Cognitive Systems Research 76 (2022) 13–25

Contents lists available at ScienceDirect

Cognitive Systems Research

journal homepage: www.elsevier.com/locate/cogsys

The impact of digital image configuration on submarine periscope operator workload, situation awareness, meta-awareness and performance

Steph Michailovs [a], Zachary L Howard [a], Stephen Pond [a], Troy Visser [a], Jason Bell [a], Gavin Pinniger [a], Jessica Irons [b], Megan Schmitt [b], Madison Fitzgerald [a], Matthew Stoker [c], Owen Carter [a], Sam Huf [b], Shayne Loft [a,*]

[a] *University of Western, Australia*
[b] *Defence Science and Technology Group, Australia*
[c] *Ergonomie, Australia*

ARTICLE INFO

Keywords:
Workload
Human–machine interface
Submarines
Periscope
Situation awareness

ABSTRACT

Submarine periscopes are moving from analog to digital, but there is a lack of empirical evidence regarding the relative advantages that different digital human–machine interface configurations might provide to the operator. We experimentally compared the effectiveness of two digital concepts for displaying and analysing periscope imagery during a simulated submarine mission. OPTIX modelled a traditional periscope by presenting a narrow (20° horizontal arc) rotating image of the horizon. OPTIMUS displayed the full 360° panoramic representation of the horizon and was augmented with digital analysis tools. OPTIMUS supported faster and more accurate performance by participants ($N = 32$) and lowered perceived workload, compared to when the same participants used OPTIX. However, time taken to respond to situation awareness queries and awareness of one's own performance (meta-awareness) was poorer when using OPTIMUS. OPTIMUS holds an advantage in that it can improve performance whilst reducing workload, but the SA and meta-awareness decrements are potentially problematic.

1. Introduction

The modern submarine optical-path periscope is the product of over a century of evolution. Periscopes pass light from above the surface to the submariner via a series of prisms and mirrors, and provide magnification and inbuilt graticules for assisting the user with conducting contact (i.e., other vessels') range and course estimates. The operator can physically manipulate this mechanism to see a portion of the horizon at a time (i.e., a view down a bearing). Periscope operators are able to garner rich information from the live periscope imagery such as the classification of visible contacts (and therefore potential threat), environmental conditions (e.g. sea state, lighting, visibility) and other contextual information. The contact imagery also enables the operator to estimate the distance of the contact as well as its course, information which is vital in order for the submarine to remain safe and undetected and to carry out specific missions (e.g. surveillance).

Analog (optical path and mechanically rotating) visual sensor systems in maritime defence platforms are increasingly being replaced with digital technology that can capture and present a video stream to the operator. Digitised imagery enables a range of human–machine interface (HMI) configurations. Some options available display full-screen imagery through a single narrow view that can be rotated, not unlike the view of traditional periscopes. However, advancements in technology provide the opportunity to innovate how the external visual environment is displayed to best support the tasks of the users. For example, it is now possible to make use of captured video to snapshot the entire horizon (Roberts et al., 2021), and future technologies may exploit multiple cameras to enable continuous 360° video. While such advancements potentially bring benefits (such as increasing the amount of horizon available at a glance), it is important to consider the differing advantage and potential cognitive burden that different emerging HMI configurations may impose.

In the present study we compared the effectiveness of two digital HMI periscope concepts for supporting the human operator completing a simulated submarine mission. One HMI, termed "OPTIX", represented a digital adaptation of the traditional "view down a bearing" periscope

* Corresponding author at: University of Western Australia, WA 6009, Australia.
E-mail address: shayne.loft@uwa.edu.au (S. Loft).

https://doi.org/10.1016/j.cogsys.2022.09.001
Received 29 November 2021; Received in revised form 30 July 2022; Accepted 6 September 2022
Available online 11 September 2022

format (Fig. 1, left panel). The other, "OPTIMUS" (OPTronics IMplementation & Usability System) represented a future-looking design simulating digital technologies to display a 360° panoramic real-time video of the horizon segmented across five panels (Fig. 1, right panel). The two HMIs also differed in the degree to which the analysis tools to support the operator's tasks were digitized (see Method for details).

We have previously demonstrated that OPTIMUS can enhance the detection of high-contrast short-range visual contacts, and considerably improve the initial estimated range and course of contacts, without creating additional perceived workload (Michailovs, Irons et al., 2022)[1]. However, these performance gains were observed in isolated contact detection and contact analysis tasks, whereas the role of the periscope operator extends beyond this. As the "eyes of the submarine", it is imperative that periscope operators maintain an accurate understanding of the tactical picture around Ownship. It remains to be seen whether a panoramic periscope HMI can preserve performance gains without increasing workload during an ongoing mission with broader goals and objectives. Further, the operator's ability to maintain *situation awareness* (SA) becomes highly relevant when the tasks involve understanding the tactical picture and the interacting behaviour of contacts.

With these concerns in mind, the current study aimed to compare the relative effectiveness of the OPTIMUS and OPTIX periscopes for supporting performance, workload, SA, and an awareness of one's own performance (referred to as meta-awareness) in a simulated submarine mission context. Below we provide an overview of performance, workload, SA, and meta-awareness in a simulated submarine control room context, before introducing the specifics of our experimental approach.

1.1. Task performance and cognitive workload

Cognitive workload is the relationship between task demands placed on an operator and the capacity of the operator to meet those demands (Parasuraman et al., 2008). If task demands exceed an operator's limited cognitive capacity task performance may be compromised (Hancock & Matthews, 2019; Howard et al., 2021). Higher workload can also compromise the capacity of the operator to respond to future task demands (Hollnagel, 2002; Loft et al., 2007; Sperandio, 1978), which is particularly an issue for performance if non-routine conditions or events are encountered. Thus, when considering HMI configurations, it is critical to consider that workload and performance are intimately linked. If a HMI benefits performance at the expense of additional workload, these relative benefits and costs need to be carefully considered in the context of HMI design/selection and control room function allocation.

Michailovs, Irons et al. (2022) found that the use of OPTIMUS did not increase perceived workload compared to OPTIX when detecting contacts or inputting one-off contact range and course estimates. Such tasks, however, represent a simplification of the periscope task set. The periscope operator also needs to estimate where each contact is (i.e., at what bearing and how far away – range), where they are going (course and speed), and how each contact may interact (i.e., the tactical picture). The degree to which the tactical picture reflects actual ground truth at any given time is a key performance measure in submarine operations. In addition submarines, not surprisingly, have a "mission-focus" and objectives (e.g. identifying contacts of interest) based around and dependent on the achieved tactical picture, and this may further increase workload.

We would expect, based on earlier reported results by Michailovs, Irons et al. (2022), that there will be tactical picture benefit when using OPTIMUS compared to OPTIX that should compound over time, and that a panoramic visual display should facilitate the periscope operator's mission-focused performance both directly as a result of a more accurate tactical picture, and because the operator can more easily interpret the spatial relationships between contacts simultaneously presented on the panoramic display.

However, in order to achieve a mission, the periscope operator must maintain awareness of all contacts, the broader context those contacts are operating within, and how Ownship fits within that tactical picture. It remains to be seen whether OPTIMUS can enhance performance without increased workload in a mission context. It is possible that using OPTIMUS to view and action multiple contacts potentially relevant to mission goals may increase perceived workload. The smaller field of view of the traditional"view down a bearing" HMI (OPTIX) may function as a natural "attentional spotlight" (Posner, 1980), reducing cognitive burden. On the other hand, maintaining tactical awareness using a traditional"view down a bearing" HMI (OPTIX) requires the operator to hold information about contacts and their relationships in working memory (Engle, 2002), potentially increasing cognitive load relative to a panoramic display (Barrouillet et al., 2007).

Further, the digital tools available in OPTIMUS, which are not provided in OPTIX (where a physical model was used instead – see Methods), may reduce workload because they allow the operator to match digital object model features directly to contact images. However, there could be workload associated with combating the potential visual clutter (Moacdieh & Sarter, 2015) that digital tools add to displays that could outweigh the potential reduction in workload associated with using OPTIMUS digital tools for contact course and range estimation.

1.2. Situation awareness

Distinct from, but related to workload and performance, SA describes an individual's understanding of the relevant elements of the task environment and how these elements will change as a function of environmental conditions or control actions (Durso & Dattel, 2004; Endsley, 1995a). SA can be positively related to individual performance in a variety of task contexts (see review by Endsley, 2021), including in simulated submarine settings (Loft et al., 2015, 2018). Although various theoretical accounts of SA exist, common threads place SA as an interface between the world and the mind, such that SA itself is "neither resident in the world nor in the person but resides through the interaction of the person with the world" (Salmon et al., 2008, pp 303; also see Chiappe et al., 2012; Vu & Chiappe, 2015).

It is critical that the periscope HMI support the periscope operator to build and maintain accurate SA because it is their sole link to the physical world. SA is of even greater importance in more complex operational contexts, in which periscope operators play a key role in collecting visual information to maintain safety and stealth and achieve mission objectives. The periscope operator needs to understand where their Ownship is located relative to other contacts and navigational entities such as land, and how contacts relate in time and space to other contacts (Stanton, & Bessell, 2014).

While OPTIMUS has been shown to enhance basic perceptual and analytic tasks (Michailouvs, Irons et al., 2022), it is unclear how well it supports SA in more demanding submarine scenarios. Of particular interest here is that OPTIMUS presents information in an unfamiliar form (as humans are not usually able to see 360° simultaneously), and thus also displays *more* information at any one time than humans are typically accustomed to. On the other hand OPTIMUS may enhance SA compared to OPTIX by allowing increased offloading of information (as all positions and contact relationships can be sampled from OPTIMUS displays). Unlike OPTIX, the OPTIMUS operator does not have to swing the view to a bearing of interest whilst holding other information (e.g., the relative location of other contacts) in memory in order to stitch them together for an integrated tactical picture.

In addition to assessing SA, we also explored participant meta-awareness of their own tactical picture accuracy. An effective periscope operator should understand which contact solutions are "tracking

[1] The Michailovs, Irons et al. (2022) manuscript is currently under review at Behaviour and Information Technology. A version of the paper is available to read at https://osf.io/5vqt6/.

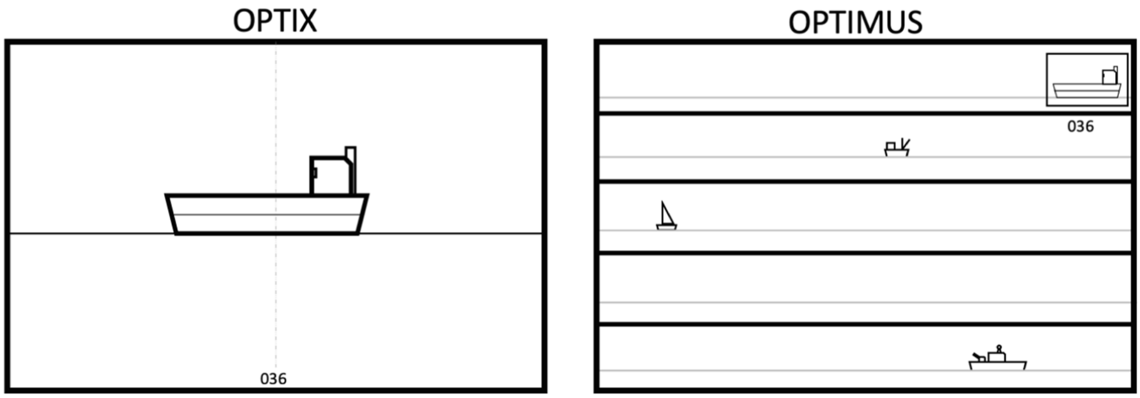

Fig. 1. Conceptual depiction of the main differences between the two digital periscope representations compared in this study. For OPTIX the periscope displayed only a portion of the horizon ("view down a bearing") and could be 'swung' across the horizon akin to a traditional periscope. For OPTIMUS the full 360° horizon was displayed across five panels (panorama view). For the purposes of the experiment the simulation displayed near-ideal visibility conditions on calm seas in daylight and thus high contrast stimuli.

well", and which are not and need updating, in order to be able to prioritise tasks and complete missions.

1.3. Current study

We designed simulated submarine scenarios in which participants utilised each HMI to not only detect contacts but track them throughout a scenario. In addition, to provide task context, we developed two "missions", one relating to Ownship safety (i.e. collision risk), and one requiring participants to track interactions between contacts. These missions, provided goal-directed relevance to the development and maintenance of SA in the simulated submarine task.

We manipulated task-load to understand the relative impact of the HMIs under different conditions (Roberts et al., 2018). In terms of task-load, we expected poorer performance and SA, and increased perceived workload, with higher task-load. We did not have strong expectations regarding interactions between HMI and task-load. Despite this, the task-load manipulation is critical because it provides a test of boundary conditions regarding the impact of the HMIs. To evaluate user experience we collected survey responses on system usability. While Michailovs, Irons et al. (2022) showed that perceived usability was higher for OPTIMUS than OPTIX, with more complex scenarios/demanding tasks it remains to be seen whether OPTIMUS will still be rated superior.

2. Method

2.1. Design

The study used a within-subject design, with HMI (OPTIX/OPTIMUS), and task-load (low/high) as independent variables. Participants completed four 24-min scenarios – two scenarios using OPTIX (low/high task-load) and two using OPTIMUS (low/high task-load). The experiment was completed over two days, with training and experimental scenarios for one HMI per day. The order of HMI over days was counterbalanced, as was the order of task-load. The dependent variables were subjective workload, SA, meta-awareness, contact solution accuracy/response time, tactical picture error (TPE), detection of mission events (rendezvous/danger sector, see 2.4.2) and system usability.

2.2. Participants

Thirty two participants (15 female) took part in the experiment and were paid $160 AUD: $70 per session, plus a bonus of $20 paid after the second session. Participants were drawn from a University undergraduate recruitment pool, and although they were extensively trained, they were naïve to submarine tasks, operations, and the maritime context in general (i.e., non-experts). The mean age was 22.75 years ($SD = 4.5$). This research was approved by the Human Research Ethics Office at UWA. Informed consent was obtained from each participant.

2.3. Periscope HMI

Each HMI concept was displayed on two 1920 × 1080, 27-inch displays, aligned vertically. The top screen displayed the visual imagery from the periscope, and the bottom screen provided information and tools required for range and course estimates. Both periscope concepts are part of the Control Room Use Simulation Environment (CRUSE; developed by Defence Science and Technology Group, see Michailovs, Pond et al., 2022).

For OPTIX, the top display presented a view down a bearing displaying approximately 20° of horizontal arc (Fig. 2). The operator could use the keyboard or mouse to move the digital periscope horizontally through 360° in order to view the entire horizon, and the view could be magnified (x12) to conduct contact range and course estimates. The lower display included elements for contact tracking such as an interactive list of all currently tracked contacts (contact tote; including tabulated contact classification, range, course and speed), a classification database, and a top-down view of the geographic area centred on Ownship and populated with contact solutions (Geoplot). Whilst the Geoplot provides a bird's eye view of the contact *estimated* position (dead reckoned from last solution) in the surrounding area, it could not be used to complete the operators' main tasks. The 'view down a bearing' provides *actual* (real time) bearing and rich information on each contact such as their aspect (to estimate course), and the vertical height and position relative to the horizon (to estimate range). For OPTIX, participants were also provided a physical manual tool to estimate Angle on the Bow (ATB; see Fig. 4).

The upper OPTIMUS display presented a 360° panoramic view (Fig. 3). Operators had access to a magnified window which could be operated by mouse to inspect a bearing at higher magnification (see Fig. 3A). In addition to a contact tote (including tabulated contact classification, range, course and speed), classification database and Geoplot, the OPTIMUS lower display included an analysis pane displaying a view of up to 12x magnification down a bearing (equivalent to the magnified view in OPTIX). The classification database could be toggled to display a 3D model of any vessel classification type to assist contact range and ATB (course) estimates. Operators made ATB estimates by rotating the 3D model (around a central vertical axis) via a scroll bar to match the contact as viewed in the analysis pane. Once the contact was classified, operators estimated range by matching the ranging boxes on the 3D model and overlayed over the contact in the analysis pane.

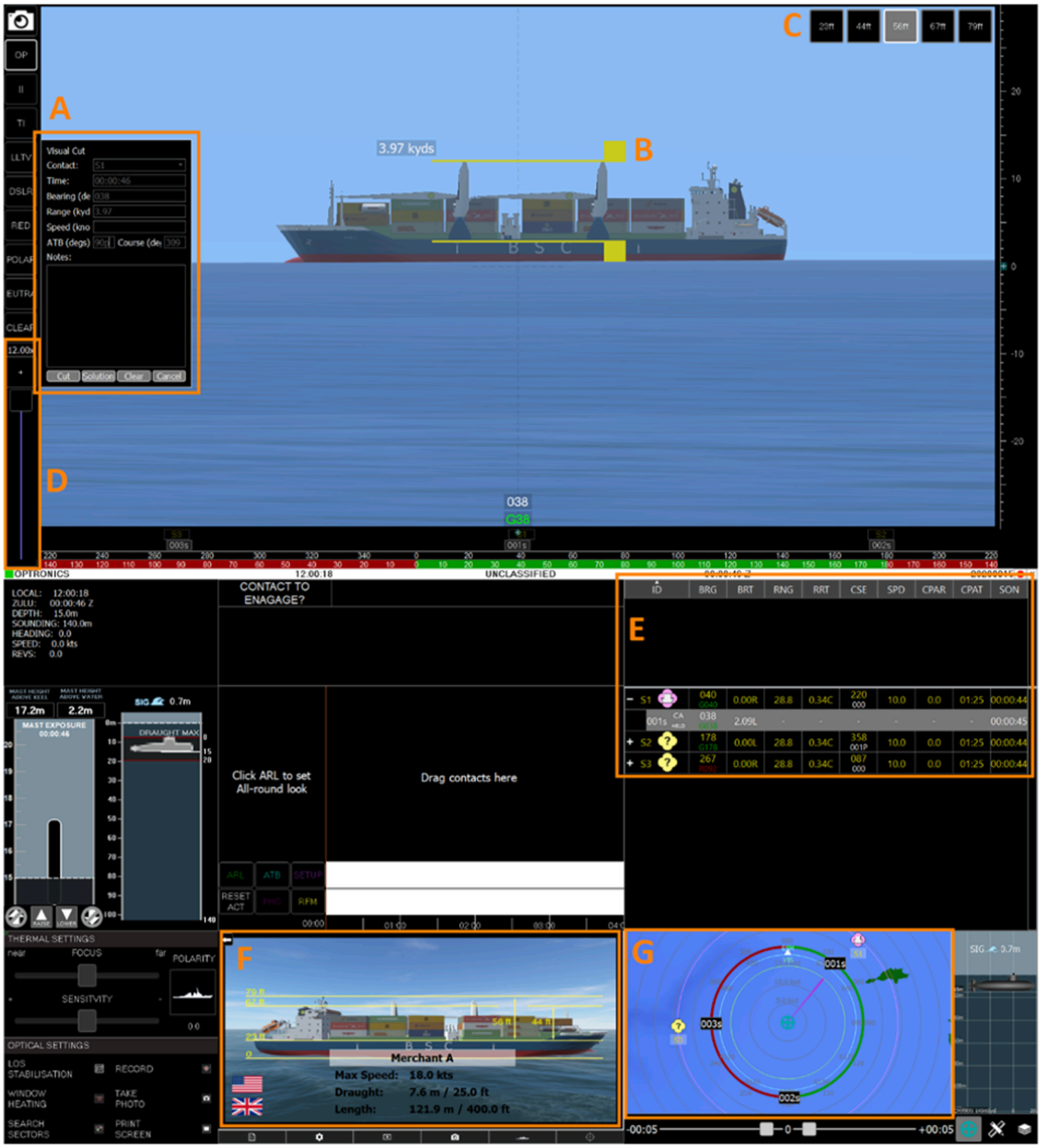

Fig. 2. The OPTIX HMI as presented on two vertically aligned screens, with key features marked. A: Visual cut window, in which operators entered range and ATB estimates. B: Range bars, which operators adjusted to match the lines on the image as displayed in the classification window (F). C: Contact height bins specific to the contact classification. D: Magnification slider for controlling magnification. E: Contact tote listing the currently tracked contacts. F: Classification database with height intelligence. Images of different vessel classification types could be viewed in this database, allowing participants to match a visible contact to a classification type. Once a classified contact was selected for analysis, the classification database provided height reference lines for ranging. G: Geoplot showing a top-down view of the location and direction of travel of the submarine (centre), and surrounding contact solutions.

Both HMI configurations included a tracking feature that used visual data to continually update the position of a contact, once it had been detected. Participants were able to initiate a visual tracker on a visible contact, and this tracker would follow the contact as long as it remained visible.

2.4. Goals and tasks

The primary goal was to ensure the safety of Ownship by developing an accurate tactical picture (having an accurate solution on detected contacts to minimize the TPE). The other goal was to complete two “missions”, described below.

Most contacts were detected automatically by the system at the start of the scenario (simulating that most contacts are detected by sonar in submarines before being detected visually) and were assigned default solutions. Additional ‘visual-only’ contacts (yachts) were presented which in the real world emit very little sound and so in our simulation were not detected by sonar, and had to be detected by participants who manually assigned trackers and solutions to them. Participants were asked to prioritise estimating solutions for contacts based on how relevant they currently were to the two missions. Once detected, participants selected a contact to classify it and develop its solution (bearing, range, course, and speed).

Bearing: The contact’s bearing was provided automatically via the (automated) sonar tracker or (manually assigned) visual tracker.

Range: To estimate range, participants adjusted two ranging lines

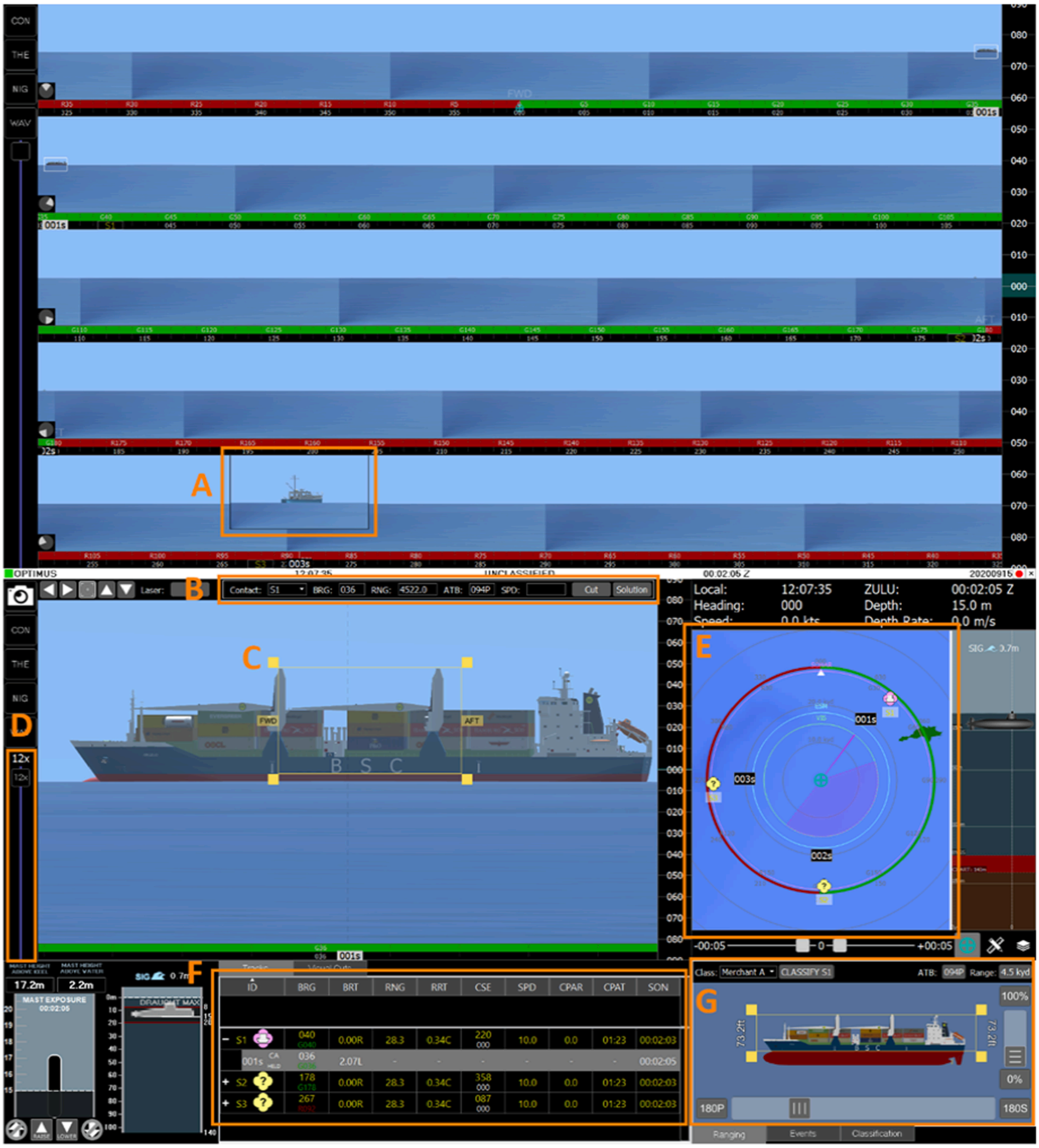

Fig. 3. The OPTIMUS HMI as presented on two vertically aligned screens, with key features marked. A: Magnified window, which could be controlled by mouse to inspect sections of the panorama. B: Visual cut window. The range and ATB (course) information in this window automatically updated as operators manipulated the yellow ranging box (C) and range/ATB tool (G). C: Ranging box in the analysis pane. Operators adjusted the position of the top and bottom lines to self-selected positions on the contact, and then matched these positions by manipulating the ranging box in the range/ATB tool (G) to generate a range estimate. D: Magnification slider for controlling magnification in the analysis pane. E: Geoplot showing a top-down view of the location and direction of travel of the submarine, and surrounding contact solutions. F: Contact tote listing the currently tracked contacts. G: Range/ATB tool showing a 3D model of the currently selected contact's classification type. In addition to the ranging box described above, the slider at the bottom of this tool could be dragged to rotate the 3D model to match the visible contact. This generated an ATB estimate which automatically populated the visual cut window. The range/ATB tool could be toggled to the classification database in order to classify contacts. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

overlaid on the display for both HMIs to select a vertical ranging height on the contact and to generate a subtended angle from Ownship position (see Fig. 2B and 3C). For OPTIX, the participant had to select the corresponding actual height from five pre-determined heights (simulating intelligence information for that class of vessel – see Fig. 3C). The height combined with the subtended angle allowed OPTIX to compute the range but sometimes required mental arithmetic. For OPTIMUS, a 3D model was available to match the vertical height bars on the digital intelligence (Fig. 3G) against equivalent height bars on the analysis pane. This procedure was expected to be less mentally effortful.

Course: The contact's course was calculated automatically from the participants' estimated ATB. With OPTIX, the participant estimated the ATB with a manual tool (Fig. 4), and the participant had to interpolate the ATB that were not indicated on the tool. By comparison, the OPTIMUS participants could rotate a 3D digital model to match the orientation of the contact and the system would determine the ATB (see Fig. 3G).

Speed: The contact's speed was fixed for each classification, and given to participants on a reference sheet. This meant that classifying each contact correctly was necessary for generating accurate speed

Fig. 4. The manual ATB tool used by OPTIX operators. Operators rotated the tool to match the orientation of a contact and used the bearing lines to inform estimates of ATB.

estimates. Nine contact classification types were used across the four scenarios, each falling into one of four groups: warships, merchants, fishing vessels and leisure (yachts). The speed of each contact type was: warship = 22 knots; merchant = 15 knots; fishing = 5 knots; and leisure = 5 knots.

2.4.1. Solutions

Once the participant was satisfied with the bearing, range, course, and speed estimates for a contact, the participant clicked on a 'solution' button (located in the panels shown in Fig. 2A and 3B), which entered the four elements combined as a contact solution. This updated the contact's solution details in the contact tote and its position on the Geoplot. It also placed a solution tracker on the bearing strip which coincided with the sonar tracker (both at the same bearing).

The solution represents both the best estimate of a contacts' position at the time it is entered, and a projection of future position. The accuracy of the solution thus became evident over time by how well the estimated solution matched the sonar/visual tracker (based on observed bearing of the contact from sonar/periscope). If they matched over time the solution was good; if they started to diverge, the solution could be updated. There was no limit to the number of solutions that could be entered for each contact, allowing for an iterative process. Each new solution for a contact overwrote the previous one.

2.4.2. Missions

The two missions given to participants were framed as *intelligence gathering*. The *danger sector* mission required participants to photograph any contact that entered any of three different "danger sectors" defined as 10° either side of port side, starboard side, or directly behind Ownship, and within a range of four kiloyards. Danger sector breaches had various durations as contacts travelled at different speeds and courses with an average breach duration of 50 s.

The rendezvous mission required participants to take a photograph of a rendezvous between any two contacts who approached each other, slowed down to 3 knots, and changed course so that they could travel side by side (simulating an information handover). Each rendezvous lasted approximately 60 secs, and the smaller contact was always closer to Ownship so that both contacts could be seen at the same time.

When participants detected a rendezvous or a danger sector breach, they centred the contact(s) in view and took a photograph (click on a 'camera button' on the display), which produced the message "VISINT LOGGED" and the time and bearing of the photograph was recorded. Only one photograph was required per event.

2.5. Scenarios

Four 24-minute scenarios were generated: two scenarios with six contacts (low task-load), and two scenarios with eight contacts (high task-load). Low task-load scenarios included one visual–only contact, while high task-load scenarios had two visual-only contacts. All contacts remained within visual range for the duration of the scenario.

Other than merchants, contacts would occasionally change course. This would require participants to reassess the contact's range and ATB and submit an updated solution. Course changes also occurred following a rendezvous when the two contacts separated, even if one of the contacts in question was a merchant.

Each scenario had two rendezvous and two danger sector breaches. All rendezvous occurred in the last third of a scenario except in one low and one high task-load scenario, in which one rendezvous occurred approximately five minutes into the scenario. Danger sector breaches could occur at any time after the first 7.5 min of each scenario. We also programmed 'close-calls' that approached but did not trigger the event criteria (i.e. two contacts converge in bearing but do not rendezvous, or a contact enters the danger sector by bearing but remains outside the danger range of 4000 yards). There were between 4 and 11 close calls per scenario.

2.6. Measures

2.6.1. Performance

The primary performance measure was TPE, a measure of the difference between a participant's contact solutions and the actual location of each contact in the simulation (measured in yards from Ownship) at any time. To compute TPE we extracted the estimated and true positions for each contact at 20 s intervals and computed the difference, scored in yards. We capped the position error at 5000 yards (which was approximately the 95th percentile of the participant-generated error distribution) to exclude extreme outliers. We also penalised both "default" solutions (that is, the solution automatically applied by the system until the participant generated their own solution) and "undetected contacts", by setting both to the position error cap of 5000 yards until a participant solution was entered for the contact. The final TPE value was an unweighted average of the individual contact position error values (and is thus capped at 5000 yards). To replicate the tasks from our earlier study (Michailouvs, Irons et al., 2022), we also measured the range error and course error of initial solutions and the time taken to generate initial solutions.

Performance on the rendezvous and danger sector missions were assessed by examining the time and bearing of events associated with photographs. When the event was visible in the frame and the photograph taken within +/- 10 s of the event period, the event was 'detected'. Multiple photographs of the same event were discarded.

2.6.2. Subjective workload

Every five minutes during the scenario participants were presented a response scale to rate their workload from 1 to 10 (the simulation continued unaffected) based on the Air Traffic Workload Input Technique (ATWIT; Stein, 1985). The scale appeared on both screens and remained visible for 10 s or until participants responded via mouse click. Missed responses were excluded from analysis.

Participants were also presented with the Modified Cooper Harper (MCH) scale three times during each scenario (Wierwille & Casali, 1983). The MCH is a unidimensional scale that assesses how effectively the system supports the workload of the user by using a decision tree to arrive at a score between 1 and 10, with lower scores more desirable. The MCH was presented during the SA pauses prior to the initial SA query (see below).

2.6.3. Situation awareness

We measured SA using a hybrid version (see Chiappe et al., 2016) of the Situation Awareness Global Assessment Technique (SAGAT; Endsley, 1995b) and the Situation Present Assessment Method (SPAM; Durso & Dattel, 2004). Consistent with SAGAT, multiple SA queries were presented to operators per task pause. However, while the task was paused, contact information was not blanked but left displayed, based on the premise of situated SA theory that operators rely on interactions with information displays to maintain SA. Eighteen SA queries were delivered in three blocks of six in each scenario, at six, 12 and 18 min into scenarios. SA questions (Table 1) probed each of the three levels of SA (perception, compression, projection) defined by Endsley (1995a). SA queries overlaid on the lower display. Participants could not interact with their displays while the questions were presented. The simulation remained paused until all questions were answered.

Included in each SA pause was a probe on confidence on a specific and recent contact solution. Phrased as "What is your Confidence in the Solution for Contact X?", these queries targeted meta-awareness of own performance (TPE). Participants responded on a 5-point Likert scale from "Very confident" to "Not confident at all". We correlated the score on each query against the TPE for the same contact.

2.6.4. *Perceived HMI usability*

Usability was measured using the System Usability Scale (SUS; Brooke, 1996), which presents 10 questions on a 5-point Likert scale. Example questions are "I thought this console was easy to use" and "I thought there was too much inconsistency in this console." Participants completed the SUS once for each HMI design.

2.7. *Procedure*

Participants completed the experiment over two days, in sessions of 3.5 and 3 h respectively, and 1–2 participants were tested in each session. In addition to behavioural data we collected physiological data (heart rate variability and electro-dermal skin conductance). These recordings were collected for exploratory analyses beyond the scope of this manuscript, and are not described further.

Participants were extensively trained. They were first given a brief overview of the tasks. They then watched an 18-minute training video

Table 1
Situation awareness questions.

o **Question**	o **SA Level**	o **Response Format**
o How many vessels are we currently tracking?	o 1	o Multiple Choice
o How many contacts are on our Port/ Starboard side?	o 1	o Multiple Choice
o Where is the X vessel?	o 1	o Multiple Choice/ Numpad
o Where is contact X?	o 1	o Multiple Choice/ Numpad
o Where is the X vessel on our port/ starboard side?	o 1	o Multiple Choice/ Numpad
o Where is the current danger sector breach?	o 1	o Multiple Choice
o Which danger sector is closest to being breached?	o 2	o Multiple Choice
o Where is the vessel that is heading towards us?	o 2	o Numpad
o Where is the vessel heading towards contact X?	o 2	o Numpad
o Where are the two vessels that are closest together?	o 2	o Numpad
o Where would you expect the next potential RV?	o 3	o Numpad
o Where will the next danger sector breach occur?	o 3	o Multiple Choice
o Where will the X vessel be in 5 min?	o 3	o Multiple Choice
o Where will contact X be in 5 min?	o 3	o Multiple Choice

Note. RV = rendezvous.

detailing concepts such as port and starboard, contact solutions, range, and ATB, as well as tasks and missions that they would perform. The video also introduced the SA and subjective workload measures. Participants then completed four practice scenarios using one HMI, with total duration one hour. Experimenters guided participants through practice, and each practice scenario introduced more task elements. The final practice scenario included a rendezvous, two danger sector breaches, and the workload and SA measures. Participants continued on to the experimental scenarios once the researchers was satisfied they understood the task. Additional coaching was given as necessary. The total training time amounted to 2.5 h per participant.

Participants then began the first experimental scenario. HMI, task-load, and whether a scenario had an early rendezvous or not were all factors that were counterbalanced, with HMI blocked so that participants changed HMIs between sessions, not within. Participants wore headphones playing ambient submarine engine noise to occlude external noise. Following the first scenario, participants then completed the usability questionnaire.

The second session followed the same procedure as the first using the other HMI, with the exception that participants did not repeat the training video. At the end of the experiment, participants were asked which HMI they preferred.

3. Results

For analyses reported throughout we use Bayesian statistics. Bayesian statistics have several advantages over Null-Hypothesis Significance Testing (Kruschke & Liddell, 2018). Bayes Factors can be interpreted as the strength of evidence for one hypothesis over another (e.g. a Bayes Factor of 10 means one hypothesis is 10 times more likely than the other). For convenience of interpretation, we always present the Bayes Factor as a number greater than one. To distinguish evidence for the Alternate vs Null hypotheses we report BF_{10} when evidence favours the alternate hypothesis (i.e. that there *is* a difference or effect), and BF_{01} when evidence favours the null hypothesis of no difference.

In all two and three-way ANOVAs we include a random effect for Subject in our ANOVAs (including in the Null model). This is an alternative to a repeated-measures approach and is useful when there are multiple observations per participant per cell (Rouder et al., 2012) and participants may differ in performance. We report 95% Bayesian Credible Intervals (distinct from Confidence Intervals) around the mean, which can be interpreted as a region of uncertainty (Wagenmakers et al., 2018).

3.1. *Correlations*

We present a correlation matrix of key dependent variables in Table 2. The values used to compute each correlation are the average for each Subject × Task-Load × HMI, and thus do not speak to differences between the HMIs but rather describe the overall relationship between variables at the mean level. The largest correlations were between ATWIT and MCH (convergent validity), SA accuracy and TPE (such that better SA accuracy was associated with lower TPE – i.e., better performance, indicating predictive validity), and number of solutions and TPE (entering more contact solutions lowered TPE). Higher perceived workload (ATWIT, MCH) was associated with higher TPE, and better SA (faster RT) with lower TPE (consistent with the SA accuracy correlation with TPE). The negative correlation between SA accuracy and SA RT indicates participants did not trade speed for accuracy when responding to SA queries. A higher number of solutions (which lowered TPE), and higher SA (more accurate/faster response to SA queries) were associated with better rendezvous and danger sector performance (d').

3.2. *Task performance*

To assess the impact of HMI and task-load on contact solution

Table 2
Correlation Matrix between dependent variables.

Variable	TPE	ATWIT	MCH	SA (acc)	SA (RT)	N Solutions	d'
TPE	—						
ATWIT	0.335***	—					
MCH	0.285**	0.732***	—				
SA (acc)	−0.625***	−0.222	−0.284**				
SA (RT)	0.389***	0.149	−0.103	−0.391***	—		
N Solutions	−0.634***	−0.223	−0.163	0.280**	−0.371***	—	
d'	−0.392***	−0.097*	−0.104	0.335***	−0.275**	0.426***	—

Note: TPE = Tactical Picture Error (higher value = poorer performance); ATWIT = Air Traffic Workload Input Technique (workload); MCH = Modified Cooper Harper (workload); SA (acc) = Situation Awareness accuracy; SA (RT) = Situation Awareness Response Time; d' = sensitivity (performance) for danger sector and rendezvous missions. *=BF_{10} > 5, **=BF_{10} > 10, ***=BF_{10} > 100, †=BF_{01} > 5.

performance, we assessed metrics of solution quality and quantity. This included comparisons of range and course error of the initial solution, time taken to generate an initial solution, magnitude and change in TPE across a scenario, and total number of solutions generated during a scenario. We also assessed performance on the danger sector and rendezvous missions.

Each of the 32 participants was presented 28 contacts (8 contacts × 2 runs in high task-load, 6 contacts × 2 runs in low task-load), yielding a total of 896 contacts to be analyzed. Of these, 30 contacts were not detected by the participant, or at least never had a solution entered (23 OPTIX, 7 OPTIMUS) suggesting that participants may have been more susceptible to missing contacts whilst conducting sweeps in OPTIX particularly under time pressure from the accompanying tasks (estimating range and ATB for the solution). A further 20 were detected but never had a participant-generated solution (15 OPTIMUS, 5 OPTIX) suggesting that when provided visual access to all contacts at once (OPTIMUS), competing priorities may have had greater salience, which potentially pushed the need to generate a solution further down the priority list and made it more likely to be forgotten. An additional 73 contacts were incorrectly classified, which interfered with accurate ranging using the tools provided by OPTIX and OPTIMUS (incorrectly classified contacts had three times higher range error ($M = 44.65\%$, 95% BCI [25.01, 64.39]) than correctly classified contacts ($M = 13.37\%$, 95% BCI [10.37, 16.359]). The total contacts that remained undetected, no solution entered, or incorrectly classified were 59 from OPTIMUS and 64 from OPTIX (14% of the total 896 contacts), and given the sum of each were approximately the same in each condition, we decided to exclude all 123 of these contacts from the analysis of the initial solution analyses, leaving 779 contacts. All contacts however were included for the TPE analyses reported further below.

3.2.1. Range error of initial solution

To determine range error, we took the absolute deviation of the initial solution range from the simulation truth range (expressed in yards) and divided it by the true range × 100 to compensate for the fact that magnitude of error was correlated with the true range (Pearson's $r = 0.247$, 95% BCI [0.179, 0.311], $BF_{10} > 1000$). The resulting percentage range error was not correlated with true range (Pearson's $r = 0.023$, 95% BCI [-0.047, 0.093], $BF_{01} = 18.10$). A Bayesian ANOVA of range error by HMI and task-load showed that the model including HMI only was preferred (posterior probability = 0.85). There was strong evidence for the main effect of HMI ($BF_{10} = 12.90$), and evidence against both task-load ($BF_{01} = 11.31$) and the interaction term ($BF_{01} = 8.66$). As indicated in Fig. 5, OPTIMUS yielded less range error ($M = 6.48\%$, 95% BCI [5.30, 7.66]) than OPTIX ($M = 9.91\%$, 95% BCI [8.12, 11.70]).

3.2.2. Course error of initial solution

To determine course error, we took the absolute deviation of the initial solution course (ATB) from the simulation truth course (expressed in degrees). A Bayesian ANOVA of course error by HMI and task-load showed that the model including HMI only was preferred (posterior probability = 0.82). There was evidence for the main effect of HMI

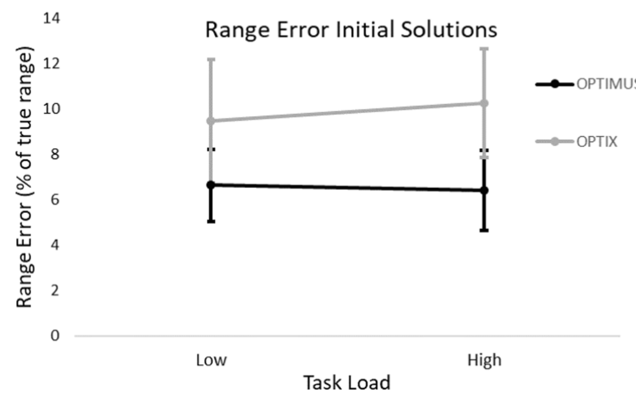

Fig. 5. Error between the initial estimated range of a contact from its true value, by HMI configuration and task-load. Range error has been scaled against the true range of the contact, such that the error is a percentage of the true range. Error bars reflect the 95% Bayesian Credible Interval of the mean.

($BF_{10} = 11.00$), and against the main effect of task-load ($BF_{01} = 9.18$) and the interaction term ($BF_{01} = 7.07$). As indicated in Fig. 6, OPTIMUS yielded lower course error ($M = 21.30$, 95% BCI [17.87, 24.72]) than OPTIX ($M = 29.35$, 95% BCI [25.53, 33.16]).

3.2.3. Time to generate initial solution

We performed a Bayesian ANOVA on Initial Solution Time (measured as the average time between the sequential classification of each contact, which included generating an initial solution) by HMI and task-load. The model including HMI only was preferred (posterior probability = 0.64). There was strong evidence for the effect of HMI ($BF_{10} > 1000$), insubstantial evidence for or against the effect of task-load ($BF_{01} = 1.96$) and evidence against the interaction ($BF_{01} = 9.05$).

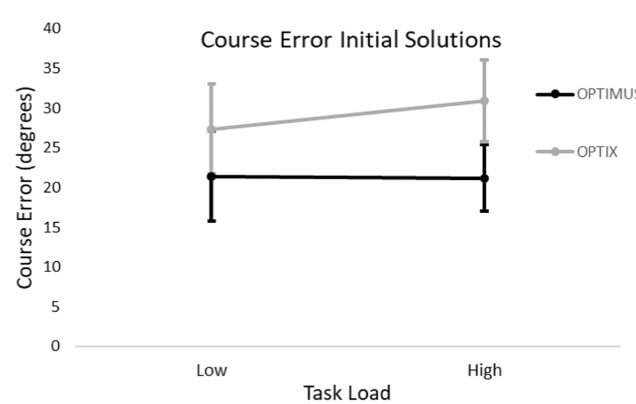

Fig. 6. Error between the initial estimated course of a contact from its true value, by HMI configuration and Task-Load. Course error is expressed in degrees. Error bars reflect the 95% Bayesian Credible Interval of the mean.

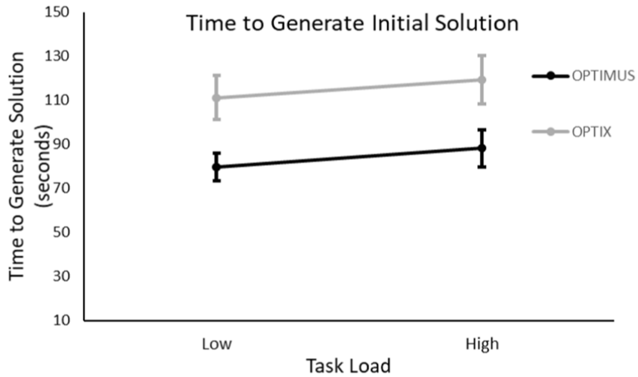

Fig. 7. Time to generate an initial solution per contact, by HMI configuration and Task-Load. Error bars reflect the 95% Bayesian Credible Interval of the mean.

As indicated in Fig. 7, OPTIMUS led to faster solutions (M = 84.49 s, 95% BCI [79.03, 89.95]) than OPTIX [M = 115.79 s, 95% BCI [108.18, 123.41].

3.2.4. Change in tactical picture error across a scenario

To the extent that operators can use a HMI to generate progressively better solutions, we would expect TPE to decline over time, with the better HMI leading to faster declines and/or more accurate final solutions. To test this, we extracted the average TPE over time for the four conditions (Fig. 8).

To assess the impact of task-load and HMI by controlling for the effect of Time (as TPE decreased as participants refined contact solutions), we fit a Bayesian Linear Regression on TPE by Time, HMI, task-load and all possible interactions using the "rstanarm" package for R (Goodrich et al., 2018). All main effects and interactions were included with posterior probability > 0.99. For most of the 24 min the conditions are ordered such that the TPE was poorer for OPTIX compared to OPTIMUS, and for high compared to low task-load, demonstrating main effects of HMI and task-load. In addition, by the end of the scenario the average TPE for OPTIMUS under high task-load fell below OPTIX when under low-task-load.

3.2.5. Total number of solutions per scenario

We performed a Bayesian Two Way Repeated Measures ANOVA on total number of solutions entered per scenario, separated by HMI and task-load. The model including HMI only was preferred (posterior probability = 0.67). There was strong evidence for the inclusion of HMI ($BF_{10} > 1000$), weak evidence against the inclusion of task-load (BF_{01} = 4.28), and insubstantial evidence for or against the interaction (BF_{10} = 1.08). On average, participants generated more solutions per scenario using OPTIMUS (M = 21.06, 95% BCI [19.29, 22.84]) compared with OPTIX (M = 15.58, 95% BCI [13.80, 17.35]).

3.2.6. Danger sector and rendezvous missions

We used Signal Detection Theory (Macmillan & Creelman, 2004) to calculate sensitivity (d') for detecting the two rendezvous and two danger-sector events in each scenario. A *Hit* was defined as the correct detection of each event while it was occurring. A *false alarm* was defined as a "photograph taken" of a distractor event. Sensitivity (d') was calculated as the difference between the hit-rate and false alarm-rate z scores. Higher d' indicates better ability to discriminate between target and noise.

A Bayesian ANOVA of d' by HMI and task-load showed that the model including HMI only was preferred (posterior probability = 0.59).

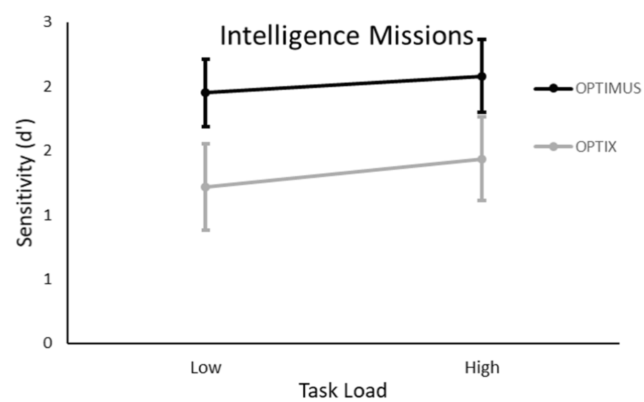

Fig. 9. The sensitivity (d') for the Rendezvous and Danger sector tasks. Error Bars represent the 95% Bayesian Credible Interval around the mean.

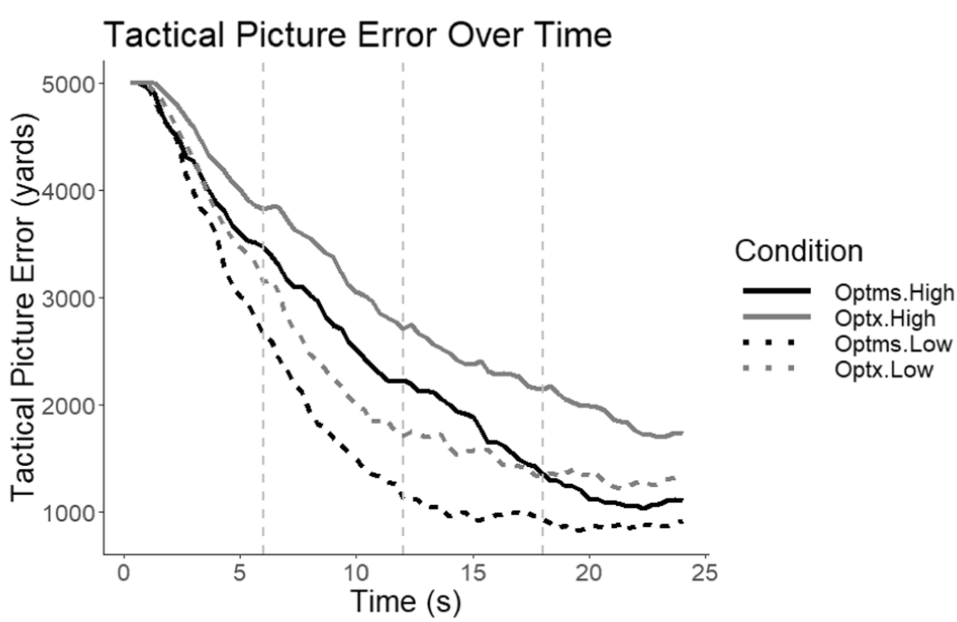

Fig. 8. Average Tactical Picture Error over scenario time, by HMI and Task-Load. Vertical dashed lines represent the commencement of each of the three simulation freezes to deliver the MCH and SA questions.

As indicated in Fig. 9, there was strong evidence for the inclusion of HMI ($BF_{10} > 1000$), such that OPTIMUS ($M = 2.02$, 95% BCI [1.76, 2.27]) showed higher d' values than OPTIX ($M = 1.33$ 95% BCI [1.07, 1.58]). There was inconclusive evidence for both the main effect of task-load on d' ($BF_{01} = 2.15$) and the interaction term ($BF_{01} = 2.64$).

3.3. Subjective workload

Overall, there were 512 assessments of subjective workload (ATWIT) (4 presentations per scenario × 4 scenarios × 32 participants). To control for the time of presentation (5, 10, 15 and 20 min into each scenario) we included a "Time Phase" factor. We performed a three-way Bayesian ANOVA on workload scores by Time Phase, HMI, and task-load (shown in Fig. 10, collapsed across Time Phase). The preferred model included the main effects of task-load and HMI, as well as their interaction (posterior probability = 0.497). There was strong evidence for the inclusion of task-load and HMI (both $BF_{10} > 1000$), such that subjective workload was greater in high task-load ($M = 6.24$, 95% BCI [6.05, 6.43]) compared to low task-load ($M = 5.59$, 95% BCI [5.39, 5.80]) scenarios, and greater for OPTIX ($M = 6.15$, 95% BCI [5.96, 6.34]) than OPTIMUS ($M = 5.69$, 95% BCI [5.48, 5.90]). There was insubstantial evidence for or against the inclusion of the HMI × Load interaction ($BF_{10} = 1.35$). There was moderate-to-strong evidence against all other effects (BF_{01} ranging from 7.42 to 18.60).

We conducted a Bayesian two-way ANOVA on MCH score by HMI and Task Load (Fig. 11). The model including only task-load was strongly preferred (posterior probability = 0.77), with strong evidence for the inclusion of task-load ($BF_{10} > 1000$) such that both HMI's had higher MCH scores (higher workload) under high-load (M = 5.43, 95% CI [5.09, 5.76]) than low-load (M = 4.43, 95% CI [4.03, 4.83]). There was moderate evidence that HMI did not influence MCH ratings ($BF_{01} = 5.02$), nor did the interaction of HMI and Task Load ($BF_{01} = 6.47$).

3.4. Situation awareness

Overall, there were responses to 1,890 SA queries (five SA queries × three SA blocks × 4 scenarios × 32 participants, minus 30 questions omitted due to technical errors). We performed a three-way Bayesian ANOVA on SA accuracy by HMI, task-load, and SA level (Fig. 12). The preferred model included the main effects of task-load and SA level, and their interaction (posterior probability = 0.45). The main effects of task-load and SA level (both $BF_{10} > 1000$) were strongly supported, and there was weak evidence for their interaction ($BF_{10} = 3.30$). There was evidence against the effect of HMI on SA accuracy ($BF_{01} = 8.06$), as well as its interaction with the other factors (all $BF_{01} > 10$). The effect of task-load and SA level was such that response to SA queries was less accurate with higher task-load and less accurate for higher SA levels. There was slight evidence for an interaction, such that there was very little

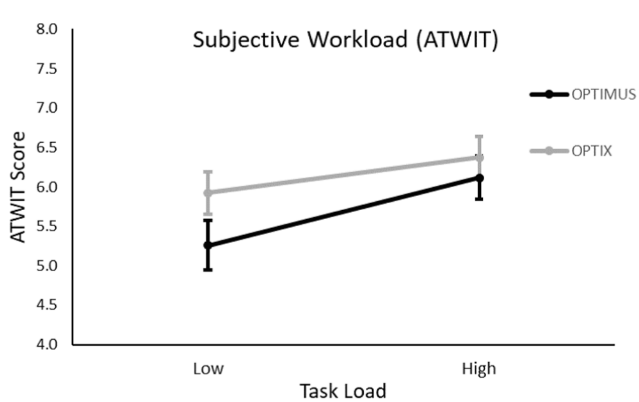

Fig. 10. Subjective workload (ATWIT) scores, by HMI configuration and task-load. Error Bars represent the 95% Bayesian Credible Interval around the mean.

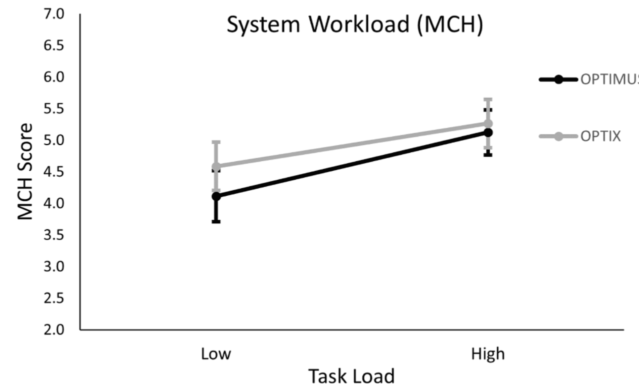

Fig. 11. Subjective workload (MCH) scores, by HMI configuration and task-load. Error Bars represent the 95% Bayesian Credible Interval around the mean.

difference in accuracy between high and low task-loads at Level 2 SA, compared with Levels 1 and 3.

We repeated this analysis for SA RT, including only correct responses. The preferred model included the main effects of HMI and SA level (posterior probability = 0.73). As seen in Fig. 13, RT increased from SA level 1 ($M = 8.96$ s, 95% BCI [8.33, 9.60]) to SA level 2 ($M = 21.02$ s, 95% BCI [19.65, 22.39]) to SA level 3 ($M = 23.01$ s, 95% BCI [20.85, 25.30]). Participants were slower to respond to SA queries at all levels when using OPTIMUS ($M = 19.52$ s, 95% BCI [18.25, 20.78]) compared to OPTIX ($M = 15.91$ s, 95% BCI [14.64, 17.19]).

3.5. Meta-Awareness

Overall, there were responses to 384 meta-awareness queries (1 query × 3 SA blocks × 4 scenarios × 32 participants) minus 30 questions omitted due to technical errors). For each participant (considering each HMI separately, but collapsed across task-load due to the relatively small number of queries), we correlated the confidence rating on each query (which assessed confidence in solution quality for a contact) with the position-error for the queried contact at the time of the query (i.e. computed an intra-individual meta-awareness value). We then subjected the correlations for each participant by HMI to a Bayesian Paired Samples T-Test. Both mean correlations were negative, such that participants were more confident in solutions that had lower position error. There was evidence for an HMI effect such that the correlation was stronger when participants used OPTIX ($M = -0.67$, 95% BCI [-0.80, −0.56]) compared to OPTIMUS ($M = -0.40$, 95% BCI [-0.58,-0.23], $BF_{10} = 7.32$).

3.6. Usability

A Bayesian paired-samples *t*-test revealed insubstantial evidence for or against a difference in system usability (as measured by the SUS) for OPTIMUS ($M = 65.47$, 95% BCI [53.82, 66.39]) compared to OPTIX ($M = 60.10$. 95% BCI [59.05, 71.89], $BF_{01} = 2.93$).

4. Discussion

4.1. OPTIMUS enhanced performance

We compared two concepts for digital periscope HMIs using a submarine simulation environment to elicit typical periscope operator tasks as well as mission-specific tasks. The impact of the two HMI concepts and low/high task-load on workload, SA, meta-awareness and performance was assessed. Our results indicated a range of task performance benefits arising from the use of OPTIMUS including faster initial contact solutions and better starting estimates for the range and course of contacts. These outcomes are in line with the perceptual (contact detection) and analytic (course and range estimates) benefits observed by

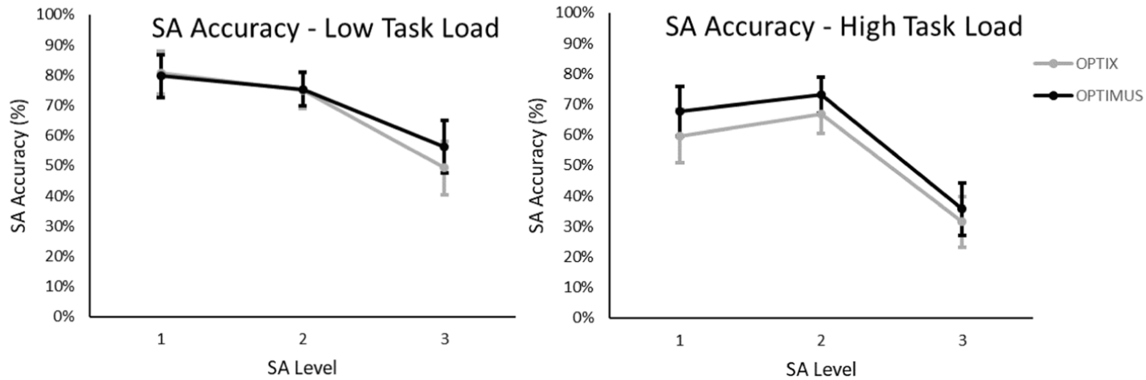

Fig. 12. Accuracy on SA questions by HMI configuration, Task Load and SA level. Error Bars represent the 95% Bayesian Credible Interval around the mean.

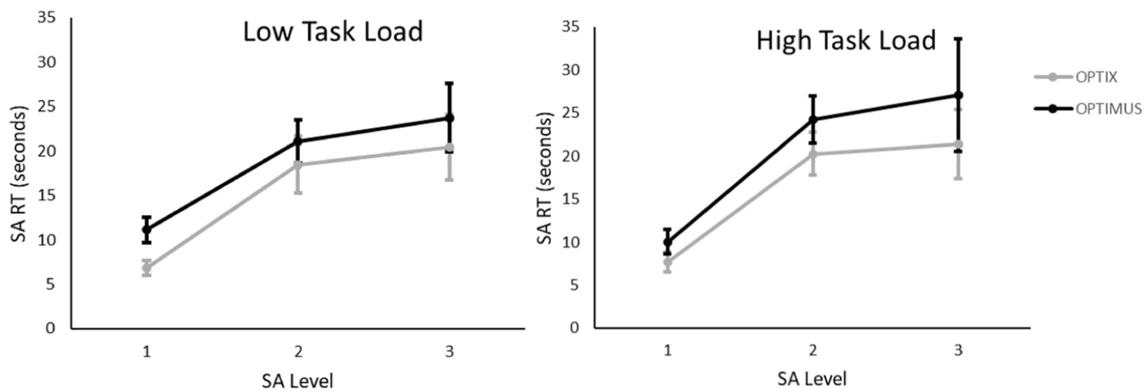

Fig. 13. Response times to correctly answered SA trials, by HMI configuration, Task Load, and SA level. Error Bars represent the 95% Bayesian Credible Interval around the mean.

Michailovs, Irons et al. (2022). These initial benefits compounded over time, resulting in lower average TPE throughout the scenarios, culminating in lower TPE when using OPTIMUS during high-task-load conditions than when using OPTIX during low task-load. This outcome demonstrates that OPTIMUS allowed participants to effectively deal with more contacts with no detriment to contact solution accuracy.

OPTIMUS also enhanced performance on the rendezvous and danger zone missions. Having a full panoramic field of view within OPTIMUS may have facilitated the detection of contacts that were converging for a rendezvous, compared to OPTIX, where the participant likely had to slew the periscope to view two contacts that could rendezvous. Additionally, while participants using both HMIs could note the bearings associated with danger sectors, only when participants used OPTIMUS could they observe contacts simultaneously entering those areas, and note which one may be within the range criteria. In addition, when using OPTIMUS, participants were able to generate a more accurate tactical picture, assisting with performance on the rendezvous and danger zone missions.

4.2. *OPTIMUS reduced perceived workload*

Access to 360° panoramic video in OPTIMUS, and the provision of the advanced digital tools, reduced perceived workload (as measured by ATWIT). Given that workload and performance are intrinsically linked (Hancock and Matthews, 2019), and workload is typically discussed in the context of the amount of cognitive resources needed to achieve a given performance level (Parasuraman et al. 2008), our findings imply that participants required fewer resources to achieve superior performance when using OPTIMUS. This indicates that while OPTIMUS displays more information than OPTIX, this did not lead to cognitive overload. The fact that participants reported lower workload when using OPTIMUS reflects that they would also be in a potentially better position to respond to future task demands or unexpected events (Loft et al. 2007; Sperandio, 1978).

The absence of costs arising from the greater information load imposed by OPTIMUS could potentially reflect a trade-off between information load and participant uncertainty. For example, while the panoramic display in OPTIMUS imposed greater information load than the narrower-focus OPTIX display, it also allowed the distance between contacts and their range and ATB to be ascertained more quickly. Given that uncertainty can contribute to higher workload (Parasuraman et al., 2005), it may thus be that the increased information load in OPTIMUS may have traded-off against lower uncertainty in OPTIMUS. This reasoning is in line with Innes et al. (2019) who showed advancements in helicopter HMIs provided pilots with additional visual information to perform tasks without concomitant workload costs.

4.3. *OPTIMUS degraded situation awareness and meta-cognitive awareness*

When collapsed across both HMI conditions, both the RT and accuracy of responses to SA queries was moderately-to-strongly correlated with performance (TPE), and on SA accuracy the HMI's did not statistically differ. SA also moderately correlated with the rendezvous and danger sector mission performance. This implies that participants with better SA also performed better in a given scenario. However, it does not directly imply an advantage to one HMI over the other. This is because the correlations reflect within-subject measures at the scenario level. So for example, if a participant performed better in scenario A than B, it's likely their workload was also lower and SA higher in scenario A than B. It can still logically be the case (as was found) that SA is lower for OPTIMUS despite better performance, as we are then dealing with different levels of statistical analysis. In addition, there are complex interactions in our between-HMI data. For example, the OPTIMUS console reduced subjective workload (which itself would suggest improved performance based on the correlations in Table 2), so considering only SA and performance in isolation does not tell the full story of the data.

When we directly compared SA in the two HMI conditions, OPTIMUS participants took longer to correctly respond to SA queries and we offer two possible explanations. The first is that the OPTIMUS display had greater *visual complexity* – that is it had more contacts on the display, displayed a greater variety of elements (e.g., different classifications of contacts), and displayed the interconnections between elements (e.g., distances between contacts). These are key characteristics that typically define display complexity (quantity, variety, and interconnections of display elements – Donderi, 2006). This increased complexity did not merely potentially make visual search more difficult (thereby slowing responses to the SA prompts), but rather the increased complexity also likely degraded the resulting quality of operator mental models, slowing participants ability to subsequently find task-relevant information on the display, and hence slowing SA RT.

The second possible explanation for the slower SA RT is the way OPTIMUS displayed this information, which may have been suboptimal for later retrieval. Many theorists contend that operators create partial representations of the world, leaving as much information as possible in the external environment to be accessed as required (Chiappe et al., 2012; Salmon et al., 2008). By displaying the entire horizon in a panorama, OPTIMUS may have encouraged/supported increased off-boarding of information and greater reliance on the interface, over the mental representation of information that may have been more necessary using OPTIX (as evidenced in the increased subjective workload of participants using OPTIX as a result of needing to hold more information in memory). Increased off-boarding using OPTIMUS would be detrimental to SA if the HMI made it more difficult for retrieval. Notable in this regard is that in order to map the 3D world to a 2D plane, OPTIMUS made some sacrifices in the geospatial representation of information (for example, contacts that are on the port [left] side of Ownship may appear on the right side of the screen). This meant that participants may have had to inhibit the contact position on the 2D display when updating their mental model of the 3D world, elevating the cognitive burden of information retrieval regarding where to find information on displays (the Simon Effect: Hommel, 1993; 2011) and potentially slowing SA responses.

Given the potential drawbacks of the OPTIMUS HMI to SA, it might appear counter-intuitive that we found that workload was reduced and performance better with OPTIMUS, but only if we consider these relationships in isolation. As noted earlier, the increased complexity of the OPTIMUS display could have increased perceived workload, but the OPTIMUS panorama view and digital analysis tools likely greatly reduced the uncertianty and perceived workload associated with developing contact solutions, outweighing the impact of increased display complexity on perceived workload. In addition, having good SA will certainly increase the probability of good performance, however it does not guarantee it, as other factors can have an effect (Endsley, 1995b). We contend that any performance degradation arising from poorer SA, likely traded-off against the advantages to performance offered by the reduced workload and the more precise nature of the OPTIMUS analysis tools. A direct cost of reduced SA to performance is most likely when some form of unexpected non-routine event occurs or there is a sudden increase in workload in which task demands exceed operator capacity (Endsley, 1995a; Vu, & Chiappe, 2015), conditions we did not include in our current study.

It should be noted that our SA measure was a unique hybrid tool (combining features of SAGAT and SPAM) which recognises that operators can both form an internal representation of the situation to be called upon when needed, as well as offload information to their environment to be retrieved as required (Chiappe et al., 2016). We recognize the extended ongoing debate on SA, and the lack of a universally accepted gold standard measurement approach (Endsley, 2015; Pritchett, 2015; Vu & Chiappe, 2015). Our decision to leave the display available to the participants whilst answering the SA probes maintains the integrity of the SA measure as it still probed the participant about what was going on (e.g. where is contact X?), whilst allowing access to the information needed. Those participants who readily accessed this information to achieve their tasks during the experimental scenarios we took to have greater SA, and would therefore know where the information was held on the HMI and access it faster following SA queries. We found no evidence of a speed-accuracy tradeoff in SA responses (see Endsley, 2021), with evidence against the effect of HMI on SA accuracy and a significant negative correlation between SA accuracy and RT, indicating that faster SA responses were more likely to be accurate. Nontheless, given that our hybrid measure is novel as compared to the more validated mainstream approaches, it would be useful for future research to replicate the current findings with SA measures such as SAGAT or SPAM to obtain convergent validity.

4.4. Meta-cognitive awareness

Participants also had lower meta-awareness when using OPTIMUS compared to OPTIX, in that participants were less able to distinguish good solutions from poorer ones. This may in part be due to longer distances between their solution bearing and the actual bearing in OPTIMUS (the width of the 360° panorama over which the bearings were 'spread' was 5 × the width of the OPTIX display and thus further apart) which could appear to inflate the distance and thus the perceived error of older solutions. Although not a perfect way of judging solution accuracy (depends on how long ago the solution was placed and a poor solution placed recently could look 'closer' than a good solution over a longer period of time), it was nonetheless used as an approximate indicator in both HMIs.

The act of physically manipulating a manual tool in OPTIX to estimate the ATB may also have helped make salient the solution accuracy for a particular contact, making it easier to recall. An alternative (or additional) explanation for reduced meta-awareness when using OPTIMUS is that the accuracy of a solution was simply not as relevant when using OPTIMUS. As participants could physically see a contact at all times it was perhaps not important to know exactly how close an estimated solution was, whereas OPTIX required a more sophisticated prioritisation strategy to manage the slewing of the periscope.

4.5. Limitations and conclusions

The near-ideal visibility and steady sea state conditions provided reasonably high contrast stimuli (contacts stood out from the sea surface and sky). While such conditions can exist for short-range contacts in submarine operations (e.g. under calm sea state in daylight), often the contrast between contacts and their surroundings is much lower (e.g. when sea state, weather or lighting conditions are poorer, for contacts on the horizon).

OPTIMUS facilitated more precise range and course estimates, consistent with our earlier work (Michailouvs, Irons et al., 2022). It remains an open question whether this benefit results primarily from the digital nature of the tool (versus the manual OPTIX ATB tool), or whether the more detailed digital contact models facilitated a match-to-sample process (Bell & Badcock, 2008) that allowed more precise estimates of contact parameters. Experimentally, this question could be tested by adapting the OPTIMUS digital ATB tool to use a generic model akin to the manual tool used in OPTIX (e.g., a digital representation of Fig. 4). Future research should consider this, as it may speak to the utility of OPTIMUS when specific intelligence is missing for some contacts.

The tasks were purposely designed to be routine and a relatively simple analogue of submariners' fundamental tasks so to be achievable for novice participants. Additionally there was limited consequence of any mistakes, unlike that for operators in complex, high-risk environments. In light of these points, it is critical to follow up this work with higher fidelity environments using expert (actual submariners) participants.

Finally, an important next step is to examine the extent to which OPTIMUS can improve performance and reduce workload in a team environment that requires communication with operators in other roles such as track management, sonar and target motion analysis (Michailouvs, Pond et al., 2022). It may the case that the extended time taken to find SA related information or the limited meta-cognitive awareness when using OPTIMUS has detrimental effects on team-work and team performance.

In conclusion, the present study extends previous evidence (Michailouvs, Irons et al., 2022) that future submarine technologies should not be limited to digital representations of a traditional"view down a bearing" periscope format. OPTIMUS supported faster and more accurate performance and lowered perceived workload, compared to OPTIX. However, SA and performance meta-awareness was poorer when participants used OPTIMUS. We conclude that while the OPTIMUS HMI concept holds a considerable advantage in that it can improve performance whilst reducing perceived workload, the SA and meta-awareness decrements are potentially problematic and warrant further investigation. Taken together however, the results support continued digital innovation in the submarine control room.

5. Author Note

This research was supported by a Research Grant (MyIP 9277) from the Defence Science and Technology Group, Australia awarded to Loft, Michailovs, Visser, Bell, & Pinniger.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 570–585.

Bell, J., & Badcock, D. R. (2008). Luminance and contrast cues are integrated in global shape detection with contours. *Vision Research, 48*, 2336–2344.

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). London: Taylor & Francis.

Chiappe, D., Morgan, C. A., Kraut, J., Ziccardi, J., Sturre, L., Strybel, T. Z., & Vu, K.-P.-L. (2016). Evaluating probe techniques and a situated theory of situation awareness. *Journal of Experimental Psychology: Applied, 22*, 436–454.

Chiappe, D., Strybel, T. Z., & Vu, K. P. L. (2012). Mechanisms for the acquisition of situation awareness in situated agents. *Theoretical Issues in Ergonomics Science, 13*, 625–647.

Durso, F. T., & Dattel, A. R. (2004). SPAM: The real-time assessment of SA. In S. Banbury & S. Tremblay (Eds.), *A Cognitive Approach to Situation Awareness* (pp. 137-154).

Donderi, D. C. (2006). Visual complexity: A review. *Psychological Bulletin, 132*, 73–97.

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

Endsley, M. R. (1995b). Measurement of situation awareness in dynamic systems. *Human Factors, 37*, 65–84.

Endsley, M. R. (2015). Situation awareness misconceptions and misunderstandings. *Journal of Cognitive Engineering and Decision Making, 9*, 4–32.

Endsley, M. R. (2021). A Systematic review and meta-analysis of direct objective measures of situation awareness: A comparison of SAGAT and SPAM. *Human Factors., 63*, 124–150.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19–23.

Goodrich, B., Gabry, J., Ali, I., & Brilleman, S. (2018). rstanarm: Bayesian applied regression modeling via Stan. *R package version, 2*, 1758.

Hancock, P. A., & Matthews, G. (2019). Workload and performance: Associations, insensitivities, and dissociations. *Human Factors, 61*, 374–392.

Hollnagel, E. (2002). Time and time again. *Theoretical Issues in Ergonomic Science, 3*, 143–158.

Hommel, B. (2011). The Simon effect as tool and heuristic. *Acta psychologica, 136*, 189–202.

Hommel, B. (1993). The role of attention for the Simon effect. *Psychological Research, 55*, 208–222.

Howard, Z. L., Innes, R., Eidels, A., & Loft, S. (2021). Using past and present indicators of human workload to explain variance in human performance. *Psychonomic Bulletin & Review., 28*, 1923–1932.

Innes, R., Howard, Z. L., Thorpe, A., Eidels, A., & Brown, S. (2019). The effects of increased visual information on cognitive workload in a helicopter simulator. *Human Factors, 63*, 788–803.

Kruschke, J. K., & Liddell, T. M. (2018). Bayesian data analysis for newcomers. *Psychonomic Bulletin & Review, 25*, 155–177.

Loft, S., Bowden, V., Braithwaite, J., Morrell, D. B., Huf, S., & Durso, F. T. (2015). Situation awareness measures for simulated submarine track management. *Human Factors, 57*, 298–310.

Loft, S., Jooste, L., Li, Y. R., Ballard, T., Huf, S., Lipp, O. V., & Visser, T. A. (2018). Using situation awareness and workload to predict performance in submarine track management: A multilevel approach. *Human Factors, 60*, 978–991.

Loft, S., Sanderson, P., Neal, A., & Mooij, M. (2007). Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications. *Human Factors, 49*, 376–399.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.

Michailovs, S. I. C., Irons, J., Howard, Z. L., Pond, S., Schmitt, M. L., Stoker, M., Visser, T., Bell, J., Pinniger, G., Fitzgerald, M., Huf, S., Carter, O., & Loft, S. (2022). *under review). Augmenting human cognition with a digital submarine periscope human-machine interface*. Behaviour and Information Technology.

Michailovs, S. I. C., Pond, Schmitt, M. L., S., Irons, J., Stoker, M., Visser, T., Huf, S., and Loft, S. (2022). The impact of information integration in a simulation of future submarine command and control. *Human Factors*.

Moacdieh, N., & Sarter, N. (2015). Display clutter: A review of definitions and measurement techniques. *Human Factors, 57*, 61–100.

Parasuraman, R., Galster, S., Squire, P., Furukawa, H., & Miller, C. (2005). A flexible delegation-type interface enhances system performance in human supervision of multiple robots: Empirical studies with RoboFlag. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans, 35*, 481–493.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making, 2*, 140–160.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*, 3–25.

Pritchett, A. (2015). Preface to the JCEDM special issue on situation awareness. *Journal of Cognitive Engineering and Decision Making, 9*.

Roberts, A. P., Stanton, N. A., & Fay, D. T. (2018). Go deeper, go deeper: Understanding submarine command and controlduring the completion of dived tracking operations. *Applied Ergonomics, 69*, 162–175.

Roberts, A. P., Stanton, N. A., Pope, K. A., & Fay, D. T. (2021). To utilize automation or not to utilize automation, that is the question: An evaluation of how drills and procedures impact optronics mast usage from a sociotechnical systems perspective. *Applied Ergonomics, 97*, Article 103543.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56*, 356–374.

Salmon, P. M., Stanton, N. A., Walker, G. H., Baber, C., Jenkins, D. P., McMaster, R., & Young, M. S. (2008). What really is going on? Review of situation awareness models for individuals and teams. *Theoretical Issues in Ergonomics Science, 9*, 297–323.

Sperandio, J. C. (1978). The regulation of working methods as a function of workload among air traffic controllers. *Ergonomics, 21*, 195–202.

Stanton, N. A., & Bessell, K. (2014). How a submarine returns to periscope depth: Analysing complex socio-technical systems using Cognitive Work Analysis. *Applied Ergonomics, 45*, 110–125.

Stein, E. S. (1985). *Air traffic controller workload: An examination of workload probe*. NJ: Retrieved from Atlantic City.

Vu, K.-P.-L., & Chiappe, D. (2015). Situation awareness in human systems integration. In D. A. Boehm-Davis, F. T. Durso, & J. D. Lee (Eds.), *APA handbook of human systems integration* (pp. 293–308). American Psychological Association.

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., … Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review, 25*, 35–57.

Wierwille, W. W., & Casali, J. G. (1983). A validated rating scale for global mental workload measurement applications. *Paper presented at the Proceedings of the Human Factors Society Annual Meeting*.