2nd run	5	4	2	4	3	4

Table 4.5: Uncertainty of the participants, and increased preparedness for the second run.

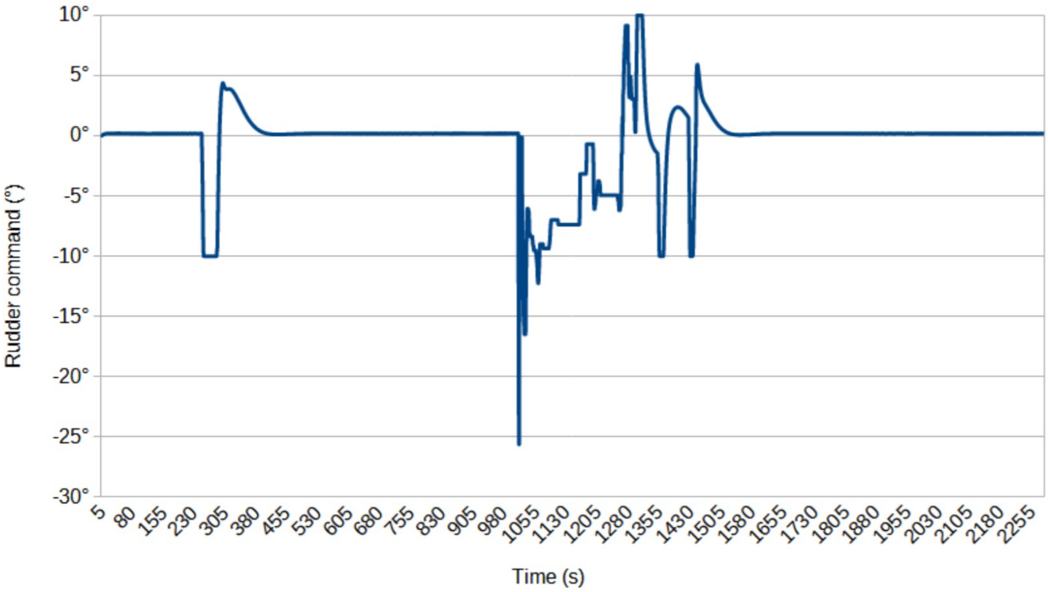
There were only two subjects that never came too close to another ship, and one subject that managed to keep the minimum distance in the morning voyage, but not the afternoon one (table 4.4). The only two who always kept the minimum distance were also the only two who disagreed with being uncertain before the voyage (table 4.5). One of them disagreed with being more prepared the second run, while the other one was neutral. All other participants felt more prepared the second run, and had been uncertain before the first run, resulting in an incredibly strong and significant correlation (r_s = 0.98, p « 0.000). This correlation is, however, self-explanatory and of little importance, as those who felt uncertain before the first run would naturally feel more prepared after their first crossing of the Strait of Dover in a long time.

Nobody was too close to more than two ships in the same simulation. The closest anyone was behind a ship was 0.33 NM, and the closest anyone was in front of another ship was 0.51 NM. Five voyages had the participant being too close behind a ship, while only two voyages had the participant too close in front of a ship.

The average voyage duration was 2518 seconds long (range 2198-3448 seconds), with the afternoon voyages being on average 107 seconds longer. The strongest correlation that voyage duration had was a weak and insignificant negative correlation with tiredness (r_s = -0.27, p = 0.40), and that was very likely simply due to the fact that the afternoon voyage had its safe passage windows slightly further away than in the first voyage. Thus, looking at the exact voyage duration does not matter much. Instead, looking at which of the three windows of safe passage a participant took tells us more about risk aversion and priorities of the navigator, with placement in the window, and thus distance to other ships, giving the greatest insight into risk aversion and performance. Five of the subjects attempted to take the first window for all voyages, while one subject attempted to pass in the vicinity to the third, and safest, window in both voyages, but did not quite manage to accurately identify the opening, and thus ended up too close behind a ship in both scenarios.



(a) Autopilot



(b) Manual steering

Figure 4.2: Comparison of autopilot and manual rudder usage of two subjects.

Average and max rudder commands were analyzed with the idea that using the rudder too much might imply that the navigator had less control and foresight. The max rudder command was on average 14.2° (range 10-35°) for both morning and afternoon voyages. Only 3 out of 12 voyages had the navigator take manual control of steering and turn the rudder more than 10°, which is the maximum for the autopilot. Figure 4.2 compares examples of rudder usage from a voyage where only autopilot was used, to one where mostly manual steering was used. The average rudder command throughout a voyage was 1.8° (range 1.3-2.4°) for morning and 1.7° (range 1.1-2.8°) for the afternoon. No correlation was found between rudder usage and any other metrics.

Visual attention

Only two participants failed to notice any of the motorboats. Both on their morning voyage, and they both noticed at least one motorboat in the afternoon. No correlation was found between noticing or not noticing a motorboat and any other metrics of significance. When we added the small motorboats to our simulated scenario, we expected the ones utilizing the binoculars and looking out the most to have a higher probability of noticing the vessels, but that was not the case.

When analyzing overall activity by looking at visits per minute, there was a moderate, but not very significant, correlation with uncertainty ($r_s = 0.50$, p = 0.095). There was no other metric that had a meaningful correlation with average visits per minute. There was, however, a strong and significant correlation between uncertainty and percentage of time spent on either radar ($r_s = 0.70$, p = 0.01). It seems logical that the more uncertain a person felt, the more time they would need on the radar to get an overview and analyze the traffic situation, and the more often they would have to switch between different AOI. The average visits per minute was 11.6 for the morning and 11.4 for the afternoon.

Comparing participants' heatmaps and gazeplots for the entire voyage found little significant change in eye activity (appendix C), apart from the outlier in figure 4.4, which had 1541 more fixations in the afternoon voyage. This can indicate more eye activity and higher mental workload during this participant's afternoon voyage (Muczyński et al., 2013).

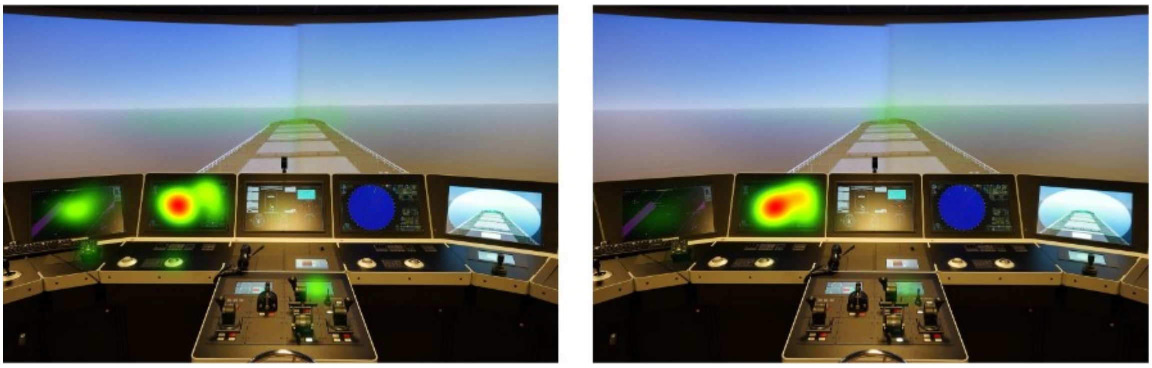


Figure 4.3: Heatmap from outlier participant, morning and afternoon.

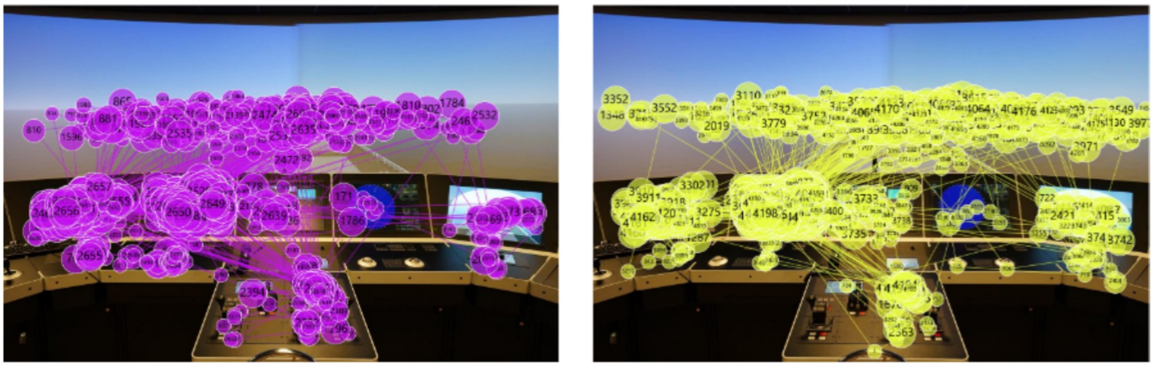


Figure 4.4: Gaze plot from the same participant as in figure 4.3 morning and afternoon.

TOI were analyzed with the idea that voyage duration would not matter, as all the TOI are defined at the same points of the voyage. The crossing to clear of traffic TOI were of specific interest, as it is the point where the simulation requires the most attention. However, that TOI had quite some variation in its duration, depending on the participant and time of day (average 1309 seconds, range 916-2237 seconds). Thus, there is little point in looking at the total number of eye fixations for the TOI. Instead, fixations per minute was the metric of choice. Figure 4.5 illustrates that no significant change in eye activity was found, apart from the outlier identified in the visual distribution analysis.

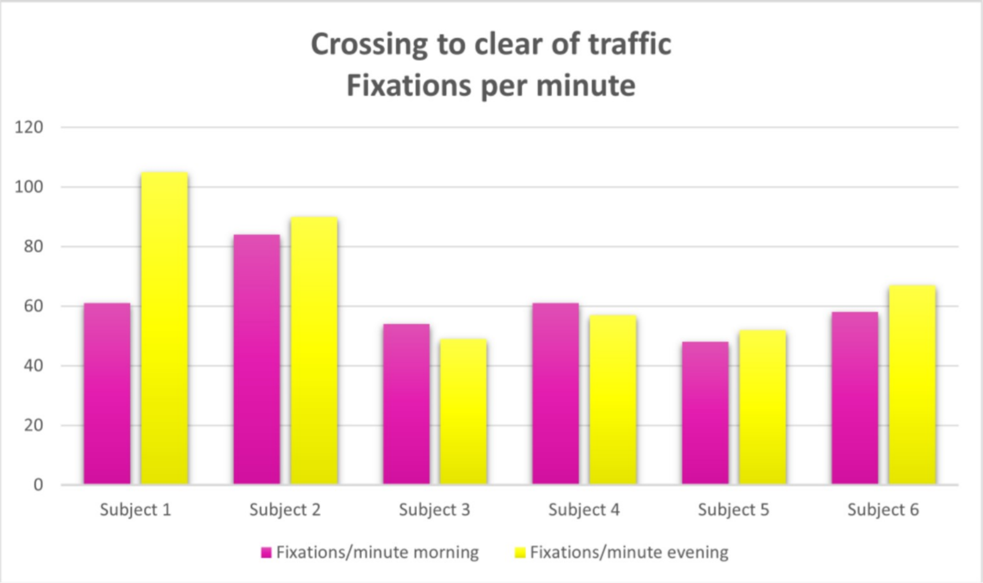


Figure 4.5: Fixations per minute for crossing to clear of traffic.

Relatively quite a bit more of the participants' time was used looking at the ECDIS in the morning than in the afternoon (17.63% compared to 11.96%). This can be explained by the fact that planning the voyage would take quite a bit less time in the afternoon, as they already did the planning once earlier in the day, even though the ECDIS was reset and the route deleted after each voyage.

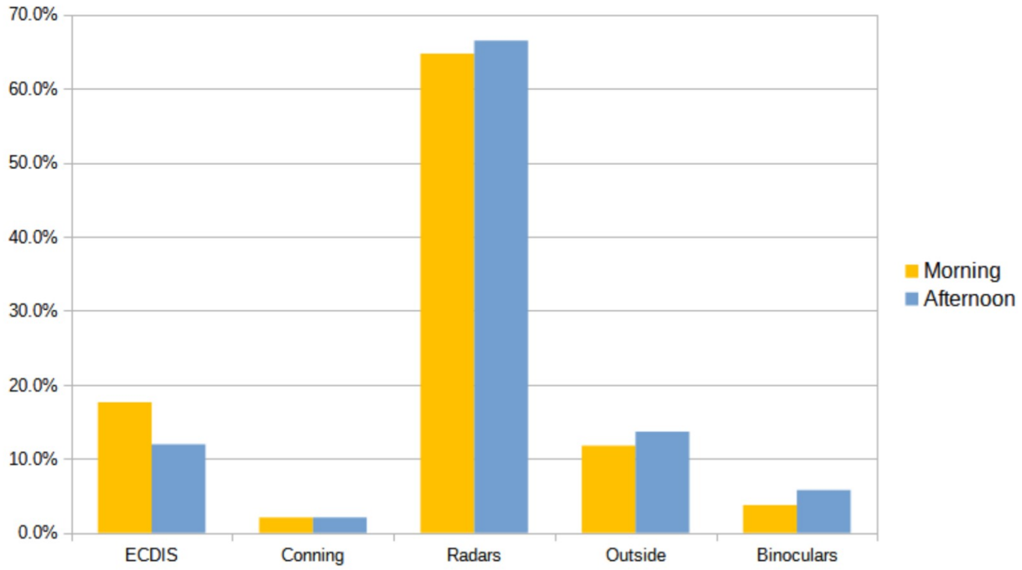


Figure 4.6: Average percentage of time spent on different AOI in the morning and afternoon.

The afternoon saw relatively more attention being focused on the binoculars, radars, and the outside view (figure 4.6). This increased visual activity can be fully explained by the decrease in attention devoted to the ECDIS, as ECDIS activity decreased by 5.67% in the afternoon, while cumulatively 5.68% more attention was directed towards the binoculars, radars, and outside view. In particular, the binoculars saw the greatest relative increase in usage between morning and afternoon, increasing from 3.76% to 5.78%. Conning usage was unchanged.

Chapter 5

Conclusion

We found that on average the level of fatigue experienced in this study by our subjects was not enough to affect their performance or risk-taking as navigators. It is hard to know whether this was due to the lack of active task-related fatigue, or if it was due to our small sample size not accurately reflecting the population. There are great individual differences in how tired someone gets after a passive day at university, and this does not accurately reflect the everyday workload experienced by seafarers.

We also found that navigators' uncertainty due to not having used certain navigational skills for a long time had a great effect on performance and risk-taking. This shows the importance of maintaining skills. Further research should be devoted to how long a navigator can be absent from navigational tasks before it starts having a negative effect on their performance. In addition, research should be devoted to how long it takes for a navigator to refresh their skills to be back on par. We also see the value of further studies looking more thoroughly into active task-related fatigue for seafarers.

Bibliography

Akoglu, Haldun (Aug. 2018). “User’s guide to correlation coefficients”. In: *Turkish Journal of Emergency Medicine* 18.

Allen, Paul, Emma Wadsworth, and Andrew Smith (Feb. 2007). “The prevention and management of seafarers’ fatigue: a review”. In: *International maritime health* 58, pp. 167–77.

– (Feb. 2008). “Seafarers’ fatigue: A review of the recent literature”. In: *International maritime health* 59, pp. 81–92.

– (July 2020). “The Relationship Between Recorded Hours of Work and Fatigue in Seafarers”. In: pp. 546–548.

American Heritage Dictionary Entry: fatigue (n.d.). `https://www.ahdictionary.com/word/search.html?q=fatigue`. (Accessed on 05/25/2022).

Atik, Oguz and Ömer Arslan (Aug. 2019). “Use of Eye Tracking for Assessment of Electronic Navigation Competency in Maritime Training”. In: *Journal of Eye Movement Research* 12.

Berman, Jules J. (2016). “Chapter 4 - Understanding Your Data”. In: *Data Simplification.* Ed. by Jules J. Berman. Boston: Morgan Kaufmann, pp. 135–187.

DNVGL-ST-0033.pdf (Jan. 2020). URL: `https://rules.dnv.com/docs/pdf/DNV/ST/2020-06/DNVGL-ST-0033.pdf`.

Ferrara, Michele and Luigi De Gennaro (May 2001). “How much sleep do we need?” In: *Sleep medicine reviews* 5, pp. 155–179.

Gillberg, Mats and Torbjörn Åkerstedt (Sept. 1994). “Sleep restriction and SWS-suppression: Effects on daytime alertness and night-time recovery”. In: *Journal of Sleep Research* 3, pp. 144 –151.

Goldberg, Joseph and Jonathan Helfman (Apr. 2010). “Comparing Information Graphics: A Critical Look at Eye Tracking”. In: *Conference on Human Factors in Computing Systems - Proceedings.*

Grech, Michelle (Jan. 2016). “Fatigue Risk Management: A Maritime Framework”. In: *International Journal of Environmental Research and Public Health* 13, p. 175.

Greenland, Sander and C. Poole (Jan. 2011). “Problems in common interpretations of statistics in scientific articles, expert reports, and testimony”. In: *Jurimetrics* 51, pp. 113–129.

Gupta, Charlotte C. et al. (2019). “Framework and Metrics for Online Fatigue Monitoring Within Submarine Teams Working in 24/7 Environments”. In: *IFAC-PapersOnLine* 52.19. 14th IFAC Symposium on Analysis, Design, and Evaluation of Human Machine Systems HMS 2019, pp. 259–264. URL: `https://www.sciencedirect.com/science/article/pii/S2405896319319391`.

Hansen, Dan and Qiang Ji (Mar. 2010a). “In the Eye of the Beholder: A Survey of Models for Eyes and Gaze”. In: *IEEE transactions on pattern analysis and machine intelligence* 32, pp. 478–500.

Hansen, Dan Witzner and Qiang Ji (2010b). “In the Eye of the Beholder: A Survey of Models for Eyes and Gaze”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.3, pp. 478–500.

Hareide, Odd Sveinung and Runar Ostnes (June 2018). “Validation of a Maritime Usability Study with Eye Tracking Data”. In: pp. 273–292.

Hauke, Jan and Tomasz Kossowski (June 2011). “Comparison of Values of Pearson’s and Spearman’s Correlation Coefficients on the Same Sets of Data”. In: *Quaestiones Geographicae* 30, pp. 87–93.

IMO - Guidelines on Fatigue (2019). `https://wwwcdn.imo.org/localresources/en/OurWork/HumanElement/Documents/MSC.1-Circ.1598.pdf`. (Accessed on 05/25/2022).

Irving, I. (1982). “Crossing the dover strait separation scheme”. In: *Journal of Navigation* 35.1, 97–102.

ISO (2020). *ISO 15007:2020(en), Road vehicles — Measurement and analysis of driver visual behaviour with respect to transport information and control systems.* `https://www.iso.org/obp/ui/#iso:std:iso:15007:ed-1:v1:en`. (Accessed on 05/16/2022).

Jensen, Olaf et al. (Oct. 2006). “Working conditions in international seafaring”. In: *Occupational medicine (Oxford, England)* 56, pp. 393–7.

Jepsen, Jørgen Riis, Zhiwei Zhao, and Wessel M.A. van Leeuwen (June 2015). “Seafarer fatigue: a review of risk factors, consequences for seafarers’ health and safety and options for mitigation”. In: *International maritime health* 66, pp. 106–117.

Komogortsev, Oleg and Javed Khan (Jan. 2008). “Eye Movement Prediction by Kalman Filter with Integrated Linear Horizontal Oculomotor Plant Mechanical Model”. In: pp. 229–236.

Krehbiel, Timothy (Jan. 2004). “Correlation Coefficient Rule of Thumb”. In: *Decision Sciences Journal of Innovative Education* 2, pp. 97–100.

Lützhöft, Margareta and Tania Dukic (2007). “Show me where you look and I’ll tell you if you’re safe: Eye tracking of maritime watchkeepers”. In.

Lützhöft, Margareta et al. (July 2010). “Fatigue at Sea in Swedish Shipping-A Field Study”. In: *American journal of industrial medicine* 53, pp. 733–40.

Martinez-Marquez, Daniel et al. (2021). “Application of Eye Tracking Technology in Aviation, Maritime, and Construction Industries: A Systematic Review”. In: *Sensors* 21.13. URL: `https://www.mdpi.com/1424-8220/21/13/4289`.

McKinley R. Andy; McIntire, Lindsey K.; Schmidt Regina; Repperger Daniel W.; Caldwell John A. (2011). “Evaluation of Eye Metrics as a Detector of Fatigue”. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 53.4, pp. 403–414. URL: `https://browzine.com/articles/4588573`.

Muczyński, B. et al. (2013). “Using eye tracking data for evaluation and improvement of training process on ship’s navigational bridge simulator”. In: *Zeszyty Naukowe Akademii Morskiej w Szczecinie* 33 (105)33 (105), 75–78.

Nettleton, David (2014). “Chapter 6 - Selection of Variables and Factor Derivation”. In: *Commercial Data Mining*. Ed. by David Nettleton. Boston: Morgan Kaufmann, pp. 79–104.

Newcombe, Robert (Oct. 1992). “Practical Statistics for Medical Research”. In: *Journal of Epidemiology Community Health* 46.

“Nicotine and health” (2014). In: *Drug and Therapeutics Bulletin* 52.7, pp. 78–81. eprint: `https://dtb.bmj.com/content/52/7/78.full.pdf`.

Oldenburg, Marcus and Hans-Joachim Jensen (Mar. 2019). “Stress and Strain among Seafarers Related to the Occupational Groups”. In: *International Journal of Environmental Research and Public Health* 16, p. 1153.

Olechno, Ewa et al. (May 2021). “Influence of Various Factors on Caffeine Content in Coffee Brews”. In: *Foods* 10, p. 1208.

Olsen, Anneli (2012). “The Tobii IVT Fixation Filter Algorithm description”. In.

Phillips, Ross Owen (2014). *An assessment of studies of human fatigue in land and sea transport Fatigue in Transport Report II.* `https://www.toi.no/getfile.php?mmfileid=39679`. (Accessed on 05/26/2022).

Phillips, Ross Owen, Tor-Olav Nævestad, and Torkel Bjørnskau (2015). *Fatigue in Operators of Land- and Sea-Based Transport Forms in Norway: Literature Review and Expert Opinion.* `https://www.toi.no/getfile.php?mmfileid=40171`. (Accessed on 05/26/2022).

Rousseau, Ronald, Leo Egghe, and Raf Guns (2018). “Chapter 4 - Statistics”. In: *Becoming Metric-Wise.* Ed. by Ronald Rousseau, Leo Egghe, and Raf Guns. Chandos Information Professional Series. Chandos Publishing, pp. 67–97.

Skvarekova, Iveta and Filip Skultety (2019). “Objective Measurement of Pilot´s Attention Using Eye Track Technology during IFR Flights”. In: *Transportation Research Procedia* 40. TRANSCOM 2019 13th International Scientific Conference on Sustainable, Modern and Safe Transport, pp. 1555–1562. URL: `https://www.sciencedirect.com/science/article/pii/S2352146519303862`.

Smith, Andrew et al. (Jan. 2003). *Fatigue Offshore: Phase 2. The short sea and coastal shipping industry.*

Snel, J. and Monicque Lorist (Dec. 2011). “Effects of caffeine on sleep and cognition”. In: *Progress in brain research* 190, pp. 105–17.

Sterne, Jonathan and George Smith (Aug. 2001). “Sifting the evidence-what’s wrong with significance tests?” In: *Phys Ther* 81, pp. 1464–1469.

Streilein, Tim et al. (2020). “Maritime Navigation: Characterizing Collaboration in a High-Speed Craft Navigation Activity”. eng. In.

Tanha, Kiarash, Neda Mohammadi, and Leila Janani (Jan. 2017). “P-value: What is and what is not”. In: *Medical journal of the Islamic Republic of Iran* 2017, pp. 377–378.

Tobii (Dec. 2019). *What is Fixation Filter?* `https://connect.tobiipro.com/s/article/What-is-Fixation-Filter?language=en_US`. (Accessed on 05/19/2022).

– (2021). *Understanding Pro Lab’s eye tracking metrics.* `https://www.tobiipro.com/learn-and-support/learn/steps-in-an-eye-tracking-study/data/understanding-Pro-Labs-eye-tracking-metrics/`. (Accessed on 05/16/2022).

Trondheim, Sveinung Odd, and Odd Sveinung Hareide (Apr. 2019). “Thesis for the Degree of Philosophiae Doctor The use of Eye Tracking Technology in Maritime High- Speed Craft Navigation lilNTNU”. PhD thesis.

Van Cutsem, Jeroen et al. (Aug. 2017). “The Effects of Mental Fatigue on Physical Performance: A Systematic Review”. In: *Sports Medicine* 47.

Wasserstein, Ronald and Nicole Lazar (Mar. 2016). “The ASA’s Statement on p-Values: Context, Process, and Purpose”. In: *The American Statistician* 70, pp. 00–00.

Watson, Nathaniel et al. (Apr. 2015). “Recommended Amount of Sleep for a Healthy Adult: A Joint Consensus Statement of the American Academy of Sleep Medicine and Sleep Research Society”. In: *Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine* 38.

What is Sleep & Why is It Important for Health? - ASA (n.d.). `https://www.sleepassociation.org/about-sleep/what-is-sleep/`. (Accessed on 05/21/2022).

Wilkinson, Robert, Robert Edwards, and Eric Haines (Dec. 1966). “Performance Following a Night of Reduced Sleep”. In: *Psychonomic Science* 5, pp. 471–472.

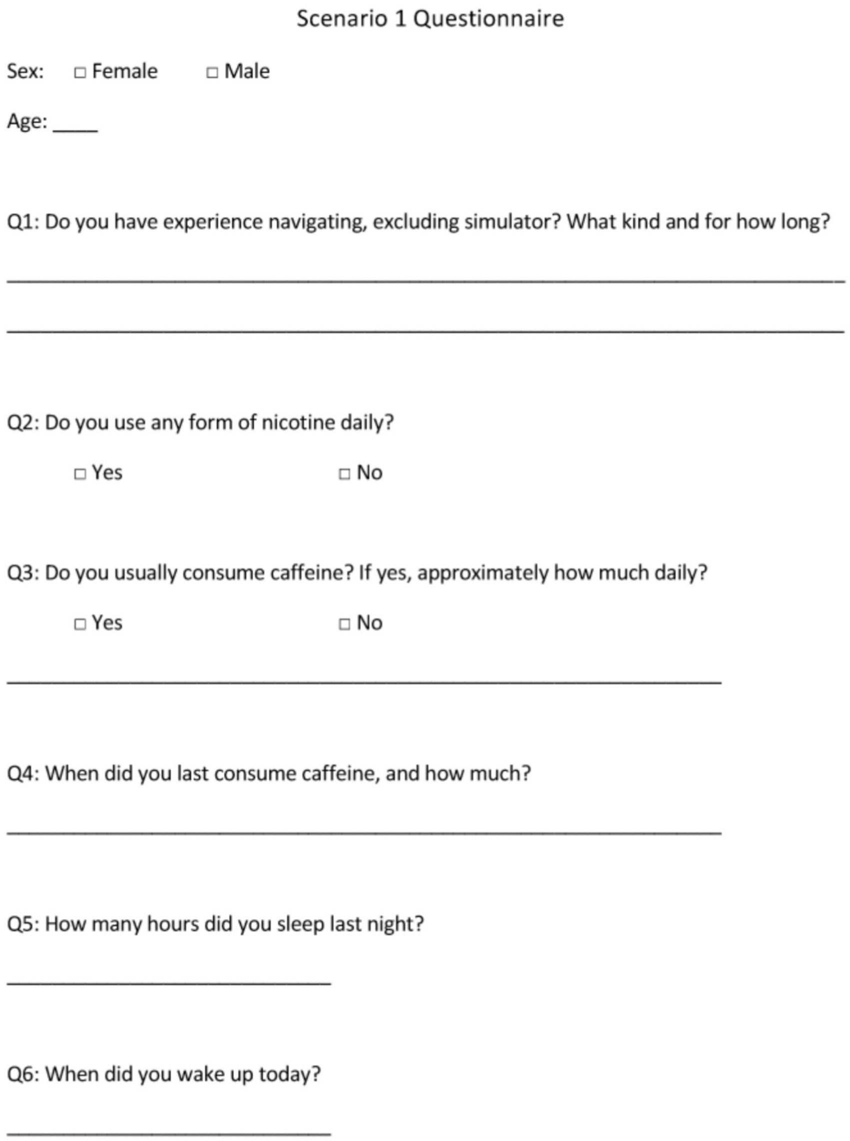
Zar, Jerrold (July 2005). “Spearman Rank Correlation: Overview”. In: vol. 5.

Zhang, Yunfeng and Anthony Hornof (Apr. 2011). “Mode-of-disparities error correction of eye-tracking data”. In: *Behavior research methods* 43, pp. 834–42.

Zhao, Luman, Guoyuan Li, and Houxiang Zhang (2020). “Multi-ship Collision Avoidance Control Strategy in Close-quarters Situations: a Case Study of Dover Strait Ferry Maneuvering”. In: *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*, pp. 1–3.

Appendix A

Questionnaires



Scenario 1 Questionnaire

Sex: ☐ Female ☐ Male

Age: ____

Q1: Do you have experience navigating, excluding simulator? What kind and for how long?

____

____

Q2: Do you use any form of nicotine daily?

☐ Yes ☐ No

Q3: Do you usually consume caffeine? If yes, approximately how much daily?

☐ Yes ☐ No

____

Q4: When did you last consume caffeine, and how much?

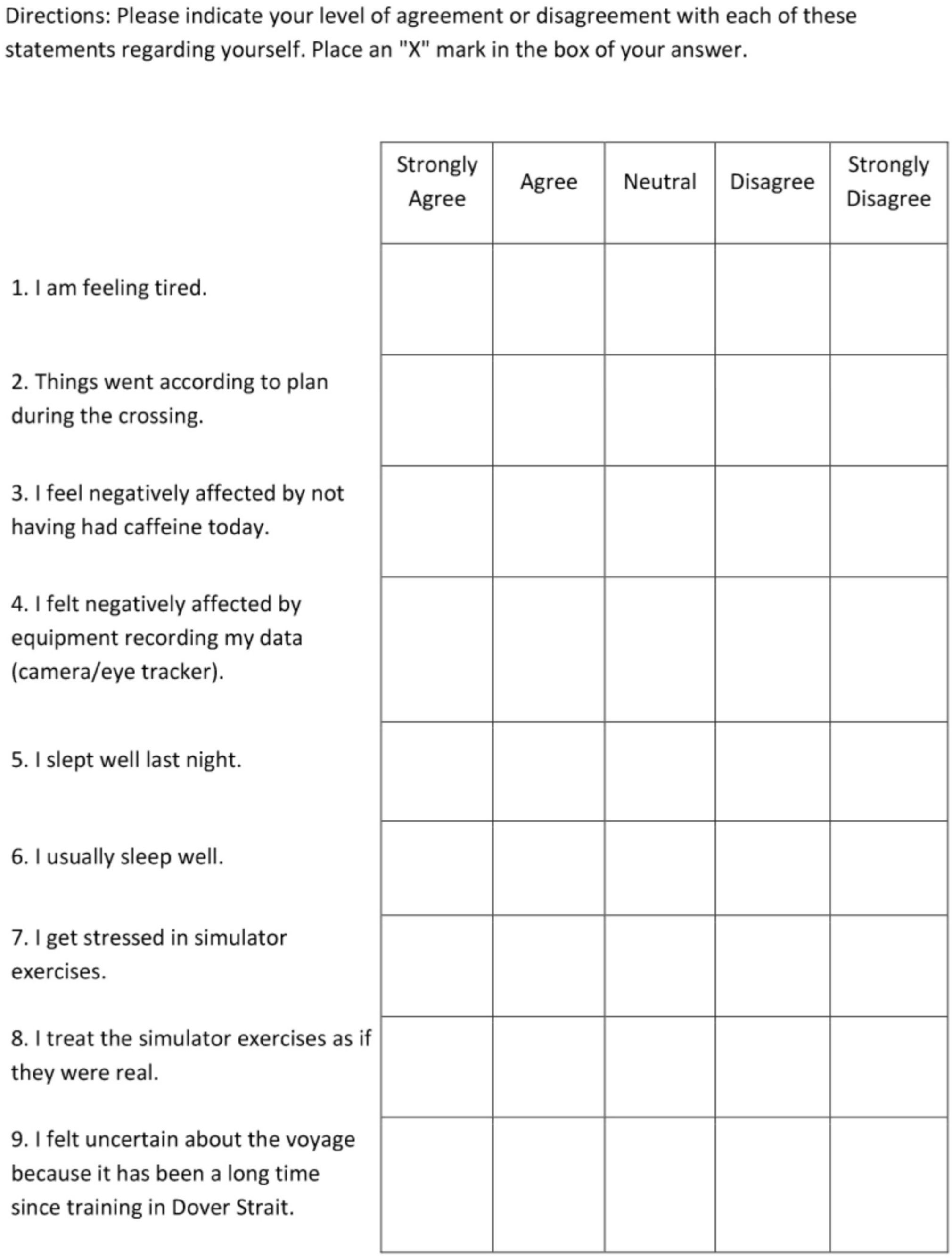
____

Q5: How many hours did you sleep last night?

____

Q6: When did you wake up today?

____



Directions: Please indicate your level of agreement or disagreement with each of these statements regarding yourself. Place an "X" mark in the box of your answer.

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1. I am feeling tired.					
2. Things went according to plan during the crossing.					
3. I feel negatively affected by not having had caffeine today.					
4. I felt negatively affected by equipment recording my data (camera/eye tracker).					
5. I slept well last night.					
6. I usually sleep well.					
7. I get stressed in simulator exercises.					
8. I treat the simulator exercises as if they were real.					
9. I felt uncertain about the voyage because it has been a long time since training in Dover Strait.					

Scenario 2 Questionnaire

Q1: Did you see any of the small motorboats in scenario 1 or 2?

____

Directions: Please indicate your level of agreement or disagreement with each of these statements regarding yourself. Place an "X" mark in the box of your answer.

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1. I am feeling tired.					
2. Things went according to plan during the crossing.					
3. I feel negatively affected by not having had caffeine today.					
4. I felt negatively affected by equipment recording my data (camera/eye tracker).					
5. I felt more prepared for the second exercise than for the first one.					

Appendix B

Caffeine usage

Participant	Usual caffeine intake
1	0.3-0.5 l of coffee or 0.5 l of Monster daily
2	2 cups of coffee daily
3	1-2 cups of coffee daily
4	2-3 times a week
5	1-2 cups of coffee daily
6	0.5-1 l of coffee daily

Participant	Last consumption	Amount consumed
1	19:00 yesterday	1 cup of coffee and 0.5 l of Monster
2	Yesterday	2 cups of coffee
3	13:00 yesterday	1 cup of coffee
4	Yesterday	330 ml of Red Bull
5	19:00 yesterday	1 cup of coffee
6	10:00 yesterday	1 cup of coffee

Appendix C

Visual attention distribution

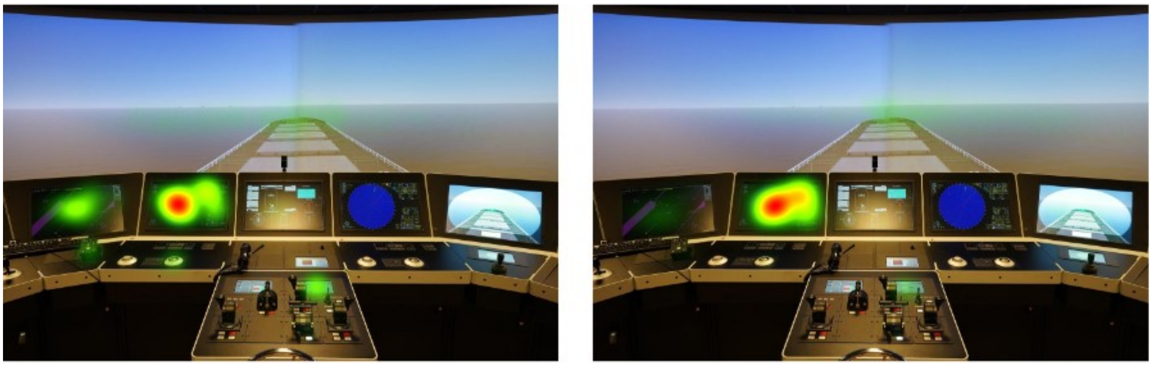


Figure C.1: Participant 1 heatmap morning and afternoon.

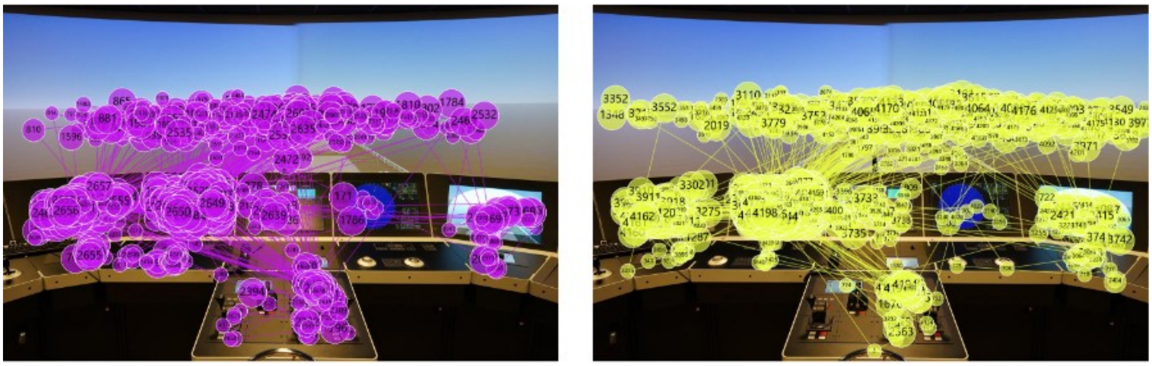


Figure C.2: Participant 1 gazeplot morning and afternoon.

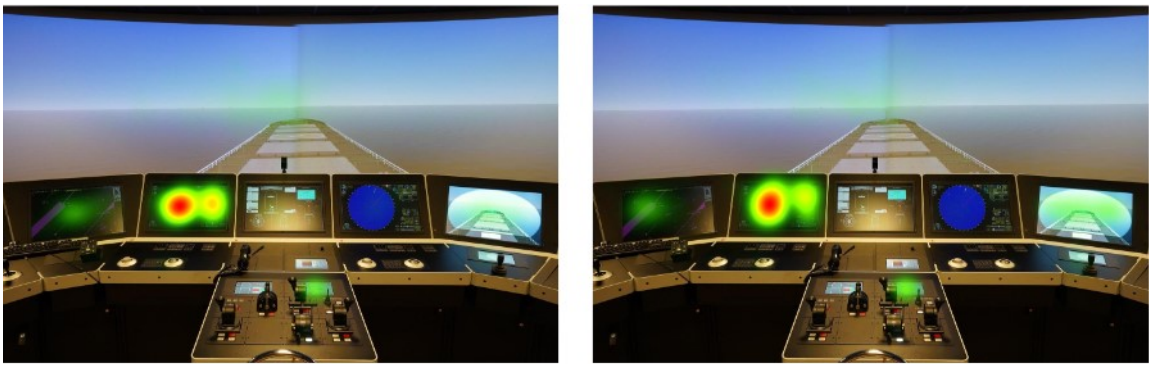


Figure C.3: Participant 2 heatmap morning and afternoon.

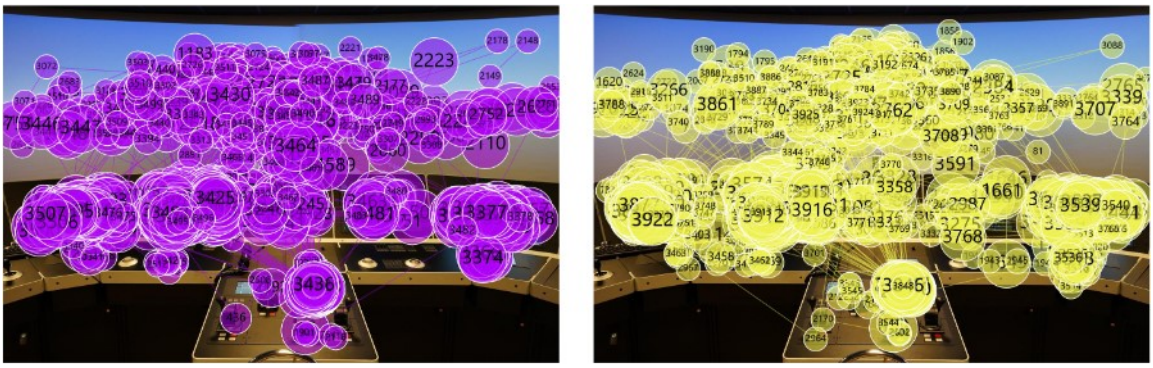


Figure C.4: Participant 2 gazeplot morning and afternoon.

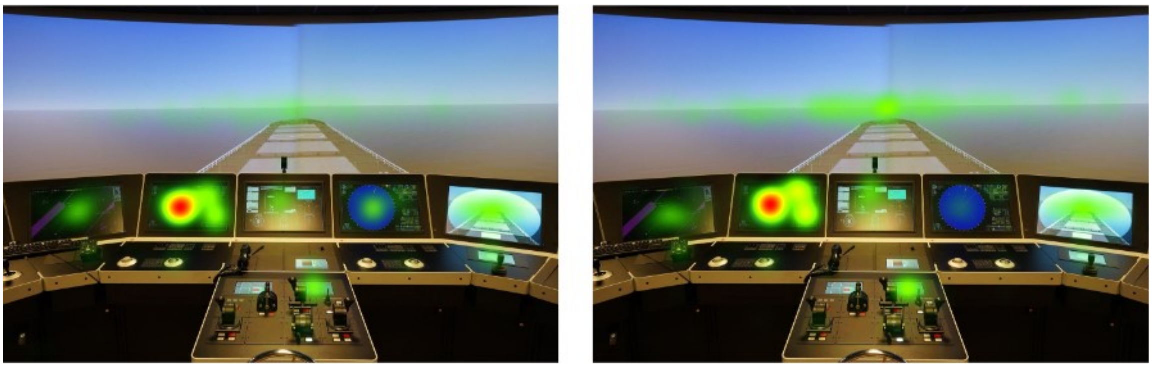


Figure C.5: Participant 3 heatmap morning and afternoon.

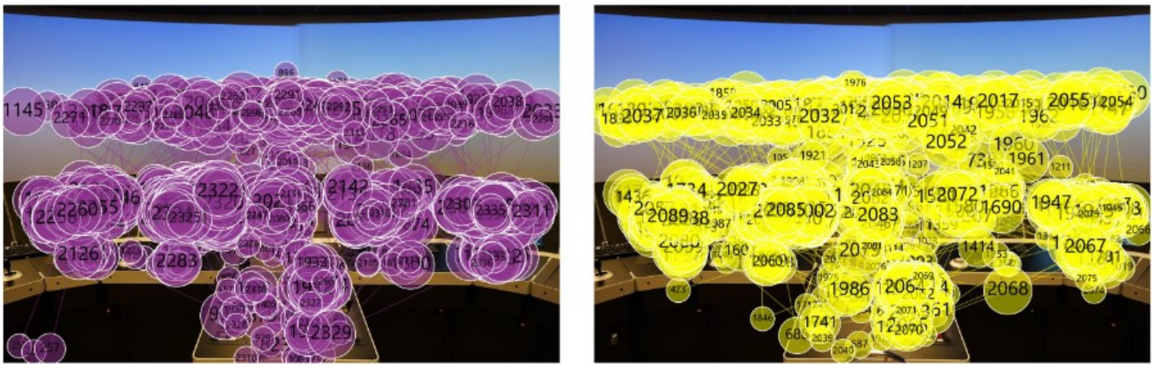


Figure C.6: Participant 3 gazeplot morning and afternoon.

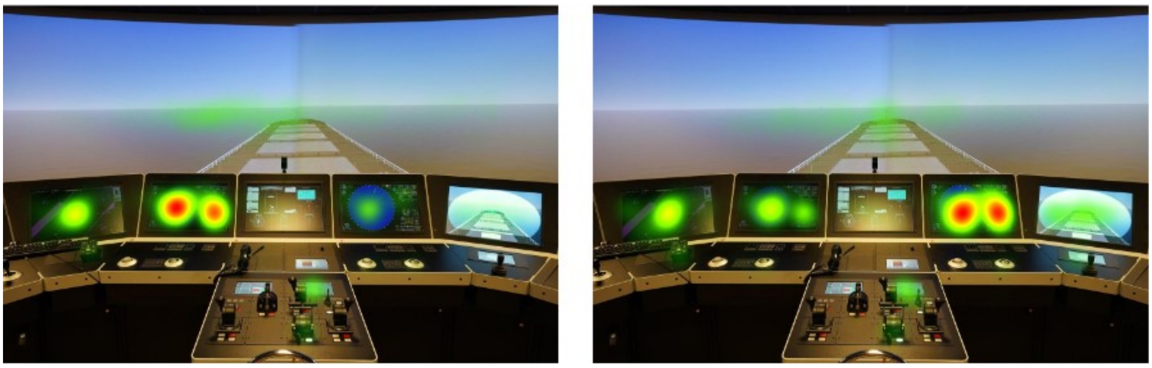


Figure C.7: Participant 4 heatmap morning and afternoon.

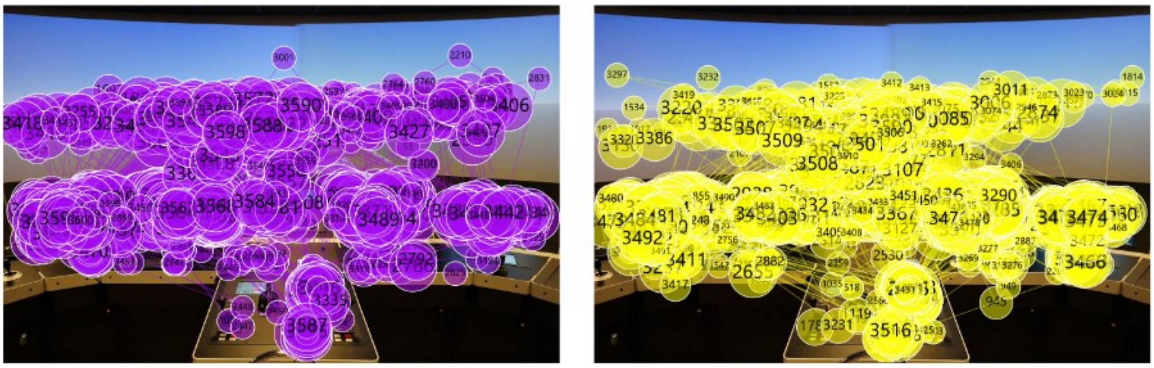


Figure C.8: Participant 4 gazeplot morning and afternoon.

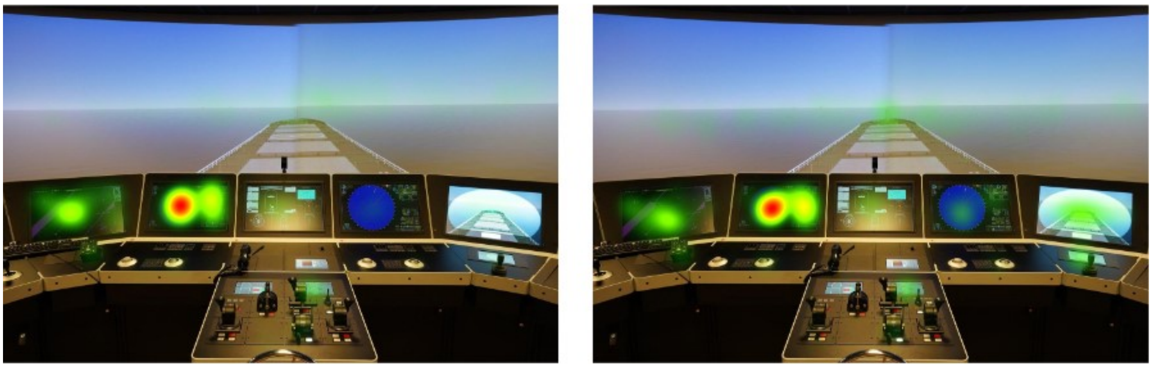


Figure C.9: Participant 5 heatmap morning and afternoon.

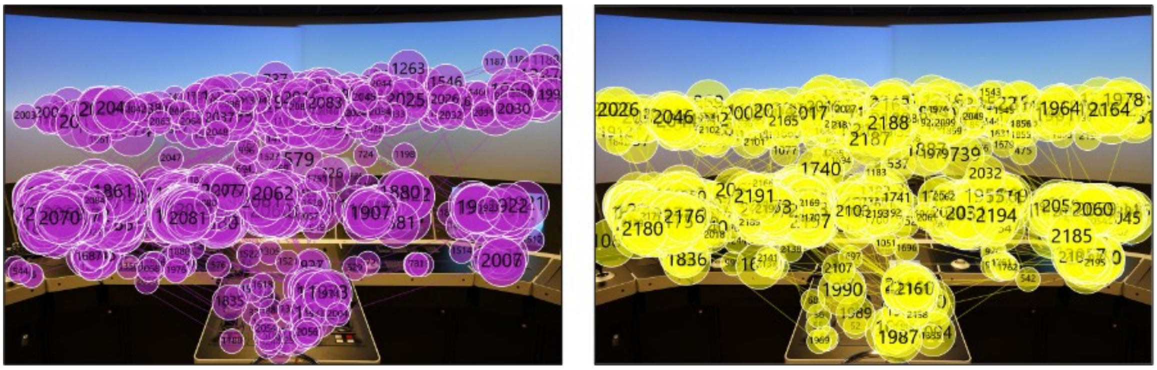


Figure C.10: Participant 5 gazeplot morning and afternoon.

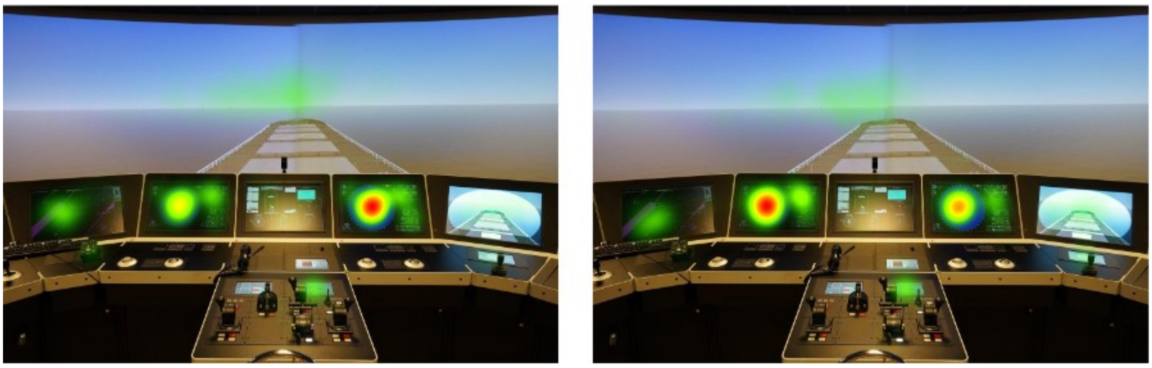


Figure C.11: Participant 6 heatmap morning and afternoon.

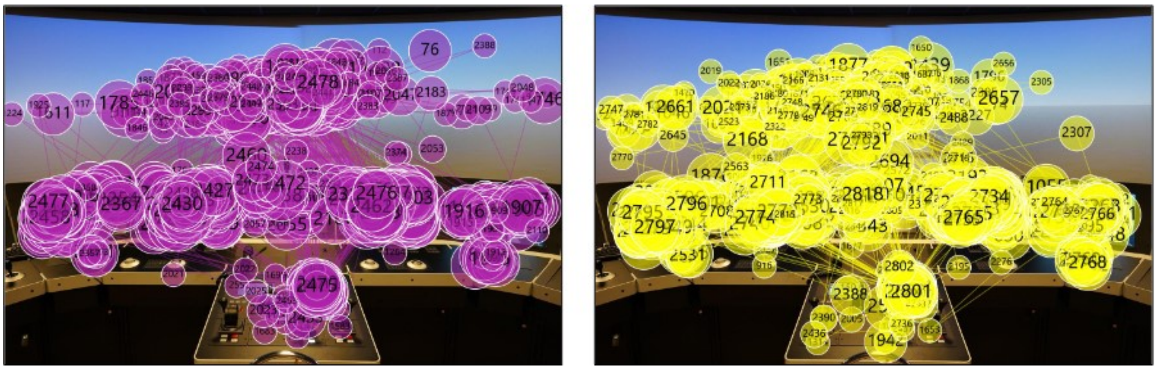


Figure C.12: Participant 6 gazeplot morning and afternoon.

NTNU
Kunnskap for en bedre verden