Effects of coherent, integrated, and context-dependent adaptable user interfaces on operators' situation awareness, performance, and workload

van Doorn, Ellemieke; Horváth, Imre; Rusák, Zoltán

DOI
10.1007/s10111-020-00642-z

Publication date
2020

Document Version
Accepted author manuscript

Published in
Cognition, Technology and Work

Cogn Tech Work (year) Vol:pp-pp
DOI xx.xxxx.sxxxxx-xxx-xxxx-x

Effects of coherent, integrated, and context-dependent adaptable user interfaces on operators' situation awareness, performance and workload

Ellemieke van Doorn • Imre Horváth • Zoltán Rusák

Abstract

Keywords User interface Situation awareness Task performance Workload

1 Introduction

Nautical traffic management in the Netherlands is shifting from local traffic control to corridor traffic management (Van Doorn et al. 2017b). Corridor traffic management operators, called nautical operational network management (N-ONM) operators, remotely manage a traffic corridor such as the main route and alternative routes between Port of Rotterdam and Germany. They need to gain and maintain situation awareness (SA) based on large amounts of information about the corridor. Previous work, however, showed that current traffic management information systems do not sufficiently support users in perceptual and cognitive processes to interpret and understand the presented information (Van Doorn et al. 2015; 2017b).

In our previous work we designed and developed three user interface (UI) concepts to overcome deficiencies of current traffic management systems in order to increase operators' SA and to improve operators' task performance (Van Doorn et al. 2017a). The three concepts built upon each other. As a first concept a coherent UI was developed, in which the UI is a logical, consistent, orderly, and harmonious interface, where the multiple UI windows form a coherent whole. The second concept, an integrated UI, furthermore uses information fusion, clustering, and interaction between UI windows. The third concept, a context-dependent adaptable UI, additionally captures context information, assesses the implications of context, and accordingly adapts the interface content and composition. The context-dependent adaptable user interface thus includes all features of the other two concepts and as such is the most elaborated UI of the three. Although this UI concept is more difficult and expensive to develop and maintain compared to the other two, the assumption in our previous study was that this user interface would also provide significantly better user support.

Usability testing was used to evaluate the three UI concepts (Van Doorn et al. 2017a). 20 traffic management operators executed N-ONM tasks in a one-hour incident management scenario in a simulator environment for each UI concept. Afterwards, the three UI concepts were evaluated through structured interviews. That study showed that all operators considered it useful, or even very useful, to redesign the current UI of the system towards a more coherent one, in contrast with the current N-ONM UI's. This applies to each of the three tested concepts. At comparing the users' satisfaction concerning the three user interface concept, none of the operators replied that they preferred the user interface that was coherent UI compared to the other two concepts. Contrary to what was expected, operators did not report a significant difference in level of support between the integrated UI and context-dependent adaptable UI. Two operators indicated that the integrated UI provided the best support of N-ONM tasks. Seven operators regarded the integrated UI and context-dependent adaptable UI interface as equally suitable. Seven operators replied that the context-dependent adaptable UI provided the best support of N-ONM tasks. This raised the question *"whether using a context-dependent adaptable UI instead of an integrated UI will improve the SA of the operators to such extent that it warrants the consideration of the first one despite of the higher efforts and overheads of implementation."* (Van Doorn et al. 2017a).

Besides the practical question discussed above, a more fundamental question evolved from the results of the previous study; is usability testing as a method to estimate subjective satisfaction with UI concepts a suitable method to evaluate the effects of UI features?. Estimating subjective satisfaction with products is a widely used technique in usability testing (Wichansky, 2010). On the other hand, several studies have reported that subjective evaluation of UI concepts by operators is not in line with objective measurements of SA and/or task performance of operators (e.g. Ben-Bassat et al. 2006, Bowden and Rusnock 2015, Kissel 1995,

Acknowledgement This research has been supported by Rijkswaterstaat, the Executive Body of the Dutch Ministry of Infrastructure and Watermanagement.

E.C. van Doorn[1] (■) I. Horváth[2], Z. Rusák[2]
[1] Rijkswaterstaat, Ministry of Infrastructure and Watermanagement, the Netherlands
[2] Faculty of Industrial Design Engineering, Delft University of Technology, Delft, The Netherlands
E-Mail: Ellemieke.van.Doorn@rws.nl

Stuut et al. 2019, Zuckerman and Gal-Oz 2013). Similar, previous literature stated that it is questionable whether operators can accurately rate SA (Bell & Lyon, 2000; Salmon et al., 2006). It is argued that user preference is more related to enjoyment in using a specific design (Zuckerman and Gal-Oz 2013) or design aesthetics (Sonderegger and Sauer 2010) instead of product usability in terms of task performance. In our case, however, the three UI concepts did not differ significantly in terms of design aesthetics. The UI concepts were all based on the same UI design, only extra features were added to create an integrated or context-dependent adaptable UI design. Consequently, our approach has been conceptualized to make it possible to evaluate whether subjective usability measures are suitable to evaluate the effectiveness of UI features in cases that differences between UI design are not related to design aesthetics.

The study presented in this paper thus has a dual objective. Firstly, we aim to objectively evaluate the effect of the three previously developed UI concepts on operators' SA, task performance, and workload in order to advice upon which UI concept to use for N-ONM tasks. Secondly, this study aims to contribute to fundamental understanding of the usefulness of subjective usability testing when evaluating the effect of UI features. This first requires to evaluate whether subjective measures are in line with objective measures. If differences are found, we aim to understanding these differences by addressing them form the perspective of the intricate relations between SA, task performance, and workload in relation to user interface design, For these purposes we used the same experiment as the previous study in terms of research set-up, scenarios, and participants. But different data, which was gathered but not analysed earlier. Instead of subjective measures obtained through structured interviews, this study uses objective performance measures in a within-subject design. Data logged by the simulator system was combined with SA assessment according to the Situation Awareness Global Assessment Technique (SAGAT) developed by (Endsley 2000). Raw NASA Task Load Index (RTLX) (Hart, 2006) was used to evaluated operators' workload.

In order to structure the reasoning about differences between subjective measures and objective measures, Section 2 provides a brief overview of current knowledge on the intricate relations between SA, task performance, and workload in relation to user interface design. In Section 3 we present the UI concepts that are evaluated in this study and how they were implemented in a N-ONM workplace simulator. Section 4 describes the methodology that is used in order to test the effect of the UI concepts on operators' SA, task performance, and workload. The results section gives an overview of the outcomes from the statistical analyses. Section 7 concludes which implication this study has on selecting a suitable UI design for N-ONM tasks. In line with our dual objectives, it also presents the relevance of the outcomes of this research to the Human Factors and UI Design community.

2 Relations between situation awareness, task performance, and workload

The concepts of SA, task performance, and workload are intricately intertwined. SA knowledge consists of the combination of perceptual knowledge (factual knowledge of elements in the current situation), comprehended knowledge (understanding of the meaning and relationships of knowledge in the current situation) and projected knowledge (insight in future activities of the elements in the environment) (Van Doorn et al. 2014). Sufficiently correct and complete SA knowledge is required for correct decision making and thus low situation awareness can have a negative effect on operators' task performance (Endsley 1995). While SA could be improved by working harder, high mental workload can negatively affect operators' SA (Endsley 1995, Vidulich and Tsang 2012). A low cognitive task load on the other hand can result in boredom and under-load, which also may negatively affect SA assessment and task performance (Edwards et al. 2017).

SA knowledge is the understanding of dynamic information associated with operator' goals and does not include more static knowledge stored in long-term memory (Endsley 2000). Command and control operators commonly require understanding of the current and prospective meaning and relationships of large amounts of information about their dynamic environment. Due to the need to have such a large amount of information mentally available, the main challenge of operators in gaining and maintaining SA is their ability to locate and process such information (Endsley 2000). As such, SA and mental workload make use of the same cognitive processes, which capacities are limited (Vidulich and Tsang 2015). A higher level of workload means that more attention is needed for performing tasks and less is left for maintaining SA.

The relation between SA and workload is especially relevant in cases of mental overload. It is argued that stress reduces working memory capacity and retrieval of information, and that overload thus negatively influences SA assessment (Endsley 1995). Under stress operators tend to focus their attention on a limited number of dominant pieces of information. They tend to (i) arrive at a decision without exploring all available information, (ii) put more attention to negative information, and (iii) have a more scattered and poorly organized scanning of stimuli (Endsley 1995). While focusing on negative information can be a positive strategy, as negative information is a cause of problems to be solved, this may also result operators missing other relevant information.

Which information operators focus on, however, also depends on how the information is presented to the user (Treisman 1985, Wolfe and Horowitz 2004). And thus, to which degree the working memory decrements affect SA also depend on the UI design of the systems used. At the same time, system design also influences operators' required SA to achieve operators' goals (Van Doorn et al. 2017b). Consider a 'System A' which simply displays long lists with all available data, and a 'System B' that processes data and only presents a small overview of derivative information. With 'System A' operators might need to use memory-based information processing strategies for timely task performance. While with 'System B' the cost of accessing information is much lower. Therefore, operators could use display-based information processing strategies with 'System B'. The information contained and presented by the technical system thus influences which information operators need to have mentally available (Mogford 1997, Patrick and Morgan 2010, Stanton et al. 2006; 2010). With 'System B' less information might be part of operators' SA, but this does not mean that operators have better SA with 'System A'. It is the men-machine interaction that determines which information is essential for operators' SA (Van Doorn et al. 2017b). This makes it difficult to predict or understand the effect of automation, such as implemented in the context-dependent adaptable UI, on operator's performance (Edwards et al. 2017, Parasuraman and Riley 1997).

In our study the three UI concepts differ in terms of how operators can access information, and additional information is displayed in the integrated and context-dependent adaptable UI. Thus the information processing strategies applied by the operators might differ among the UI's. Considering the above, different information might be part of operators' required SA. This makes it less straight forward to evaluate the effect of the different UI concepts. While the UI concepts were designed to better support operators SA accessment, more information as part of SA might not result in better task performance. Instead, combined measures of SA, task performance, and workload are required to understand the effect of UI concepts and thus to evaluate the added value of objective measures in evaluating UI concepts.

3 The tested user interface concepts

In our previous work we presented three UI concepts to better support N-ONM SA and task performance; (i) a coherent UI, (ii) an integrated UI, and (iii) a context-dependent adaptable user interface (Van Doorn et al. 2017a). The coherent UI concept is the concept that is closest to current practice. In current practice, however, there is not yet a uniform UI concept for N-ONM operators. Each traffic management control room uses different workplaces and partly different information systems. In this study we aim to measure the differences between the concepts. For this purpose it's important that all other variables, such as the workplace design and design aesthetics, are the same in our research set-up. The only difference between the UI's to be compared needs to be the features of the concept. It, however, was impossible to present a current N-ONM UI on the same hardware set-up and with the same design aesthetics as the three newly developed UI concepts. Besides, features such as using the same UI interactions and style guides everywhere in the UI is part of what distinguishes a coherent UI from current N-ONM UI's. We therefore did not include a current UI in our comparison. Instead, the coherent UI is used as best practice for current system set-up, developed in a way that we can measure the effect of the other concepts. The other two concepts were based on this coherent UI, with extra features implemented to overcome identified deficiencies. See Table 1. Paragraph 3.1 provides an overview of the coherent UI, which also formed the basis of the two other UI concepts. Paragraph 3.2 describes the features implemented to create an integrated UI. The features implemented to create a context-dependent adaptable user interface are explained in paragraph 3.3.

3.1 Coherent user interface

The implemented coherent UI consists of six user interface windows, which were coherent in use of colors, buttons, and menu structures, as well as consistency in interactions. For example, double mouse click in all windows is used to display more information about the clicked information element. The different windows all use the same data source in case they present the same information. For example, all windows use the same vessel information database. If this information is adapted by the user in one window, then this information also changes in the other windows. In terms of features, the coherency of the UI is summarized as feature 1;

1. *UI windows together form a coherent whole (logical, consistent, orderly, and harmonious).* The total coherent UI concept consists of the following windows (See Figure 1):

1. Area of Focus window with static geographic information.
2. Area of Focus window with dynamic vessel traffic information
3. Information overview window, which lists all available information elements. This window contains a tab per information cluster (vessel traffic information, nautical object information, event information, hydro-meteo information, etc.).
4. Information detail window, which provides an overview of all detailed information about one object of interest (one vessel, one nautical

Table 1 Overview of which feature is implemented in which UI concept

Feature	**Coherent UI**	**Integrated UI**	**Context-dependent adaptable UI**
1. UI windows together form a coherent whole (logical, consistent, orderly, and harmonious)	X	X	X
2. Present all geographic information that is needed for the same task(s) in the same map		X	X
3. Support filtering of vessel information by human operators		X	X
4. Interactions between UI windows and visualized relations between windows / elements		X	X
5. Context-dependently show relevant location in extra Area of Focus map			X
6. Automatically show available alternative routes in case of obstruction on main route			X
7. Context-dependently show traffic prognoses information if traffic intensity exceeds limit			X

object, one event, one hydro-meteo location, etc.).

5. Area of Control window which displays the entire area under control of the operator.
6. Notices-window, which displays the top priority notices relevant for the specific operator role.

3.2 Integrated user interface

The implemented integrated UI is a coherent UI with three extra features, see also Table 1 and Figure 2.

2. *Present all geographic information that is needed for the same task(s) in the same map:* Two different types of N-ONM tasks require geographic information presentation. One set of tasks is related to handling local traffic management events, such as incidents. This requires detailed information about the area of focus. Another set of tasks is related to corridor management, which requires an overview of the entire corridor, or area of control. For both set of tasks, operators need both static geographic information and dynamic vessel and event information. The integrated UI therefore consists of two windows containing a geographical information system (GIS), instead of the three that are present in the coherent UI. The Area of Focus displays detailed information about both static and dynamic elements, such as anchorage type, vessel course, and event type. The Area of Control window only displays location information of most elements. Only of events, the Area of Control window also displays detailed information, as this information is needed for both sets of tasks. This in contrast to the coherent UI, were event information was only displayed in the 'Notices-window', 'Information overview window', and 'Information detail window'.

3. *Support filtering of vessel information by human operators:* Operators are able to filter the information that is displayed in the vessel information overview window by selecting a location in the corridor. Only vessels that will pass this location then will be displayed, in order of projected arrival at this location. The selected location is also displayed in both the Area of Control and Area of Focus map.

4. *Interactions between UI windows and visualized relations between windows and elements:* The following interactions between UI windows are implemented: (i) Highlight location of object (vessel, lock, event, hydro-meteo station, etc.) on both maps by clicking on this object in the information overview window. (ii) Open detail window of object (vessel, lock, event, hydro-meteo station, etc.) by double clicking this object on the map. (iii) Click object in Area of Focus window to highlight location of this object in the Area of Control Window. (iv) A blue rectangle in the Area of Control Window visualizes which area (location) is shown in the Area of Focus Window. (v) Click notification window to open the notification information overview window. (vi) Double click a notification in the notification window to open the detail window of this notification. (vii) Type, status, and location of events / notifications also visible on both maps.

3.3 Context-dependent adaptable user interface

The implemented context-dependent adaptable UI is an integrated UI with three extra features, see also Table 1 and Figure 3.

5. *Context-dependently show relevant location in extra Area of Focus map:* The system assesses context to automatically display an extra Area of Focus window. The coordinates of the center of the map visualized in this window are the coordinates of the event. If there are multiple events, then the event type (priority) and event start time determine which coordinates are taken as the center of the map. In our research set-up, this feature was implemented as a 'Wizard of OZ' method (Green and Wei-Haas 1985). This means that the participants believed that it was the system who opened this window, but actually it was the test leader that opened the extra Area of

Focus window. Since the events were part of the script, and not initiated by the participants, the test leader could do so without the need to understand user' actions.
6. *Automatically show available alternative routes in case of obstruction on main route:* If there is no obstruction of the main route, then all waterways are visualized in the color blue. If there is an obstruction on the main route, then the main route is visualized in grey blue and if at that moment there is no obstruction on an alternative route, then this alternative route is visualized in violet. Since the participants influenced which routes were available, we programmed the system to automatically carry out this feature. Thus this feature did not depend on accurate understanding of the test leader of participants' actions.
7. *Context-dependently show traffic prognoses information if traffic intensity exceeds limit:* The system constantly showed a simple bar with prognoses information below the Area of Control map. Only in cases of traffic intensities that would hinder traffic flows, the system automatically also displayed more detailed prognosis information as a layer on top of the waterways in the Area of Control window. Since the participants influenced prognosis information and prognosis information was complex to calculate, we programmed the system to automatically carry out this feature. Thus this feature did not depend on accurate human calculations and understanding of the test leader of participants' actions.

4 METHOD

The methods used in this study need to support our dual objective. Firstly, we aim to validate the results of previously conducted subjective usability testing. The usability testing resulted in a ranking of which UI concept operators preferred most. To validate this subjective ranking, this study needs to provide an objective ranking of the effects of the UI concepts to be compared with the subjective ranking.

Secondly, our aim is to evaluate whether the costs of implementation of the different UI concepts is worth the effort. This means that methods to evaluate the UI concepts need to quantify the effects in terms that are meaningful for practice. The research set-up therefore needs to be realistic. It is not useful to test the effects of the UIs in extreme situations that will never occur in practice. Besides, the values to be measured need to be meaningful. For traffic management operation it is relevant to evaluate task performance in terms of speed, accuracy, and order. Effects not only need to be statistically significant, but also meaningful in practice. For example, a difference in speed of task performance of a second is not meaningful, while a difference of minutes is very relevant in incident situations.

Only feature 1, the coherency of information content and presentation in all UI windows, was specifically designed to better support operators' Level 1 SA. This feature was present in all UI concepts. Consequently, no differences between operators' Level 1 SA was expected. Accessing information, however, differed between the coherent UI and the other two interfaces. This could influence information processing strategies and therefore SA knowledge. Measures were needed to evaluate whether this affected operators Level 1 SA. We assumed that features 2, 3, 4, 5, and 6 would support operators in gainng Level 2 SA, as these features visualized relation between information elements. These features were not implemented in the coherent UI. Features 5 and 6 were only implemented in the context-dependent adaptable UI. Level 1 and Level 2 SA is required for gaining Level 3 SA. Additionally, we assumed that feature 7 that showed prognosis information, only implemented in the context-dependent adaptable UI, would support gaining Level 3 SA.

We assumed that the features 2, 3 and 4 of an integrated UI, which were implemented in the integrated UI and context-dependent adaptable UI, would help operators to more quickly access information that is required to gain SA about an incident situation. We assumed that increase of speed of gaining SA would also result in quicker task performance and more support for gaining SA would result in more accurate task performance. In order to understand the effects of the UIs on operators' SA and task performance, it is nessecary to reflect on the intricate relations between workload, SA and task performance.

4.1 Participants

Twenty traffic management operators were randomly selected to participate in the experiment. Data from one participant was not available because of errors made by the test leader. Four subject-matter experts (SMEs) were involved as test leader, responsible for imitating communication using scripts. One SME, however, had only limited training prior to the experiments and only participated once. Data from this experiment was also excluded. Data from four participants only included test results for experiments with the coherent and integrated UI, because of bugs in the simulator system. Counter balancing required six orders of treatment, therefore the number of participants had to be a multiple of six. Consequently, the effect of the context-dependent adaptable UI was evaluated using a dataset of twelve operators, while the difference between the coherent UI and integrated UI was evaluated using a dataset of eighteen operators.

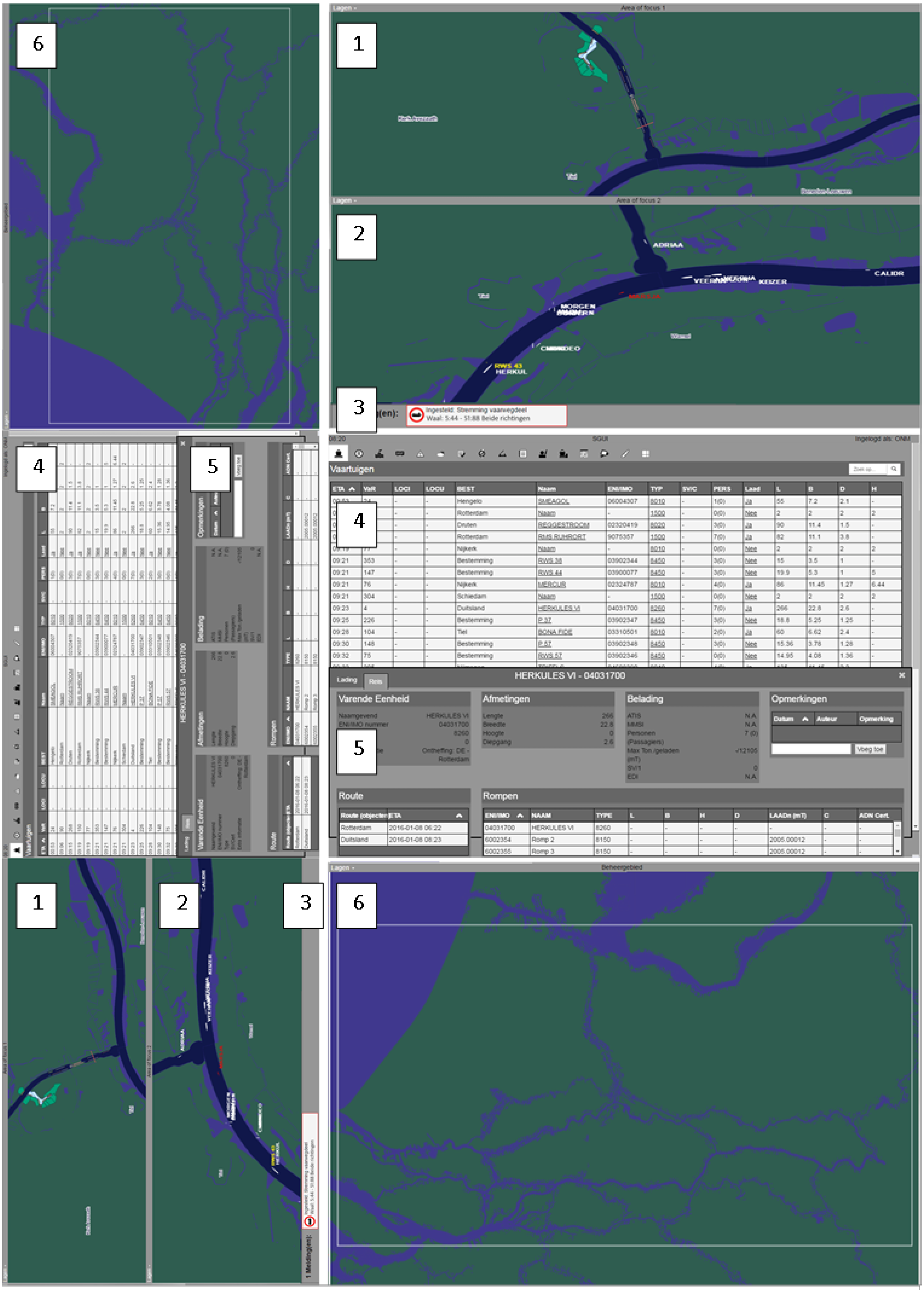

Figure 1: Screenshot of Coherent UI (left) and windows enlarged (right). The different windows are numbered in line with paragraph 3.1

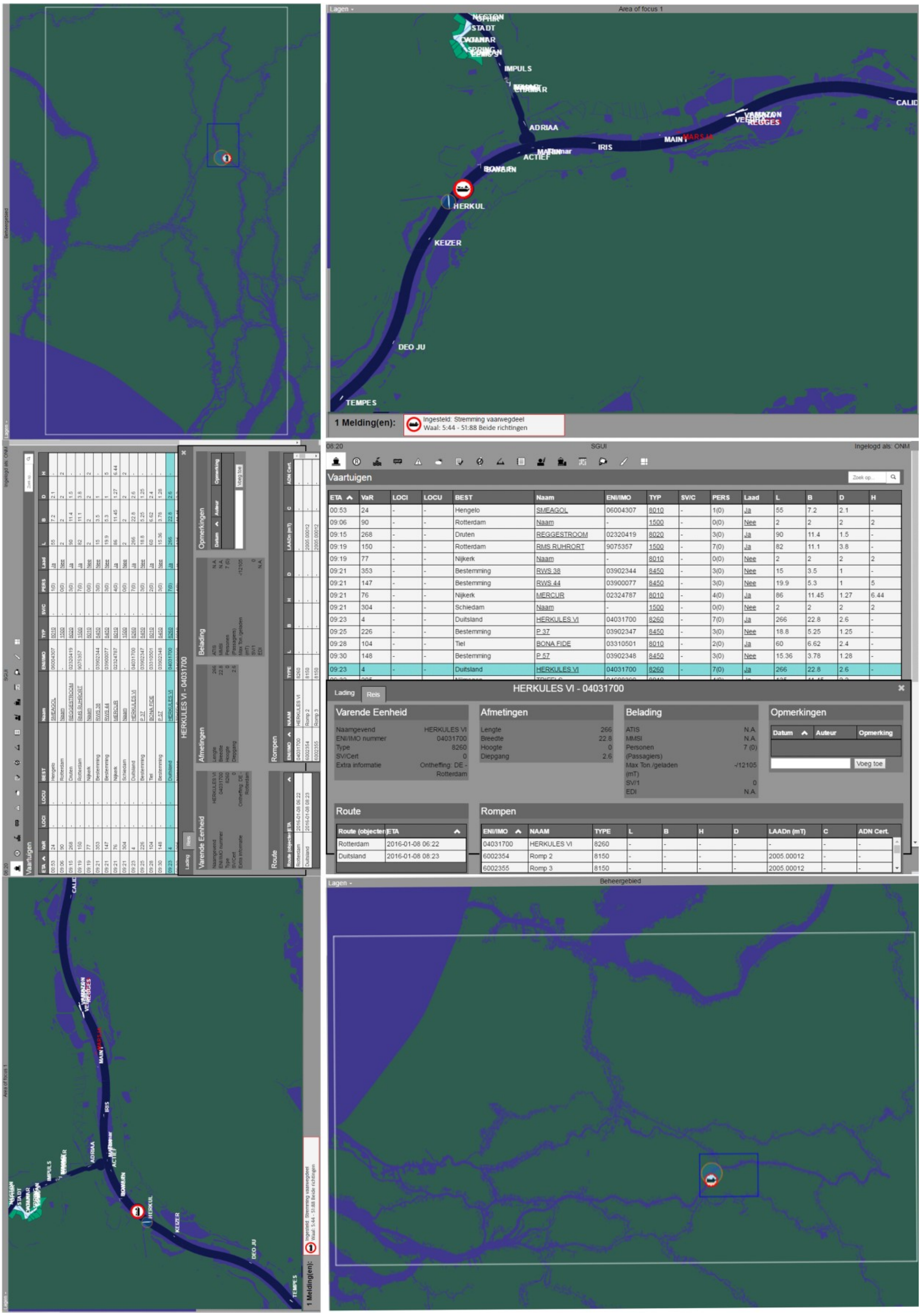

Figure 2: Screenshot of Integrated UI (left) and windows with implemented features enlarged (right).

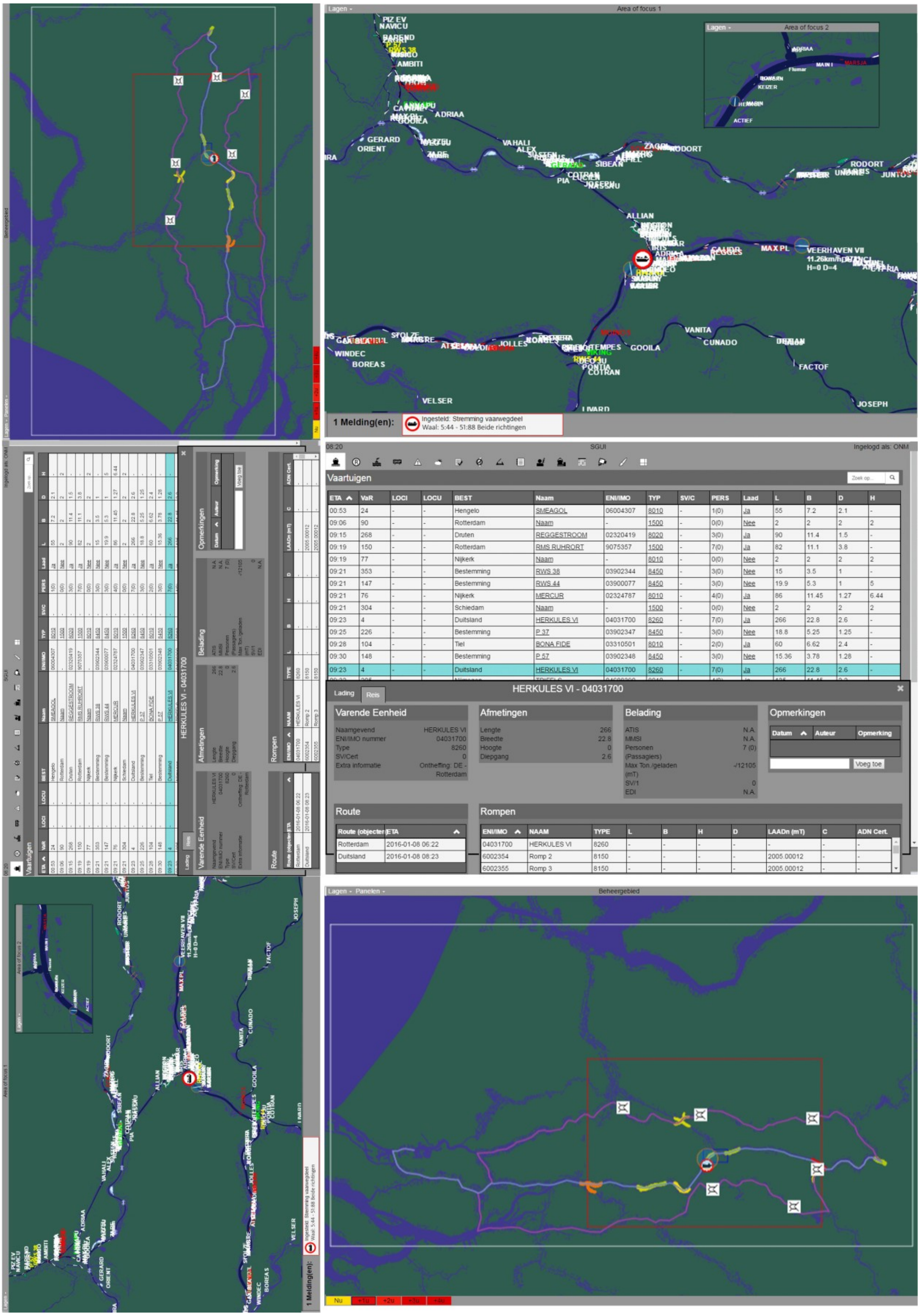

Figure 3: Screenshot of Context-dependent adaptable UI (left) and windows with implemented features enlarged (right).

The majority of the involved operators were highly experienced and had prior experience as steersman and/or skipper, see Table 2. This is consistent with the entire population of N-ONM operators working in the Netherlands.

Table 2: Participants' experience

	Years of traffic management experience			
	0 - 3	3 - 6	6 - 9	> 9
Frequency	1	1	3	13

	Experience as steersman / skipper	
	No	Yes
Frequency	2	16

4.2 Test environment and scenarios

The research set-up, as presented in (Van Doorn et al. 2017a) used a nautical traffic management workplace simulator, which consisted of an operator desk, a test leader desk, and an observant desk, see Figure 4. The simulator software logged all operator's actions. Communication was logged by both the test leader and an observant. They could log foreseen communication by clicking items in a script. Unforeseen communication was logged as typed text. Three realistic challenging traffic management scenarios were developed together with four SMEs. We instructed the SMEs to aim for highly similar scenarios in terms of structure, duration, traffic intensity and level of difficulty. The content, however, differed, see Table 3. Each scenario included communication to handle the events which were part of that scenario. Communication was imitated by a SME using scripts. Additionally, each scenario included communication scripts for the test leaders to initiate questions from skippers who were not involved in those events. The simulator software controlled when which script needed to be activated.

4.3 Procedure

Prior to the experiments, participants were sent a description of the research background, including which tasks were part of the experiment. This information was repeated at the beginning of the experiments. Participants read and signed the informed consent form and filled in a survey about their work experience. After that, the three UIs were explained to the participants and they completed a 10-minute tutorial scenario for each UI. The UIs were referred to as UI1 (coherent UI), UI2 (integrated UI) and UI3 (context-dependent adaptable UI). The participants then performed the N-ONM tasks in the three traffic management scenarios, in a counterbalanced manner. Counterbalancing was used both to ensure that each scenario was equally often played with each UI concepts, and that the order of UI use was evenly distributed among the experiments. Each scenario took approximately one hour and was followed by a short break.

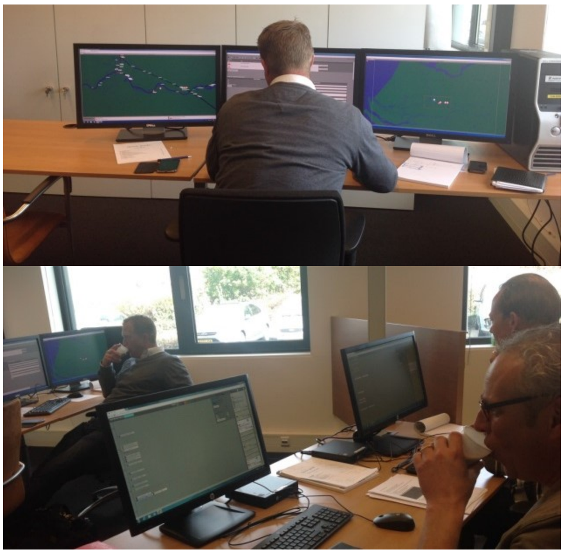

Figure 4: Experiment set-up with N-ONM workplace simulator (at the top image and left side of lower image) and test leader and observant desk (at the bottom picture on the right side)

Table 3: Scenarios

Scenario A – Collision near Houten	Time
UI displays information about planned blockage of Lock Beatrix starting a 18:00 hours	16:00
Phone call about malfunction of Lock Bernardsluis	16:04
Freeze 1: SAGAT + RTLX	16:08
VHF communication about collision between Calidris and Fueltrans	16:11
Freeze 2: SAGAT + RTLX	16:21
Freeze 3: RTLX	end

Scenario B – Fire near Culemborg	Time
UI displays information about planned blockage of Lock Hagestein starting a 17:00 hours	16:00
VHF communication about fire on board of Presco	16:04
Freeze 1: SAGAT + RTLX	16:16
Phone call about malfunction of Lock Prinses Irenesluis	16:21
Freeze 2: SAGAT +RTLX	16:34
Freeze 3: RTLX	end

Scenario C – Vessel aground at Waal	Time
UI displays information about anchorage Ravenswaaij not available	07:00
VHF communication about Hercules VI run aground	07:05
Freeze 1: SAGAT + RTLX Time jump communicated (to 09:30)	07:26
VHF communication about Hercules VI loose, release waterway	09:38
Freeze 2: SAGAT + RTLX	09:42
Freeze 3: RTLX	09:49

4.4 Measurements

As explained in Section 2, we needed to measure operators' SA, task performance, and workload in order to understand the differences between the three UI concepts. Measures were based upon the assumptions about the effects of the different UI concepts. For SA we looked both at what information was part of operators'SA knowledge during freeze probes, as at the speed of gaining SA. For task performance we looked at speed of task performance and accuracy of task performance. For workload we measured subjective workload.

The Situation Awareness Global Assessment Technique (SAGAT) developed by (Endsley, 2000) was used to measure the quality of operators' SA at two moments (see Table 3) in each scenario. The SAGAT included queries about perception of data (Level 1 SA), comprehension of meaning (Level 2 SA) and projection of the near future (Level 3 SA). These queries were developed together with the four SMEs after analysis of operators' required SA. The SMEs unanimously agreed on the desired answers for each freeze. Table 4 lists the SA queries that were used.

Operators required to perform two actions in their UI to gain SA in cases of an incident with one or more vessels involved: they needed to search for the vessel(s) on their Area of Focus map and they had to open the vessels' detail information window. The data logged by the simulator system was used to calculate how quickly operators carried out these actions. An operator was assumed to have identified the vessel involved in an incident when the operator for at least four seconds did not adapt the location and/or zoom level of the Area of Focus window while the vessels location was displayed in the Area of Focus window with a zoom level that allowed to read the vessels' name.

For each scenario we defined which actions were required for accurate task performance. The SME, observer, and simulator software logged execution of required actions and the speed of actions. These actions included communication with stakeholders, such as skippers, emergency services, and the officer of duty. Additionally, operators needed to send notices to skippers through VHF radio, and they had to activate and release traffic measures using their computer system. In each scenario operators additionally had to answer questions of skippers who were not involved in the incident. In each scenario an equal amount of questions required perception of data (related to Level 1 SA), comprehension of meaning (related to Level 2 SA), or projection of the near future (related to Level 3 SA).

Table 4: SAGAT queries used (Original in Dutch)

1	Click on the map to enter the location of all current events. Provide a short description for each event.
2	Which of the following vessel types are the vessel type of vessels involved in an incident?
3	Which of the following names are the names of vessels involved in an incident?
4	Which of the following cargoes are the cargo of the vessels involved in an incident?
5	Which of the following names are of vessels with wounded persons on board?
6	Which of the following names are of vessels leaking fuel or cargo or that make water?
7	Which of the following locks are currently obstructed, or to a limited degree available
8	For which of the following locks do skippers over an hour need to take into account that there will be extra crowds and possible longer delays as a result of blockages or restrictions elsewhere on the waterway?
9	Which of the following service vessels is currently the closest to an incident?
10	How long will it take for the closest service vessel to be on site of the incident?
11	Which of the following vessels need to take into account that there are obstructions on their current route?
12	Which of the following restrictions apply to a motor cargo (length 85.00 m, width 9.60 m, height 7.90 m, depth 1.30 m) that is currently at lock Weurt, when she wants to arrive at the Port of Amsterdam as quickly as possible?
13	Which of the following routes is best advised to a motor tanker (no cones, length 109.00 m., width 11.40 m., height 6.00 m., depth 2.25 m.) which plans to depart in one hour from the Port of Rotterdam towards Enschede?
14	Which of the following routes is best advised to a container vessel (length 135.00 m., width 17.40 m., height 10.30 m., depth 2.10 m.) that plans to depart in one hour from the Port of Rotterdam towards Duisburg in Germany?

In our analysis we evaluated if there is a significant relation between the used UI and the (i) execution of required actions, (ii) the speed of executing required actions, (iii) the accuracy of executed required actions, and (iv) the order in which required actions were executed.

RTLX (Hart, 2006) was used to measure subjective workload at three moments in each scenario, see Table 3. This method requires participants to respond to six questions about their workload. Since our operators were all native Dutch speakers, we translated the questions to Dutch. We compared the observed workload scores with ranges and percentile ranks found in similar studies. Because no scores of other studies concerning nautical traffic management have been found, we compared our scores to the scores reported by Grier (2015). Her analysis of 1173 reported workload scores in 237 publications showed that 80% of the reported scores are between 26.08 and 68.00. Of those task environments which were taken into consideration by Grier, process control is most relevant when comparing our workload scores. For 38 process control test cases, the reported percentile ranks were: 25th: 31.91, median: 42.00, and 75th: 51.83 (Grier, 2015).

4.5 Data analysis

For data analysis, we firstly evaluated whether (and how big) there was an effect of the UIs on operators' SA, task performance, and workload. To answer this question, the effect size was calculated. Data were analyzed by using a within-subject design. The data could not be considered normally distributed due to the relative small sample size. Two within-subject tests are commonly used for testing differences between conditions in human factors research if the assumption of normally distributed data is violated; Friedman's ANOVA is used for more than two categories and Wilcoxon Signed Ranks Test is used for two categories (Willages 2007; Field 2009). Friedman's ANOVA only shows whether there is difference between the tested conditions, but does not shows where this difference occurs. For that purpose, a post hoc analysis is required. Wilcoxon Signed Ranks Test is commonly used as post hoc analysis for Friedman's ANOVA.

Since our sample size (n = 18) is large relative to the population (N = 60), it is needed to apply a correction to the formulas used to compute standard error (SE). This correction is called the finite population correction (FPC), which is calculated by $FPC = \sqrt{(N-n)/(N-1)}$ (Ramachandran and Tsokos 2009). The standard error must be corrected by multiplying it with FPC. To calculate the significant of the test statistic (T), Wilcoxon Signed Rank Test looks at the mean ($\bar{T}$) and standard error ($SE\bar{T}$) by the formula $Z = (T-\bar{T})/SE\bar{T}$ (Field 2009). To apply FPC in case of Wilcoxon Signed Ranks Test therefore means that the test statistic Z needs to be divided by FPC. The formula used to calculate Friedman's ANOVA test statistic does not include standard error. Consequently, it is not possible to correct Friedman's test statistic with FPC. Therefore, we used Wilcoxon Signed Ranks Tests only in our analysis.

Pearson's correlation $r = Z/\sqrt{N}$ is commonly used as an effect size for Wilcoxon Signed Ranks Tests. Here Z is the test statistics as defined by the formula above and N is the number of observation. Cohen (1988; 1992) gives guidelines for evaluating effects sizes for Wilcoxon Signed Ranks Test; An effect size between 0.10 and 0.29 is considered a small effect. An effect size between 0.30 and 0.49 is considered to be a medium effect. An effect size of 0.50 or more is considered a large effect.

If an effect was found, then the second question was: how likely is it that there is a true effect in the entire population of N-ONM operators? In line with common practice we considered the effect statistically significant if $p \leq 0.05$. In cases where an effect is found, but this effect cannot be considered significant, than we cannot be sure at the 95% level that what we see is not due to a random fluctuation. It can be that there indeed is an effect, but than our sample was too small for statistically significant results.

5 RESULTS

5.1 Speed of gaining situation awareness

Speed of gaining SA was assumed to be influenced by features 2, 3, and 4. These were features of an integrated UI, and thus not implemented in the coherent UI. With an integrated UI (UI 2) operators were significantly quicker in identifying the involved vessels in the Area of Focus Window than when using a coherent UI (UI 1) (effect size = -0.34 and p = 0.04). The difference in speed is not only statistically significant, but with a difference of up to minutes, also significant in terms of relevance for N-ONM tasks, see Table 5. The same effect was expected when comparing the coherent UI (UI1) with the context-dependent adaptable UI. Our results, however, do not show a significant effect when comparing these interfaces. This could be due to the small sample size (n=12).

Table 5: Descriptive statistics of speed of opening an incident in the area of focus window (in sec.) with UI1 = coherent UI, UI1 is integrated UI, and UI3 = context-dependent adaptable UI.

	Percentiles		
Variable	**25th**	**50th**	**75th**
Speed. UI1 (n = 18)	33	138	475
Speed. UI2 (n = 18)	28	57	209
Speed. UI1 (n = 12)	32	107	525
Speed. UI3 (n = 12)	27	100	226

With an integrated UI operators were not significantly quicker in opening the detail information window of vessels involved in an incident (effect size = -0.22 and p = 0.19). With context-dependent adaptable UI operators even seemed slower in opening this window compared to coherent (effect size = -0.38 and p = 0.03). Operators, however, did significantly more often opened a detail information window with a context-dependent adaptable UI instead of a coherent UI (effect size = -0,34 and p = 0.05). An incident can be handled without opening this window. Instead, operators can ask the skipper about this information. SMEs, however, agreed that opening the detail information window is the quickest and most accurade way to access this information. Besides, evaluation of the communication scripts revealed that those operators that did not open this window did not interrogate the skippers about this information. Indeed, several operators mentioned during the evaluation that they forgot to use the feature that allowed to quickly open a detail information window by clicking the element of interest in the Area of Focus window, while they did consider it useful or very useful. They expected to commonly use this feature once they are used to it. A ten-minute tutorial might have been too little to change the way in which they search for detail information.

5.2 Accuracy of situation awareness knowledge

SAGAT query 11 was significantly more often answered correctly when operators used a coherent UI instead of an integrated UI (effect size = -0.52 and p = 0.00). A similar trend was found when comparing the coherent UI with the context-dependent adaptable UI (effect size = -0.31 and p = 0.06). There was no significant difference in how well operators answered the other SAGAT queries when comparing the coherent UI and integrated UI. When comparing the coherent UI with the context-dependent adaptable UI, results show that for several queries operators more often answered correctly when using a coherent UI; query 3 (effect size = -0.42 and p = 0.02), query 4 (effect size = -0.52 and p = 0.01), query 12 (effect size = -0.33 and p = 0.05), and query 13 (effect size = -0.38 and p = 0.03).

One SA query showed striking results which needs special attention. In Scenario A (collision near Houten), several operators reported the wrong location of the collision. Incorrect understanding of the location of an incident has major impact: service vessels, the officer of duty, and emergency services are send to the wrong location, traffic measures are wrongly placed, and skippers get incorrect advice. Since this occurred in just one scenario, our data is not sufficient to find statistical differences between UIs in how often this occurred. The trend, however, is serious enough to be mentioned. With a coherent UI, 33% of the operators thought that the incident took place at different waterway section. With an integrated and context-dependent adaptable UI, 17% of the operators made the same mistake. SMEs reported that operators working with a coherent UI were not able to identify their mistake. With an integrated or context-dependent adaptable UI, operators were able to identify their mistake when the officer of duty arrived at the wrong location and contacted the operator.

5.3 Execution of required actions

Wilcoxon Signed Rank Tests showed a difference in which of the required actions operators were more likely to execute depending on which UI concept was used. Operators were more likely to report an incident with a marine VHF radio when using an integrated UI instead of a coherent UI (effect size = -0.37 and p = 0.03). This trend was the same, although not significant, when comparing the context-dependent adaptable UI with the coherent UI (effect size = -0.24 and p = 0.12). However, the opposite was found for sending out notices to skippers using their traffic management information system. Operators were more likely to send out a notice to skippers using a coherent UI instead of an integrated UI (effect size = -0.34 and p = 0.04) or context-dependent adaptable UI (effect size = -0.42 and p = 0.02). When looking at the total amount of required actions that were executed, no significant difference was found; using a coherent UI instead of an integrated UI (effect size = -0.15 and p = 0.37) or context-dependent adaptable UI (effect size = -0.13 and p = 0.26).

5.4 Speed of executing required actions

Of all required operators' actions, only a significant difference in speed of executing required actions was found for communication to priority stakeholders. Wilcoxon Signed Rank Tests shows that operators are up to minutes quicker (median = 154 seconds quicker) in speed of communication with priority stakeholders when using an integrated UI instead of a coherent UI (effect size = -0.44 and p = 0.01). A small effect was found when comparing the coherent UI with the context-dependent adaptable UI, but the data cannot confirm that this effect is not due to a random fluctuation (effect size = 0.19, p = 0.18). This result might be due to the small sample size (n = 12) in combination with an extreme outlier (Speed = 1528) in the data of an operator using the context-dependent adaptable UI, see Table 6. Due to the already small sample size (n = 12) and the need of counter balancing we were not able to repeat this analyses after removing the extreme outliers. An evaluation of only six experiments would not be meaningful.

Table 6: Descriptive statistics of measured speed of communication (in seconds) with priority stakeholders for the different UI concepts (UI1 = coherent UI, UI2 = integrated UI, and UI3 = context-dependent adaptable UI)

	Percentiles		
Variable	**25th**	**50th**	**75th**
Speed. UI1 (n = 18)	381	485	657
Speed. UI2 (n = 18)	307	331	459
Speed. UI3 (n = 12)	309	384	545

5.5 Accuracy of executing required actions

The data analysis shows that with all UI prototypes most operators were able to correctly answer the skippers questions related to Level 1 SA and Level 2 SA, see Table 7. Apparently, all UI's sufficiently supported answering these questions. Several operators, however, were not able to correctly answer the skippers' questions related to Level 3 SA. The analysis shows a medium and significant effect of the used UI on accuracy in answering questions related to Level 3 SA in favor of UI3 (effect size = -0.33, p = 0.05) compared to UI1. Although the data showed a similar trend when comparing UI1 with UI2, no significant difference was found (effect size = -0.13, p = 0.44).

5.6 Accuracy of order of task execution

Operators significantly more often execute the necessary actions in the required order when using an integrated UI instead of a coherent UI (effect size = -0.32, p = 0.03). The same trend, although not significant, is found when comparing the coherent UI with the context-dependent adaptable UI (effect size = -0.25, p = 0.12).

5.7 Workload

Based on the results of the Wilcoxon Signed Rank Test, we can conclude that the UI prototypes do not differ in their impact on operators' workload. The workload measures are all at the lower end of the range found by Grier (2015), see Table 8. There was no statistically significant difference in measured workload between the different UIs.

Table 8: Average of RTLX scores and reference data of process control test cases taken from Grier (2015)

	Percentiles		
Variable	**25th**	**50th**	**75th**
RTLX. UI1	12,54	24,50	38,42
RTLX. UI2	14,86	25,44	28,58
RTLX. UI3	15,96	24,36	28,71
Reference data (Grier, 2015)	31.91	42.00	51.83

6 DISCUSSION

In this study we tested the effects of three UI concepts on operators' SA, task performance, and workload. We assumed that the coherent UI would provide the least support for N-ONM tasks, but our results at first sight do not entirely confirm that. In terms of SA, the results showed that operators had more information as part of their SA knowledge when using a coherent UI instead of an integrated UI or context-dependent adaptable UI. On the other hand, operators were slower in gaining SA with this UI concept compared to the others. No significant difference was found in how likely operators were to execute the required tasks. Speed of task execution was lower when operators' used a coherent UI compared to the other two interfaces. And finaly,

Table 7 Descriptive statistics of scores for answering skippers' questions, in which answer correct = 1 and answer wrong = 0.

	N	Percentiles		
		25th	50th (Median)	75th
Question.LevelSA1.UI1	18	1,0000	1,0000	1,0000
Question.LevelSA1.UI2	18	1,0000	1,0000	1,0000
Question.LevelSA1.UI3	12	1,0000	1,0000	1,0000
Question.LevelSA2.UI1	18	1,0000	1,0000	1,0000
Question.LevelSA2.UI2	18	1,0000	1,0000	1,0000
Question.LevelSA2.UI3	12	1,0000	1,0000	1,0000
Question.LevelSA3.UI1	18	,0000	,5000	1,0000
Question.LevelSA3.UI2	18	,2475	,8350	1,0000
Question.LevelSA3.UI3	12	,3725	1,0000	1,0000

operators were more accurate in answering questions related to Level 3 SA when using a context-dependent adaptable UI instead of one of the other two interfaces. No significant difference was found when comparing operators' workload.

In evaluating these contradicting results of this study, we need to consider the intrinsic relations between SA, task performance, and workload. As discussed in section 2, a too high or too low mental workload could negatively influence operators' SA or task performance. In case of our study, this could have explained our findings if the workload of operators' working with a coherent UI was all right, while the workload with the other two UI was either too high or too low. Since no significant difference in workload was found, this apparently was not the case. Another possible explaination that followed from Section 2 is that operators used a different information processing strategy when using a coherent UI compared to the other two UI concepts. Indeed, information access with a coherent UI was more difficult and time consuming then with the other two interfaces. This makes it plausible that operators were more likely to use memory-based information processing when using the coherent UI, while they used a display-based information processing strategy with the other two interfaces. This line of reasoning is confirmed by our findings. More information was part of operators' SA when using a coherent UI, but this did not result in better task performance.

7 CONCLUSION

Our previous study showed that according to operators' subjective opinion, there was a significant advantage of the integrated UI and context-dependent adaptable UI compared to the coherent UI. A minimal advantage of the context-dependent adaptable UI compared to the integrated UI was found. These results were confirmed by the objective measures conducted in this study. This shows that operators were able to evaluate whether UI features are useful. Subjective measures, however, were insufficient to understand the effect of UI features on operators SA, task performance, and workload. While subjective-measures give meaningful insights, objective measures are required to evaluate the effects of UI features.

In our case of three UI concepts for N-ONM tasks we conclude that the difference between a coherent UI and integrated UI is sufficiently significant to conclude that an integrated UI better supports operators' SA and task performance. The largest effects were found in relation to speed of task performance, especially speed of communication. Implementation of an integrated UI thus proved to be beneficial in dynamic task environments, where operators require SA for time-constrained decisions and actions.

Our results show that the advantage of a context-dependent adaptable UI is only confirmed for tasks in which operators require Level 3 SA. Further research on and implementation of context-dependent adaptable UIs therefore should focus on task environments were operators' use trends and prior knowledge to project out the current situation to predict future states of their task environment.

References

Ben-Bassat, T., Meyer, J., Tractinsky, N. (2006) Economic and subjective measures of the perceived value of aesthetics and usability. ACM Transactions on Computer-Human Interaction 13(2):210-234

Bowden, J.R. and Rusnock, C.F. (2015) Impact of display design individual preferences on process control performance. In Proceedings of the 2015 Industrial and Systems Engineering Research Conference, Nashville, Tenn, USA, 1278-1287

Cohen, J. (1988) Statistical power analysis for the behavioral sciences. Erlbaum, Lawrence & Associates.

Cohen, J. (1992) 'A Power primer', in Psychological Bulletin, 112(1):155-159

Edwards, T., Martin, L., Bienert, N. and Mercer, J. (2017) The relationship between workload and performance in air traffic control: exploring the influence of levels of automation and variation in task demand. In Long, L. and Leva, M.C. (Eds.) Human mental workload: models and applications, Springer, Ireland, 120-142

Endsley, M.R. (1995) Toward a theory of situation awareness in dynamic systems. Human Factors 37(1):32-64

Endsley, M.R. (2000) Direct measurement of situation awareness: validity and use of SAGAT. In Endsley, M.R. and Garland D.J. (Eds.) Situation awareness analysis and measurement, Lawrence Erlbaum Associates, Mahwah, NJ, 147-173

Field, A. (2009) Discovering statistics using SPSS, Third Edition. SAGE Publications Ltd.

Green, P. and Wei-Haas, L. (1985) The rapid development of user interfaces: experience with the Wizard of OZ method. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 29(5):470-474

Grier, R.A. (2015) 'How high is high? A meta-analysis of NASA-TLX global workload scores. In Proceedings of the Human Factors and Ergonomics Society 59th annual meeting, Los Angeles, USA, 1727-1731

Kissel, G.V. (1995) Effect of computer experience on subjective and objective software usability measures. In Proceedings of the Conference on Human Factors in Computing Systems, Part 2, Denver, CO, USA, 284-285

Mogford, R.H. (1997) Mental models and situation awareness in air traffic control. International Journal of Aviation Psychology, 7(4):331-341

Parasuraman, R. and Riley, V. (1997) Humans and automation: use, misuse, disuse, abuse. Human Factors, 39(2):230-253

Patrick, J. and Morgan, P.L. (2010) Approaches to understanding, analysing and developing situation awareness. Theoretical issues in Ergonomics Science, 11(1 – 2):41-57

Ramachandran, K.M. and Tsokos, C.P. (2009) Sampling distributions, in Ramachandran, K.M. and Tsokos, C.P. (Eds.) Mathematical Statistics with Applications, Elsevier, 183-224.

Sonderegger, A. and Sauer, J. (2010) The influence of design aesthetics in usability testing: effects on user performance and percieved usability. Applied Ergonomics 41(3):403-410

Stanton, N.A., Salmon, P.M., Walker, G.H. and Jenkins, D.P. (2010) Is situation awareness all in the mind? Theoretical Issues in Ergonomics Science, 11(1 – 2):29-40

Stanton, N.A., Stewart, R., Harris, D., Houghton, R.J., Baber, C., McMaster, R., Salmon, P., Hoyle, G., Walker, G., Young, MS., Linsell, M., Dymott, R. and Green, D. (2006) Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology. Ergonomics, 49(12 – 13):1288-1311

Stuut, R., Van Doorn, E.C., De Jong, K (2019) The use of automatic object detection signals from smart cameras to enhance bridge operators' situation awareness. In Proceedings of the Human Factors and Ergonomics Society 63rd annual meeting, Seattle, WA, USA

Treisman, A. (1985). Preattentive processing in vision. Computer Vision, Graphics, and Image Processing, 31(2):156-177

Van Doorn, E.C., Horváth, I., Rusák, Z. (2014) A systematic approach to addressing the influence of man-machine interaction on situation awareness. In Proceedings of The Tenth International Symposium on Tools and Methods of Competitive Engineering, Budapest, Hungary, Vol. 1:109-120

Van Doorn, E.C., Horváth, I., Rusák, Z. (2015) Combined use of cognitive task analysis and observational research data to identify deficiencies of support for situation awareness. In Proceedings of the Human Factors and Ergonomics Society 59th annual meeting, Los Angeles, USA, 1717-1721

Van Doorn, E.C., Horváth, I., Rusák, Z. (2017a) Information engineering for developing and testing coherent, integrated and context dependent user interfaces. Cognition, Technology & Work 19(2-3):375-397

Van Doorn, E.C., Rusák, Z., Horváth, I. (2017b) A situation awareness analysis scheme to identify deficiencies of complex man-machine interaction. International Journal of Information Technology and Management 16(1):53-72

Vidulich, M.A. and Tsang, P.S. (2012) Mental workload and situation awareness. In Salvendy, G. (Ed.), Handbook of Human Factors and Ergonomics: Fourth Edition, John Wiley and Sons, 243-273

Vidulich, M.A. and Tsang, P.S. (2015) The confluence of situation awareness and mental workload for adaptable human-machine systems. Journal of Cognitive Engineering and Decision Making, 9(1):95-97

Willages, R.C. (2007) Human factors experimental design and analysis reference. Army Research Laboratory. Aberdeen Proving Ground, MD 21005-5425.

Wolfe, J.M. and Horowitz, T.S. (2004). What attributes guide the deployment of visual attention and how do they do it? Nature Reviews Neuroscience, 5(6):495-501

Zuckerman, O. and Gal-Oz, A. (2013) To TUI or not to TUI: Evaluating performance and preference in tangible vs. graphical user interfaces. International Journal of Human-Computer Studies 71(7):803-820