



Universiteit Leiden

Psychologie
Faculteit der Sociale Wetenschappen



Enhancing operator situational awareness through enhanced design of Shore Control Centre interface

Name: Lars Berntzen Arholm

Student number: s2313553

Date: 21.08.19

Supervisor: Guido Band

Second reader: Roy de Kleijn

Word count: 11350

Cognitive Psychology

Thesis Msc Applied Cognitive Psychology

Acknowledgement

A master thesis may be written by one person, but you never walk alone. I would like to acknowledge all those people who helped me along the way and made this project possible.

First and foremost, I would like to thank my university supervisor Guido Band and my MARIN supervisors Gerrit van der Want, Colin Guiking and Hans Huisman. Your supervision, feedback, ideas and support have been invaluable.

I would like to thank all the other maritime experts at MARIN who took time out of their day to give me input on my project; Egbert Ypma, Jaap Koolmees, Yvonne Koldenhof, Erwin van Iperen, Dimitri van Heel. I would also like to extend a warm thanks to the software team; Ionut Plesa, Arjan Lamaker, Flip Uithol for turning my ideas and drawings into a fully functional SCC.

I would also like to express my sincerest gratitude to the incredibly hospitable, helpful external parties, who met me with open arms, offering me experience, advice and let me visit their facilities to help me gain a broader overview for my project. To the whole crew at Windea de la Cour, especially Alex Meier and Hendrik Busshoff for inviting me. Alec Steevels and Marliene Pattipawaej who shared their process with me for how they developed completely new software for the Coastguard, and Marnix Bras for suggesting it in the first place. Hans van den Broek and Jaco R. Griffioen at Hogeschool Rotterdam for sharing their research on autonomous vessels. Karsten Seven at the Port of Rotterdam, the VTS personnel and remote pilots who took the time to speak with us. Koos Smoor at Kotug for showing us the remote-control interface and RT Borkum. ECDIS-developer Mads Friis Sørensen for providing insightful experience about the importance of user-centered design in maritime technology development.

I would like to thank all the participants, without your participation there would be no study.

Last, but absolutely not least, I would also like to thank all those who kept me sane through this process. Floris van der Oever, a tremendous officemate for exchanging ideas, discussing solutions and learning from each other's successes and mistakes. Michiel Bicker and Giovanni Trabace, the poor engineer-students who put up with my endless psychology rants, as well as all the other interns at MARIN. A special thanks go to Divyaa Balaji and Radosław Wincza for proof-reading my thesis. Finally, I would like to thank my parents, for supporting me through thick and thin of five years of studying abroad. I would not have been where I am today without you.

Abstract

In order to safely monitor and control autonomous vessels, it is essential for a Shore Control Centre (SCC) operator to achieve an adequate level of situational awareness (SA). To achieve this, an SCC interface must be created to support SA. However, there is a lack of empirical research on SA-supportive design for SCC interfaces, in addition to how many autonomous vessels an operator can monitor. To research this, an exploratory study was conducted with 13 participants, who were divided into three experimental groups and tasked to monitor a different number of vessels (1, 3 or 6) in the first setup of MARIN's SCC. The participants completed four simulated scenarios, three of which contained deliberate incidents, to see if they could become adequately SA within the current system and if this would differ between experimental conditions. Quantitative measures such as perceived SA, perceived workload, reaction time, usability and performance (CPA) were collected, in addition to qualitative feedback from the participants. The results indicate that the current system fails to support SA for more than 1 vessel, as the participants in the 3 and 6 vessel condition had lower perceived SA and higher perceived workload than the 1 vessel condition. Qualitative feedback from the participants was collected and related to relevant literature to create user-centred recommendations for how to improve the interface SA support. An SCC interface built employing these recommendations should enhance safe monitoring and controlling of autonomous vessels through improved SA support.

Introduction

Autonomous vessels have the potential to revolutionize the maritime industry. Porathe, Prison and Man (2014) mention several potential benefits, such as cost reduction, reduction of emissions, more attractive work environment and increased safety. Creating an attractive work environment is important, as it is expected that recruitment of marine personnel to long haul voyages will become more difficult (Ottesen, 2014; Porathe et al., 2014). Cost reduction in the crew will not be high enough to justify autonomous vessels, but the modified design built for cargo and not humans makes it economically viable (Kretschmann et al., 2017). Further, Rolls Royce estimate autonomous vessels to reduce fuel costs and operation costs by 20% and 40% respectively (Mooney, 2015). In Norway, it has been suggested that autonomous vessels could relieve congested roads by transporting goods by sea (Rødseth, 2017) and be a more economical

alternative to new, expensive bridges and tunnels (Rødseth, 2018). Wrobel et al. (2017) claim autonomous vessels would have caused fewer accidents in the past than manned vessels did.

Autonomous vessels would likely not sail without human supervision, however. Instead, they will be monitored and/or controlled by human operators in SCCs (MacKinnon, Man and Baldauf, 2015). The concept is quite new, and therefore there is only a handful of articles on the topic. In the final report by MUNIN (Maritime Unmanned Navigation through Intelligence in Networks), a joint European project to research autonomous vessels, they suggested that SCCs should be set up as a chain of centres around the world, distributing vessels amongst centres to only work in daylight hours (MacKinnon et al. 2015). As for the individual operator, MUNIN suggests that each operator would be responsible for 6 vessels (MacKinnon et al. 2015) so that 5 minutes is spent actively monitoring each ship, and then 30 minutes is spent passively monitoring all vessels (Man et al., 2018). However, the report does not elaborate on how this number of vessels or this monitoring strategy was tested or created, and the results are marked as inconclusive.

Automation

Although the idea of autonomous technology is to function independently, without external intervention, humans must usually still supervise it due to brittleness (Endsley, 2017). In this context, brittleness refers to automation not being able to solve situations it was not specifically designed to solve (Woods & Cook, 2006). Therefore, human operators would monitor the automation, and intervene if the automation could not handle a specific situation. However, while humans have the cognitive skills to solve these situations, they only have limited abilities when it comes to monitoring. When it comes to attentive limitations, Mackworth (1950) had participants monitor a clock which would skip a second on random intervals, and the participants had to make a note of every time it happened. He found that the participants were only able to monitor it for 30 minutes before performance decreased drastically.

One of the greatest problems with human monitoring of automation is over-reliance and complacency. Multiple authors have theorized that humans put too much faith in automation (e.g. Wiener and Curry, 1980; Parasuraman & Riley, 1997; May, 1999; Endsley, 2017). Because automation seems to perform perfectly, humans do not feel the need to closely monitor it. Therefore, when errors do occur in automation, human operators tend not to notice it (Endsley, 2017). Further, Endsley (2017) says that when automation is correct, people are more likely to

make a correct decision, however, when it is incorrect, people do worse than they would do without any advice at all. This was empirically supported by Skitka, Mosier, and Burdick (1999), who compared an automated and a non-automated flight task. They trained the participants not to trust the decision aid 100%, yet the participants in the automated condition committed over-reliance errors when the automation failed, and in total performed worse than the non-automation condition. In a different study by Riley (1994), plane pilots experienced this automation bias to a higher degree than student participants, even though the experimental task was not related to aviation. This suggests that experience does not decrease automation bias, but rather increases susceptibility (Riley, 1994). May, Molloy and Parasuraman (1993) found that detection of automation failure varies inversely with automation reliability, which was validated by Oakley, Mouloua & Hancock (2003) using more levels of reliability. This effect has also been empirically validated in a maritime context by Pazouki, Forbes, Norman and Woodward (2018). In their small-sample study (n=12), all but one participant who did not receive specific training failed to recognise a failure in the automation which was not announced by an alarm. Pazouki et al. (2018) believe this to be due to ineffective monitoring and over-reliance.

As MacKinnon et al. (2015) wrote in the MUNIN report on SCCs, SCC operators must *“be able to quickly identify operational abnormalities, unexpected threats and errors quickly and efficiently in a highly automated context”*. However, due to the mentioned automation-complacency, operators’ risks falling “out of the loop” (OOTL). This challenge is referred to as the “automation conundrum” (Endsley, 2017). With operators being OOTL, they are slow to realize that their intervention is required, and when they do intervene, their performance is reduced (Endsley, 2017). Endsley and Kris (1995a) found that OOTL issues are caused by a loss of SA in the operator.

SA

The casual definition of SA is *“knowing what’s going on”* (Endsley, 1995a), but Endsley (1995a) also defines SA more formally as *“the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future”*. The Endsley model (Endsley, 1995a) describe SA as states, divided into three different levels: perception (level 1), comprehension (level 2) and projection (level 3). Perception is the most basic level, with activities like monitoring and simple recognition. This process depends largely on working memory and long-term memory, selectively guiding limited

attentional resources towards the most relevant areas, by knowing which areas should be prioritized (Endsley, 2001; Johannsdottir & Herdman, 2010). The second level, comprehension, goes beyond simple awareness of environmental elements, to understand the significance of these perceived elements. The third level, projection, is the ability to project future actions based on these elements. The processes in level 2 and 3 rely on working memory, with the operator combining new information with existing knowledge (Endsley, 1999; Endsley, 2001). The importance of SA in a maritime context has also been tested empirically by Cordon, Mestre and Walliser (2016). They performed a factor analysis on a large list of words connected to seafaring, which they had marine officers rate depending on how important they were to do a good job as a marine officer. The analysis revealed underlying factors very similar to the Endsley SA model.

SA is the primary way to measure the effectiveness of an interface system in an SCC (see MacKinnon et al., 2015; Endsley, 2017; Ramos et al., 2019). To explain the SA model in an SCC context, an ECDIS (Electronic Chart Display and Information System, an electronic sea map) will be used as an example. To reach level 1 SA, the operator needs to perceive, recognise and monitor the different elements on their ECDIS. To reach level 2, the operator must fully comprehend what these elements mean (i.e. the dots on my ECIDS represent real-life vessels in a specific area). To reach level 3, the operator must understand how these elements can project a future situation, for instance, an emergency situation where the operator must intervene (i.e. a traffic vessel is crossing my route from starboard, I should monitor the situation to make sure that my autonomous vessel gives way).

There are several important factors that cause loss of SA (Endsley, 2017), like cognitive engagement, complexity and workload. Endsley (2017) notes that it is inherently challenging to make an operator stay cognitively engaged in a task when they are not actively participating in the task. Further, SA is negatively affected by increased system complexity, as it reduces the system predictability since it is difficult for the operator to make an accurate mental model of how the system works (Endsley, 2017).

Not only is SA affected negatively by task complexity, but also by display complexity (Endsley, 2017). Often interfaces for automated systems do not provide sufficiently salient information about the state of automation nor does it provide much feedback about the state of the system (i.e. system assessment, ability to handle the current/upcoming situation) it controls (Endsley, 2017). This low

level of transparency can make it difficult to comprehend the system and project future actions. Endsley (2017) notes that this can be a challenge for a designer, as removing unnecessary information is often needed to highlight information the designer regards as important, but this may also cause other important information to be removed by accident. Automation that gathers and present information (level 1 SA), and then integrates that into the specific context to help comprehension and projection (level 2 and 3 of SA) enhance SA (Endsley, 2017). However, making information too salient, or automating an information-filtering system to help the operator find important information can significantly degrade SA, as the operator could lose overview over the full situation (Endsley, 2017). If these challenges are solved adequately, Endsley (2017) believes that SA in an automation context can be significantly improved. As Endsley said in (2001): *"Success will go to the developers who understand how to combine and present the vast amounts of data now available from the many technological systems present in order to provide true situation awareness"*.

The informational presentation also affects workload. While onboard crew would have received sensorial information from the ship, giving them "ship sense" (Prison, Dahlman & Lundh, 2013), an SCC operator must gain their SA through information on a screen. If too much information is perceived, the operator will not have enough attentional capacity to pay enough attention to each of them and will suffer from "informational overload" (Galbraith, 1974). This refers to a task having information-processing requirements that exceed the capacity of the individual (Galbraith, 1974), and has been recognised as a problem in human-computer interaction for a long time (Jones & Kelley, 2018). Informational overload increases the general cognitive workload for an operator, which in turn decreases SA (Endsley, 1999). In SA terms, this occurs when the operator is not able to fully perceive all information on their displays, and therefore not able to fully comprehend nor project future actions. A simple solution could be to drastically reduce the information presented to the operator on the displays. However, this must be done with care, as the operator still needs to have enough information to achieve adequate SA. With too little information, the operator cannot fully comprehend the situation of the vessel, thus making the operator unavailable to project future actions. Further, too little information would cause cognitive underload, in which attentional resources decrease to adjust for reduced demand for attention (Young & Stanton, 2002). This may even be a bigger concern than cognitive overload, as it too decreases performance, but it is harder to detect (Young & Stanton, 2002).

SCCs

There is little knowledge of the informational requirements of an SCC. In the only known study on this topic, Porathe et al. (2014) interviewed 6 bridge officers with the intention of finding the most important information that is needed to gain situational awareness in an SCC. They came up with 165 pieces of information, split into 9 different categories: voyage plan, sailing, observations, safety and emergencies, safety, cargo, technical, shore control centre and administrative. However, these suggested informational requirements have not yet been empirically researched and implementing all 165 pieces of information could potentially evoke informational overload.

To develop a system which integrates this information in an optimized way, it is crucial to use a human-centred design approach. According to Billings (1997), through human-centred automation, one can ensure that the human operator is able to monitor a system, that they receive sufficient feedback and that the automation works in a predictable way, so that high levels of performance can be reached, and which would support SA. Human-centred design has been an essential principle in developing the new interface software for the Dutch Coastguard (A. Steevels, M. Pattipawaej, personal communication, May 9th, 2019). Furthermore, according to ECDIS developer M. F. Sørensen, the lack of human-centred design was what caused the first generations of ECDIS to be so complex and hard to use that they were practically useless (M. F. Sørensen, personal communication, May 13th, 2019).

To address all these mentioned challenges, the design of an SCC interface is therefore of utmost importance. Various authors within various fields of automation, also some specifically working with SCCs, believe that SA issues with autonomous interfaces can be resolved partly or fully through improved design (Arrabito, 2010; Parasuraman & Manzey, 2010; Westrenen & Praetorius, 2014; Ottesen, 2014; Endsley, 2017; Pazouki et al. 2018; Man et al., 2018).

However, there has been little research about SAs implication on interface design for SCCs (Man et al., 2018). Such research is of high importance, as an SCC cannot simply copy the design of the bridge of a ship. This was tested by Man et al. (2015), but it was not effective. Instead, Man et al. (2015) notes, SCCs has to be designed specifically to support the operators need to develop SA. Man et al. (2018) also underline the importance of ecological design when creating SCCs, taking into account both the environmental changes related to an SCC, as well as cognitive limitations.

The scope of this study is to evaluate MARINs first version of an SCC interface by measuring how well it supports SA, workload, performance and usability. Thus, this study aims to answer the main research question R1:

- *R1: What information should be presented to an SCC operator in an optimal SCC interface, and how should this information be presented, so that the interface is usable and supports the creation of SA, ensuring the operator's ability to safely monitor and control an autonomous vessel, in this specific operational context?*

By answering the sub-research questions R2, R3 and R4:

- *R2: Conceptualising SA as perceived SA (measured with self-report questionnaires) and objective SA (measured with eye-tracking), to what degree are the participants able to become situationally aware with the current interface?*
- *R3: Considering their navigational performance and which feedback they gave concerning informational requirements; which parts of the interface need to be changed to enhance situational awareness and performance?*
- *R4: How does the number of vessels to monitor and control affect the operator's performance, workload, and situational awareness?*

The hypothesis for the main research questions is:

- *H1: The necessary information for the latter will be; heading, RPM, position, course over ground, rate of turn, speed over ground, planned route, mode information, engine/rudder status, a security parameter and a cargo parameter, as well as weather information.*

And the hypotheses for sub-research questions are:

- *H2: The participants will experience low levels of perceived SA, have low levels of relevant fixations and high entropies, as the system does not currently support SA sufficiently.*
- *H3: Most parts of the interface must be changed to some degree to enhance SA and performance, as a user-centred design is necessary to achieve high levels of SA and performance.*
- *H4: The operator will perform better and be more SA when monitoring fewer vessels, as fewer vessels should produce less workload, and low workload is associated with high SA.*

Method

Design

This exploratory study employed a between-subjects, mixed-methods design. The independent variables are the number of vessels the participants monitored (1, 3 or 6) and purposefully caused incidents in the scenarios. The distance at the start time of incident between the incident vessel and the autonomous vessel was measured as a covariate. The conditions were counterbalanced by randomizing scenario 2, 3 and 4, to avoid learning effects. Scenario 1 was not randomized, as it was important to have this as the first scenario, to build trust in the automation, which as previously mentioned affects SA and OOTL problems through automation-complacency (Parasuraman & Riley 1997; Endsley 2017). The dependent variables are SA (measured by eye-tracking reaction time and SART inspired questionnaire items), perceived workload (adjusted NASA-TLX, Hart & Staveland, n.d.), usability (measured by SUS scale, Brooke, 1996) and performance (measured by CPA).

Participants

The study was conducted with 13 male participants with a mean age of 43.3 years (range 21-72, SD = 18.04), thus including younger participants than what previous studies have done (Man et al. 2015; 2018). The participants were eligible for the study if they had some experience with maritime navigation and were; maritime students (graduated/nearly-graduated) (n=3), prior mariners (n=3), current mariners (n=3), prior mariners experienced with MARIN simulator software (n=2) or Vessel Traffic Services (VTS) personnel (n=2). The inclusion criteria were purposely broad to capture the entire spectrum of possible future SCC operators, including both experienced and novice seafarers. This was important in order to design for the relevant end-users, in accordance with user-centred design principles. The recruitment was conducted with opportunity sampling through MARINs and the researcher's contact network. All student and some prior mariners were recruited by contacting the following Dutch maritime institutes: Maritiem Instituut Willem Barentsz, Hogeschool of Amsterdam and Hogeschool of Rotterdam. The participants with VTS background were recruited from the VTS at Port of Rotterdam. All current and some prior mariners were recruited from the semi-autonomous vessel Windea de la Cour. The prior mariners with Dolphin experience were recruited internally within MARIN. In accordance with MARINs internal guidelines for participant compensation, students were compensated with 150 euros and

professionals with 500 euros. The study was ethically approved by the ethics committee of Leiden University, Institute for Psychological Research. The study was conducted in accordance with GDPR rules concerning privacy and storage of data.

Procedure

Before the day of their participation, each participant received a written briefing (appendix A) to understand the basic principles about SCCs, autonomous vessels and the task of SCC operators. On the day of the experiment, the participant completed the informed consent form (Appendix B), which informed about the general nature of the study, anonymisation of the data, the measurements which would be taken and the participant's possibility to withdraw at any moment without consequences. The participant also completed a questionnaire collecting demographic details (Appendix C), as well as completing a questionnaire measuring their attitudes to automation (appendix D). The participant was then theoretically briefed for 30 minutes to explain the topics mentioned in the written briefing more in-depth, as well as explaining the general idea of the interface. Finally, the participant was given 20 minutes for practical familiarization with the SCC setup. The instructions were identical between the experimental conditions.

The study was conducted in MARIN's facilities in Wageningen, in a room dedicated to SCC simulation. A maximum of two participants was done per day to ensure a reliable quality of the experiment. The duration of the full experiment (including briefings and breaks) was 5 hours for one participant and 7 hours for two participants. The same interviewer conducted both the interviews and experiment observation, to ensure there was no inter-experimenter variability.

During the debriefing (Appendix H) the participants were explained more thoroughly the specific objectives of the study, how automation had been simulated and how to contact the experimenters in case of questions or discomfort due to having been part of the experiment.

Tasks

The participants acted as SCC operators and were tasked to monitor autonomous vessels and intervene when necessary. The participant did 4 sessions of 30 minutes, during which the participant was exposed to an incident in scenario 2, 3 and 4. The deliberate incidents involved violations of COLREGS, a vessel suddenly appearing on their electronic sea chart and a rudder failure. The full details of each incident can be found in Appendix I. The participants were

responsible for medium-sized shipping vessels, sailing outside the coast of the Netherlands and Barcelona. The routes for the maritime simulation was created with inspiration from real-life maritime tracking from marinetraffic.com (pictures in Appendix J). Experimental conditions were randomized within the occupational group to ensure that experimental conditions were equally distributed among groups. The order of scenarios was randomized between all participants. All randomization was done using Randomizer.org (Urbaniak & Plous, 2013).

Each scenario was created to measure the SA level of the participant in two general situations; with or without an incident occurring. As previously mentioned, the biggest challenge for SCC operators is to stay SA (MacKinnon et al., 2015; Ramos et al, 2019), mainly due to OOTL problems (Endsley & Kiris, 1995). By exposing them to scenarios with low levels of stimuli for prolonged periods of time, they were hypothesised to become OOTL, thus lowering their SA. The scenarios would, therefore, measure their ability to quickly regain SA to efficiently solve the incidents. Furthermore, the workload was measured as it is an important part of SA, and especially relevant for the last research question, being the part of SA hypothesized to cause a difference between the experimental conditions. All tasks were created specifically for this study, in cooperation with maritime experts at MARIN, to ensure validity and feasibility.

Apparatus

An SCC simulator was developed based on the existing MARIN simulation software “Dolphin”. The final set-up consisted of five displays. Three of the displays are traditional maritime features: an ECDIS display, a RADAR display and a conning display. An ECDIS plots the position of the operator’s vessel and nearby traffic on a maritime chart, the RADAR plots any object of sufficient size in close proximity and the conning display informs of vessels navigational information (i.e. speed, heading, rate of turn, weather). The other two, the vessel status display and the manual control window, are specific for an SCC. The vessel status display was supposed to inform of the health status of the vessel(s) based on Porathe et al. (2014) suggested informational requirements, in addition to the mode of automation. However, due to technical difficulties, it could only change colour to indicate a general error, which the participant had to locate specifically themselves. The manual control window let the participant change from autonomous to manual control, by selecting either waypoint (autonomous, track following), course (custom course and speed) or “manual command” (direct, remote control with virtual telegraph, rudder and bowtruster). The inspiration

for the SCC setup came from MUNINs suggested setup from 2015 (Rødseth & Baumeister), as it was considered to be a good presentation of the necessary information to become situationally aware. However, MUNIN did not elaborate on how the interface was created, nor how it was tested. Therefore, part of this study is exploring empirically how well an SCC inspired by their interface works. The MUNIN setup was chosen as it seemed to provide sufficient information to become aware of the external conditions through ECDIS and RADAR, and internal conditions through the conning and the vessel status display. This information was hypothesized to support higher levels of SA, supporting the participant comprehending their own vessel and its environment (SA level 2) and thus enable projection of future states (SA level 3), which is important in an autonomous interface (Endsley, 2017). The system also supports low level of SA (level 1) by presenting the functions and information in the interface in an intuitive manner. The alarms in the SCC setup was kept to a minimum, with no audible alarms or pop-up notifications. This was done to prevent the operator from relying too much on alarms and thus relying too much on the system, as this could make the operator too complacent, lowering SA and performance due to automation bias (Parsuraman & Riley, 1997). The full set-up is available in Appendix K. The full experiment was tested with 1 pilot tester. Based on the experience from this pre-testing, errors in routes and interface were corrected before the participants took part in the experiment.

Measures

Demographical information, appendix C: Collected the following data: Age, nationality, current and prior maritime occupations, previous experience with maritime navigation, previous experience with maritime simulators, previous experience with MARIN software. It took an average of 5 minutes to complete.

Attitudes to automation (quantitative measure), appendix D: A questionnaire was created to measure attitude towards maritime automation, as this could have had caused cognitive overhead (Parasuraman & Riley, 1997) and confound performance. The questionnaire was inspired by the Technology Acceptance Model (Davis, 1989), and took an average of 5 minutes to complete.

Eye-tracking: Eye-tracking was measured with the intention of providing an objective measurement of perceived SA, as SAGAT had been found too intrusive past studies at MARIN. Further, stopping a participant doing a task and asking them about awareness could have artificially increased their awareness.

Post-scenario questionnaire (quantitative measure), appendix E: After each scenario, the participant completed a questionnaire measuring perceived workload, SA and usability. Perceived workload (items 1-6) was measured using the entire NASA-TLX questionnaire, which was measured due to its relation with SA (Endsley, 1999) and to possibly further explain differences between experimental conditions. Perceived SA (7-10) was measured using relevant items from the SART (Selcon & Taylor, 1990), leaving out workload items. Usability (items 11-12) was measured to look for differences between scenarios. It took 5 minutes to complete on average.

Post-interview (qualitative measure), appendix G: A semi-structured interview was conducted to measure participants overall perceived SA, perceived workload, perceived usability and perceived realism of scenarios. Items 1-6 were inspired by the SART (going more into depth than post-scenario questionnaire), item 7 was created to gain feedback on the realism of scenarios (inspired by Man et al., 2018), items 8-10 and 13 measured overall perceived usability, and items 11-12 measured overall perceived workload. The interview took 30 minutes on average.

SUS usability questionnaire (quantitative measure), appendix F: The SUS measures the perceived usability. The major advantage of using SUS for this study is that it produces a global score which can be compared to a global benchmark (Lewis, 2018), which was especially useful due to the lack of control group/baseline. The questions were answered on a Likert scale from 1 (strongly disagree) to 5 (strongly agree). The questionnaire was presented in English, but with Dutch translation present, created by a native Dutch speaker at MARIN who used the questionnaire in his work (van den Oever, 2019). It took 5 minutes on average to complete.

Performance (quantitative measure): Performance was measured by CPA (closest point of approach), which indicates the closest point a vessel will be to approach another vessel. The measure is given in nautical miles to the incident vessel and was automatically computed by the simulation software. CPA is a common performance measure in a maritime context, which would usually be supplemented by a subjective expert rating, however, there are no SCC operator experts yet. Therefore, the measure was simplified to be that a higher CPA indicated safer sailing.

Analysis

The participants were split into three experimental condition; monitoring 1 (n=4), 3 (n=4) or 6 (n=5) autonomous vessels. This was important to gain an understanding of how the number of

vessels affects perceived SA and perceived workload, thus indicating how the current interface is able to support safe monitoring and control of several vessels. Two recordings of eye-tracking data disappeared while transferring files to the computer with the eye-tracking software. Although an effort was made to retrieve the missing files, this was unsuccessful. The analysis where this could have had an affected makes a mention of this. Due to the small sample size and purposefully high variability in participants, there were several outliers. However, as none of these were due to measurement error, and likely reflected the true performance or opinion of the participant in question, these outliers were not removed. This made the analysis more difficult but removing or changing the outliers could have caused artificial results. The raw eye-tracking data was analysed in Imotion/Tobii software.

Attitude to maritime automation analysis

The questions were answered on a Likert scale from 1 (strongly disagree) to 5 (strongly agree). A higher score indicated a more positive attitude to automation.

Post-scenario questionnaire analysis

For workload, lower scores indicate lower perceived workload. For SA, lower scores indicate higher perceived SA. For reaction time, higher scores indicate higher reaction time.

SA eye-tracking analysis

Fixations on relevant Areas of Interest (AOIs) and entropy would have been used to calculate SA level 1 and 3, respectively, based on research by Moore & Gugerty (2010) and Merwe et al. (2012). However, due to low variability in fixations, a lack of a unified SCC operator performance measure and not being able to extract scanpath from the eye-tracking software, this was not possible to calculate. As reaction time has been found to be a possible measure of SA (Endsley, Sollenberger & Stein, 2000), it was therefore used as the replacement measure for “objective” SA.

SUS questionnaire analysis

A score was calculated for each participant by subtracting 1 from the score of each odd number items, subtracting the value of the even-numbered items from 5, and multiplying the total score with 2.5 (Brooke, 1996). Then, a global average for all participants was calculated. Total scores can range from 0 to 100. The industry average is 68, but the industry standard is 80 (Lewis, 2018).

Qualitative data analysis

The qualitative data was assessed inspired by grounded theory (Glaser & Strauss, 1967). Grounded theory is a qualitative research approach where you do not start with specific theory, but instead develop it based on the data whilst analysing it (Glaser & Strauss, 1967). In this context, this was done by taking notes of common themes of suggestions and actions between participants and between groups and creating a “theory” of what an ideal SCC would need to contain.

Performance analysis

CPA is a continuous measure which was correlated with other relevant continuous variables. Because the participants could freely change their routes, the scenarios had to be dynamically adjusted in some cases in scenario 2 and 3. This caused some differences in the starting position between the incident vessel and autonomous vessel between participants, however, this was taken into account in all relations calculated with the CPA, using either partial correlations or ANCOVA.

Results

To test R2, regarding adequate level of SA, perceived SA and reaction time was compared to perceived workload, performance and perceived performance between the scenario and the experimental conditions. To test R3, regarding the usability of the system, the global SUS score was compared to perceived SA, qualitative data was collected and compared to reaction times. To test R4, about differences between monitoring a different number of vessels, differences in means between experimental conditions were compared in perceived SA, perceived workload, performance and reaction time.

Perceived SA and perceived workload

The mean values and the SD's for perceived SA and perceived workload are presented in table 1. The data met the assumption of normality using a Shapiro-Wilks test, $p = .882$ for workload and $p = .531$ for SA. The assumption for homogeneity was met using Levene's test, $p = .298$ for workload and $p = .915$ for SA. The assumption of independence was met for ANOVAs and correlations.

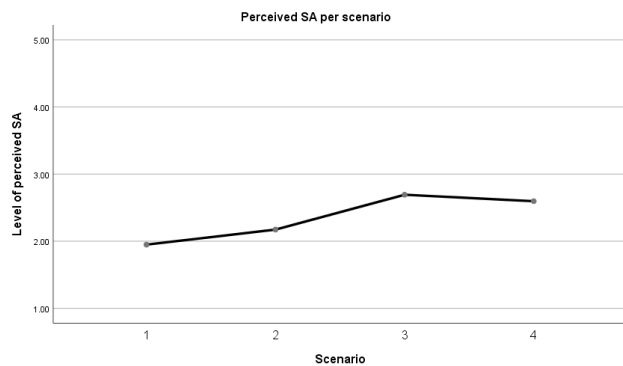
Table 1: *Mean and SDs, perceived SA and perceived workload, between experimental conditions*

	Perceived SA		Perceived workload	
	Mean	SD	Mean	SD
1 vessel	1.82	.40	1.80	.46
3 vessels	2.55	.45	2.58	.54
6 vessels	2.61	.41	2.66	.26

To see if there was a general relation between perceived SA and perceived workload, a correlation was conducted. A very strong, significant, positive correlation was found, $r = .937, p < 0.001$. This indicates that as workload increase, situational awareness decreases (a higher score means lower perceived SA).

To see if perceived SA and perceived workload differed between scenarios, a repeated-measures ANOVA was carried out. Mauchly’s W test did not indicate sphericity violation in either test, $p = .755$ for workload and $p = .543$ for SA respectively, and therefore sphericity was assumed. The difference in perceived workload was not significant $F(3, 36) = .856, p = .473$, however, for SA a significant difference was observed $F(3, 36) = 3.648, p = .021, \eta^2 = .233$. The LSD posthoc test indicated that perceived SA was significantly higher in scenario 1 compared to scenario 3, $p = .007$, and 4, $p = .01$. As presented in figure 1, this suggests that the participants perceived SA was significantly higher in the non-incident scenario (1) than in two of the incident scenarios 3 and 4.

Figure 1: Number of fixations on each display, between scenarios.



To see if there was a relation between the number of monitored vessels, the perceived SA and perceived workload, a one-way ANOVA was conducted. A significant main effect difference was observed for perceived SA, $F(2,10) = 4.480, p=.041, \eta^2 = .473$, and for perceived workload, $F(2,10) = 5.270, p = .027, \eta^2 = .513$. The posthoc test indicated a significant difference between 1 and 6 vessels for both perceived SA, $p = .048$, and perceived workload, $p = .038$. This suggests

that when a participant monitors more than 1 vessel, perceived SA decreases, and perceived workload increases.

Objective SA

The mean reaction time was the highest for scenario 3 (65.44 seconds), followed by scenario 4 (54.85 seconds) and 2 (16.13 seconds), which is presented in Table 2. The data met the assumption of normality using a Shapiro-Wilks test, $p = .338$, the assumption for homogeneity using Levene's test, $p = .843$, and the assumption of independence.

Table 2: Mean reaction time to incidents, between scenarios

	Mean	SD
Scenario 2	16.13	24.06
Scenario 3	65.44	45.32
Scenario 4	54.85	76.25

To see if the number of vessels monitored affected the reaction time of the participant, a one-way ANOVA was conducted between reaction time and experimental condition. However, there was no statistically significant differences, $F(2, 11) = .918, p = .434$. This comparison could have been affected by the missing recording mentioned previously, which in this case affected the 6-vessel condition, resulting in all conditions having 4 participants. The mean values and SD's between experimental conditions are presented in Table 3.

Table 3: Mean reaction time to incidents, between experimental conditions

	Mean	SD
1 vessel	36.94	20.89
3 vessels	24.55	28.88
6 vessels	51.64	33.65

Performance

The means and SDs for the CPA are presented in table 4. The data met the assumptions for normality with a Shapiro-Wilk test, $p = .812$, homogeneity with a Levene's test, $p = .645$, and independence. All correlations met the assumption of homoscedasticity and linearity.

Table 4: Mean and SDs for CPA, between experimental conditions and scenarios

	Scenario 2		Scenario 3		Scenario 4	
	Mean	SD	Mean	SD	Mean	SD
1 vessel	.50	.35	.11	.09	.29	.08
3 vessels	.69	.39	.03	.05	.72	.39
6 vessels	.58	.29	.02	.03	.35	.16

To see if performance was affected by the number of vessels monitored, a one-way ANCOVA was conducted with CPA and the experimental condition, with start distance in scenario 2 and 3 as covariates. No statistically significant differences were found, $F(2, 12) = 1.873, p = .220$, however, an inverted U-trend was observed. The highest CPA in the 3 vessel condition, while the 1 and 6 vessel condition had almost identical overall means.

Several correlations were performed to see how SA was related to performance, both perceived and objectively. To measure perceived performance, item 4 of the workload scale was included, as this item asked the participants to rate their own performance. This data met the assumption of normality with a Shapiro-Wilk test, $p = .925$, and independence.

To see if objective performance related to objective SA, a partial correlation was performed between CPA and reaction time, with start distance in scenario 3 as a control variable. This was done specifically for scenario 3, as reaction time had no impact on CPA in scenario 2 and 4. It revealed a significant, strong, negative correlation, $r = -.817, p = .025$, which indicates that lower reaction times was related to a higher CPA in scenario 3.

Then, a correlation was performed between perceived performance and perceived SA. A strong, significant, positive correlation was observed, $r = .614, p = .025$. Lastly, a partial correlation was performed between CPA and perceived SA, with start distance in scenario 3 as a control variable, to see if objective performance related to perceived SA. However, there was no significant correlation, $r = .332, p = .365$. This indicates that when perceiving themselves to be more SA, the participants perceived themselves to perform better, but this did not relate to performing better objectively. The correlations and partial correlations are presented in Table 5.

Table 5: *Correlations between perceived SA, reaction time, CPA and perceived performance*

	CPA	Perceived performance
Perceived SA	.332	.614*
Reaction time	-.667*	.094

* = Correlation is significant at 0.05 level. ** = Correlation is significant at 0.01 level.

Usability

After completing all four scenarios, the participants completed the SUS usability questionnaire (Brooke, 1996). The global score of all participants was 68.65 (range 52.5 - 95, SD = 11.7), which is just above the industry average of 68, although well below the industry standard of 80 (Lewis, 2018). The data had met the assumption of normality with a Shapiro-Wilks test, $p = .296$, linearity and homoscedasticity.

To see if usability was perceived differently between experimental conditions, a one-way ANOVA was performed, however, no significant differences were observed, $F(2,10) = .127$, $p = .882$.

To see if the perception of SA related to the perception of usability, a correlation was performed between perceived SA and the SUS score. A strong, significant, negative correlation was found, $r = -.566$, $p = .044$, indicating that as the perceived level of usability increases, the level of perceived SA increases as well.

Qualitative data

To investigate how to improve the usability of the system, qualitative feedback was collected during the sessions, after the sessions and during the post-interview. Following is a summary of the qualitative data.

Autonomy specific

Specific to the autonomy, two main issues were raised; the mode of automation is unclear and there is no feedback on the next action.

Currently, the only way to check the mode of automation of a vessel (autonomous/manual) is to check each vessel individually in the control window. This was mentioned by several participants to be too cumbersome, mostly by participants controlling more than one vessel. Three participants

had a vessel go off track as they forgot/did not realize they were in manual mode, two of which grounded their vessel (not related to a planned incident).

Another issue, mentioned by all participants, was “when is the vessel going to act, and what is it planning on doing?”. They had been informed that the vessel would act to keep a safety distance of 1 nautical mile to other vessels. However, the participant did not know if the autonomous vessel had seen the other vessel and therefore spent time to monitor if the vessel actually would take action. This was mentioned particularly often in scenario 2, where an autonomous vessel approaches a manned vessel head-on. In this scenario, only two participants waited long enough to let the autonomous vessel act, while the rest took manual control earlier.

Prompts

In addition to the vessel status display with colour-coded notifications, the participants also wanted other types of prompts. A close-distance prompt was suggested by 4 participants. They imagined this to be a prompt integrated into the vessel status display which gave a prompt if another vessel came too close to one of the autonomous vessels. The specifics of this they did not agree on, however. One participant (monitoring 6 vessels) suggesting a prompt for any vessel with a CPA of less than 3 nm, while another (monitoring 1 vessel) suggested a safety distance of 0.25 nm.

Participant 7 (monitoring 6 vessels) suggested a notification if an operator had not monitored a specific autonomous vessel for a specific amount of time. During the experiments, it was noticeable that all participants who monitored more than 1 vessel, for some period of the experiment focused too much on a particular vessel. The consequence was that many vessels were left without supervision leading to many groundings/collisions, particularly in scenario 3.

Lastly, a suggestion was made for a prompt appearing if a vessel goes heavily off course, without being under the active supervision of an operator. The idea was that this would make errors in the vessels track-following capacities apparent at an early stage.

Display marking

Only some participants suggested enhancing the marking of the displays, however, several participants seemed to be affected by this, particularly for the RADAR and the conning. Currently, only the control window is labelled with which autonomous vessels is currently selected. They

Enhancing operator situational awareness through enhanced design of Shore Control Centre interface

would like this information to also be available in the RADAR, conning and vessel status display, as well as more salient marking in the controlled display.

Physical changes

Three participants would like an increased size of the ECDIS, and one would like an increased size of RADAR display. One of the participants suggested keeping the ECDIS the way it is but having a large separate ECDIS display. This, they said, would enable the operator to have some overview of all vessels, but still, focus on one particular vessel of interest.

Furthermore, four participants mentioned the placement of the vessel status display, which was placed in the bottom-right corner of the monitor on the left. When the vessel status display turned red in scenario 4, most participants reacted late, as they did not notice the change. According to them, it was out of their line of sight, and they only noticed something was wrong by checking the RADAR or the ECDIS. This related well with their eye-tracking recordings.

Functionality

Six participants, distributed among all experimental conditions, asked to be able to control the speed of an autonomous vessel, without taking it off route mode. This is currently not available. Such a semi-autonomous feature allows them to control the speed, while not having to spend so much time monitoring the heading of a vessel, they said. On two occasions, participants used manual control to reduce the speed of a vessel, left the vessel momentarily to monitor other vessels, and when they came back, the manually controlled vessel had grounded. Both occurred in scenario 1 (non-incident), in the 6-vessel condition.

Currently, when selecting a specific vessel through the vessel status display, only the RADAR and the conning changes, while the controlled display did not. This inconsistency caused confusion, as some participants forgot to change the controlled display, and ended up taking control of the wrong vessel, causing at least one grounding. This was mentioned as a problem by 7 participants, almost all of which monitored more than 1 vessel. As one of the participants said: “One-click should change all of them, so you don’t forget.”

Another frequent mention was CPA information. While only four participants mentioned it explicitly, all participants seemed to find it cumbersome to use, as accessing the CPA information

required a total of 8 clicks. As participant 5 said, “You want this to be available with a single click”. One participant “solved” this issue by covering his ECDIS-screen with multiple CPA screens, but he was only monitoring 1 vessel, and even then, it took up quite a lot of his screen.

Some participants would also like a list of nearby vessels with basic information like the vessels name, speed, heading, ROT and destination. The information is available in the current interface but requires the operator to click manually on each vessel he would like information from. The information was suggested either to be shown on the ECDIS itself, or on the RADAR with ARPA.

When asked in the post-interview if they would like a live visual feed, 6 participants said yes, 5 participants said maybe, and 3 participants said no. The yes-side mentioned benefits like double-checking their displays the same way they use windows in manned vessels. The maybe-side said that it is only useful when going into port or looking at a vessel close up. The no-side thought it would just add additional workload to check another display, without giving much-added value.

When asked in the post-interview if any of the information they were provided with was not useful, 8 said no. Of the remaining five, 2 found the latitude/longitude presented with the CPA information not useful, 1 said the conning on the control window, 1 mentioned the conning in general and 1 mentioned the vessel status display.

ECDIS and RADAR

There were quite a lot of suggestions specifically for the ECDIS and the RADAR, some quite detailed. For the ECDIS, four participants said it was too hard to distinguish one autonomous vessel and its route from the other vessels and routes. They suggest solving this by colouring each autonomous vessel and corresponding route in a different colour than the other autonomous vessels. One participant also suggested representing the vessels by triangles, so that the operator would instantly know which direction they were going. Further, seven participants would like clear marking on the ECDIS which vessel was currently selected. The participants would like the currently selected vessel to either increase in size (participant 2 and 7), centring the ECDIS on this vessel (participant 3 and 10) or have the label constantly showing (participant 7). As for more detail-level suggestions, two participants would like past waypoints/route to disappear to have a cleaner ECDIS, and one suggested presenting the scale of the ECDIS.

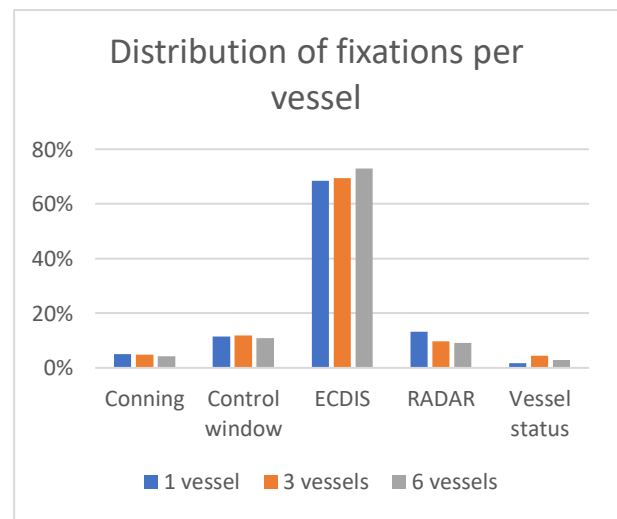
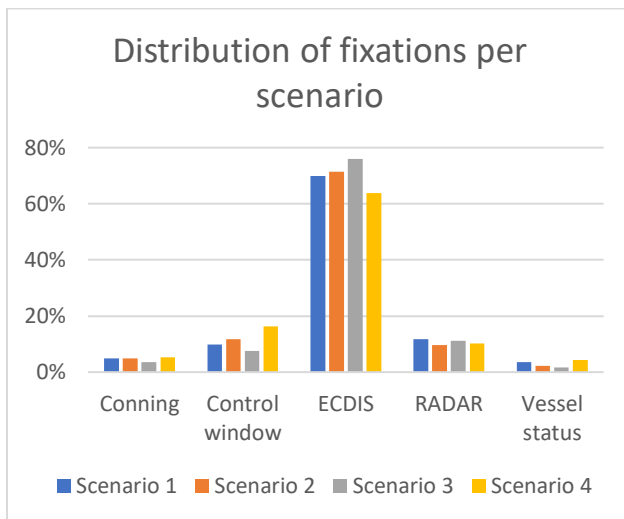
Concerning the RADAR, participants mainly proposed quite detailed changes and would like the following features to be included: CPA, targets, ARPA, bow-crossing time, bearing points and parallel lines. One participant also mentioned that setting a range for the RADAR for specific vessels would be helpful, as this varies greatly for vessels in open sea, confined waters and port.

Six participants would like either partial or full RADAR integration on ECDIS. This relates well with the post-interview, where participants were asked about which information was particularly useful in the interface, and 11 mentioned a RADAR or an ECDIS feature. 7 mentioned a RADAR features (vectors, bearings/range or the whole RADAR) and 8 answered an ECDIS feature (CPA, predictions, AIS, waypoints, measuring lines). Therefore, integrating the two most perceived most useful features would make sense.

Due to issues with recording lengths, which affected the total amount of fixations, statistical tests could not be carried out to check for significant differences. However, descriptive statistics for the distribution of fixations on displays indicate that ECDIS was fixated on the majority of the time, with a mean of 70% between experimental conditions and between scenarios. Any clear difference between the RADAR and the other displays is not visible in either, however.

Figure 2: *Fixations between scenarios.*

Figure 3: *Fixations between number of vessels.*



Feasible number of vessels to monitor

When asked explicitly in the post-interview to name a feasible number of autonomous vessels an SCC operator could monitor and control safely, the answers varied greatly, ranging from to 2 to more than 10. The modal number was 4 but only mentioned by 3 participants.

Discussion

The main research question for this study was: *What information should be presented to an SCC operator in an optimal SCC interface, and how should this information be presented, so that the interface is usable and supports the creation of SA, ensuring the operator's ability to safely monitor and control an autonomous vessel, in this specific operational context?*

This was split into three sub-research questions, which will be discussed separately.

- **R2:** *Conceptualising SA as perceived SA (measured with self-report questionnaires) and objective SA (measured with eye-tracking), to what degree are the participants able to become situationally aware with the current interface?*

The hypothesis was that the operators would experience low levels of perceived SA and low levels of objective SA, as the system would not support SA sufficiently.

Perceived SA was found to increase as the perceived workload increased, in accordance with previous findings (Endsley, 1999), thus serving as basic form of validation that the selected SART-inspired items did, in fact, measure a construct similar to perceived SA. Perceived SA also related to perceived performance but did not correlate with the objective performance measure (CPA). This is in accordance with previous studies, which has found perceived SA to relate with perceived performance, but not their objective SA (Xu et al., 2018). Instead, perceived SA seems to be a separate construct (Salmon et al., 2009). While this construct should not be underestimated, as a system without perceived SA would make an operator feel uncomfortable, it does not necessarily indicate objective SA. A correlation was however found between reaction time and CPA, indicating that the objective SA was associated with the objective performance. In other words, while the SA participants perceived to have had little to with their objective performance, their objective SA did relate to their objective performance.

However, when comparing perceived SA between scenarios, a significant difference was found, with perceived SA being significantly higher in the non-incident scenario 1 than in two of the incident-scenarios, 3 and 4. This pattern follows the general, though non-significant pattern, of the objective SA. In scenario 2, the mean reaction time was low, which was likely due to the potential incident being observed by participants right from the start of the scenario, as well as being the slowest incident to unfold. Scenario 3, which required the fastest intervention to avoid

incident had the highest mean reaction time, explaining why most participants experienced groundings/collisions in this scenario. This scenario tested if participants would notice and manage to take action with an incident that happened suddenly due to failure in the automation. As previously cited, MacKinnon wrote: “*SCC operators must be able to quickly identify unexpected threats and errors quickly and efficiently*”. However, while that is the goal, the result of this study is in accordance with previous studies (Pazouki et al., 2018), in which only 1 out of 12 participants recognised a failure in automation without receiving specific training. Pazouki et al. (2018) believe the error is due to ineffective monitoring and over-reliance. This over-reliance likely causes an OOTL situation, which reduces SA, which in turn delays and reduces the quality of performance of the operator (Endsley, 2017). Scenario 4 was used to test the alarm functionality in the vessel status display, by turning on a red light to indicate technical failure. The idea of the alarm was to inform the operator straight away that something is wrong. However, the high mean reaction time indicates that this was not successful. The alarm was made less salient on purpose to avoid automation bias, however, this result indicates that the alarm was not salient enough to support enhanced SA of the situation. Further, the high variability indicates that the system is not reliable enough between participants.

Thus, the participants perceived themselves to have a SA which match their self-perceived performance, but this did not relate to their objective performance. Between scenarios, the participants perceived their SA to be lower during the non-incident scenario than in incident scenarios 3 and 4. Similarly, for objective SA, the reaction time in scenario 3 was generally so high that the simulated incident could not be avoided, and so high in scenario 4 that the alarm functionality in the vessel status display did not support enhanced SA. Therefore, during the incidents nor the participants perceived SA nor their objective SA was sufficiently supported by the current interface, and thus adequate SA was not achieved.

R3: *Considering their navigational performance and which feedback they gave concerning informational requirements; which parts of the interface need to be changed to enhance situational awareness and performance?*

The hypothesis was that since the first version of the system was not developed with a user-centred approach, most parts of the interface would have to be changed to enhance SA and performance.

The global SUS score seems to support the potential for improvement. The system achieved a score of 68.65, which is just above the industry average but well below the industry standard. Further, the SUS score strongly correlated with perceived SA. This indicates a potential for improving perceived SA by improving perceived usability. That would also be in line with the authors which believe that problems concerning SA could be solved through better interface design (see Parasuraman & Manzey, 2010; Endsley, 2017; Pazouki et al. 2018; Man et al., 2018).

To investigate further the specific parts of the interface which should be changed, qualitative data was collected from the participants, in line with a human-centred design approach (Billings, 1997).

Related to the autonomy itself, the participants mentioned two main issues; unclear mode of automation and lack of feedback of the next action. Clear mode of automation and high system transparency are both features regarded as highly important in an interface for an autonomous system according to Endsley (2017). Specifically, for an SCC interface, mode of automation could be highlighted in a larger, salient position on the vessel status display, and feedback on next action could be given as simple text information about next planned action(s).

For prompts, the current system only has a simple prompt-system that works by giving a colour-coded message to indicate internal failure. This was somewhat due to limitations on the complexity of the simulator system which was used, but also to not make the operators dependent on these prompts, as this could increase automation bias. However, several of the prompts requested by the participants seems reasonable. A close-distance prompt would likely increase reaction time in incidents like scenario 3, thus increasing the creation of rapid SA, and a prompt for a vessel going heavily off course would be a good fail-safe for track-mode failures. One participant (monitoring 6 vessels) suggested being prompted if a vessel had not been monitored for a specific amount of time. In general, building an interface to support the tasks of the operator is positive, and could reduce workload, as well as help the operator stay aware of the possible future incidents (SA level 3). However, all of these must constantly be evaluated with the trade-off of increased possibility for automation bias (i.e. if the prompt does not work, the operator might perform poorly without it). This forms part of the automation conundrum, with more robust automation causing more complacency (Endsley, 2017).

Several of the suggestions made and problems faced by the participant related to the same issue; a confusion about which ship they were operating. Inspired by classic automation term “mode

confusion”, used to describe confusion about mode of automation, it seems appropriate to name this “ship confusion”.

Regarding functionality, the internal inconsistency when switching screens (RADAR and ECDIS switching with vessel status display, but conning did not), was an issue most participants experienced. This was an example of where poor usability leads to lowered situational awareness comprehension (level 2), as this lack of comprehension led to ship confusion on several occasions. Another usability issue was the lack of saliency for CPA information. This feature frequently used by all participants and cited by many participants as the most important piece of information in the interface. This feature was important for participants to gain awareness of the relative positions of the traffic vessels to the autonomous vessel, thus supporting SA level 2 (comprehension) and level 3 (projection). Instead of the 8 clicks which are currently required to present this information, it should have a salient place in the SCC interface, for instance forming part of the conning display or an integrated ECDIS/RADAR. The participants did not agree about whether to include a visual feed. It is a technical issue, as it is a lot of data to transmit without delay, but it is also a psychological issue, as it could increase operator awareness, but they could also become dependent and become very vulnerable to technical failure. This trade-off should be subject to further research to see if the added value of a visual feed outweighs the risks.

For the ECDIS, one of the suggestions was to colour waypoints to make it clearer which belonged to which vessel. Such a feature was found to be the most useful added feature in a study on UAV interfaces (Fuchs, 2014). One participant also suggested representing the vessels with triangles, to clearly indicate which way vessels are going. Both these suggestions seem likely to increase SA, as it would facilitate quick recognition of the situation (level 1 SA), which in turn should support comprehension and projection of the situation (SA level 2 and 3).

When asked about the most important information in the current interface, the ECDIS and/or RADAR was mentioned by almost all participants. However, looking at the descriptive statistics for share of fixations, the ECDIS was indeed fixated on heavily, but the RADAR did not seem more fixated than any other display. Since the RADAR was subjectively regarded as important, but not highly fixated on, it could indicate that the RADAR was not saliently enough placed, which was also mentioned by a few participants. The RADAR was placed on the left monitor, meaning the participant had to turn their head to see it. The other display on this monitor, the vessel status

display, was the least fixated display of all, even in scenario 4 which tested this display specifically. Therefore, one could argue that the RADAR was, in fact, fixated on quite frequently considering the placement.

Thus, most parts of the interface need to be changed to increase usability and thus enhance SA and performance. The vessel status display and RADAR need to be placed more saliently, and the following features need to be integrated into the current interface: clear mode of automation, feedback on next action, additional prompts and more consistent screen functionality. All the suggestions relate well to Endsley's (1995a) SA framework and to general HCI guidelines.

- *R4: How does the number of vessels to monitor and control affect the operator's performance, workload and situational awareness?*

The hypothesis was that a higher number of vessels would increase the cognitive workload, thus decreasing situational awareness and performance, relative to a lower number of vessels.

Both perceived SA and the perceived workload were found to be significantly different depending on the number of vessels an operator was responsible for. Both cases followed the same pattern. Once a participant had to monitor more than 1 vessel, SA decreased and workload increased, while there were no significant differences between monitoring 3 or 6 vessels.

On one hand, this could indicate that monitoring more than one vessel simply is not feasible, as it increases perceived workload and decreases perceived SA. However, these results can also be interpreted to indicate that this is not feasible within the current system. In the latter case, performance could be increased by improving the usability of the system. That would align with comments made by some participants, which mentioned that they thought they could handle more vessels in an improved system.

One potential issue with monitoring only one ship is boredom. Boredom is a danger to SA, with higher levels of boredom found to relate to slower reaction times (Thompson et al., 2006). Although not formally measured, the participants in the 1 vessel condition showed several signs of boredom which other participants did not, like tapping their fingers, whistling and taking phone calls. In this study, sessions were kept to a maximum of 30 minutes, which is the limit of a human's ability to monitor monotonous stimuli according to Mackleworth (1950). Therefore, one could hypothesise that in a real-life setting, for the duration of a full working day, one vessel would

become difficult to monitor simply because it is not stimulating enough to sustain the attention of the operator. This would align with comments made by participants, as four participants said they did not think they could do this for more than one hour and another participant said he could not do it for a full working day. This is not an issue on traditional manned vessels, as the operator in this context is actively steering the vessel, while in an SCC setting the participant is reduced to a passive observer, which should only intervene when necessary. As Endsley (2017) notes, it is inherently challenging to make an operator stay cognitively engaged in a task when they are not actively participating in the task. This could lead to cognitive underload, which as mentioned could be a bigger concern than cognitive overload, as it also decreases performance, but it is harder to detect (Young & Stanton, 2002). When asked explicitly about how many vessels one operator could safely monitor, the answers were very varied, but all were above one. In general, participants in the 1 vessel condition indicated a number between 2 and 4, participants with 3 vessels a number between 3 and 4, and participants with 6 vessels generally indicated lower numbers than 6. While all participants seem to believe they could monitor more than 1 vessel, they disagree on how many, indicating a general tendency for this study – an interface which does not support participants reliably, as many measures are riddled with outliers.

Thus, the results indicate that perceived SA and perceived workload are better when monitoring one vessel. However, considering factors like boredom, cognitive underload and the subjective opinion of the participants, 1 vessel does not seem ideal either. Therefore, while the results meet the hypothesis, future research should be conducted to investigate if more than 1 vessel is feasible in an improved SCC interface.

Significance, limitations and conclusion

The findings in this study both have large significance for the research of SCCs, both theoretically and practically. Theoretically, this research supports that general theory about SA and automation are applicable also for SCCs. This means the maritime field can borrow knowledge from other fields, and thus do not need to “re-invent the wheel”. Furthermore, this study provides a blueprint for which psychological dimensions to include when researching SCC interfaces in an evidence-based manner. Further, it also comments on the limitations of some measures, like using perceived SA to measure objective SA, and the challenges using eye-tracking to measure actual SA. Practically, this study provides specific recommendations, created with a user-centred approach,

for how to create an SCC interface that should support higher usability, and thus higher SA and performance than what the current system offers. These can be applied directly in any SCC interface and should be researched further.

However, there are several limitations to this study. Although the sample is large compared to other SCC studies (Man et al., 2015; Man et al., 2018), the groups within the sample were small (range, 2-3), meaning that between-groups comparisons were not feasible. While participants in the prior/current mariner group were relatively easy to find, students and VTS personnel were not. However, this was positive for the quality of the qualitative feedback and the exploratory nature of this study, as it was given from a broader range of perspectives. Further, since objective SA could not be calculated, and perceived SA does not relate to objective performance measures, it reduces the study's ability to comment on the level of objective SA in the participants. As for eye-tracking, differences between fixations on different displays could not be statistically analysed calculated due to too high variability in eye-tracking recording length. Also, analysing eye-tracking fixations within the ECDIS would also have given a more complex picture of participant behaviour but was not possible to analyse with the utilised eye-trackers and the eye-tracking software. Further, there was no inter-experimenter variability since once experimenter did all observations and analysis of the qualitative data. However, this also means that the presented summary of this data could have been subject to selection and phrasing bias of that experimenter. As for the simulation, although routes were inspired by real maritime traffic, the traffic density was judged to be too light by most participants, possibly reducing workload and making it easier to become SA. Lastly, while the participants did express increased trust in the *interface* as they completed more scenarios, most of them reported a continuous doubt in the autopilot as well as for our simulated collision-avoidance automation. Therefore, it is possible that they did not trust the system fully and became less complacent than they would have been with better automation, possibly causing higher SA and higher performance (Hogg and Gosh, 2016).

Conceptually, the main issue is the lack of a unified performance measure. In other words, what constitutes good performance for an SCC operator? For traditional sailing, measures like CPA, rudder and telegraph control is normally combined with a subjective evaluation to create a performance measure. Therefore, CPA was included, but manual sailing is only part of an operator's performance. The decision-making process about when to (not) intervene seems equally

important. Some participants almost tried to remote control all vessels, which is not the intention. This should be subject to a discussion in the maritime community, new IMO regulations and further research. This also relates to the second main question; how should an SCC operator be trained? For instance, should they be taught a specific monitoring pattern? With a standard training established, performance could be measured by how well the operator follows their training.

Future research should empirically test the recommendations made in this study and verify if they do increase SA, performance and optimize workload. Once a sufficiently reliable system is created, this opens the possibility to look at an almost endless array of more specific research questions. How are these same variables affected by traffic density and weather? Further, when testing whole SCC teams, how can team SA be supported? Can vessels be transferred from one operator to another in case of an emergency? How much redundancy is necessary yet still cost-effective?

To conclude, with the current system participants perceive their SA to match their perceived performance. However, the reaction times indicate that the system did not support sufficient objective SA, and the perceived SA decreased in incident scenarios. Separating by the number of vessels, the participants perceived themselves to be more SA and to have lower workload monitoring just one vessel. While this could indicate that it is only feasible to monitor this vessel, there is a risk of boredom and cognitive underload. Therefore, this study recommends future systems to be created for monitoring more than 1 vessel. However, for this to be feasible, the interface must support higher SA. This can be achieved through increasing the usability with the SA-supporting features such as: more salient placement of important features, a clearer mode of automation, feedback on next action, additional prompts and more consistent screen functionality. By employing these user-centred recommendations, which relates well to existing literature, a more SA-supportive interface can be created, thus increasing an operator's ability to safely monitor and control autonomous vessels.

References

- Arrabito, G., Ho, G., Lambert, A., Rutley, M., Keillor, J., Chiu, A., . . . Hou, M. (2010). *Human Factors Issues for Controlling Uninhabited Aerial Vehicles*(Vol. 043, pp. 1-87, Rep.). Toronto: Defence R&D Canada.
- Billings, C. E. (1997). *Aviation Automation: The Search for a Human-Centered Approach* (Mahwah, NJ: Lawrence Erlbaum Assoc).
- Brainbridge, L. (1983). Ironies of automation. *Automatica* 19, 775-780
- Brooke, J. (1996). SUS - A quick and dirty usability scale. Retrieved from <https://hell.meiert.org/core/pdf/sus.pdf>
- Cordon, J. R., Mestre, J. M., & Walliser, J. (2017). Human factors in seafaring: The role of situation awareness. *Safety Science*,93, 256-265.
- Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*,13(3), 319.
- Endsley, M. R & Sollenberger, R & Stein, E. (2000). Situation awareness: A comparison of measures. *Proceedings of the Human Performance, Situation Awareness and Automation: User-Centered Design for the New Millennium*, Savannah, GA.
- Endsley, M. R. (1995a). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*,37(1), 32-64.
- Endsley, M. R. (2001). Designing for situation awareness in complex systems. *Proceedings of the Second international workshop on symbiosis of humans, artifacts and environment*, Kyoto, Japan.
- Endsley, M. R. (2017). From Here to Autonomy. *Human Factors: The Journal of the Human Factors and Ergonomics Society*,59(1), 5-27.
- Endsley, M. R., (1995b). Measurement of situation awareness in dynamic systems. *Human Factors* 37, 65–84
- Galbraith, J. R. (1974). Organization design: An information processing view. *Interfaces*, 4(3), 28–36.

- Geiselman, E. E., Johnson, C. M., & Buck, D. R. (2013). Flight Deck Automation. *Ergonomics in Design: The Quarterly of Human Factors Applications*, 21(3), 22-26.
- Glaser, Barney G. and Strauss, Anselm L. (1967) The discovery of grounded theory: strategies for qualitative research. Chicago.: Aldine.
- Hareide, O. S., & Ostnes, R. (2017). Maritime Usability Study by Analysing Eye-Tracking Data. *The Journal of Navigation*, 70, 927-943.
- Hareide, O. S., Ostnes, R., & Mjelde, F. V. (2016). Understanding the Eye of the Navigator.
- Hart, S. G. (2006). Nasa-task load index (NASA-TLX); 20 years later. *PsycEXTRA Dataset*.
- Hart, S. G., & Staveland, L. E. (n.d.). NASA Task Load Index. Retrieved from <https://humansystems.arc.nasa.gov/groups/tlx/downloads/TLXScale.pdf>
- Hogg, T., & Ghosh, S. (2016). Autonomous merchant vessels: examination of factors that impact the effective implementation of unmanned ships. *Australian Journal of Maritime & Ocean Affairs*, 8(3), 206-222.
- Johannsdottir, K. R., & Herdman, C. M. (2010). The Role of Working Memory in Supporting Drivers' Situation Awareness for Surrounding Traffic. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 52(6), 663-673. doi:10.1177/0018720810385427
- Jones, S. L., & Kelly, R. (2018). Dealing With Information Overload in Multifaceted Personal Informatics Systems. *Human-Computer Interaction*, 33(1), 1-48.
- Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science*, 5(2), 113-153.
- Kretschmann, L., Burmeister, H., & Jahn, C. (2017). Analyzing the economic benefit of unmanned autonomous ships: An exploratory cost-comparison between an autonomous and a conventional bulk carrier. *Research in Transportation Business & Management*, 25, 76-86.
- Lewis, J. R. (2018). Item Benchmarks for the System Usability Scale. *Journal of Usability Studies*, 13(3), 158-167.

- MacKinnon, S. N., Man, Y., & Baldauf, M. (2015). *D8.8: Final Report: Shore Control Centre*(Rep.). MUNIN.
- Mackworth, N. H. (1951). Researches on the Measurement of Human Performance. *Southern Medical Journal*,44(11), 1087.
- Man, Y., Lundh, M., Porathe, T., & Mackinnon, S. (2015). From Desk to Field - Human Factor Issues in Remote Monitoring and Controlling of Autonomous Unmanned Vessels. *Procedia Manufacturing*,3, 2674-2681. doi:10.1016/j.promfg.2015.07.635
- Man, Y., Weber, R., Cimbritz, J., Lundh, M., & Mackinnon, S. N. (2018). Human factor issues during remote ship monitoring tasks: An ecological lesson for system design in a distributed context. *International Journal of Industrial Ergonomics*,68, 231-244.
- Marangunić, N., & Granić, A. (2014). Technology acceptance model: A literature review from 1986 to 2013. *Universal Access in the Information Society*,14(1), 81-95.
- May, M. (1999). *Cognitive aspects of interface design and human-centered automation on the ship bridge: The example of ARPA/ECDIS integration*(pp. 394-399, Rep. No. 463). Bath: L. Proceedings of An International Conference on Human Interfaces in Control Rooms, Cockpits and Command Centers.
- May, P., Molloy, R., & Parasuraman, R. (1993). *Effects of automation reliability and failure rate on monitoring performance in a multitask environment*. Paper presented at the Annual Meeting of the Human Factors Society, Seattle, WA
- Merwe, K. V., Dijk, H. V., & Zon, R. (2012). Eye Movements as an Indicator of Situation Awareness in a Flight Simulator Experiment. *The International Journal of Aviation Psychology*,22(1), 78-95.
- Mooney, T. (2015). Study finds autonomous ships unlikely within next decade. Retrieved June 10, 2019, from https://www.joc.com/maritime-news/ships-shipbuilding/autonomous-ships-unlikely-next-decade_20151201.html?destination=node/3257076
- Moore, K., & Gugerty, L. (2010). Development of a novel measure of situation awareness: The case for eye movement analysis. *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting*,1650-1654.

- Oakley, B., Mouloua, M., & Hancock, P. (2003). Effects of Automation Reliability on Human Monitoring Performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*,47(1), 188-190. doi:10.1177/154193120304700139
- Ottesen, A. E. (2014). *Situation Awareness in Remote Operation of Autonomous Ships*(pp. 1-12, Rep.). Trondheim: Norwegian University of Science and Technology.
- Parasuraman, R., & Manzey, D. H. (2010). Complacency and Bias in Human Use of Automation: An Attentional Integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society*,52(3), 381-410. doi:10.1177/0018720810376055
- Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors*, 39(2), 230–253.
- Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance Consequences of Automation-Induced Complacency. *The International Journal of Aviation Psychology*,3(1), 1-23. doi:10.1207/s15327108ijap0301_1
- Pazouki, K., Forbes, N., Norman, R. A., & Woodward, M. D. (2018). Investigation on the impact of human-automation interaction in maritime operations. *Ocean Engineering*,153, 297-304.
- Porathe, T., Prison, J., & Man, Y. (2014). Situation awareness in remote control centres for unmanned ships. *Human Factors in Ship Design & Operation*,1-9.
- Prison, J., J. Dahlman, & M. Lundh. (2013). *Ship Sense—Striving for Harmony in Ship Manoeuvring*. WMU Journal of Maritime Affairs 12: 115–127.
- Ramos, M. A., Utne, I. B., & Mosleh, A. (2019). Collision avoidance on maritime autonomous surface ships: Operators' tasks and human failure events. *Safety Science*,116, 33-44.
- Riley, V. (1994). *A theory of operator reliance on automation*. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Recent research and trends* (pp. 8-14). Hillsdale, NJ: Erlbaum.
- Rødseth, Ø, & Burmeister, H. (2015). *D10.2: New ship designs for autonomous vessels* (pp. 1-36, Rep.). MUNIN.
- Rødseth, Ø. (2017). Førerløse skip kan erstatte mange vogntog. Retrieved July 8, 2019, from <https://www.sintef.no/siste-nytt/forerlose-skip-kan-erstatte-mange->

[vogn tog/?fbclid=IwAR21dGmUrzihDrQ_93EeIjsUdwVWu4DqoodwUF9FecW0unwoskK05fPMRzc](https://www.sintef.no/siste-nytt/avskaff-fergekoene-ikke-fergene/?fbclid=IwAR21dGmUrzihDrQ_93EeIjsUdwVWu4DqoodwUF9FecW0unwoskK05fPMRzc)

- Rødseth, Ø. (2018). Avskaff fergekøene, ikke fergene. Retrieved July 8, 2019, from <https://www.sintef.no/siste-nytt/avskaff-fergekoene-ikke-fergene/?fbclid=IwAR1Z8zwMWQkCEynISIVe3X42eOzVjQco5X5e2U6o1SQO23GxVQhH2B9V-Hc>
- Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D., Ladva, D., Rafferty, L., & Young, M. (2009). Measuring Situation Awareness in complex systems: Comparison of measures study. *International Journal of Industrial Ergonomics*, 39(3), 490-500. doi:10.1016/j.ergon.2008.10.010
- Sauer, J., Wastell, D. G., Hockey, G. R., Crawshaw, C. M., Ishak, M., & Downing, J. C. (2002). Effects of display design on performance in a simulated ship navigation environment. *Ergonomics*, 45(5), 329-347.
- Selcon, S.J. & Taylor, R.M. (1989). Evaluation of the Situational Awareness Rating Technique (SART) as a tool for aircrew systems design. *Proceedings of the AGARD AMP Symposium on Situational Awareness in Aerospace Operations, CP478. Seuilly-sur Seine, France: NATO AGARD.*
- Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies*, 51(5), 991-1006.
- Thompson, W., Lopez, N., Hickey, P., DaLuz, C., Caldwell, J. L., & Tvaryanas, A. P. (2006). *Effects of shift work and sustained operations: Operator performance in remotely piloted aircraft (OP-REPAIR)* (Report No. HSW-PE-BR-TR-2006-0001). Brooks CityBase, TX: United States Air Force, 311th Human Systems Wing.
- Urbaniak, G. C., & Plous, S. (2013). Research Randomizer (Version 4.0) [Computer software]. Retrieved on June 22, 2013, from <http://www.randomizer.org/>
- Westrenen, F. V., & Praetorius, G. (2012). Situation awareness and maritime traffic: Having awareness or being in control? *Theoretical Issues in Ergonomics Science*, 15(2), 161-180.
- Wiener, E. L., & Curry, R. E. (1980). Flight-deck automation: Promises and problems. *Ergonomics*, 23(10), 995-1011.

- Woods, D., & Cook, I. (2006). *Incidents: Markers of resilience or brittleness?* In E. Hollnagel, D. Woods, & N. Leveson (Eds.), *Resilience engineering: Concepts and precepts* (pp. 69–76).
- Wróbel, K., Montewka, J., & Kujala, P. (2017). Towards the assessment of potential impact of unmanned vessels on maritime transportation safety. *Reliability Engineering & System Safety*, 165, 155-169.
- Xu, J., Anders, S., Pruttianan, A., France, D., Lau, N., Adams, J. A., & Weinger, M. B. (2018). Human performance measures for the evaluation of process control human-system interfaces in high-fidelity simulations. *Applied Ergonomics*, 73, 151-165.
- Young, M. S., & Stanton, N. A. (2002). Malleable Attentional Resources Theory: A New Explanation for the Effects of Mental Underload on Performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 44(3), 365-375.

Appendices:

Table of contents

Appendices:.....	36
Appendix A – Briefing.....	37
Appendix B – Informed consent form.....	38
Appendix C – Demographics interview questions.....	39
Appendix D – Attitudes to automation questionnaire.....	40
Appendix E – Post-scenario questionnaire.....	40
Appendix F – SUS questionnaire.....	42
Appendix G – Post-interview questionnaire.....	43
Appendix H – Debriefing.....	44
Appendix I – Tasks.....	44
Appendix J – Screenshots marinetraffic.com.....	46
Appendix K – Full interface setup.....	46

Appendix A – Briefing

Shore Control Centre study briefing

Dear participant,

Thank you for deciding to participate in our study. To allow you to prepare for the tasks you will perform during the experiment, this briefing will give you some basic information about Shore Control Centres and your task as a participant. On the day of the experiment you will be briefed more in-depth and you will also have the opportunity to ask questions if something is unclear.

What is a Shore Control Centre?

A Shore Control Centre is where autonomous vessels will be monitored and/or controlled in the future by a Shore Control Operator. Ideally, an autonomous vessel should just control itself and sail independently, however, it is difficult to perfect the automation. Therefore, the ship could find itself in a situation it does not know how to solve and will need help from a human operator. In the future, these human operators will likely work in cooperation with other operators in a Shore Control Centre, but in this experiment we are only looking at one operator.

Experimental context

You will be tasked to act as the human operator in a Shore Control Centre. The autonomous vessels you are responsible for are sailing in a traffic situation with manned vessels, there are no other autonomous vessels besides the ones your supervising.

You will be performing your task as an operator working for fictional SCC international. They work with shipping within Europe, with bases in the Netherlands and in Spain. The autonomous ships you will work with will be a medium-sized shipping vessel.

The autonomous vessels will follow pre-determined routes. They have been set by the company, as they are regarded to be the most economical and efficient. The company policy is to stick to this route as much as possible.

Your task as a Shore Control operator will be to continually monitor and control the autonomous vessels and to make sure that they get to their destination safely.

To do so, you must follow the Shore Control operator decisions process. This consist of four main phases:

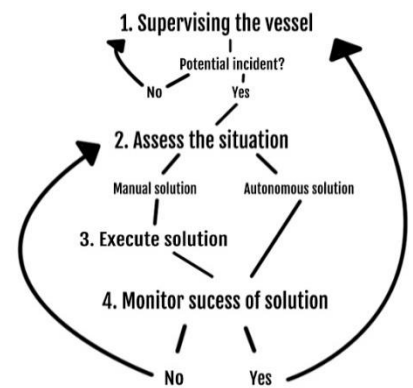
1. Supervising safety status of the ship

If a potential incident is observed:

2. Assess the situation and decide on a strategy

If deciding on a manual solution:

SCC operator decision process



3. Executing the manual solution.

After a solution (manual or autonomous) has been executed:

4. Monitoring the situation and ensuring that the solution was successful. If successful, return to the first phase, if unsuccessful, return to the second phase.

The Shore Control Centre interface

The interface is built up by 5 displays. It contains an ECDIS display, a RADAR, a conning display, a vessel-specific status display and an overview display. Some of these are probably familiar to you, and others not, but all of them will be described in further detail on the day of the experiment.

Autonomous vessels

If a traffic vessel is crossing the route of an autonomous ship, the autonomous ship is programmed to reduce speed substantially to ensure that the other ship can cross with a safe distance of margin.

Context of study within MARIN

Shore Control Centres is one of many subjects within MARIN's larger research program on autonomous shipping. The larger objective is to create a full test bed for Shore Control Centre and autonomous shipping which can be used for training and simulations. The interface you will be using is the first version of MARIN's Shore Control Centre. The objective of this specific study is to further enhance the user-friendliness of the interface, enabling the operator to monitor and control autonomous vessels in the safest possible manner.

Appendix B – Informed consent form

Titel van het onderzoek: Enhancing situational awareness through improved design of interfaces for operators in Shore Control Centres

Verantwoordelijke onderzoeker: Lars Berntzen Arholm, Gerrit van der Want

In te vullen door de deelnemer

Ik verklaar op een voor mij duidelijke manier op de hoogte te zijn van de aard, methode, doel en [indien aanwezig] de risico's en de belasting van het onderzoek.

Ik weet dat de gegevens en resultaten van het onderzoek alleen anoniem en vertrouwelijk zullen worden gebruikt en gepubliceerd. Mijn vragen zijn naar tevredenheid beantwoord.

Ik begrijp dat eye-tracking, schermopnames en video-opnames alleen worden gebruikt voor analyse en / of presentatie van resultaten.

Enhancing operator situational awareness through enhanced design of Shore Control Centre interface

Ik stem vrijwillig in met deelname aan deze studie. Hoewel ik mij het recht voorbehoud om mijn deelname aan deze studie te beëindigen zonder opgave van reden.

Naam deelnemer:

Datum: **Handtekening deelnemer:**

In te vullen door de verantwoordelijke onderzoeker

Ik heb een gesproken en geschreven uitleg gegeven van het onderzoek. Ik zal de resterende vragen over het onderzoek naar macht beantwoorden. De deelnemer zal geen nadelige gevolgen ondervinden in het geval van vroegtijdige beëindiging van deelname aan dit onderzoek.

Naam onderzoeker: Lars Berntzen Arholm

Datum: **Handtekening onderzoeker:**.....

Appendix C – Demographics interview questions

Demographical information

Please fill in the following personal information:

Age: _____ Gender: _____ Nationality: _____

What is your (maritime) occupation? _____

Please answer the following questions about your experience:

1. Have you been in a maritime simulator before? Yes No
2. Have you been in a maritime simulator at MARIN before? Yes No
3. Do you have experience using an ECDIS map? Yes No

Enhancing operator situational awareness through enhanced design of Shore Control Centre interface

4. Do you have experience using a RADAR? Yes No
5. Do you have experience remote controlling a maritime vessel? Yes No
6. Are you already familiar with the concept of Shore Control Centers? Yes No
7. How many years of experience do you have with maritime navigation? _____

Appendix D – Attitudes to automation questionnaire questions

Please answer the following questions from the perspective of your current maritime experience and occupation, and how you think autonomous vessels will affect the maritime industry in general.

1 = Strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly agree

Statement	1	2	3	4	5
1. Autonomous vessels would enable the maritime industry to accomplish tasks better					
2. Autonomous vessels would increase safety in the maritime industry					
3. It is worth sacrificing some traditional maritime jobs to achieve the potential benefits of autonomous vessels					
4. It would be easy to do long haul shipping with autonomous vessels					
5. It would be easy for the maritime industry to implement autonomous technology					
6. I would like to work in a Shore Control Centre in the future					
7. In general, autonomous vessels would have a positive influence on the maritime industry					

Appendix E – Post-scenario questionnaire

Participant number:

Scenario number:

1. How mentally demanding was this session?.

1 2 3 4 5

Very low Very high

2. How physically demanding was this session?

1 2 3 4 5

Very low Very high

3. How rushed was the pace/speed of the session?

	1	2	3	4	5	
Very low	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very high

4. How successful were you in accomplishing what you had to do?

	1	2	3	4	5	
Perfect	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Failure

5. How hard did you have to work to achieve this level of performance?

	1	2	3	4	5	
Very low	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very high

6. How insecure, discouraged, irritated, stressed and annoyed were you?

	1	2	3	4	5	
Very low	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very high

7. Were you able to concentrate on the other vessels during the incident (high) or only the incident vessel (low)?

	1	2	3	4	5	
Very high	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very low

8. How complicated was the session?

	1	2	3	4	5	
Very simple	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very complex

9. How much mental capacity did you have to spare during the incident?

“Very much” indicates that you had a lot of capacity left, and so you could easily have done more

tasks/more complicated tasks. "Very little" indicates that you had very little capacity left and so you would have struggled to do any more tasks.

	1	2	3	4	5	
Very much	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very little

10. How much time did you feel you had to solve the incident?

	1	2	3	4	5	
More than enough	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Not enough

11. How helpful was the system at supporting you to achieve the task in this scenario?

	1	2	3	4	5	
Very helpful	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very unhelpful

12. Did you find the information needed to make informed choices in this scenario?

	1	2	3	4	5	
Very informative	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very uninformative

Appendix F – SUS questionnaire

	Strongly disagree					Strongly agree
1. I think that I would like to use this system frequently Ik denk dat ik dit systeem vaak zou willen gebruiken	1	2	3	4	5	
2. I found the system unnecessarily complex Ik vond het systeem onnodig complex	1	2	3	4	5	
3. I thought the system was easy to use	1	2	3	4	5	

Enhancing operator situational awareness through enhanced design of Shore Control Centre interface

Ik denk dat het systeem makkelijk te gebruiken was

1	2	3	4	5

4. I think that I would need the support of a technical person to be able to use this system
Ik denk dat ik hulp van een technisch persoon nodig zou hebben om dit systeem te gebruiken

1	2	3	4	5

5. I found the various functions in this system were well integrated

1	2	3	4	5

Ik vond dat de verschillende functies van dit systeem goed waren geïntegreerd

6. I thought there was too much inconsistency in this system

1	2	3	4	5

Ik denk dat er teveel inconsistenties in dit systeem waren

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

1	2	3	4	5

Ik denk dat de meeste mensen vrij snel zouden leren hoe ze dit systeem kunnen gebruiken

8. I found the system very cumbersome to use
Ik vond het systeem moeizaam was om te gebruiken

1	2	3	4	5

9. I felt very confident using this system
Ik voelde me zeker in het gebruik van dit systeem

1	2	3	4	5

10. I needed to learn a lot of things before I could get going with this system
Ik moest veel dingen leren voordat ik met dit systeem aan de gang kon

1	2	3	4	5

Appendix G – Post-interview questionnaire

Post-experiment interview questions

1. How complicated did you find it to monitor the vessels?
2. How complicated did you find it to control the vessels?
3. How concentrated were you when no incident was happening?
4. How concentrated were you when an incident was happening?
5. Did you feel you had enough time to solve the incidents?
6. Did you find the scenarios realistic? Why/why not?
7. If you could add one feature to the interface, what would it be, and why?
8. What information did you find particularly useful?

9. Was any of the information provided not useful?
10. Was there any information you wish you did have, which was not provided?
11. Would live visual feed have been helpful to monitor and control the vessels?
12. How mentally demanding was the experiment?
13. How frustrated did you feel during the experiment?
14. If you could change one feature to the interface, what would it be, and why?
15. What is a feasible number of ships a SCC operator can monitor and control safely?

Appendix H – Debriefing

Debriefing

Thank you for agreeing to participate in this study! The experiment you recently undertook was done in order to test the first version of MARIN's Shore Control Center interface setup. Based on your feedback, the researchers will make recommendations for how this system should be designed, and which information it needs to have, in order to ensure safe monitoring and controlling of autonomous vessels.

The participants invited for this experiment were: students, current seafarers, Dolphin software instructors and VTS personnel.

For any discomfort you might feel as a result of having participated in this study, or if you have any further questions about the research, please contact the researchers at:

Guido Band – BAND@fsw.leidenuniv.nl - +31 71 5273998

Lars Berntzen Arholm – L.Berntzen@academy.marin.nl - +47 920 95 120

Gerrit van der Want - G.J.v.d.Want@marin.nl - +31 317 47 99 19

Kind regards,

Guido Band, Lars Berntzen Arholm and Gerrit van der Want

Leiden University Faculty of Social and Behavioral Sciences

Wassenaarseweg 52, 2333 AK Leiden

Appendix I – Tasks

In scenario 1, no incident occurred, and the participant needed only to monitor the vessel(s).

In scenario 2, an incident occurred in which an approaching traffic vessel violated COLREGS rules, passing the autonomous vessel on the wrong side, thus creating a potential collision. The operator needed to notice the other ship making an error and take over manual control to steer the

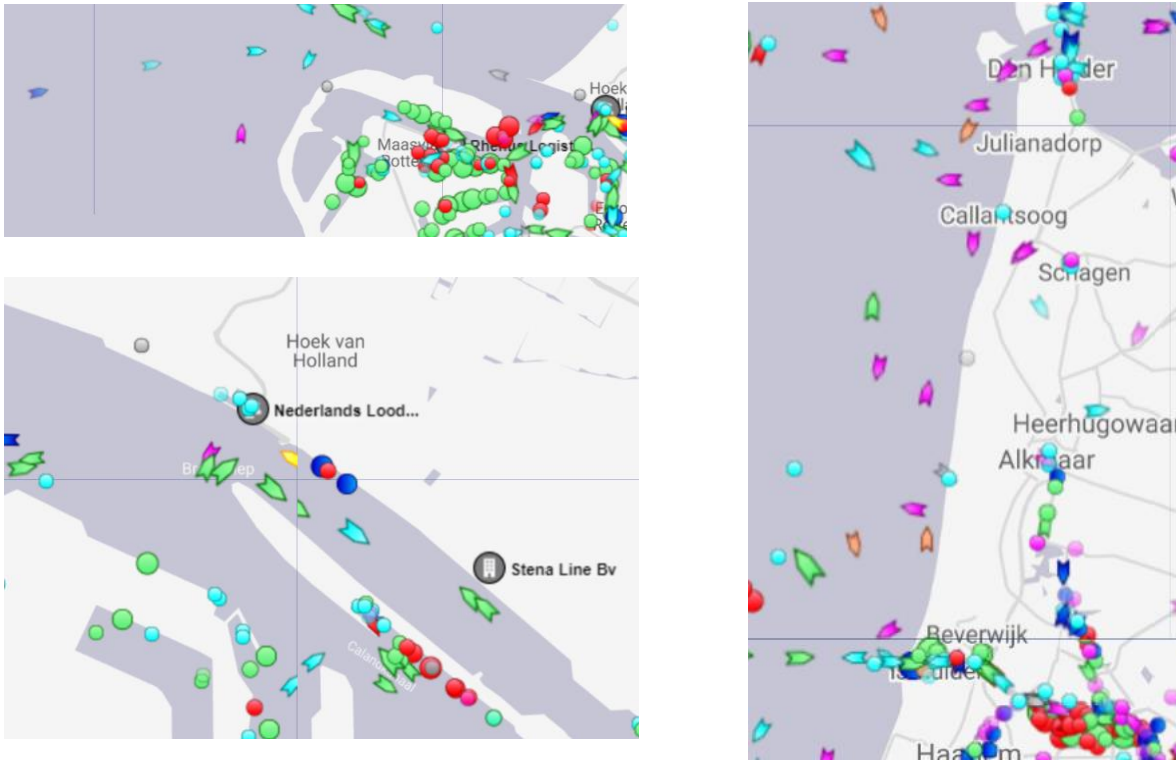
autonomous vessel to safety. 56% of collisions at sea are caused by violations of COLREGs (Ramos et al., 2019).

In scenario 3, a fishing vessel suddenly appears close to one of the autonomous vessels, as it suddenly turns on its AIS system. The operator needed to notice the vessel suddenly having appeared, take manual control and steer the ship to safety. This incident was chosen as it is a realistic occurrence, confirmed by the MARIN's internal Traffic & Safety team. Furthermore, it exposes one of the weaknesses of not having crew onboard, namely the high reliance on the ship's sensors, and what happens when they are unable to give a full picture of the traffic situation.

In scenario 4, one of the autonomous ships starts deviating of course. By checking the vessel status display, the status display for autonomous vessel 1 had changed to red. The solution to this was to call the "engineer" for him to "fix" the rudder. This incident was chosen partly because technical problem is another large source of debate around the feasibility for long haul autonomous shipping (reference), but more importantly it tests the operator's attention to the non-traditional display specific to SCC. The operator's ability to efficiently monitor these will likely be crucial in the absence of "ship sense", as this display will inform of any status changes in any of the internal systems.

Enhancing operator situational awareness through enhanced design of Shore Control Centre interface

Appendix J – Screenshots marinetraffic.com



Appendix K – Full interface setup

