Effects of change detection automation and task interruption on monitor accuracy, situation awareness and workload during monitoring.

Tijdeman, M. D.

14-08-2016

Student number: 3991164
Internal supervisor: Dr. Rick van der Kleij, TNO
External supervisor: Jelle A. van Dijk

Abstract

Dynamic positioning operators supervise an automated system that keeps a ship at a specific position. However, decreased alertness, limited attention, and especially loss of situation awareness (SA), can lead to minor or major incidents. Our research focusses on the effects of a change detection support on situation awareness, monitor performance and workload. We also investigate if it is possible for an operator to leave his monitoring task, and perform a secondary task when support is provided on the monitoring task. We conduct an experiment with 23 students. The experiment consists of a task in a Dynamic Positioning monitoring environment, and the Visual Elevator task as interruptive secondary task. All subjects complete four conditions: Control, Control with support, interruption, and interruption with support. We show that a support only improves performance when subjects are interrupted during the monitoring task and SA levels are low. When already monitoring, subjects experience more attentional demand when support is present. Furthermore, we show that with support provided after an interruption, subjects make the same number of errors as when they are not interrupted. This suggests that operators could perform other tasks in addition to monitoring, without a significant increase in error rate. This was not the case with workload and response time, however, with subjects being significantly slower after the interruption compared to the static monitoring condition due to higher workload. Hence, it would not be wise to let operators perform a secondary task with only a change detection support.

1.Introduction

Automation is on the rise, with more and more industries using it every day. The human component, however, is still present in some of these industries in the form of a human supervisor, referred to as human supervisory control (HSC). This is the process in which a human operator interacts with a computer, receiving and providing feedback to a machine controlled process (Sheridan, 1992). Research has shown that HSC is not an error-proof concept, with most of the problems associated with deficiencies in human operator states, including decreased alertness, limited attention, and especially loss of situation awareness (SA) (Kaber & Endsley, 2004 ; Parasuraman, Cosenzo & De Visser, 2009). Extensive research has been carried out to find ways to support the human supervisor or operator (Fortmann, Brauer, Müller & Boll, 2014; Gartenberg, Breslow, McCurry & Trafton 2014; Gómez, Díez, Díaz, & Aedo, 2013; John, 2013; John & Smallman, 2008; Pfaff, Klein, Drury, Moon, Liu & Entezari, 2013). However, research involved with the HSC environment in Dynamic Positioning (see section 2.1) is limited, while this HSC-type is gaining popularity in the maritime domain, showing an increase in size and operational complexity (Kleij, Te Brake & Broek, 2015). With the increasing complexity also comes an increase in incidents, with 620 incidents reported between 2000 and 2010 (IMCA Reports). In the worst case an incident of loss of position can lead to a collision between seagoing vessels and fixed installations such as oil platforms, bridges, or quays. Such collisions can have major consequences for human and economic assets. Moreover, damage to pipes used for oil drilling can lead to severe oil spills, thereby threatening ocean life and ecosystems (Kleij et al., 2015; Sandhåland, Oltedal & Eid, 2014). The most important cause of incidents are operator errors (Tjallema, van der Nat, Grimmelius & Stapersma, 2007). Because the monitoring task is largely passive, the operator's SA is often low, also referred to as being 'out-of-the-loop'. Loss of SA with Dynamic Positioning Operators (DPOs) and subsequently missing critical information has been found to be the leading cause of incidents (Sandhåland et al., 2014). Despite these incidents and known causes, research regarding concepts to support the DP operator is scarce. The goal of this research is to find a support concept that will restore SA in DP operators and quickly gets them back into the loop again, decreasing the amount of human error in future operations.

2.Background

2.1 Dynamic Positioning system

A DP system is defined as a set of components used to keep a floating vessel at a specific position, or to make said vessel follow a pre-defined path by means of propeller action. DP is used in shuttle tanker operations, deep water drilling, dredging, rock dumping, pipe laying, cable laying and repair operations, and military operations. The human operator is called the Dynamic Positioning Operator, referred to as the DPO (Fossen, 1994). The DPO mostly supervises the DP control system, but also has the ability to take over manually and to give instructions to the system, such as a desired position. To manoeuvre to the desired position or to stay stationary, the control system combines information provided by position references and sensors, with the exact numbers of systems and sensors varying between system brands and versions (Hauff, 2014). The IMCA (International Marine Contractors Association) has been publishing incident reports and analyses about DP operations over the course of a decade. This has given a good insight in the cause of incidents. As Tjallema and colleagues (2007) describe, the DP incident data shows that DP operators cannot prevent a substantial part of DP incidents. However, for major loss of position incidents, operator errors are the leading cause. These are not just active errors, for example pushing the wrong button or wrongfully taking over manual control are considered causes. The majority of the errors is associated with a failure to notice indications that an incident is about to happen, or noticing them too late. Although a cause could be a faulty sensor, it is the failure of noticing the defect, finding the cause of the defect, and the absence of the DPO handling appropriately to solve the problem that is causing problems. This indicates that in a great number of DP incidents, human error is involved somewhere in the process. This process can be described in five phases: fault detection, fault identification, generation of solution strategy, solution implementation, and system's reaction (*see figure 1* for a schematic model). Because the operator has to act in three or four of these stage (the system not always notices problems, but will react to the solution), the time taken to complete the event chain is mostly influenced by the operator's actions (Tjallema et al., 2007).

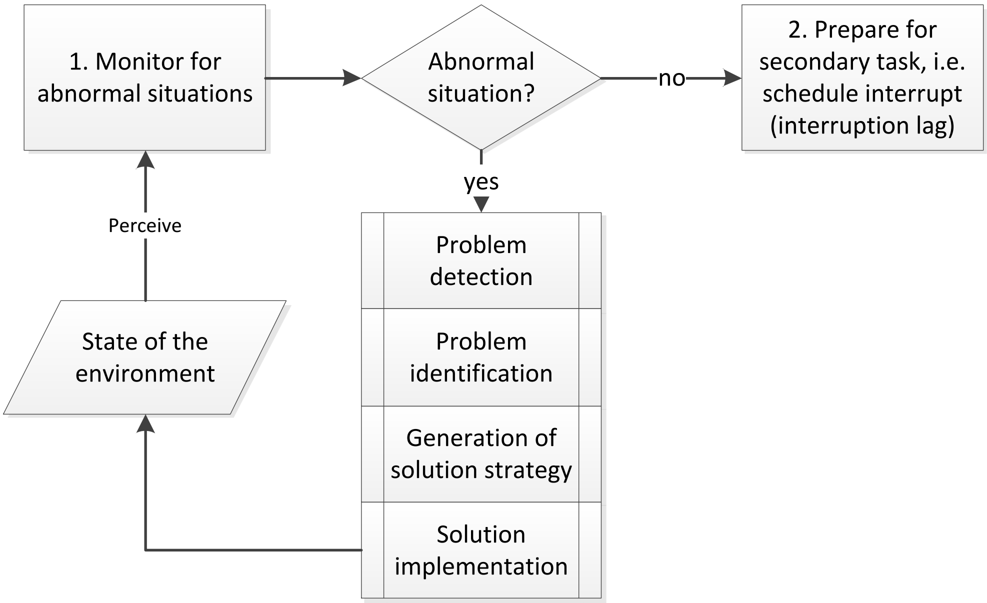


Figure 1. The process that a DPO runs through doing his primary monitoring task.

2.2 Factors influencing the decision process

As described earlier, a factor often contributing to human errors in HSC environments is loss of Situation Awareness (SA). According to Endsley (1988) SA is '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'. In the same article Endsley described the decision model of an aircrew, showing the importance of SA in decision making. This model can also be placed in the perspective of other professions involved with human computer interaction, such as DPOs (See figure 2).

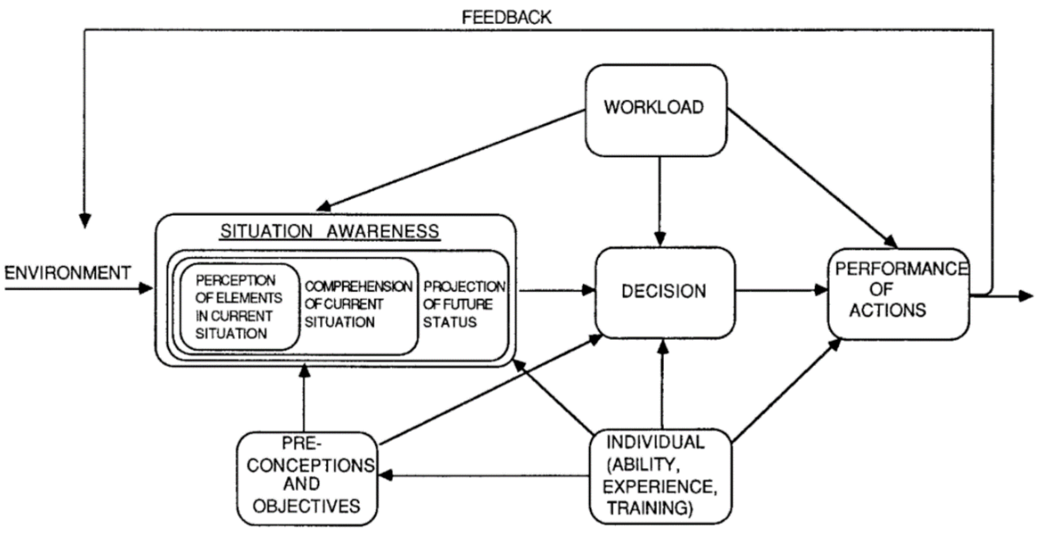


Figure 2: Aircrew Decision model by Endsley (1988).

According to this model, not only SA but also workload plays an important role, affecting situation awareness, decision making, and performance of actions directly. Problems found in HSC environments and human-automation interaction include unbalanced mental workload, reduced SA, decision biases, mistrust in the system, overreliance on the system, and bad compliance with the system (Parasuraman, Cosenzo & De Visser, 2009). Looking at the first two problems and the decision model of Endsley, a support concept improving SA and workload would be the most beneficial in optimizing the decision process.

Most of the time, the DP-system is fully automated and the primary task of the DPO is just monitoring the screen(s) for abnormal situations. The consequence of this boring task is the so-called 'out-of-the-loop' problem, characterised by a low SA and an unbalanced workload level: the operator is not actively part of the process. If an automation problem occurs that demands the input of the operator, he suddenly has to regain SA; a prerequisite for quick and good decisions (Endsley, 1995). This process often takes too long, resulting in a situation in which the available time to solve a problem is shorter than the time needed to solve the problem, leading to an incident (Tjallema et al., 2007). It therefore is of utmost importance for the DPO to remain in the loop and to keep SA high and workload at an optimal level. This can be done by recovering SA, called situation awareness recovery (SAR) introduced by Gartenberg et al. (2014).

2.3 Theorizing a support concept.

SAR is defined as the process of restoring SA after SA has been reduced. SA reduction can for example be caused by loss of attention, interruptions, or multitasking. Gartenberg and colleagues (2014) state that since interruptions have a harmful effect on SA, the operator automatically engages in SAR after an interruption. John & Smallman (2008) describe four stages of maintaining and recovering SA after an interruption: (1) Real Time change detection, (2) preparing for interruptions, (3) Reorienting and retrieving task goals, and (4) Post hoc change detection.

The future goal is to let the DPO perform a secondary task besides the monitoring task, which also can be seen as a form of interruption or multi-tasking. If no abnormal situation has been detected during the primary task (monitoring) the DPO can prepare for the secondary task (pre-interrupt preparation). After the interruption he has to retrieve the information from before (post-interrupt reorientation) and has to detect changes in that information (Post hoc change detection).

Gartenberg and colleagues used the memory for goals theory (MFG) as devised by Altmann & Trafton (2002), as a framework for the process of SAR. The MFG theory is described as raising the activation of previous goals and plans, so that they (old objects) are acting as contextual cues that promote the recovery of SA.

The goal of a DP operator is to monitor certain values given by sensors and other output signals (level 1 SA). When the DPO is performing his secondary task, he has to memorise his goals (values), so that after completing the task, or after an alarm, he knows which values have changed (level 2 SA) and what these changes mean for the future state of the system (level 3 SA). Only then, problems can be detected. *Figure 3* shows the dynamic positioning model, when the theory of SAR and memory for goals are implemented.

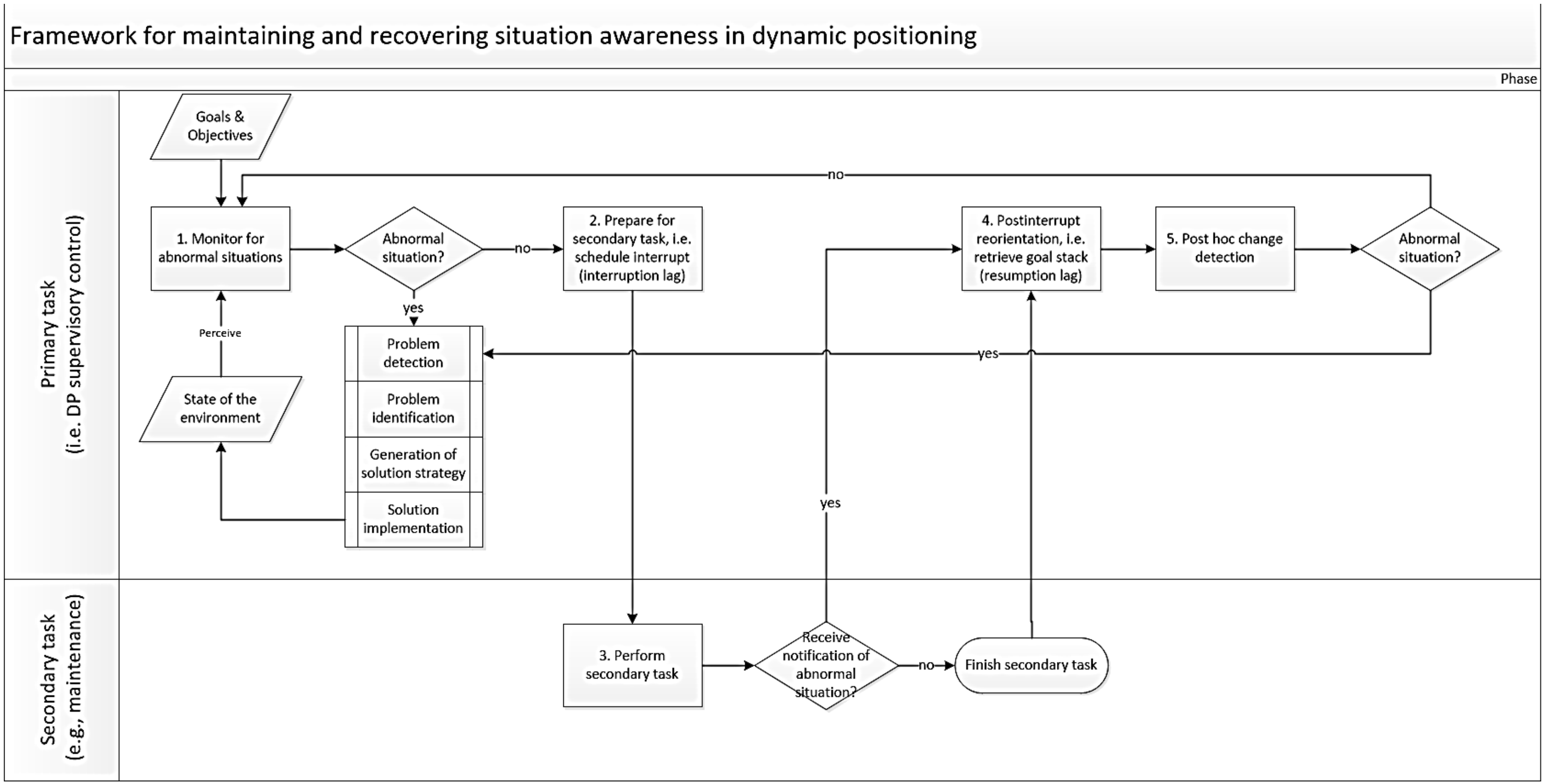


Figure 3: The process that a DPO runs through during his monitoring task, in addition with the steps taken when performing a secondary task.

With low SA, steps 4 and 5 of the model are delayed. This is consistent with the findings by Gartenberg et al. (2014), who showed that participants experienced a longer resumption lag after an interruption. They also found that interruptions lead to increased scanning (quicker and more fixations), presumably to rebuild SA. To speed up the course of action of a DP operator after an interruption, the solution lies in accelerating SAR. This can be done by a support system that gives updates to the operator about his goals and objectives (e.g. sensor values) to help him perform change detection. This can be done during the monitoring task or during a secondary task. In this way the activation of his previous goals is increased, which then shortens the resumption lag. After the interruption, the operator will be more familiar with the environment and therefore the scanning procedure in the post hoc change detection is faster (level 2 SA) because he no longer has to perform change detection all by himself. This way the duration of the problem detection phase is shortened and in the solution generation phase, the right decision will be made more often, leading to fewer errors (level 3 SA).

2.4 Existing support concepts and design principles

With the concept of SAR in mind, John & Smallman (2008) tried to help users detect and interpret changes, either to maintain SA or to recover it following interruptions. They outlined 4 principles for developing a support concept, based on their predetermined stages of SAR mentioned in section 2.3 . These design principles should make the support effective for change detection. The first and most important one is *automated change detection, where the system automatically detects and shows changes.* This is expected to augment user's natural change detection ability. According to the article, automated change detection should significantly improve the user's ability to maintain and recover situation awareness in dynamic tasks (John & Smallman, 2008). The problem, however, is not implementing automated change detection, but whether it helps the users, and how it is presented to them. This brings forward design principle 2: *Unobtrusive Notification,* which states that users should be notified of changes in a relatively unobtrusive manner so that they are minimally distracting from ongoing tasks. It should not interrupt other important tasks. On top of that the notification should not obscure or clutter other task relevant information like a pop-up does. Principle 3 says to *help users prioritize the changes in order of importance*. You can for example show all the changes, but some may be more important than others. For example, a large change of values is often more important than a minor change. The 4th principle states that *for busy, cluttered displays the change information should only be available on demand by the user*. In that way the user chooses when he wants to see it. It could be argued that in stressful situations or situations where quick handling is necessary, it is doubtful whether this is a useful principle. By the time the user has decided whether he wants to see the change information, it might be too late, mainly because their situation awareness has to be high in order for him to decide. These principles were guidelines for designers and other researchers to use in developing support concepts and have not actually been tested by John and Smallman in an experiment.

Some years later, John (2013) developed a new mini-dashboard display, called a Janus display. This supports supervision, and rapid scanning. Both are important skills for an operator in monitoring the Dynamic Positioning System. Here the design principles are visualisations of the display and not just theoretical. A Janus display is build up out of layers of information, increasing in amount of information presented (see Figure 4). John states:

> 'Janus displays focus on providing level 2 SA, the meaning of the data with regard to the current situation, in the sense that they provide a high level view of system status' (John, 2013).

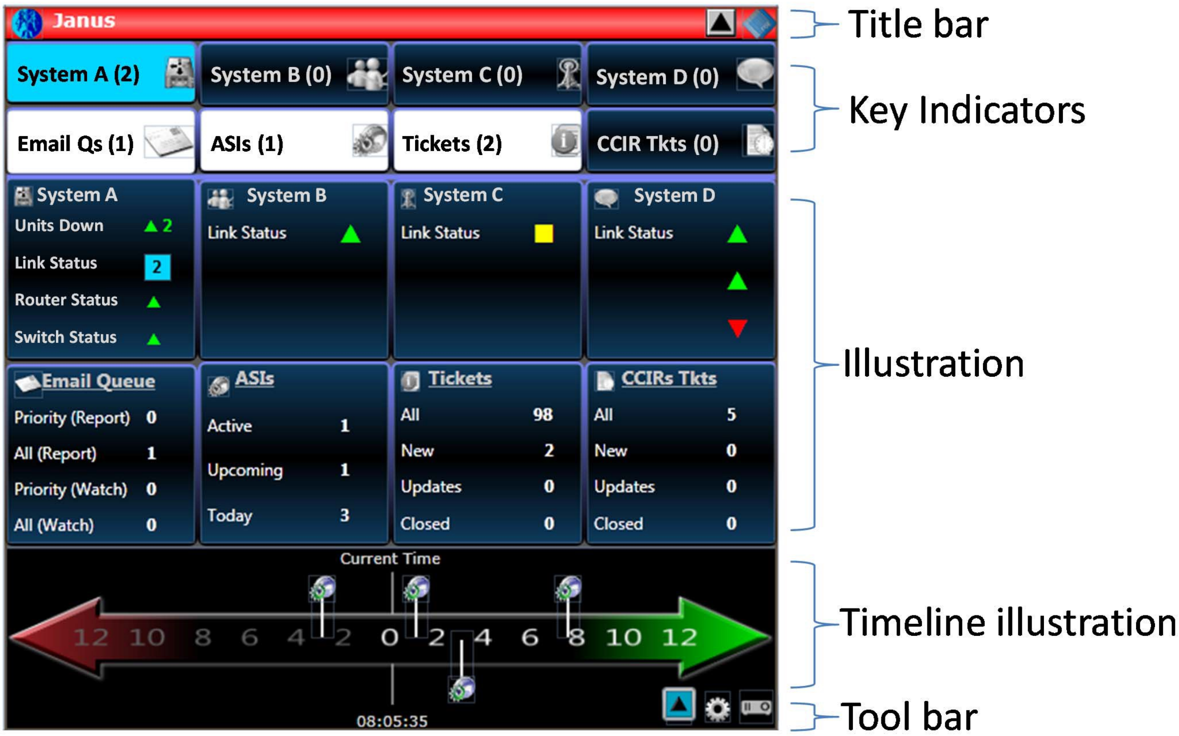


Figure 4: A Janus display as published by John (2013) supporting a telecommunication network.

The display is divided into three components: key indicators, giving the most important information about the status of the systems; illustrations to supplement the indicators with additional information; and a timeline providing further detail regarding the timing of scheduled events. Janus displays are designed to detect changes and enhance change awareness in users. They have to be constantly visible in the corner of a user's primary screen. Results showed a decline in response time with using a Janus display and an increase in performance (John, 2013). It was, however, tested in a far more active supervisory role with lots of decisions, than the more passive monitoring task of a DPO. We will adapt this paradigm to the case of the DPO.

2.5 Hypotheses

The designed SAR support system is expected to lead to an increase in accuracy, decreased response times on the monitoring task, and a reduction of perceived work load levels. Moreover, we predict that an interruption causes loss of SA and thereby elicit SAR, and that this interruption will cause an increase in perceived workload. We expect the support concept to be less beneficial when no interruption happens, as no loss of SA is expected in that situation.

3. Methods

To investigate the benefits of the support system, participants performed a monitoring task in a micro-environment based on a Dynamic Positioning Simulator. It was manipulated whether a support concept was available to participants, and if there was a secondary task during the monitoring task.

3.1 Participants and experimental design

The participants were 23 students (12 male, 11 female) aged 18-28 years (M = 22.1, SD = 2.4). All participants filled in an informed consent before participating. Participants were asked to report their gaming experience beforehand. No correlation was found between gaming experience and performance. The session lasted approximately 2.5 hours and participants received financial compensation for their time. The best and second best participant received a financial bonus. This incentive was introduced to increase motivation. The design was 2x2 factorial, with interruption condition (present/absent) and support condition(present/absent) as within factors. The sequence of conditions was counterbalanced.

3.2 Task environment

We developed a DP simulation that was based on an actual DPO training simulator. This training simulator is designed according to a complex decision model, creating realistic sea environments and scenarios. Because of the complexity of the simulator and the use of students (not experts) as participants we decided to use a simplified version of the simulator. To measure cognitive constructs like SA and workload, an environment as complex and realistic as the simulator could influence the desired effects found, because the task would be too hard to complete. Besides, it would take months of training and numerous courses for a subject to get an understanding of the simulator system (The Nautical Institute, 2016). The simplified version displaying the start of a scenario, is depicted in Figure 5a. Nine numeric parameters or variables had to be monitored while the virtual ship went through twelve task scenarios, and ten training scenarios based on realistic DP-events.. The parameters were: Wind angle (degrees), Azimuth angle (angle of the Azimuth thruster in degrees), Azimuth thruster (% power), Bow thruster (% power), Wind strength (m/s), Setpoint error (distance from setpoint in meters), Wave height (m), Roll angle (absolute degrees), and wave period (seconds/wave). Scenarios included sudden wind increase, thruster failure, drifting, wave height increase, change of wind angle and various combinations.

This setup retains the ecological validity, as the scenarios were based on realistic failure modes, derived from the simulator and interviews with two experienced DP operators. This way the subjects were operating in a micro-world of a DP-operator. Performance can be easily measured and the variables and the number of variables could be easily manipulated to increase or decrease the difficulty, which makes it suited for a controlled experiment.

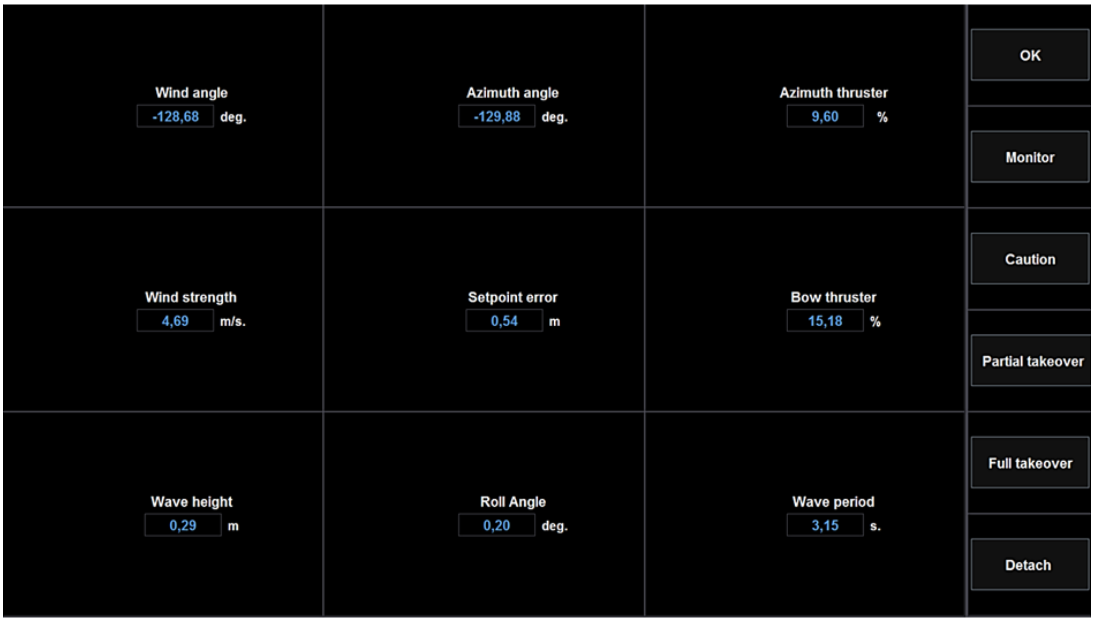


Figure 5a: Simplified monitoring environment based on the DP-simulator at the start of the scenario.

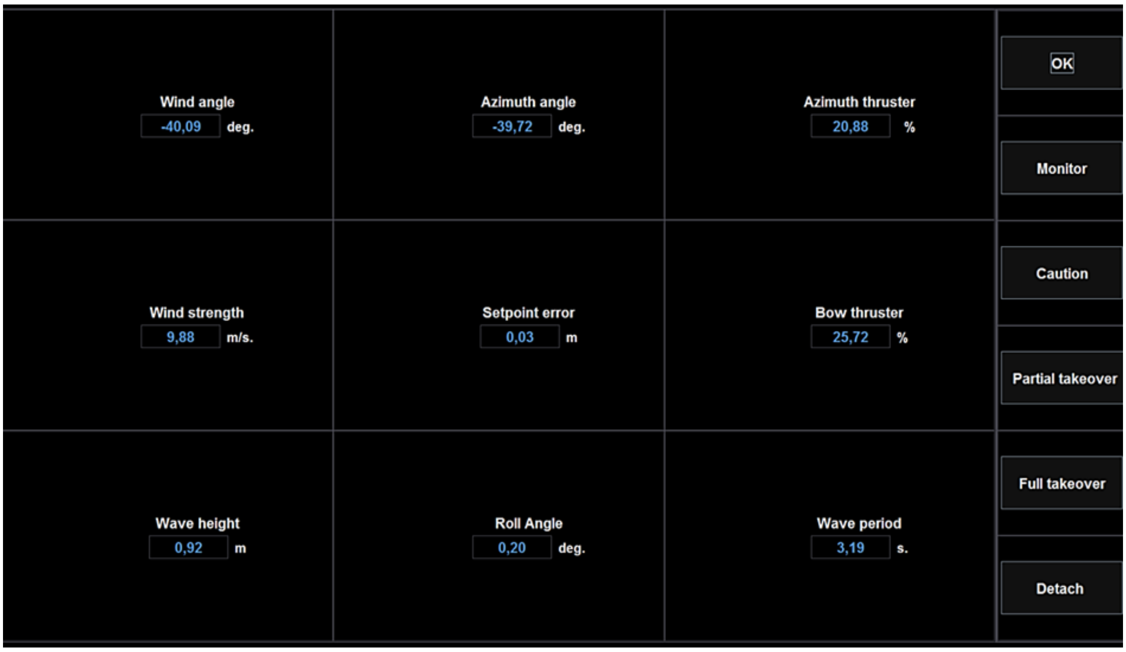


Figure 5b: The same monitoring environment as *figure 4a* at the first alarm. Comparing the values to the values at the start of the scenario, wind angle and azimuth angle changed with more than the Delta. Also wind strength increased. No thresholds are exceeded. Because the azimuth angle still has the same angle as the wind, this change is no abnormality. However, the change of wind angle and the wind strength are both abnormalities. The right button to choose would be: Monitor (1 or 2 abnormalities).

3.3 Monitoring task description

Subjects impersonated a DP operator on a virtual ship. They had to monitor the nine parameters for thresholds that could be exceeded, as well as two forms of abnormalities: linear increase or decrease of variables -called deltas-, and freezing of variables. Beforehand they learned information about thresholds, delta's and interplay between the variables (Figure 6). The 9 variables fluctuated around a value in the baseline condition. At predetermined time intervals, a maximum of 6 values could begin to increase or decrease. The task of the subjects was to react to an alarm signal indicating a response had to be given. An alarm indicated that a response window of 50 seconds started and that subjects had to make a decision as quickly as possible. Scenarios were 240 seconds long, with the first alarm sounding between 65 and 100 seconds from the start of the scenario. The second alarm sounded between 160 and 185 seconds from the start. A response was given by clicking on one of the six action buttons (Figure 7), increasing in extremity of the situation: (1). "OK", (2) Monitor, (3) Caution, (4) Partial Takeover, (5) Full Takeover and (6) Detach. According to the learned rules and a mental model of the situation the subjects had to decide which action was the correct one. For an example, see figure 4b. The tables were printed and placed under the computer monitor and therefore were also available to subjects during the task (See Figure 8). For more pictures of the experimental setup, see appendix 1).

Variable	Min Threshold	Max Threshold	Delta	Delta only an abnormality when:
Wind angle	-	-	20 degrees	Changed (+/-)
Azimuth angle	-	-	20 degrees	Different from the wind angle
Azimuth thruster	5%	95%	15%	Decreased/frozen when Setpoint error increased
Wind strength	0 m/s	15 m/s	3 m/s	Increased
Setpoint Error	-10 m	10 m	2 meters	Increased or decreased from 0
Bow thruster	5%	95%	15%	Decreased/frozen when Setpoint error increased
Wave height	0 m	8 m	2 meters	Increased
Roll Angle	0 gr	20 gr	5 degrees	Increased
Wave period	-	-	5 seconds	Changed (+/-)

Figure 6: Translated table with rules about abnormalities and thresholds. When a variable was below the minimum or above the maximum threshold at the sound, it had to be reported as exceeding the relevant threshold. When a variable differed by one delta or more compared to the start of the scenario, subjects had to judge with use of the last column if the delta was an abnormality. If so, it had to be reported as an abnormality.

#	Action	Decision rule
1	OK	No abnormalities or exceeded thresholds
2	Monitor	1 or 2 abnormalities
3	Caution	1 threshold exceeded <u>or</u> 3 abnormalities
4	Partial takeover	1 threshold exceeded <u>and</u> abnormalities
5	Full takeover	2 thresholds exceeded
6	Detach	2 thresholds exceeded <u>and</u> abnormalities

Figure 7: Action buttons with decision rules. With use of the learned rules (or table 1) about abnormalities and thresholds subjects made a decision by clicking on one of the six action buttons. They had a 50 second window to respond.

Figure 8: The experimental setup with the tables (Figure 5 and 6) provided below the pc-monitor.

3.4 Independent variables & conditions

3.4.1 Support

In the support condition a s*upport concept* was added. This was a change detection- and comprehension support in the form of an arrow, displayed in Figure 9b, indicating a change (detection) and a direction (understanding the change). The arrow appeared when a variable had in- or decreased during the past 5 seconds, exceeding a predefined threshold. The arrow remained present after it appeared, giving information that a value had been increasing but also remained if the variable was stable again. Hence, the arrow depicted whether variables had changed significantly over a fixed time interval. This was especially important in the interruption condition because real time changes could not be perceived. Participants were told beforehand in the support condition that the change detection tool would be switched after the first response window ended. The support concept helped with perceiving deltas, but subject still had to decide if these changes were also abnormalities or exceeding thresholds. For examples, see Figure 9a and 9b.

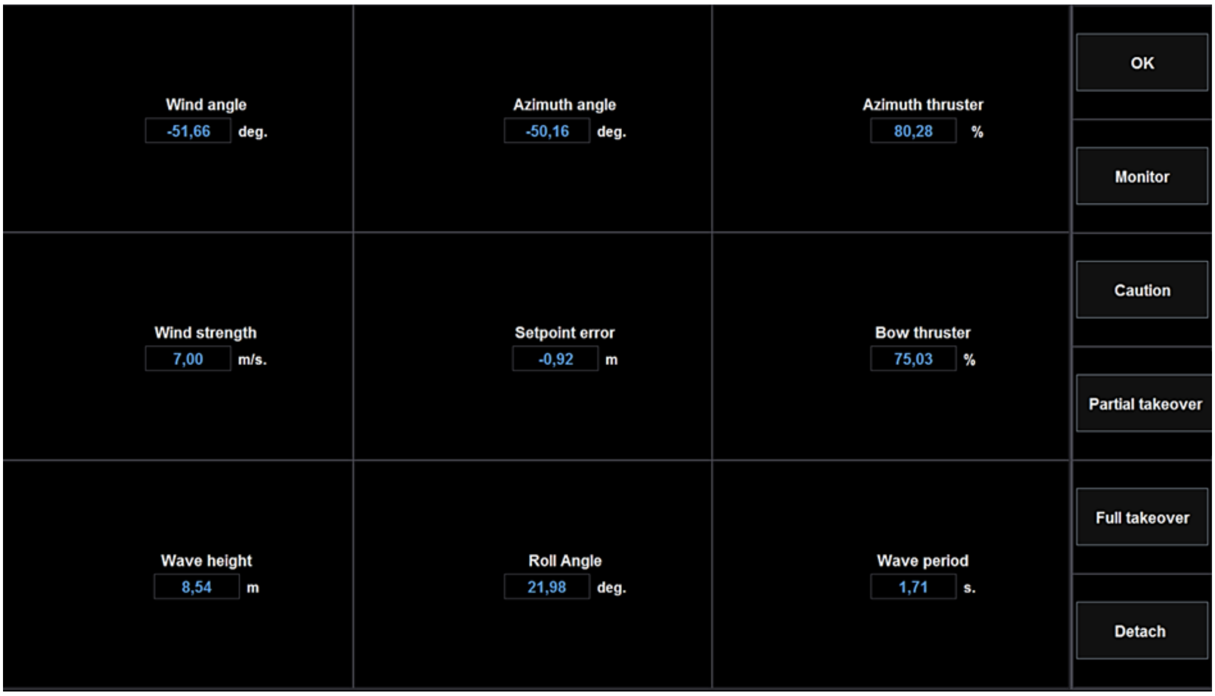


Figure 9a: The task environment at the first alarm in the support condition. The wave height is above the threshold as is the Roll angle.

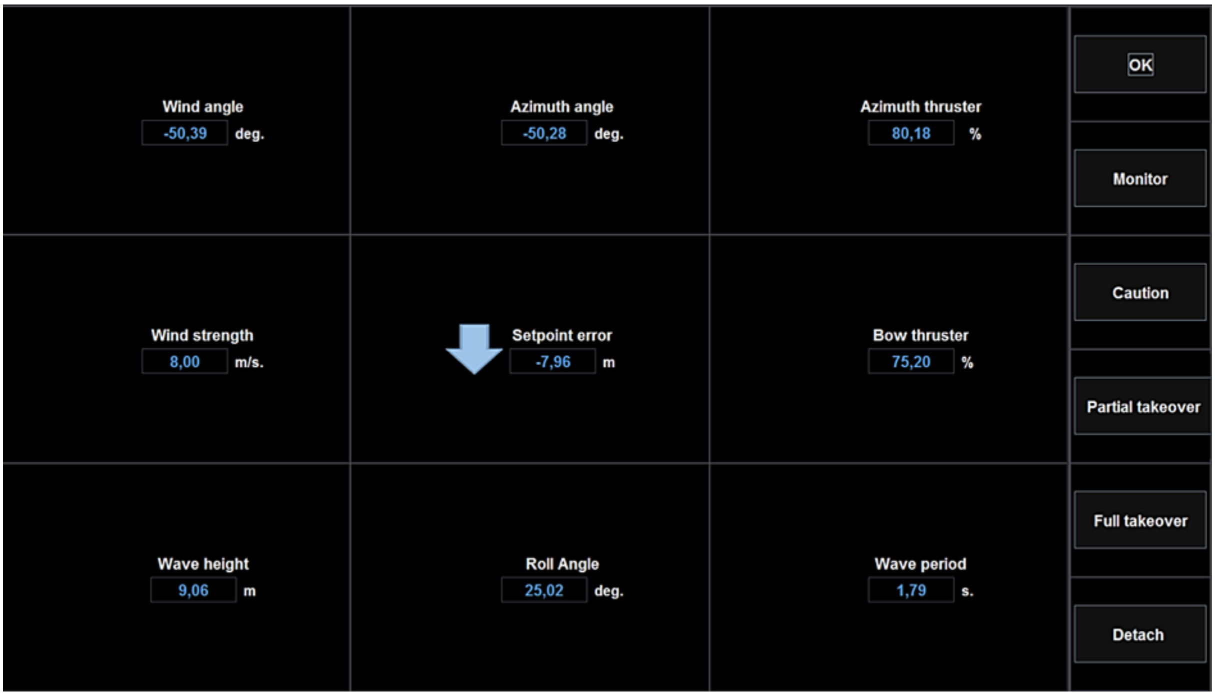


Figure 9b; The monitoring environment at the second alarm, showing the support concept in the form of an arrow. In this example, the setpoint error changed with a delta or more, being an abnormality. Wave height and Roll angle are still above thresholds. This gives a total of 1 abnormality and 2 thresholds. The correct decision would therefore be: Detach (2 thresholds exceeded + abnormalities)

3.4.2 Interruption condition (secondary task)

In the interruption condition subjects had to perform a secondary task to induce the loss of SA: the Visual Elevator Task (VET). This is a subtask of the Test of Everyday Attention developed by Robertson, Ward, Ridgeway and Nimmo-Smith (1996). This task was chosen because it causes interference with the monitoring task as it also uses numeric trials and arrows, just like the monitoring task has numeric variables andthe support concept is an arrow. This way it was attempted to create a big loss of SA, hopefully causing a stronger SARreaction than with using a task not appealing to those constructs. Subjects were told to leave the monitoring task and perform the secondary task as soon as the first response was given. Hence, each scenario consisted of two parts. It was made clear that subjects were not allowed to look at the screen while performing the VET. When the sound indicating the start of the second response window sounded, subjects immediately had to stop with the VET and return to the monitor screen to give a response. They performed the VET on paper on another desk, making it impossible to watch the variables simultaneously with the task. The experimenter kept an eye on the subject to assure instruction compliance.

The VET task requires subjects to count up and down as they follow a series of visually presented ‘doors’ of an elevator. This task is normally a measure of cognitive flexibility and loads on the same factors as the number of categories on the *Wisconsin Card Sorting Test,* appealing to cognitive functions as memory and attentional switching. Participants were asked to count a series of drawings of elevator doors that were presented in rows on the pages of a presentation booklet. A number of rows were combined and form a string. The drawings of the elevator doors are alternated with large up- and down-pointing arrows, indicating that the direction of counting should change in line with the arrow (so counting up or down). At the beginning of a row, a number is presented. This is the number of the ‘floor’ the participant begins on (see Figure 10 for an example of the Visual Elevator Task). The participant always began with counting up. At the end of a string the participant had to decide at which floor he/she had arrived. The task was self-paced, which means that the subject could turn the pages at their own speed. However, they were told that they should perform as many trials as possible, as accurately as possible.

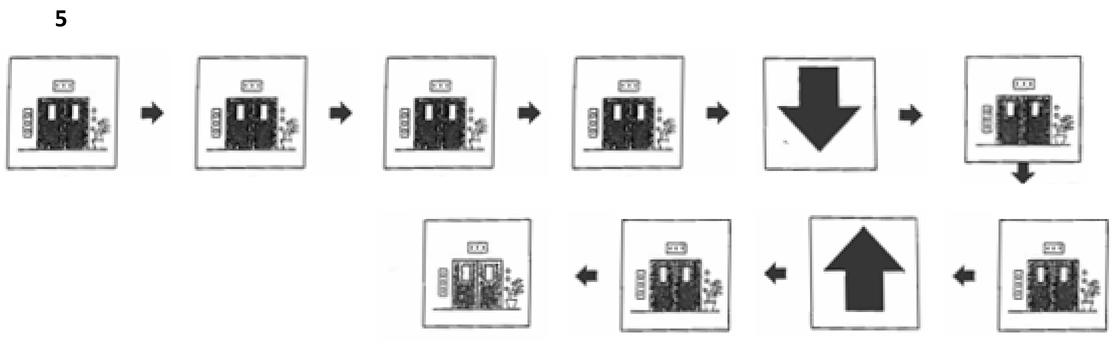


Figure 10: Scan of Visual Elevator Task practice trial 2. This elevator begins on floor 5. The correct counting is: 5-6-7-8-7-6-7-8, with 8 being the final floor of this string and the correct answer.

3.5 Dependent variables

3.5.1 Monitoring task performance

The performance on the monitoring task was measured in two ways: the percentage of correct decisions (*accuracy*) and the response time on the correct responses(*speed*). Only the response time of correct responses was analysed because the time of the incorrect responses also include guessed responses, being very fast, but wrong. This would lead to non-representative results. After the response window had expired, the next phase of the scenario began. When participants failed to choose an action when the response window expired, the response was s marked as incorrect. It could also be that nothing happened and that they were still in the baseline condition. Then subjects still had to choose an action, being the action "OK".

3.5.2 Secondary task performance

The secondary task score was calculated by dividing the total time a subject had to perform the task (time of the first response till the start of the second response window), divided by the amount of correct trials.

3.5.3 Situation Awareness

To measure SA, the Situation Awareness Rating Technique (SART) was used (Taylor, 1990). This is a subjective measure. An objective measure that is often used, is the SAGAT, however for our experiment it was not suitable because of its obtrusive nature, i.e. freezing the screen. Another limitation is that the SAGAT cannot be administered in short trials of 4 minutes. We chose to use a subjective measure of SA for a number of reasons. Firstly, self-rating techniques are non-intrusive to task performance because they are completed post-trial. Moreover, it is easy to use, and requires little training. Of the subjective measures, the SART is the most popular technique and is validated by 10+ studies (See Salmon, Stanton, Walker and Green, 2006 for a review). The SART measures three components: *Attentional demand, Attentional supply* and *Understanding.* The three components have a total of 10 dimensions, resulting in 10 questions that are answered on a 7- point scale: the 10-D SART questionnaire (Kennedy & Durbin, 2005). For the Experiment, the components were translated into Dutch questions (*see Appendix 2*).

The scores of the subjects on the scale can be converted into percentages (see formula 1).

$$Vn = \frac{v - vmin}{vmax - vmin} * 100\% \quad (1)$$

Whereas *Vn* is the value in percentages, *v* is the true value, *vmin* is the minimum value of the scale and *vmax* is the maximum value of the scale (Satuf, Kaszkurewicz, Schirru, de Campos, 2016).
To calculate the situation awareness, the mean score per component is converted to a percentage. Then from this values SA can be calculated with formula (2).

$$SA = U - (D - S) = U + S - D \quad (2)$$

With *SA* being the Situation awareness score, *U* the understanding score, *D* the score for attentional demand and *S* the score for attentional supply.

3.5.4 Workload

To measure workload, we used the Rating Scale Mental Effort (RSME) (Zijlstra, 1993). The scale consists of Dutch verbal anchors expressing different degrees of effort expenditure: 'ontzettend inspannend' (very effortful), 'een beetje inspannend' ('a bit effortful') etc., posited along a continuum ranging from 0 to 150 (see Appendix 3). The scale values of the anchor points range from 3 'helemaal niet ontspannend' (not at all effortful) to 115 'ontzettend inspannend' (tremendously effortful). The RSME has proven to be an adequate indicator of psychological costs of task performance, and has demonstrated to be a valuable instrument in both the field and experimental studies. The 0-150 ratings can be calculated into percentages on a 0-100 scale (see formula 3).

$$Workload\ percentage = \frac{RSME\ score}{150} * 100\% \quad (3)$$

3.6 Procedure

Subjects were welcomed at the entrance and guided to the experiment room where they got information about the work of a DP operator and a brief explanation of the task. After this, they filled out an informed consent and a short questionnaire about their education and gaming experience. Thereafter they got a verbal and written explanation of the monitoring task, and decision rules, which they studied for five minutes. After this the training started with 10 practice scenarios. The experimenter operated as coach to give subjects feedback on their responses. The same information was given to every subject. They were also told that no questions would be answered during the training. Only when something was not clear, with the result that a subject would answer everything wrong, information was given. Subjects ran through the practice scenarios until 70% of the responses were correct. Only then subjects were qualified to begin the real experiment with the monitoring task.. When not qualified after 10 practice scenarios, the participant was excluded from further participation in the experiment.

After a short break participants began with the first condition. Before entering, participants were informed about the condition they entered. The experimenter left the room and observed the participant from behind a window (see Appendix 1). The total experiment consisted of 4 conditions with 3 scenarios each, lasting 4 minutes per scenario. A scenario consisted of two parts, with both parts having a response window. Only the second part contained the manipulation (secondary task and/or support). When the second response window and thereby the scenario ended, the screen blacked out with the instruction to perform the SART on paper.

When the SART was completed, one scenario had ended and the task was reset to another scenario. Between conditions participants performed the RSME to determine their workload during the previous condition. A time-out screen was shown in between conditions with a message to take a 5-minute break and with the option to continue the task when ready. The continue-button was only activated by a directed mouse click to avoid accidental resumption of the task by pressing a key. The control condition followed the same course as the training scenarios. In the support condition the support was invoked after the first response window. In the interruption condition a subject left the monitoring task and began to perform the secondary task (the Visual elevator) after they had given the first response. The second alarm indicated that the second response window started. This was also the sign for participants to leave the secondary task and return to the primary task to take action as quickly as possible. In the condition with support and interruption the subject performed the secondary task after his first response and after the first response window the support concept was activated so on return of the secondary task the arrows gave deltas that occurred while the participant was absent from the screen. When all four conditions were run through, the task ended and the monitoring environment was closed.

4. Results

Repeated measures analysis of variance (RM-ANOVA, Bonferroni corrected) were conducted on the data. Results were obtained from 22 subjects. One subject did not reach the required performance rating in the training to continue with the experiment. Subject completed all conditions (Control, Control + support, Interruption and Interruption + support) in a Latin-square balanced design. Analysis was only conducted on the performance of the second response. This is because the manipulation (interruption task, support or both) was only present in the second half of the scenarios. Task performance was measured in task accuracy (percentage correct responses) and response time (RT) of correct responses.

4.1 Monitoring task Performance

Results for monitoring task accuracy scores revealed no main effect for support, $F(1,21) = 1.05$, $p = .318$, partial $\eta^2 = .05$, and no main effect for interruption, $F(1.21) = 1.30$, $p = .266$, partial $\eta^2 = .06$. There was a significant interaction between support and interaction, $F(1,21) = 8.26$, $p < .01$, partial $\eta^2 = .28$. Post-hoc paired sample *t*-tests revealed that participants performed significantly worse in the interruption condition (*M* = 56.%, *SD* = 30%) than in the control condition (*M* = 79%, *SD* = 26%), $t(21) = 3.07$, $p < .01$. No difference was found between the control condition (*M* = 79%, *SD* = 26%) and the control + support condition (*M* = 68%, *SD* = 30%), $t(21) = 1.23$, $p = .231$. However, when comparing both interruption conditions, results show that participants performed significantly better with support (*M* = 77%, *SD* = 24%) than without support (*M* = 56%, *SD* = 30%), $t\ (21) = -3.31$, $p < .001$. Participants had very similar accuracy scores in the interruption + support condition (*M* = 77%, *SD* = 24%) as in the control condition (*M* = 79%, *SD* = 26%), $t(21) = 0.24$, $p = .815$ (see Figure 11 for an overview of the effects).

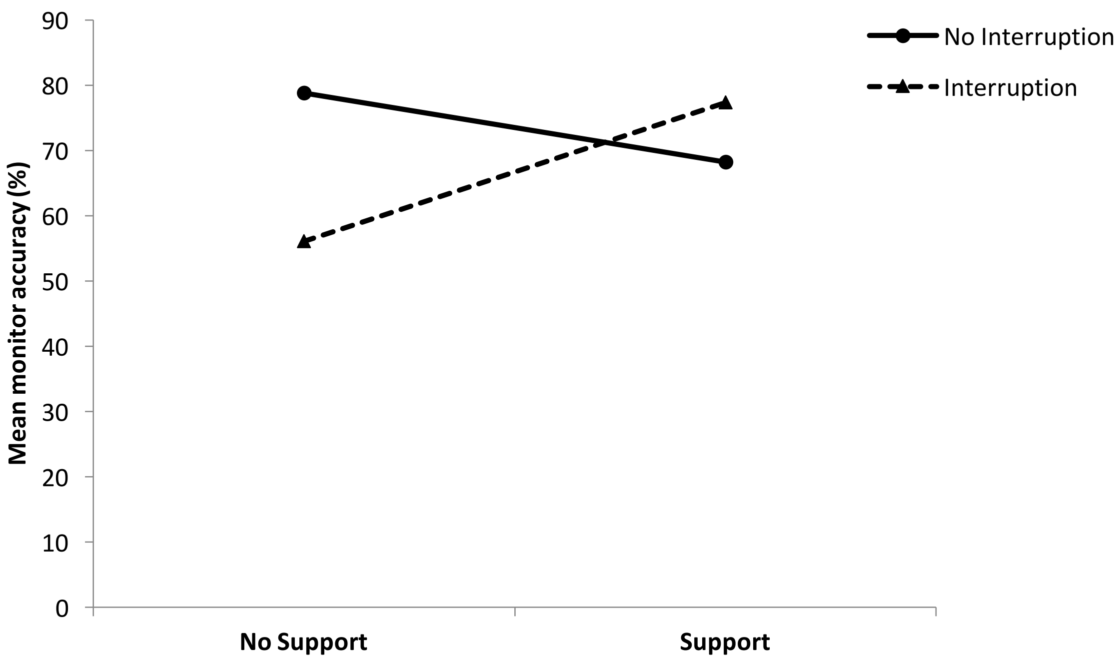
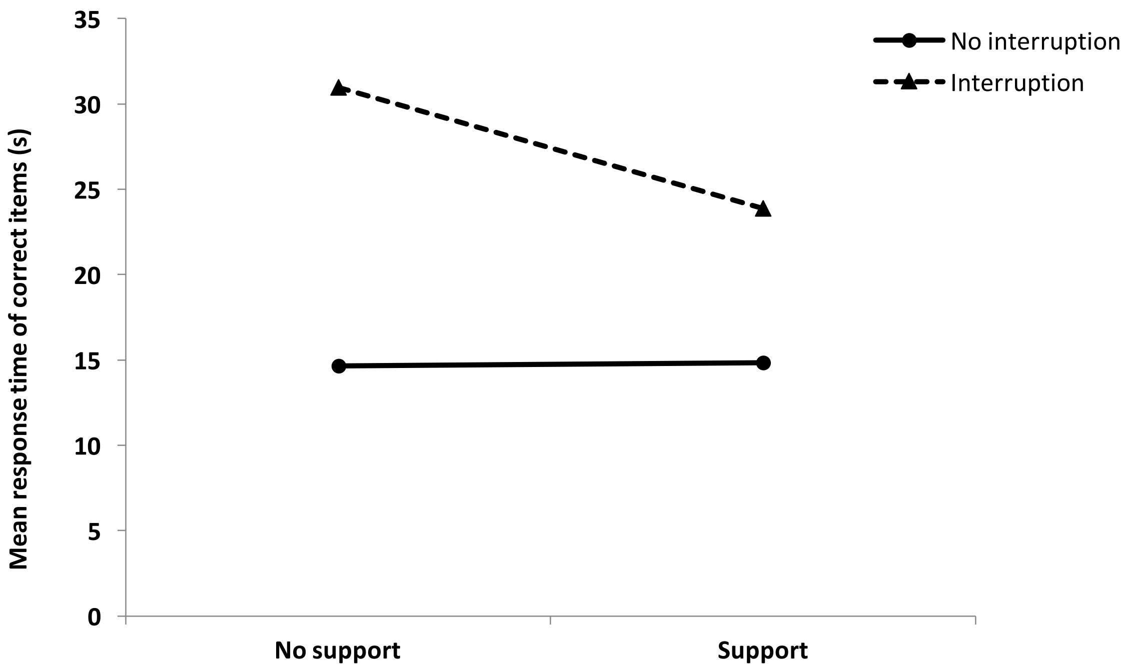


Figure 11: Effects of support and interruption on monitoring task accuracy.

When looking at response times of correct items, analysis showed a significant main effect of interruption, $F(1,21) = 33.07$, $p<.001$, partial $\eta^2 = .61$. Paired samples *t*-tests revealed that participants responded significantly slower in the interruption condition (*M* = 30.98s, *SD* = 17.67s) than in the control condition (*M*=14.64s, *SD*=7.52s), $t(21) = -4.12$, $p<.001$. Also, a significant difference was found between the support condition and the support + interruption condition, $t(21) = -7.23$, $p < .001$.

No main effect of support was found, $F(1,16) = 3.35$, $p = .081$, partial $\eta^2 = .14$. No significant interaction was found, $F(1,21)$, $p = .074$, partial $\eta^2 = .144$. However, because it was close to significance we chose to conduct a post hoc t-test (one-tailed) to compare both interruption conditions. A significant difference was found between the interruption condition ($M = 30.98$s, $SD = 17.67$s) and the interruption + support condition ($M = 23.88$, $SD = 6.49$), $t(21) = 1.96$, $p < .05$. Subjects responded significantly faster when support was given, but only after an interruption was present. The results are displayed in figure 12.



with

Figure 12: Effects of support and interruption on response time (s). There was a main effect of interruption.

4.2 Secondary task performance

Performance scores were calculated by dividing the total time (s) taken for the task by the amount of correct trials. The lower the score, the better the performance.

There was no significant difference in secondary task performance between the interruption condition ($M = 15.44$ s, $SD = 4.9$ s) and the interruption + support condition ($M = 15.21$ s, $SD = 5.38$ s), $t(21) = .184$, $p = .856$.

4.3 Situation Awareness

There was a significant main effect of support on situation awareness, $F(1,21) = 6.862$, $p < .05$, partial $\eta^2 = .246$. Paired samples *t*-tests showed that the SA percentage of subjects did not differ between the control ($M = 63.99\%$, $SD = 9.75\%$) and the control + support condition ($M = 63.21\%$, $SD = 9.84\%$),

$t(21) = .625$, $p = .539$. In the interruption conditions, however, the interruption + support condition showed a significantly higher SA percentage (M = 65%, SD = 11%), compared to the SA percentage when no support was given (M = 58%, SD = 9%), $t(21) = -4.22$, $p < .001$.

No main effect of interruption was found on situation awareness, $F(1,21) = 2.30$, $p = .144$, partial $\eta^2 = .10$. We found a significant interaction effect, $F(1,21) = 18.83$, $p < .001$, partial $\eta^2 = .47$. The significant interaction effect allows further contrast analysis. Paired samples t-tests were used to compare the control condition with the interruption condition. Results showed that SA was significantly lower in the interruption condition (M = 58%, SD = 9%) than in the control condition (M = 64%, SD = 10%), $t(21) = 3.42$, $p < .01$. It was also found that subjects had similar SA percentages in the control condition (M = 64%, SD = 10%) as in the interruption + support condition (M = 65%, SD = 11%), $t(21) = -.27$, $p = .789$. In figure 13 the results are displayed.

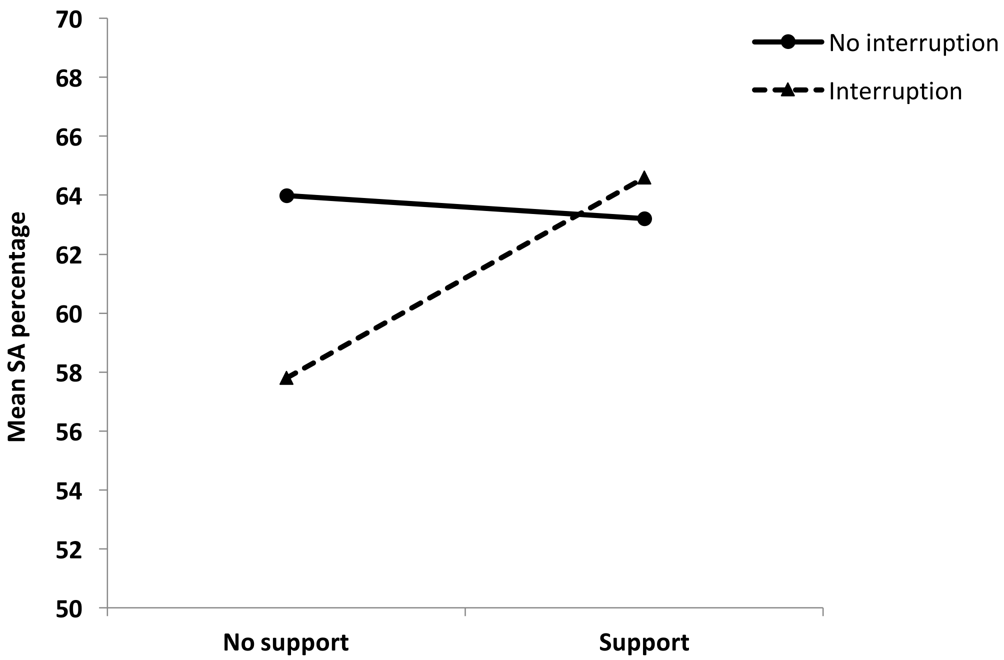


Figure 13: Effects of support and interruption on SA. There was a main effect of support.

4.4 SA Components

We further analysed the different components of SART (attentional demand, attentional supply and understanding). For **attentional demand**, no main effect of support was found, $F(1,21) = .24$, $p = .631$, partial $\eta^2 = .01$ and neither for interruption, $F(1,21) = .55$, $p = .466$, partial $\eta^2 = .03$. There was a significant interaction between support and interruption, $F(1,21) = 18.26$, $p < .001$, partial $\eta^2 = .47$ (see Figure 14). Post hoc t-tests revealed that attentional demand was significantly higher in the interruption condition (M = 14.77, SD = 2.11) than in the control condition (M = 13.71, SD = 2.42), $t(21) = -2.24$, $p < .05$. In other words, more attention was demanded when a participant was interrupted. This was opposite for the two support conditions. Attentional demand was significantly higher in the control + support condition (*M* = 14.91, *SD* = 3.09) compared to the support + interruption condition (*M* = 13.15, *SD* = 2.28), *t* (21) = 2.67, $p < .05$. Looking at the interruption conditions, attentional demand was significantly higher in the condition without support (*M* = 14.77, *SD* = 2.11) than with support (*M* = 13.15, *SD* = 2.28), *t*(21) = 3.30, $p < .01$. The demand was practically the same for the control condition (*M* = 13.71, *SD* = 2.42) compared to the interruption + support condition (*M* = 13.15, *SD* = 2.28), *t* (21) = .86, *p* = .402.

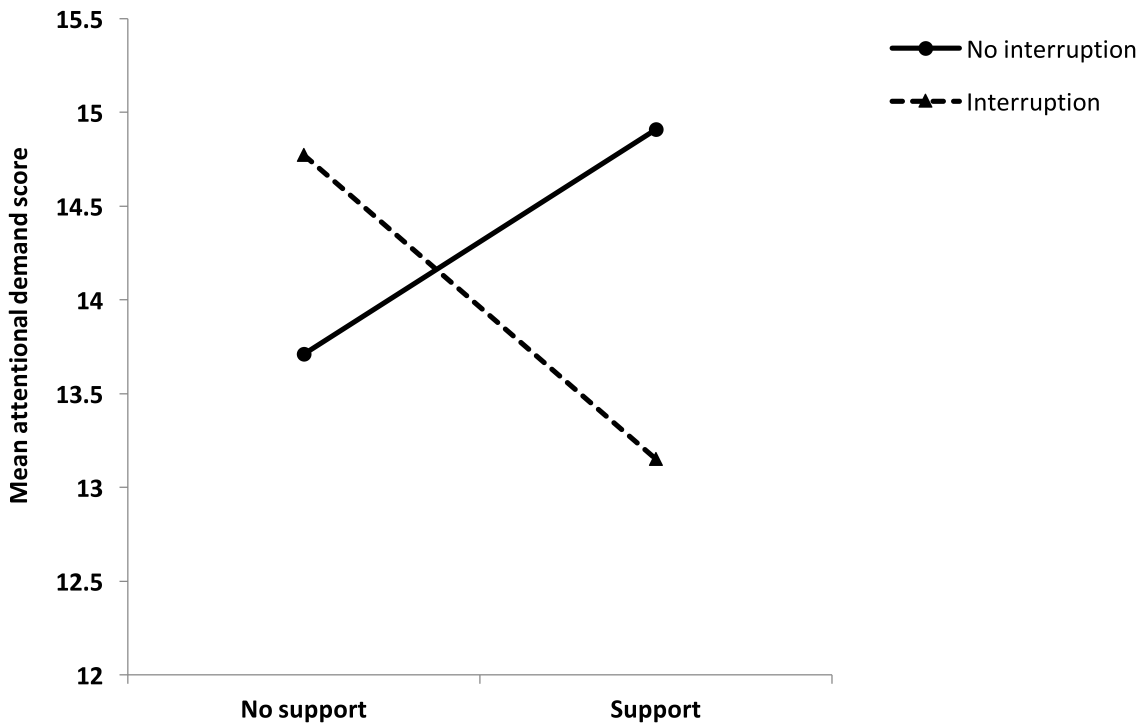


Figure 14: Effects of support and interruption on the attentional demand component of SA. An interaction effect is present.

No main effect was found of support on **attentional supply**, *F*(1,21) = .51, *p* = .483, partial η^2= .02. Likewise, no main effect was found for interruption, *F*(1,21) = 3.95, *p* = .060, partial η^2= .16. Also no interaction effect was found between the two, *F* (1,21) = 3.71, *p* = .068, partial η^2= .15.

When looking at the last component, **understanding**, we found a significant main effect of support, *F*(1,21) = 24.49, $p < .001$, partial η^2= .54. Paired samples *t*-tests revealed that subjects had a significantly higher understanding in the control + support condition (*M* = 16.76, *SD* = 2.12) than in the control condition (*M* = 15.41, *SD* = 2.33), *t*(21) = -3.57, $p < .01$. Besides, the interruption + support condition (*M* = 16.59, *SD* = 2.61) showed a significantly higher understanding score than the interruption condition without support (*M* = 14.38, *SD* = 2.28), *t*(21) = -4. 90, $p < .001$.

Furthermore, a significant main effect of interruption was found, $F(1,21)$ =8.47, $p < .01$, partial η^2= .29. Paired samples *t*-tests showed that subjects had significantly less understanding in the interruption condition (M = 14.38, SD = 2.28) than in the control condition (M = 15.41, SD = 2.33), $t(21)$ = 3.44, $p < .01$. Subjects had a significantly higher understanding in the interruption + support condition (M = 16.59, SD = 2.61) than in the control condition (M = 15.41, SD = 2.33), $t(21)$ = -2.61, $p < .05$. Apparently the benefit of the support was so high, that even with interruption the subjects still had a better understanding. No interaction effect was found, $F(1,21)$ = 4.22, p = .053, partial η^2= .17 (see Figure 15).

.

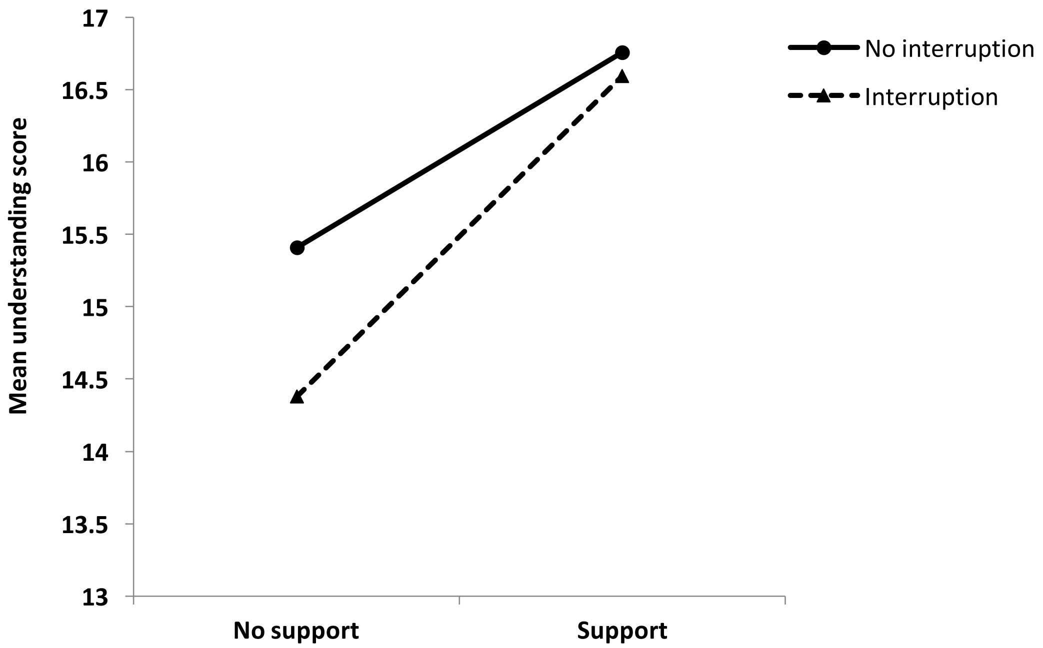


Figure 15: Effects of support and interruption on the understanding component of SA. There was a main effect of support.

4.5 Workload

There was a significant main effect of interruption on workload, $F(1,21)$ = 39.547, $p< .001$, partial η^2= .653. Paired samples t –tests showed that workload was significantly higher in the interruption condition (M = 78.55, SD = 22.90) compared to the control condition (M = 49.60, SD = 20.15), $t(21)$ = -6.316, $p < .001$. This was also the case when looking at the two support conditions: workload was significantly higher in the support + interruption condition (M = 67.29, SD = 17.58) compared to the support condition (M = 52.69, SD = 19.87), t (21) = -3.396, $p < .01$

No main effect of support was found on workload, $F(1,21)$ = 2.045, p = .167, partial η^2= .089. However, we found a significant support x interruption interaction, $F(1,21)$ = 6.652, $p < .05$, partial η^2= .241. Paired samples *t*-tests revealed that workload was significantly higher in the interruption condition (M = 78.55, SD = 22.90) than in the interruption + support condition (M = 67.29, SD = 17.58), $t(21) = 2.851$, $p = .01$. Although workload was lower in the interruption + support condition, it was still significantly higher than in the control condition (M = 49.60, SD =20.15), t (21) = -4.733, $p <$ .001. Workload was still higher compared with the control conditions (see Figure 16).

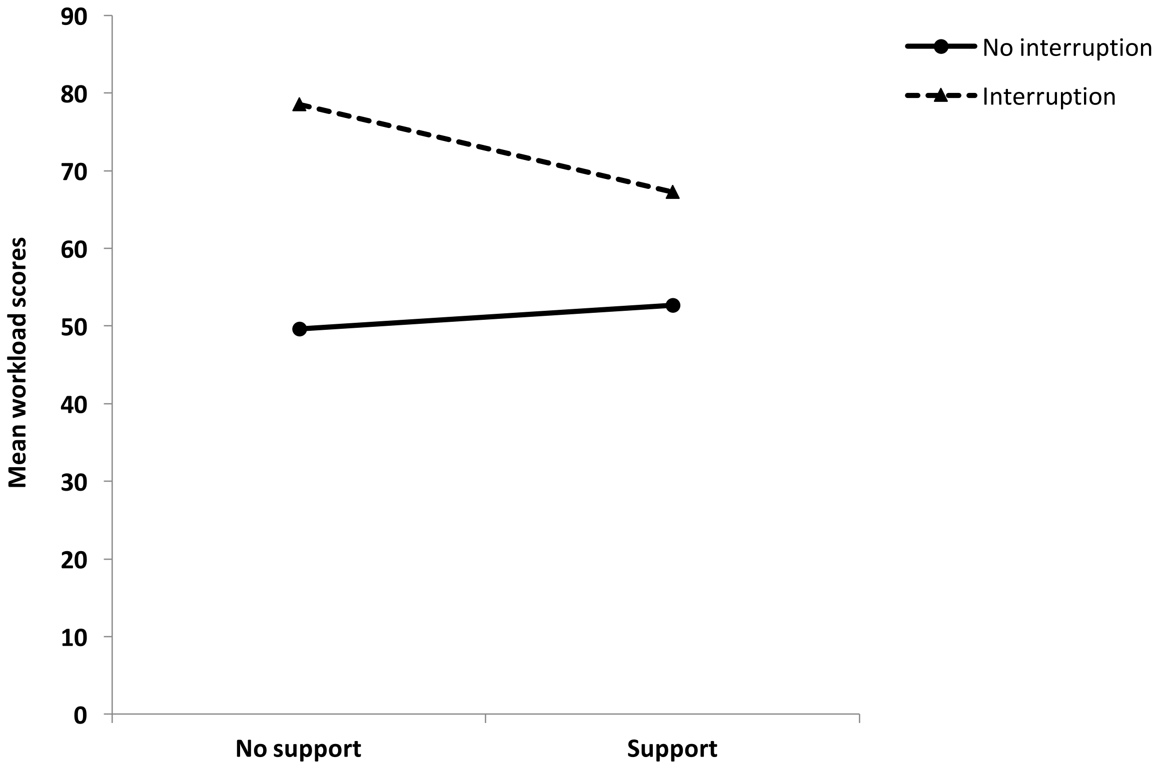


Figure 16: Effect of support and interruption on workload.

4.6 Relationship between monitoring accuracy, situation awareness, and workload.

The relationship between monitoring accuracy (%correct), situation awareness, and workload are shown in Figure 17, which plots mean values for these measures. The figure shows that situation awareness levels are similar to accuracy and that workload logically follows an opposite curve. The higher the workload, the lower the SA and the lower the accuracy.

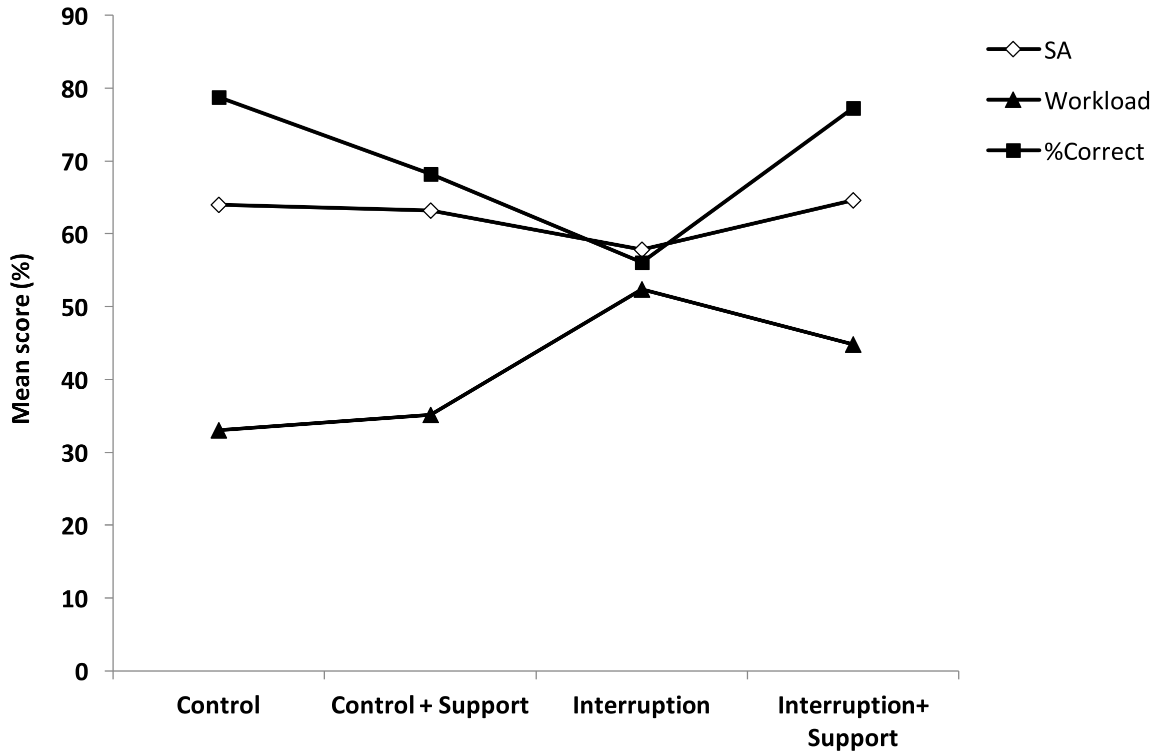


Figure 17: Interrelationships between effects of support and interruption on response accuracy, SA and workload. All scores are show in percentages, therefore the workload scores are transformed from a 0 -150 rating to a 0-100% scale.

5. Discussion

The primary goal of this experiment was to investigate the effects of a change detection support, and interruption on SA, workload, and performance in a monitoring task. We used Dynamic positioning as our HSC monitoring environment. The concept was designed according to the principles by John & Smallman (2008). The visual elevator task was used as an interruption to maximize the loss of SA in the hope that participants would engage in SAR after this task. We expected the support to speed up the process of SAR and decrease workload after the secondary task, and therefore enhance performance. The benefit of the support system in the control condition was considered to be minimal. Results showed that this was the case for the two performance measures, workload and SA percentages. However, when zooming in on the different components of SA, some surprising outcomes were found. Subjects experienced higher attentional demand when support was on in the control condition. Whereas support had a negative effect on attentional demand, the understanding of participants was better with support, together leading to no main effect of support on SA as a whole. This is the danger of considering SA as one variable, while instead it is a combination of multiple cognitive features.

The presence of a support system was especially useful after the interruption, as predicted. Subjects' SA levels increased with the presence of the support system, and they did perform better in terms of correct responses. Moreover, the SA levels and monitor accuracy were on the same level as in the control condition. Presence of the support system nullified the negative effect of the interruption on SA levels and monitoring accuracy. These results suggests that firstly, support is only useful when SA is low. This is the case when the operator leaves his monitoring task or when an operator is monitoring for long shifts and gets out of the loop. Secondly, this supports the idea of a DP environment where the operator could perform a secondary task besides his monitoring task, as the presence of a support system will help to achieve the same level of performance.

However, despite the support, subjects were still significantly slower after the interruption compared to the static monitoring situation. The process of SAR was accelerated, but not enough. The same was the case for workload. Workload levels decreased with support, but were still much higher than when subjects were only monitoring. These results show a strong relationship between SA and accuracy on one hand, and workload and response time on the other hand. Low SA is linked to more errors and high workload is linked to slower responses. This suggests that the possibility of a DPO performing secondary tasks should be treated carefully when only given a change detection support system. The solution generation process of a DPO is a matter of seconds and results show that a lot of seconds are wasted by the process of SAR after an interruption.

An explanation for the unexpected results regarding response time and workload, might be that the VET was interfering too much with the monitoring task. Due to the use of numbers and arrows, it possibly was too difficult for subjects to perform the secondary task, resulting in high workload levels. As is shown in the decision model of Endsley (1988) (Figure 2), workload influences the performance of action. Our experiment revealed that workload influences the performance in terms of response speed and that situation awareness influences performance in terms of making the right decision.

Future support concepts should adapt to the SA level of the user, hence only be switched on when SA levels are low. This could be established with real time eye tracking data. Gartenberg and colleagues (2014) have shown that an interruption resulted in increased scanning due to quicker fixations, more fixations and refixations (when a subject fixates twice on the same item in a short time period). Their theory is that subjects show this kind of scanning-pattern in service of SAR. Future research should look at eye tracking data as a measure to determine operator SA, instead of a questionnaire (which would be very impractical during a monitoring job). Ideally, when the eye tracker sees a pattern linked to loss of SA, the support could be switched on. Also, when the operator leaves his monitoring task shortly the eye tracker gets no data and as a result the support will be switched on. Furthermore, future research should focus on finding additional support concepts that focus on bringing workload down after a cognitively stressful interruption. That way operators will be able to leave their monitoring task and still retain appropriate SA and workload levels. This will keep the operators busy with other tasks so that the out of the loop problem caused by the boring monitoring job, will not be an issue any more. When eventually the adaptive automation based on SA, the workload based support, and change detection support will be a reality, the DPOs can take over other tasks on the ship, resulting in less required bridge personal, and eventually leading to lower cost. Of course the more important consequences will be fewer operator related errors, and therefore a decrease of incidents, resulting in a safer HSC environment where not only automation is understood by the operator, but also the operator is understood by the automation.

References

Endsley, M. R. (1988, May). Situation awareness global assessment technique (SAGAT). In *Aerospace and Electronics Conference, 1988. NAECON 1988., Proceedings of the IEEE 1988 National* (pp. 789-795). IEEE.

Endsley, M. R. (1995). 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., Selcon, S. J., Hardiman, T. D., & Croft, D. G. (1998, October). A comparative analysis of SAGAT and SART for evaluations of situation awareness. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 42, No. 1, pp. 82-86). SAGE Publications.

Fossen, T. I. (1994). *Guidance and control of ocean vehicles*. John Wiley & Sons Inc.

Fortmann, F., Müller, H., Brauer, D., & Boll, S. (2014, October). Supporting situation awareness with peripheral feedback on monitoring behavior. In*Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational* (pp. 895-898). ACM.

Gartenberg, D., Breslow, L., McCurry, J. M., & Trafton, J. G. (2013). Situation awareness recovery. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 0018720813506223.

Gómez, R. R., Díez, D., Díaz, P., & Aedo, I. (2013). Situation Awareness-Oriented Alarm Visualizations: A next Step in HSC Environments. In*GRAPP/IVAPP* (pp. 483-488).

Hauff, K. S. (2014). Analysis of Loss of Position Incidents for Dynamically Operated Vessels.

John, M. S., & Smallman, H. S. (2008). Staying up to speed: Four design principles for maintaining and recovering situation awareness. *Journal of cognitive engineering and decision making*, *2*(2), 118-139.

John, M. F. (2013, September). Janus Design Principles and the Acquisition Process for a Supervisory Situation Awareness Display. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 57, No. 1, pp. 2002-2006). SAGE Publications.

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.

Kleij, R., te Brake, G. M., & Broek, J. (2015). Enabling swifter operator response in the event of a fault initiation through adaptive automation. In*MTS DYNAMIC POSITIONING CONFERENCE,(DPC) October 13-14, Houston, Texas*.

Parasuraman, R., Cosenzo, K. A., & De Visser, E. (2009). Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. *Military Psychology*, *21*(2), 270.

Pfaff, M. S., Klein, G. L., Drury, J. L., Moon, S. P., Liu, Y., & Entezari, S. O. (2013). Supporting complex decision making through option awareness.*Journal of Cognitive Engineering and Decision Making*, *7*(2), 155-178.

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.

Sandhåland, H., Oltedal, H., & Eid, J. (2015). Situation awareness in bridge operations–A study of collisions between attendant vessels and offshore facilities in the North Sea. *Safety science*, *79*, 277-285.

Sheridan, T. B. (1992). *Telerobotics, automation, and human supervisory control*. MIT press.

Satuf, E. N., Kaszkurewicz, E., Schirru, R., & de Campos, M. C. M. M. (2016). Situation awareness measurement of an ecological interface designed to operator support during alarm floods. *International Journal of Industrial Ergonomics*, *53*, 179-192.

Tjallema, A., van der Nat, C., Grimmelius, H., & Stapersma, D. (2007, October). The road to eliminating operator related dynamic positioning incidents. In *Dynamic positioning conference* (Vol. 2).

Taylor, R. M. (1990). Situational Awareness Rating Technique(SART): The development of a tool for aircrew systems design. *AGARD, Situational Awareness in Aerospace Operations 17 p(SEE N 90-28972 23-53)*.

The Nautical Institute (2016). Dynamic Positioning Operator Training Scheme Accreditation Standard, revised version 1.2.

Zijlstra, F. R. H. (1993). *Efficiency in work behaviour: A design approach for modern tools*. TU Delft, Delft University of Technology.

Appendix 1: The experimental setup


smarteye

Appendix 2 Situation Awareness Rating Technique

The original concepts of SAR in English, for the experiment translated in to Dutch questions.

1 Instability of situation

Hoe veranderlijk vond je het scenario? Was deze onstabiel en kon deze snel veranderen (Hoog) of was het erg stabiel en eenvoudig (laag)?

Laag 1 2 3 4 5 6 7 Hoog

2 *Variability of situation*

Waren er veel variabelen die veranderden dit scenario (hoog) of waren er weinig variabelen die veranderen (laag)?

Laag 1 2 3 4 5 6 7 Hoog

3 C*omplexity of situation*

Hoe ingewikkeld was het scenario? Is deze complex met veel samenhangende componenten (hoog) of is het simpel en eenvoudig (laag)?

Laag 1 2 3 4 5 6 7 Hoog

4 *Arousal*

Mate van alertheid. Was je alert en klaar voor de taak (hoog) of was je helemaal niet alert (laag)?

Laag 1 2 3 4 5 6 7 Hoog

5 *Spare mental capacity*

Hoe veel mentale capaciteit had je in dit scenario? Had je genoeg capaciteit om alle variabelen in de gaten te blijven houden (hoog) of had je te weinig om op alle variabelen te blijven letten (laag)?

Laag 1 2 3 4 5 6 7 Hoog

6 C*oncentration of attention*

Hoe goed kon je je concentreren tijdens het scenario? Kon je je goed concentreren (hoog) of was je concentratie slecht (laag)?

Laag 1 2 3 4 5 6 7 Hoog

7 *Division of attention*

Hoe goed kon je je aandacht verdelen? Kon je je focussen op veel verschillende variabelen of maar op een of twee?

Laag 1 2 3 4 5 6 7 Hoog

8 Information quantity

Hoe beoordeel je de hoeveelheid informatie die je kreeg om dit scenario goed te beantwoorden? Slechte hoeveelheid (Laag) of een goede hoeveelheid (Hoog)

Laag 1 2 3 4 5 6 7 Hoog

9 *Information quality*

Hoe goed vond je de gegeven informatie? Erg duidelijk (hoog) of niet duidelijk (laag)?

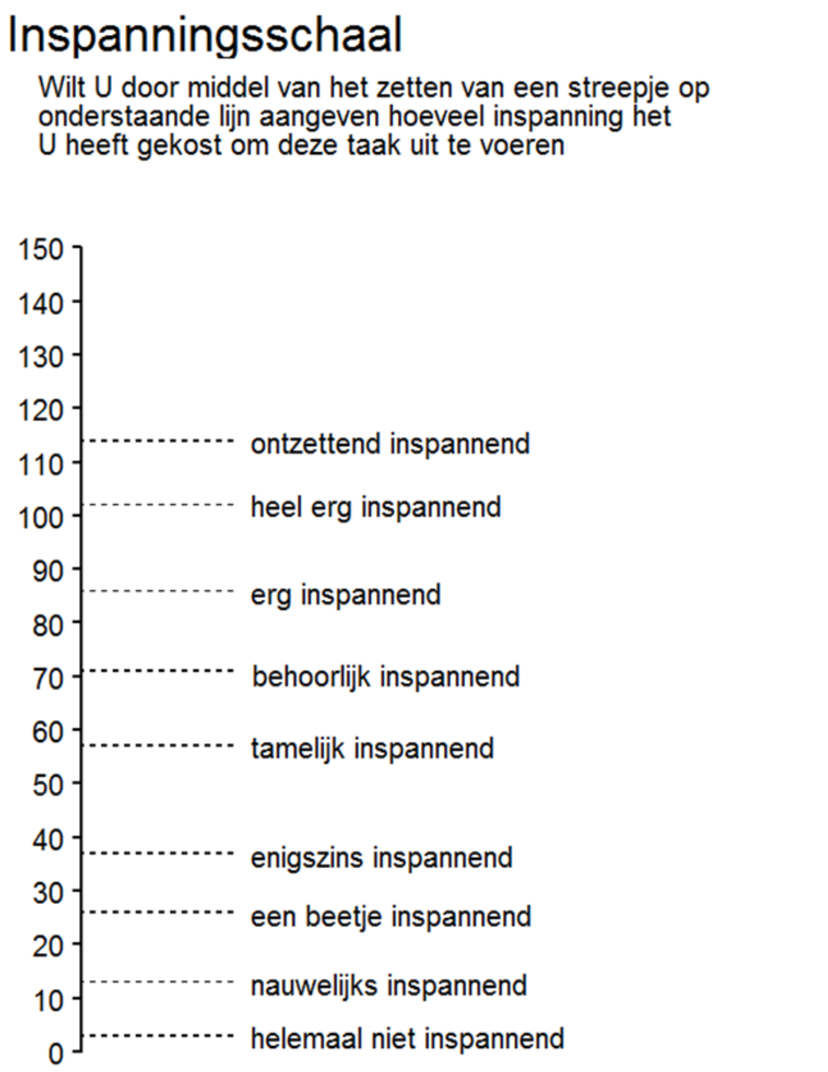
Laag 1 2 3 4 5 6 7 Hoog

10 *Familiarity*

Hoe bekend was je met de situatie die ontstond in dit scenario, de bepaalde variabelen die gingen stijgen etc.? Had je er al genoeg ervaring mee (hoog) of was het compleet nieuw voor je (laag)?

Laag 1 2 3 4 5 6 7 Hoog

Appendix 3: The Rating Scale Mental Effort.



The Rating Scale Mental Effort. Participants had to draw a line on the scale to indicate their workload in the concerned condition.