Using the situation present assessment method to measure situation awareness in simulated submarine track management

Shayne Loft* and Daniel B. Morrell

School of Psychology,
The University of Western Australia,
35 Stirling Hwy, Crawley, Western Australia 6009
E-mail: Shayne.Loft@uwa.edu.au
E-mail: MorrellDaniel@gmail.com
*Corresponding author

Samuel Huf

Submarine Human Systems and Analysis,
Maritime Operations Division,
Defence Science and Technology Organisation,
A51, HMAS Stirling,
P.O. Box 2188, Rockingham, Western Australia 6967
E-mail: Samuel.huf@dsto.defence.gov.au

Abstract: Extant query-based measures of situation awareness that require interruption of dynamic tasks can be inappropriate in certain work contexts. A candidate real-time query-based method, the situation present assessment method (SPAM), was used to examine participants’ situation awareness in simulated submarine track management. The simulation and situation awareness queries were based upon a goal-directed task analysis conducted with expert submariners. Participants decided whether to engage enemy vessels using rules based on clear firing corridors for submarine torpedoes, and responded to SPAM queries that assessed their awareness of the current and future display situation. Participants were less accurate and slower to respond to SPAM queries that assessed awareness of the future display situation compared to the current display situation. Consistent with theoretical frameworks of visual sampling and attention-situation awareness integration, the relative predictive utility of current- and future-SPAM queries depended on the nature of the display information relevant to performance goals.

Keywords: situation present assessment method; SPAM; situation awareness; submarine track management; attention; visual sampling.


Biographical notes: Shayne Loft is an Associate Professor at the University of Western Australia. He received his PhD in 2004 from the University of Queensland.
1 Introduction

1.1 Situation awareness in submarine track management

Arguably, all of the sensor, navigation, and communication systems in the submarine control room are designed to assist tactical decision making. However, assessing the degree to which such systems optimally support the submariner is difficult because often there is little overt variation in performance between expert submariners in routine operations. A more sensitive method is to examine changes in the quality of situation awareness (SA) maintained by operators to perform task goals. Although a number of theoretical definitions of SA have been proposed, Endsley’s (1995a) definition, “the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future”, is the most widely accepted. The likelihood of human error increases as the operators’ understanding of the current situation, and their ability to anticipate the future consequences of observed events or their actions, becomes degraded. SA has been shown to provide the foundation for safe and efficient performance in work systems as diverse as combat aviation (Vidulich et al., 1995), anaesthesiology (Gaba et al., 1995), and air traffic control (Durso et al., 1998). Increased SA is assumed by naval technology developers to be provided by advances in submarine sensor and combat systems. However, given that SA only really exists in the minds of operators, this assumption is impossible to verify unless the SA requirements of submariner tasks are accurately defined and measured.

One prototypical submarine control room task where it is essential to measure SA is track management (Kirschenbaum, 2011). Track management requires a console operator to make use of the output from various sensors to create, monitor, and update a digital representation of the outside environment (a tactical picture). The accuracy of the tactical picture that is communicated by the track manager to the Officer on Watch (OW) affects the submarine’s ability to complete missions such as anti-surface or submarine warfare, intelligence gathering, or blockading. While a handful of papers have reported broad task analyses of submarine personnel (Kirschenbaum, 2011; Ehret et al., 1997, 2000), no study has been published to date that has defined or measured SA in simulated submarine track management.

There are a variety of subjective methods used to measure SA, including asking operators to rate their own SA using measures such as the Situation Awareness Rating Technique (SART), or having Subject Matter Experts (SMEs) evaluate the degree to
which other operators exhibit good SA. However, objective query-based techniques are likely to provide more accurate SA measurement. The Situational Awareness Global Assessment Technique (SAGAT) is the most widely used and validated query-based SA measure (Endsley, 1995b). SAGAT focuses on the product of SA using recall techniques, uncovering information of which the operator is consciously aware. SAGAT employs timed freezes in a task, during which elements in the display are blocked, and operators recall what was happening at time of the freeze. However, the fact that SAGAT disrupts the operator from the tasks at hand limits its use in submarine control rooms. This is because much of the cognitive work of the submarine control room team involves the management of uncertainty stemming from the fact the submarine’s passive sensors do not yield veridical range information. Temporarily removing access to the display would lead to operators’ mental picture of the situation being quickly lost. Hence, assessing SA in real-time without pausing the task operation appears the most advantageous in a practical sense for submarine track management.

The situation present assessment method (SPAM; Durso and Dattel, 2004) is potentially suited to the submarine control room because it measures SA in real-time without freezing task operations. SPAM distinguishes workload from SA by warning the operator that a question is in the queue, and waiting until the operator accepts the question. SPAM query accept time is measured as the time between when the experimenter asks the participant whether they were ‘ready’, and the time that the participant accepts the question. Following this, the SA question is asked and SPAM query response time is measured as the time between when the experimenter completes asking the question and the time the participant responds (Durso and Dattel, 2004). The logic underlying SPAM is that operators who have better SA will know where to find appropriate information and thus be able to respond faster or more accurately. It is arguable that the SPAM method simulates the command team interactions typical of real-time submarine operations (i.e., live unsolicited requests for information from the OW to the track manager). SPAM has been effectively used to measure SA and to predict performance in the simulated work domains of air traffic control and driving (Durso et al., 2006, 1998; Durso and Dattel, 2004). However, the utility of SPAM, or in fact the utility of any measure of SA, has yet to be evaluated in a simulated submarine track management context.

1.2 The current study

We conducted a goal-directed task analysis by observing Royal Australian Navy submariners in combat system track management simulations at the Defence Science and Technology Organisation, Australia. We also interviewed three submariners (all male, with > 5 years’ experience) as SMEs. During the interviews, the SMEs were asked to describe the task goals of submarine track management, the decisions that must be made to achieve these goals, and the dynamic information requirements needed to support the decisions. This task analysis was generally consistent with what Marr (1982) referred to as the computational level (inputs and outputs of task goals), in that it provided a characterisation of the behaviour that the track manager is trying to achieve (Loft et al., 2009).
Figure 1  A screenshot from the SIM(sub) fremantle condition

Notes: Six merchant vessels are visible within the merchant shipping lane, six friendly vessels are patrolling around the ran patrol area and six enemy vessels are patrolling around the area of operations. Ownship is represented as the cross in the middle of the display with black circles radiating out from it representing 5 km intervals. An indication of how far each type of vessel can move in 5 minutes is given in the top right corner; 1.25 km for friendly and enemy vessels, and 2.5 km for merchants. Enemy vessels four and six are clear to be fired upon; while enemy vessels one, two and three are in violation of the 30° arc rule. Enemy vessel five is in violation of the two kilometre proximity rule.

We then used the data from the goal-directed task analysis to develop a task simulation that emulated the core tasks of the submarine track manager. The role of the submarine track manager is to manage the localisation and tracking of all vessels (contacts) detected by the Ownship’s sensors. Contacts can include enemy, friendly, and neutral vessels (such as commercial or recreational vessels). Track managers monitor the location and behaviour (bearing, range, speed) of contacts in relation to Ownship and to strategic landmarks. The OW often asks the track manager questions regarding aspects of the tactical picture (e.g., is the enemy submarine manoeuvring into position to attack a
friendly?), or regarding the ability of the Ownship to engage the enemy (e.g., will firing on an enemy endanger other vessels?). The interface in Figure 1 shows a number of contacts (friendly, neutral, and enemy vessels). Using the display information and tools provided, participants were required to make decisions about the Ownship’s ability to engage enemy vessels using two rules of engagement based on clear firing corridors for submarine torpedoes:

1. Ownship was not permitted to fire upon an enemy vessel if any other vessel (including enemy) was within a 30° arc from Ownship’s line of sight to the target enemy vessel.

2. Ownship was not permitted to fire upon an enemy vessel if that enemy vessel was within 2 km of another vessel.

The performance measures were designed to emulate how submarine track managers communicate engagement decisions to the OW in naval operations. One performance measure was the speed and accuracy of enemy engagement responses when the participant was asked by the OW (the experimenter) whether it was safe to fire on a particular enemy vessel (OW task). A second performance measure was the speed and accuracy with which participants alerted the OW when an enemy vessel became clear or unclear to fire upon (state change task). This localisation and communication of enemy engagement information is an overt performance action; except that the information is vocally passed on at the required time (either on request for the OW task, or proactively for the state change task) rather than the trigger actively being pulled by the track manager.

The SPAM queries, also developed from the goal-directed task analysis, enabled measurement of the operators’ SA of the current and future location, as well as the direction (bearing) of different vessels in relation to both each other and the Ownship. Specifically, these SPAM queries tapped into the display information which our SMEs rated as being critical to the operators’ ability to make accurate and fast enemy engagement decisions on the state change task and the OW task. Current-orientated SPAM queries required participants to report on events that were currently occurring on the track management display, and future-orientated SPAM queries required participants to project the current state of the task environment and report on events that would occur in the short-term future.

This study is the first, to our knowledge, to have examined SA in simulated submarine track management, and makes three specific contributions to the literature. The first is that we examine whether participants are slower or less accurate to respond to future-oriented than to current-oriented SPAM queries. While such a finding would make intuitive sense, past research has been mixed. While Durso et al. (2006) and Dao et al. (2009) reported that participants were slower or less accurate to respond to future-oriented compared to current-oriented SPAM queries in simulations of air traffic control and piloting, other studies have reported null effects (e.g., Strybel et al., 2008). In addition, the updating of display elements evolves comparatively more slowly in simulated submarine track management than in air traffic control or piloting. That is, the vessels do not change position greatly over time, which may render the distinction between current and future SA less relevant in submarine track management than in other
complex control tasks. Thus, it was unclear whether we would replicate the Durso et al. (2006) and Dao et al. (2009) findings.

The second contribution is that we directly test the validity of current-oriented and future-oriented SPAM queries for predicting performance on the OW and state change tasks. To our knowledge, only two prior studies (Durso et al., 1998, 2006) have examined whether variability in the speed or accuracy of SPAM responses can predict unique variance in performance. For example, Durso et al. (1998) reported that current-orientated SPAM queries predicted SME subjective ratings of air traffic controller performance, whereas future-orientated SPAM queries predicted remaining controller action counts. Our statistical analysis method is more rigorous than that of Durso et al. (1998, 2006) because we control for SPAM query accept time (which is a SPAM proxy for workload) in our regression models when using SPAM query accuracy and SPAM query response time to predict performance. This statistical control is essential because in dynamic tasks online SA questions can potentially arrive when the operator is unusually busy, or unusually idle. It is possible then that latency of response to the SA question will reflect workload as well as SA. Durso and Dattel (2004), and Durso et al. (2006) argue that the delay in accepting the SPAM query should be interpreted as a measure of workload, and that the subsequent time taken to then answer the SPAM query should be interpreted as a measure of SA, thereby allowing the effects of workload and SA to be distinguished. If this assumption is true, SPAM query accept time must be controlled for before entering SPAM query accuracy and SPAM query response time into the regression model to predict operator performance.

The third contribution of the current study is that we make predictions based on theory regarding the differential effect of current-orientated and future-oriented SPAM queries on the submarine track management performance measures. We expected participants to make accurate (near ceiling) and quite rapid OW task responses, due to the fact that the experimenter would prompt them when specific enemy vessels required an engagement decision. OW task response times should be most dependent on participants’ ability to access and interpret information related to the current state of the display. On this basis, we predicted that OW task response time would be most uniquely predicted in our regression analyses by current SPAM queries. In contrast, we expected more variation in accuracy, and considerably longer response times, in the state change task because it would require more proactive control behaviour. Participants would need to divide graded resources of attention between multiple evolving display events (Bundesen, 1990). According to the attention-SA model of McCarely et al. (2002), see also Wickens et al. (2003), and models of optimal visual sampling theory (Senders, 1964; Sheridan, 1970), the likelihood that participants sample display events related to a particular enemy vessel will depend on the expectancy that those events have relevance to the state change task. The McCarely et al. (2002) model assumes that information collected by an attention module is fed forward to a cognitive SA-updating module where it is integrated within the operator’s mental representation. Information from this representation is then fed back to the attention module to guide future information sampling. Thus, in order to prioritise when and where to allocate attention in order to alert the OW of state changes, participants will need to project how the relationships between vessels on the display will evolve. On this basis, we predicted that performance on the state change task would be most uniquely predicted in our regression analysis by future SPAM queries.
Using the situation present assessment method

2 Method

2.1 Participants

Fifty-five participants from the University of Western Australia (29 males, 26 females) with mean age of 23.62 years participated for either course credit or remuneration of AUS$20. Participants were recruited using an online research management system.

2.2 SIM(sub) paradigm, materials, and procedure

2.2.1 SIM(sub) paradigm

SIM(sub) was programmed in MATLAB (Mathworks Inc., 2010). As illustrated in Figure 1, the display showed an area of operation that had a radius of 15 km, with black circles radiating out from the centre representing 5 and 10 km intervals (range rings). There were three types of vessel; enemy vessels (red squares), merchant vessels (white triangles) and friendly vessels (green triangles). These vessels moved constantly and entered and exited the screen within scenarios. Each vessel had a displayed number between 1 and 6, a history marker indicating movement for the previous five minutes, and a yellow line protruding from its icon indicating heading. Merchant shipping lanes were represented as white lines and Royal Australian Navy bases and patrol areas as green circles. The Ownship position was represented by a cross at the centre of the screen. The Ownship was stationary, reflecting operational circumstances requiring stealth where the submarine is remaining relatively idle to gather intelligence. The top right corner of the display showed the distance each type of vessel could move in five minutes (1.25 km for friendly and enemy vessels, and 2.5 km for merchant vessels). Three 12.5 minute scenarios were presented. The scenarios were named Fremantle, Exmouth, and Broome, and were designed to have equated task demands.

Participants were required to apply rules of engagement to make decisions regarding the Ownship’s ability to fire upon enemy vessels. The rules of engagement were that:

1. Ownship was not able to fire upon an enemy vessel if any other vessel (including an enemy) was within a 30° arc from Ownship’s line of sight to the targeted enemy vessel (including the arc extending behind the target vessel).

2. Ownship was not able to fire upon an enemy vessel if that enemy was within 2 km of another vessel (including other enemy vessels).

Participants were issued with a Perspex 30 degree arc tool and a ruler representing 5 km that they could place over the display to take accurate measurements.

2.2.2 Objective performance measures

For the OW task, participants were asked by the OW (the first experimenter) whether it was safe to fire on a particular enemy vessel. Three OW requests were interspersed at varying intervals over each scenario. An example of the probe/response format for the request questions was: OW: ’Are we clear to fire on enemy vessel 2?’ Participant: ‘Yes’ or ‘No’. A second experimenter used a laptop (capable of millisecond timing) to record when the first experimenter finished asking the OW task question, and to record when the participant finished answering that question. The difference between these two key
presses was the OW task response time. The state change task required participants to alert the OW when an enemy vessel became clear or unclear to be fired upon (referred to as a ‘state change’). Participants identified a state change by informing the experimenter of the enemy vessel number and its new state (for example, ‘Enemy 2 clear’). Five enemy vessels changed state per scenario. The second experimenter used the laptop to record when the participant identified the state change. The time in the task that the vessel first changed state was pre-recorded in the scenario script. These two event times were then used to calculate state change response time.

2.2.3 SPAM queries

Four SPAM queries were constructed for each scenario. Participants were first asked by the first experimenter whether they were ready for a SPAM question. Participants responded ‘yes’ when they were ready to accept the question. SPAM query accept time was measured as the time between when the first experimenter asked the participant whether they were ‘ready’ and the time the participant accepted the question. The second experimenter used a laptop to record when the experimenter asked the participant if they were ready for a SPAM question, and when the participant responded ‘yes’. Following this, participants were asked the SPAM question by the first experimenter, with correct answers being yes/no or numeric responses. SPAM response time was measured as the time between when the experimenter completed asking the SPAM question and the time that the participant finished answering that question (Durso and Dattel, 2004). The second experimenter used the laptop to record when the SPAM questions were asked and answered.

Two current and two future SPAM queries were presented within each of the three scenarios. These were designed to assess participant awareness of relative location and course of different vessels. As a full illustrative example, in the Fremantle scenario the two current SPAM queries were ‘How many merchant vessels are heading toward Ownship?’ and ‘How many friendly vessels are within 30 degrees of Ownship to enemy vessels?’; and the two future SPAM queries were ‘In 5 minutes, how many friendly vessels will be within 5 km of Ownship?’ and ‘In 5 minutes, how many enemy vessels will cross Ownship’s line of sight with Friendly vessels?’ Queries varied in content enough to reduce the ability of participants to anticipate SPAM queries (Durso et al., 2006). All queries were evaluated by our SMEs for their relevance to enemy engagement decisions. For example, clearly it was crucial for participants to have SA of the relative heading and distance of vessels to Ownship, or SA of which display elements were contained within or would cross Ownship-enemy 30 degree arcs in the future. Note that we did not expect participants to be actively ‘counting events’; rather, the logic of SPAM is that operators who have better SA will know where to find appropriate information to answer the SPAM query and thus be able to respond faster and more accurately when prompted.

2.2.4 SART measure

Following completion of each scenario participants completed a ten-dimension SART questionnaire (Selcon and Taylor, 1990). As it is most common for SART ratings to be given by participants only once at the end of scenarios (see Rousseau et al., 2011), we
followed this practice. Participants provided ratings on a number of scales regarding the degree to which they perceived:

1. the demands on their resources
2. their supply of resources
3. their understanding of the situation.

A combined SART score, derived from the combination of the three subscales, was used to assess overall subjective SA; Combined SA = Understanding – (Demand-Supply).

Participants first watched a 15-minute narrated PowerPoint presentation. This presentation introduced simulated submarine track management and detailed the tasks that participants would perform. Participants then completed a seven-minute practice trial. Participants then completed the three 12.5-minute experimental scenarios, the presentation of which was counterbalanced across participants. Performance feedback was not provided during the experimental scenarios. Participants were provided 5-minute breaks between each scenario, during which they also completed the SART measure.

3 Results

An alpha level of $\alpha = .05$ was adopted for all statistical tests. There were no significant differences in SA or performance across the three experimental scenarios. Thus, for simplicity, the data for the SPAM and performance measures was averaged across the three scenarios (see Durso et al., 2006). However, for SA and performance response time, for each scenario we first removed incorrect raw responses and raw response times that were more than 3 $SD$s above the mean (1.25% of raw response times were trimmed as outliers). Shapiro-Wilk tests indicated that each SPAM response time measure and each performance response time measure was normally distributed. Descriptive statistics and correlations among SPAM (query accept time, accuracy, and response time), SART, and the two performance measures are presented in Table 1.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Descriptive statistics and intercorrelations among the SART, SPAM, OW task and state change task variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD) 1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>1 OW task mean response time</td>
<td>3.91 (1.35) (sec) -</td>
</tr>
<tr>
<td>2 State change mean response time</td>
<td>21.63 (8.31) (sec) .20 -</td>
</tr>
<tr>
<td>3 State change accuracy</td>
<td>0.92 (0.10) - .29* - .29* -</td>
</tr>
<tr>
<td>4 SPAM current response time</td>
<td>9.16 (3.13) (sec) .41** -.01 -.03 -</td>
</tr>
<tr>
<td>5 SPAM future response time</td>
<td>13.12 (5.82) (sec) .39** .21 -.06 .59** -</td>
</tr>
</tbody>
</table>

Note: *$p < 0.05$; **$p < 0.01$. 

Table 1  Descriptive statistics and intercorrelations among the SART, SPAM, OW task and state change task variables (continued)

<table>
<thead>
<tr>
<th>Mean (SD)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 SPAM current accuracy</td>
<td>0.73 (0.15)</td>
<td>.12</td>
<td>.09</td>
<td>.01</td>
<td>.15</td>
<td>.04</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 SPAM future accuracy</td>
<td>0.63 (0.20)</td>
<td>.02</td>
<td>.39</td>
<td>.06</td>
<td>.19</td>
<td>.20</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 SPAM current accept time</td>
<td>0.92 (0.52)</td>
<td>.08</td>
<td>.12</td>
<td>.09</td>
<td>.13</td>
<td>.03</td>
<td>.02</td>
<td>.01</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>9 SPAM future accept time</td>
<td>1.09 (0.67)</td>
<td>.01</td>
<td>.05</td>
<td>.20</td>
<td>.01</td>
<td>.15</td>
<td>.53</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 SART</td>
<td>5.98 (1.47)</td>
<td>.17</td>
<td>.01</td>
<td>.15</td>
<td>.10</td>
<td>.08</td>
<td>.10</td>
<td>.03</td>
<td>.04</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: *p < 0.05; **p < 0.01.

3.1 SPAM measure

On average it took participants one second to indicate that they were ready to accept a SPAM query. There was no difference in the time taken for participants to accept future SPAM queries ($M = 1.09$ sec, $SD = 0.67$ sec) compared to current SPAM queries ($M = 0.92$ sec, $SD = 0.52$ sec), $t(54) = 1.66$, $p = .10$. These two measures were correlated ($r = .53$), but neither significantly correlated with any SA measure or any performance measure.

Participants made less accurate responses to future SPAM queries ($M = 0.63$, $SD = .20$) compared to current SPAM queries ($M = 0.73$, $SD = .15$), $t(54) = 3.32$, $p < .01$, $d = .57$, and these two SPAM accuracy measures were not significantly correlated. Participants also made slower responses (for correct responses) to future SPAM queries ($M = 13.12$ sec, $SD = 5.82$ sec) compared to current SPAM queries ($M = 9.16$ sec, $SD = 3.13$ sec), $t(54) = 6.21$, $p < .01$, $d = .85$. As indicated in Table 1, response times to current and future SPAM queries were correlated ($r = .59$), but neither of these two SPAM response time measures were correlated with current or future SPAM accuracy.

3.2 SART measure

The SART measure of SA was not significantly correlated with SPAM accuracy, SPAM response time or the performance measures.

3.3 Performance

As predicted, accuracy on the OW task was at ceiling (99% correct). Participants successfully reported 92% of state changes, and accuracy on the state change task was negatively correlated with correct response time on both the state change task ($r = -.29$) and correct response time on the OW task ($r = -.29$); participants who were faster at the state change and OW task were more accurate on the state change task.
We conducted hierarchical multiple regressions to examine the extent to which the SPAM measures predicted submarine track management performance. In the first block we entered SART, response times to accept SPAM queries, and any other control variable that significantly correlated with the criterion (see Table 1). Thus, in contrast to past research using SPAM (Durso et al., 1998, 2006), we controlled for SPAM query accept time (a proxy for workload) by entering it in block 1 of the hierarchical regression models. The four SPAM measures (current accuracy, current response time, future accuracy, and future response time) were then entered in the second block to examine if they predicted criterion variance over and above the control variables. Predictors in each step of the regression model were entered simultaneously. This provided a direct test of our theoretically-driven hypotheses that current and future SPAM queries would be differentially related to the performance tasks, because it allowed determination of the unique contribution of each predictor; that is, the relationship between each predictor and criterion when the contributions of other predictors in the hierarchical model were statistically removed (Meyers et al., 2006). The power to detect a medium-to-large effect (Cohen's $f^2 = .25$) of the SPAM predictors in a regression model with eight predictors was adequate at .82 (alpha = .05; G*Power; Faul et al., 2007).

### 3.4 Predicting performance on the OW task

As indicated in Table 1, mean correct response time for the OW task was significantly correlated with both current ($r = .41$) and future ($r = .39$) SPAM response time, and state change accuracy ($r = -.29$). We conducted a hierarchical regression where response time (correct trials only) for the OW task was entered as the criterion. In the first block we entered SART, current and future SPAM query accept time, and state change task accuracy. These predictors accounted for 11% of variance in OW task response time but the model was not significant, $F(4, 54) = 1.49, p = .22$. In the second block, the four SPAM measures were entered, and these accounted for an additional 23% of variance in OW task response time, $F_{\text{Change}}(4, 46) = 4.14, p = .01$, and the final model was significant, $F(8, 54) = 2.95, p < .01$. Current SPAM response time ($\beta = .35, p < .05$) and accuracy on the state change task ($\beta = -.31, p < .05$) were both significant predictors of OW task response time, and current SPAM accuracy approached significance ($\beta = -.22, p = .08$). Future SPAM response time ($\beta = .14, p = .38$) and future SPAM accuracy ($\beta = .15, p = .29$) did not predict unique variance in OW task response time. No other predictors in the final model approached significance (smallest = .36).

Stepwise regression has been used in the past to examine the predictive validity of SPAM (e.g., Durso et al., 1998, 2006). The disadvantage of stepwise regression is that the order in which predictors are entered (ultimately influencing which predictors are included in the final model) is based purely on statistical grounds (Meyers et al., 2006). We preferred hierarchical regression because it provided a more stringent test of our hypotheses. However, to further increase the power to detect a contribution of future SPAM response time (which was correlated $r = 0.39$ with OW task response time), we also conducted a stepwise regression. On the basis of semi-partial correlations, the stepwise regression model entered future SPAM response time as the fourth predictor, but it remained non-significant ($p = .16$). Thus, future SPAM response time did not predict unique variance in OW task response time, even when there were only three other predictors in the regression model.
3.5 Predicting performance on the state change task

As indicated in Table 1, accuracy on the state change task was significantly correlated with future SPAM accuracy ($r = .39$) and state change response time ($r = -.29$). We conducted a hierarchical regression where state change task accuracy was entered as the criterion. In the first block we entered SART, current and future SPAM query accept time, and state change response time. These predictors accounted for 11% of variance in state change task accuracy but the model was not significant, $F(4, 54) = 1.50, p = .22$. In the second block, the four SPAM measures were entered and they accounted for an additional 17% of variance in state change task accuracy, $F_{\text{change}}(4, 46) = 2.62, p < .05$, and the final model was significant, $F(8, 54) = 2.16, p < .05$. Future SPAM accuracy ($\beta = .44, p < .01$) was a significant predictor of state change accuracy, and state change response time approached significance ($\beta = -.26, p = .06$). Current SPAM accuracy ($\beta = -.11, p = .42$) and current SPAM response time ($\beta = .05, p = .77$) did not predict unique variance in state change task accuracy. No other predictors in the final model approached significance (smallest $p = .24$). In a stepwise regression, current SPAM accuracy was entered as the fourth predictor but remained non-significant ($p = .37$).

4 Discussion

To our knowledge, the current study is the first published in the literature to examine SA in simulated submarine track management. The first contribution of the current paper is the replication of the Durso et al. (2006) and Dao et al. (2009) findings that participants were less accurate and slower when responding to future-oriented compared to current-oriented SPAM queries. This finding was not obvious a priori, because the dynamic elements on the submarine track management display evolve slowly, and thus do not change position greatly over a five-minute time period. This finding is important because replication is vital for building reliable knowledge that can be effectively used by human factors practitioners (Jones et al., 2010). Replications are even more worthwhile when they are combined with theoretically-relevant extensions because this aids in determining boundary conditions and the appropriate level of generalisation. Our demonstration that the predictive utility of current- and future-oriented SPAM queries depended on the nature of the display information relevant to task goals, as discussed further below, provides such a theoretical extension. In addition, we demonstrated that SPAM can predict unique variance in performance when SPAM query accept time has been controlled for, thereby providing empirical support to the claim of Durso and Dattel (2004), and Durso et al. (2006) that the SPAM method allows the effect of workload and SA on task performance to be distinguished.

Participants made enemy engagement decisions accurately and quickly (< 4 secs.) when prompted by the experimenter, and response time was predicted by the speed at which participants accessed/interpreted information relating to the current state of the display. SA for the future display did not predict OW task response time. In contrast, the state change task required participants to strategically monitor the display and to predict how it would evolve over time in order to alert the OW when an enemy vessel became clear or was no longer clear to fire upon. This strategic, tactical type of activity would have required more proactive behaviour (see Hollnagel, 2002; Loft et al., 2007), and it is not surprising then that participants were less accurate and slower to detect state changes.
compared to OW task decisions. State change detection accuracy was predicted by the accuracy with which participants accessed/interpreted information relating to the future state of the display. SA for the current display did not predict state change detection. This dissociation between current and future SA in predicting performance is crucial because it indicates that the SPAM queries measured the unique dynamic information requirements (inputs) underlying decision making (outputs) for each performance task (Loft et al., 2009; Marr, 1982).

These data provide empirical support for theoretical models of optimal visual sampling (Senders, 1964; Sheridan, 1970) and attention-SA integration (McCarely et al., 2002; Wickens et al., 2003). According to these models, an individual’s visual sampling of information in a multi-item dynamic display depends on their expectancy regarding the relevance of display events to task goals. Prioritising the visual sampling of different vessels in the track management display would have depended on assessment of the relative importance of display elements to the state change task goal, thereby requiring projection of the future relationships between vessels. The McCarely et al. (2002) model of attention-SA assumes that after a display element enters the attention module, its attentional weight in the cognitive SA-updating module exponentially decays. In our task, participants needed to divide attention among multiple evolving events and would have needed to defer state change decisions. The likelihood or speed at which participants returned to display information relating to a deferred state change may have been related to levels of activation in memory for that state change event (Anderson and Lebiere, 1998). Thus, it would have been important for participants to correctly project the future state of the display in order to determine the relative importance level of enemy vessels to the state change task, and to assign appropriate memory activation levels.

It was important that SPAM outperformed SART in predicting performance because there is no doubt SART is the easier measure to prepare and administer. SART was not correlated with SPAM or performance. Participants may have rated their SA based on how well they thought they were doing at the task, rather than on their perceived SA. As in past research, our SART assessments were post hoc at the end of each scenario. While it is common for post hoc SART ratings to correlate with performance (e.g., Rousseau et al., 2011), it is possible that completion of the SART scale was more reliant on participants’ memory than on the completion of our immediate SPAM queries, thereby impacting the predictive utility of SART. One potential problem with SPAM is that the queries may be distracting or impose cognitive loading. Alternatively, SPAM may facilitate performance by cueing operators to information they might otherwise have missed, or perhaps even functions as a warning signal that increases participant vigilance (which is usually helpful in sustained attention tasks; see Helton et al. (2011)). However, the majority of past research has reported no performance differences between SPAM and control conditions (e.g., Durso et al., 2006, 1998; but see Pierce, 2012). It is unlikely SA benefited from the time that participants waited to accept SPAM queries because SPAM query accept times did not correlate with SPAM accuracy/response time or task performance, and we controlled for SPAM query accept time in our regression models.

One must take care generalising the current findings to naval operations given the use of the medium fidelity task, the student participant sample, and the limited task duration. There are undoubtedly differences in domain-specific cognitive skill, and in motivation, between expert submariners and students. However, given that no prior studies have examined SA in simulated submarine track management, the current work takes a critical first step towards demonstrating that SPAM can potentially be used to assess operator SA
in submarine track management. Furthermore, a significant advantage of the current project is that our task simulation and SA questions were informed by observation of real submarine combat systems and by interviewing SMEs. This notwithstanding, future research will need to replicate these findings using naval personnel in simulations of Collins Class Submarine T102 or T106 combat systems and that is our intent. The practical end-point of this research is a valid, minimally intrusive measure of SA.

Gaining an unobtrusive, real-time measure of SA is a priority for Australian Defence researchers attempting to design new information handling systems and to evaluate the utility of off-the-shelf technologies. Research to date in the submarine context is limited; the unique requirements of the submarine, together with the specific command team procedures in the submarine control room, need to be better understood so that emerging technology might be exploited effectively. It is expected, for example, that the volume and variety of sensor data may increase dramatically with technological change in coming years. How the data are translated and communicated to human decision makers (i.e., the cognitive realm) needs to be carefully managed, filtered, and processed to ensure operational performance and safety. The capability of the submarine will depend, as it always has, on the human achieving SA appropriate to the task at hand. Interface design must begin with identification of the SA requirements of the submarine operator. The problem with relying purely on cognitive interview or behavioural observation methods is that the information used to make task decisions is not always readily available to experts via introspection. SA measurement can be used to complement these other knowledge elicitation techniques.

Acknowledgements

This research was supported by Research Agreement 353520 from the Defence Science and Technology Organisation Australia. We would like to thank Peter Henley, Kate Ponton, and Megan Schmidt for their advice during various stages of the project. We also like to thank Janelle Braithwaite for task programming, and Dannielle Finnerty and Louise Delane for assistance in data collection and analysis. We also thank Frank Durso for commenting on earlier versions of this paper.

References


Using the situation present assessment method


