Adapting Situational Awareness Measures for Hydropower Display Evaluations

Li, X.,[1] Sanderson, P.,[1] Memisevic, R.,[2] Wong, W.[3]
[1] School of ITEE, The University of Queensland, St Lucia, Australia
[2] Powerlink Queensland, Australia
[3] Middlesex University, UK

This study used Situation Awareness (SA) as a measure to evaluate two new functional displays supplementing existing monitor displays in a large hydropower system control room. Because it was impractical to use traditional SA measures this paper proposes a novel SA measurement framework, in which controllers' SA levels are derived from their in-the-loop utterance and viewing patterns, their context-specific reports of the situations, and their overall SA reflections. Results indicate that the SA measures not only support and complement one another, but also are consistent with performance results. This study offers a novel approach of using convergent lines of evidence to assess SA in the situations that involve a whole control room or command centre or in situations constrained by time and resources.

INTRODUCTION

The goal of this paper is to describe how we developed a practical solution to the problem of assessing hydropower controllers' situation awareness (SA) despite some complex constraints in testing new displays for a control room.

Research suggests that current monitor displays in the hydropower control room are unable to support hydropower controllers' need to work in multiple domains across different time frames (Sanderson et al., 2004). Two new ecological-inspired displays, called "Functional Displays", were developed, intended to integrate high level scheme properties over appropriate time frames, as a supplement to the current monitor displays in a large hydropower system control centre (Memisevic et al., 2005). Figure 1 shows an evaluation trial of the displays in progress with professional controllers.

The goal of this paper is not to analyse details of the two functional displays themselves but instead to describe the methods used to evaluate them. Briefly, however, the leftmost display in the back row of Figure 1 is intended to support short term (5 to 30 min) coordination of power generation by revealing mechanical and operational constraints and by providing an integrated overall scheme snapshot from moment to moment. The display third from left supports support longer term (30 min to 24 hours) coordination of water planning with the plan for the day's generation, as determined by the market. It shows water storages and predicted levels into the future.

The evaluation goal was to see whether the new displays would lead to a better outcome in the controllers' control performance, strategies and situation awareness (SA). SA measures such as SAGAT and SART have been commonly used in display evaluation in many domains (Vidulich, 2000). However, these measures are impractical to use in the hydropower domain because of the following constraints.

First, traditional SA measures such as SAGAT require an investigator to randomly blank displays and suspend the simulator while requiring participants to respond quickly to SA queries (Endsley, 2000). However, because of the way it was implemented technically, with thousands of state variables in a dynamic relationship, our hydropower scheme simulator could not freeze at the selected time as required by SAGAT.

Figure 1 Experimental setup showing controller (foreground) and coordinator (background at right)

Second, we only had access time of only around 5 hours for each participant pair. It would have reduced the time that participants had to interact with the displays if we had spent a lot of time asking participants detailed questions, as most SA measures require. Moreover, our testing scenarios, each around 30 minutes, combined representative unexpected contingencies inside and outside the hydropower plants, emergent storage problems, and different market contexts for the purpose of requiring controllers to act in all different domains. Only by placing controllers in such challenging and time pressured situations could any expected benefits of the new display be realized from the various performance measures (Li et al., 2006). As a result, using self-rating standard questions like SART (Taylor, 1990) may not be sensitive in such situations. It was important to develop context-specific SA questions to capture the critical dynamics contained in each scenario, without disrupting the scenario by asking for full details as SAGAT would have done.

Third, as shown in Figure 1, our testing environment is a multiple-display-multiple-controller environment, more like a real control room. The *controller* (left) is the team leader who is responsible for configuring the scheme to meet ever-changing energy targets during the shift. The *coordinator* (right) supports the controller while managing interactions with scheme planners. In the simulation there were over 38 displays that could be put on the six computer screens available to the control team, including nine displays intended for monitoring system status. The displays show thousands of data points updated from seconds to hours, so it would be impossible to test reliably an individual's SA at the level of display elements as SAGAT or SART do.

Finally, commonly used SA measures are all based in the theory of defining SA as three distinct constructs in the head of an individual: perception, comprehension, and projection. However, this approach has weaknesses in representing SA as a dynamic concept arising from the interface between users and their environment (Smith et al., 1995). This is the most crucial feature offered by the new Functional Displays; they reveal the temporal constraints and task constraints in the hydropower domain. Therefore, to elicit the expected benefit of the new displays, it was important that the dynamic, interactive, adaptive relationship between the users and the system should be reflected in SA measures used.

Thus, we need to look for a richer model of SA to evaluate SA in the hydropower domain. A more general perspective regards SA as a measure of the degree of dynamic coupling between a user's cognitive status and the content of the situation in which they are interacting (Flach, 1995). This view emphasizes the inseparability of situations and awareness, as well as the dynamic, adaptive, interactive relationship between them, which exactly matches the goal of our display design and the purpose of our evaluation. However, no formal SA measures based on this view have been developed.

As a result, we developed a novel SA measures framework to represent hydropower controllers' degree of cognitive coupling with different time frames and with the different properties of the hydropower scheme (see Figure 2). The desired measures should address controllers' perception and cognition as traditional SA measures do, and should also reflect their work performance as they interact with the system. In the next section we provide an overview of the experiment we ran and the measures we used to test whether the Functional Displays promote better SA.

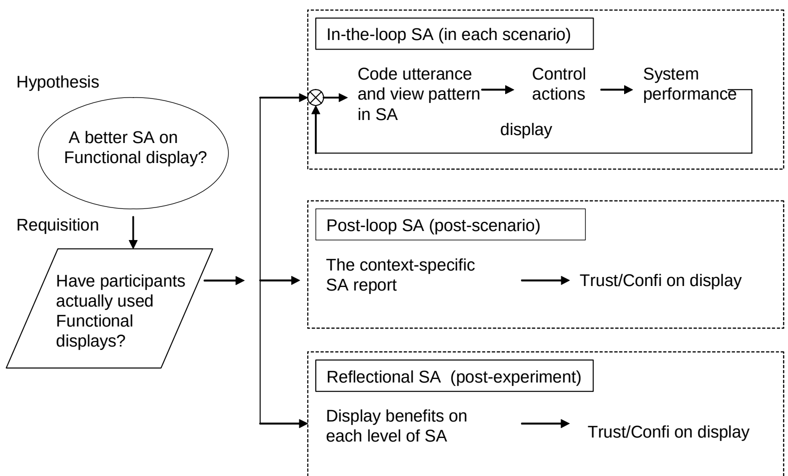

Figure 2. The proposed multiple SA measurement framework

METHOD AND ANALYSIS

A first evaluation of the two Functional Displays was done with three professional controller-coordinator pairs (the total population of controllers is seven) We used with a medium-fidelity hydropower simulator installed on two connected laptops (Memisevic et al., 2007). The controller used one laptop and the coordinator used another (see Figure 1). The control teams operated the scheme, starting and stopping generators, controlling storages and water diversions, and bidding into the market if needed.

Participants used a subset of the current displays (Current Displays) for two scenarios and then the current displays supplemented with the two new displays (Functional Displays) for two more scenarios. The order of scenarios was counterbalanced across control pairs but the order of displays, was not counterbalanced. Given our limited time with controllers and the primary need to elicit thoughtful reactions to, and comments about, the new displays, we decided to give all controllers the opportunity to become familiar with using the simulator with their normal Current displays before adding the novel Functional displays. The control teams had about 30 minutes of practice with the Current Displays before doing the first formal scenarios with the Current Displays.

The simulation had an embedded data capture program that every four seconds captured simulated hydropower system status and recorded the displays that had been accessed. In addition, two video cameras captured the control team's use of different displays and their utterances. As Figure 1 shows, the controller also wore a small head-mounted camera to track the general direction of his gaze during the scenario (Omodei et al., 1997). All three video sources were used for coding. Because the resolution of the head-mounted camera did not let us identify the direction of the controller's gaze, we also referred to the other video files and we could identify which screen the controller was looking at with any surety only when the controller was talking and gesturing.

Our multiple SA measures framework is shown in Figure 2. It indicates how we evaluated the impact of the Functional Displays. First it was necessary to confirm whether participants actually looked at the new displays during each experimental scenario. This is a logical requirement of our experiment. If we are to attribute participants' better SA to the Functional Displays, then it is important to show that participants actually used the Functional Displays. Because our experimental set up included many different displays, somewhat like the real control room, participants could have ignored the new Functional Displays if they had wished. As shown in the rightmost part of Figure 2, the multiple SA measures included the three kinds of measures outlined in the next three sections.

In-the-loop SA. Participants were encouraged to have natural task-oriented conversations during all trials. Their utterances are therefore the best resource for reflecting their mental picture from moment to moment. The head-mounted camera worn on the controller's head together with two other video cameras behind the control team captured their use of displays when they were talking. As a result, a measure of the control team's SA could be elicited from their utterances and viewing patterns while they managed the hydro-scheme with the displays. The control team's actions and system performance were also collected to help interpret their SA.

It was often difficult to separate the three levels of SA (Endsley, 2000) in a single utterance. Sometimes, participants just used one sentence to address all three levels. Therefore, the utterance data should at least partly represent this. The utterance data indicated that control teams were concerned mainly about either of two things: (1) the accuracy of individual data points, such as an energy target for a power station, or (2) overall issues about system status or future states that were difficult to separate. As a result, utterance occurrences fell most readily into two categories: data-focused awareness (SA1-related) and goal and/or knowledge-driven awareness (SA2-3 related).

Following Endsley (2000), SA at level 1, SA1, was mainly derived from utterances related to an observation or to a question about the particular data value and status of an individual system component or a control action of interest. SA at levels 2 and 3 came from utterances of assessing and reasoning about the current overall scheme status, predicting the future scheme status, planning future actions, estimating the interplay impact on other domains and the mapping toward the control goals, prioritizing the control tasks etc. The utterances between control pairs and the associated viewing pattern of controllers were coded using the MacSHAPA (Urbana, IL) video analysis software tool.

Post-loop SA. Immediately after each scenario trial was finished, the control team reported their experience, prompted by context-specific SA questions. The questions were intended to test different aspects of the control team's overall situation awareness when they were coupled to a particular situation.

- The depth of the coupling. Participant teams were asked about their awareness of the unexpected system contingencies, any emergent water storage problems and their options for handling such problems.
- The breadth of the coupling. These questions were related to less-critical control issues in the scenario, such as how much reserve capacity the generators had at any moment.
- The temporal aspect of the coupling. These questions asked about the long term market impact of continuing to generate power as in the situation just experienced.

Moreover, participants were asked to rate their level of trust in whether the displays provided a complete and accurate picture of scheme status, and their level of self-confidence in their ability to control the scheme with the displays. Responses were collected individually from the controller and the coordinator and were coded with MacSHAPA,

Reflective SA. After finishing all scenarios, participants individually rated the overall benefits of the Functional Displays at each level of SA. They also reported their overall trust and self-confidence with the Functional Displays.

RESULTS AND DISCUSSION

Frequency of Use of Functional Displays

When each scenario trial was finished, the participants were asked to rank their usage frequency order of the nine available monitor displays. Because the video files only captured participants' use of displays with complete certaintly when the participants were talking about the displays, we also collected participants' retrospective report of display use. The "Energy Flow" (EF) display, one of the two new Functional Displays, was ranked as the most frequently used display in 100% of controllers' answers and in 60% of coordinators' answers, while the remaining 40% of coordinators' answers ranked it as the second most frequently used display.

The "Water Efficiency" (WE) display, another new display, was apparently used less often. No controllers ranked it among the top three displays and only 20% of coordinators ranked it as their third most often used display. This might be because the WE display was intended for the coordinator's use in planning and so was among the rightmost displays, where controllers had less access to it. However the WE display is intended to support reasoning in much longer timeframes than the EF display, so intermittent use by the coordinator is expected in the timeframe of our experiment. It was needed less often than, for example, the displays needed to rebid into the electricity market. We will return to this issue later.

In-the-loop SA

From Figure 3 it is evident that participant teams tended to talk more often about goal related, or knowledge-driven (SA23 related) topics when using the Functional Displays. The result of a within-subject ANOVA analysis indicated that there was no significant effect of Display (display type), but there was a significant effect of SA, $F(1,2)=36.19$, $p=0.027$, and a significant interaction between Display and SA, $F(1,2)=28.52$, $p=0.033$. There were more SA23speech utterances with the Functional Display than in any other condition.

Moreover, the video analysis indicated that, on average, 51% of SA1 utterances and 48% of SA23 related utterances from control pairs were obtained on using the new EF display, because controllers were looking at the EF display when they were talking.

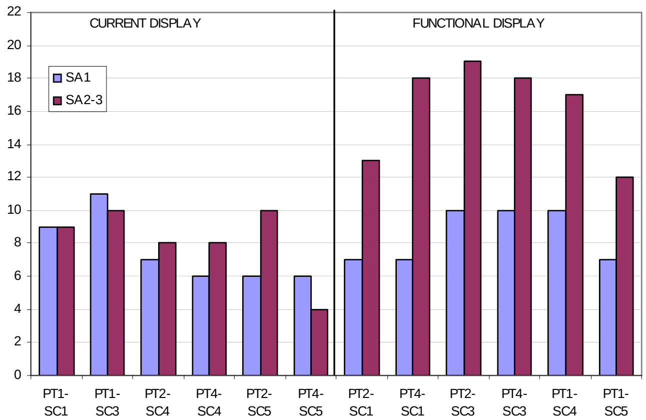

Figure 3 The number of SA1 and SA23 utterances in each experimental trial .Current vs. Functional Display). Note: each trial is named with participant team number (PT) and scenario number (SC).

The fact that SA23 utterances were more frequent with the Functional Displays than with the Current Displays may reflect some tangible benefits of control performance with the Functional Displays, as follows (Li, in preparation).

- Control teams set up their control directions more quickly with the Functional Displays as indicated by the fact that they responded sooner to the initial configuration of the hydropower scheme and the market situation
- Control teams using the Functional Displays made more changes to their bids into the electricity market and achieved a larger allocation of energy production from the market in more scenarios than with the Current Displays.
- Control teams using the Functional Displays had better performance at meeting the allocation of energy production given to them by the market operator.

Discussion of in-the-loop SA

The above results suggest that the Functional Displays may have allowed participants to think more readily about the overall system state, the future impact of changes and the interplay between the subsystems than the Current Displays, so leading to a better system performance, a more immediate response, and more effective activities.

Results for post-loop SA

Control teams using the Functional Displays appeared to have better situation awareness; 70.8% of their answers were correct, whereas for control teams using the Current Displays only 30.8% of answers were correct. When using Functional Displays the control teams answered 100% of the questions about unexpected system contingencies correctly, whereas when using the Current Displays they answered only 20% of those questions correctly. However with the Functional Displays the control teams did not show a broader awareness of less-critical issues or a more accurate market prediction. From participants' comments, this may due to the intensity of the scenarios—they had no time to consider these issues.

On a 7-point Likert scale (1 = much less; 7 = much more), participants reported a marginally better trust in getting a more complete and accurate picture of the scheme when the Functional Displays were added (mean rating = 4.83) compared with just using the Current Displays (mean rating = 4.08); t (5) = -2.09, p=0.091 (one-tailed tests were used as we have a directional hypothesis).

Participants also had a marginally better self-confidence in their ability to control the scheme with the Functional Displays (mean rating =5.25) than with Current Displays (mean rating = 4.33); t (5) = -2.2, p=0.079.

Discussion of post-loop SA

The context-specific SA reports of scheme situations may reflect some of the findings with control performance. For example, no significant differences were found in controllers' ability to keep each generator's reserve within the boundary between the two interface conditions, possibly because they did not recognize the changes while their attention was on tasks with higher priorities, as shown from their SA report.

Moreover, the result for trust and self-confidence may suggest that participants were inclined to talk more often about SA23-related topics when they had more trust in their display support. Alternatively, the more frequent SA23 utterances may result from a better self-confidence in their ability to control the scheme when they felt that they had full observation of the scheme. This provides some rough but interesting insights into the interactions between SA and trust/self-confidence.

Result for Reflective SA

A 7-point Likert scale was used with the mid-point rating of 4 again indicating indifference between displays. Ratings significantly higher than 4.0 were not found for SA1 (mean rating =4.75); t(5)=1.69, p= 0.150; but were found in favour of the Functional Displays for SA2 (mean rating =4.83); t(5)=2.19, p=0.079; and for SA3 (mean rating =5.3); t(5)= 6.5, p=0.003.

It seems that participants considered that the Functional Displays had a greater effect on their comprehension of system state and on their ability to predict future states than on their straightforward perception of system information.

However the participants' overall ratings of trust and self-confidence were not significantly different between the two displays conditions. From participants' comments, this may have been caused by the "simulator noise". There were some anomalies in the simulation data compared with real scheme data, caused by either some of the approximation techniques used, or the constrained computational power of the two laptops, or our limited access to more commercially sensitive scheme data for use in the simulator.

CONCLUSION

The experimental findings for SA with the Functional Displays are summarized in Figure 4. Participants reported that they frequently used the EF display, one of the two Functional Displays, in the experiment. The other Functional Display, the WE display, was not used as often as the EF display, but participants nonetheless ranked it as one of the five wallboard displays they would like to have in their control room, where it was in the third place, just after the EF display.

The result of in-the-loop SA suggests that better SA23 was reflected in more frequent SA23 utterances, ultimately leading to more effective activity patterns as well as better control performance. The post-loop SA report indicates that a better awareness is rooted in a deeper cognitive coupling with the most emergent, dynamic issues within a particular situation. However, the benefits for the breadth and the temporal aspect of the coupling of a particular situation were not easily observed. In addition, participants' better trust of Functional Displays may lead to more frequent SA23 utterances, which in turn may promote better self-confidence. The overall reflective SA rating demonstrates that Functional Displays have more benefits for SA2 and SA3. However, no significant convergent evidence for better trust/self-confidence was found, possibly due to simulator noise.

Overall, these SA results are informative, and are consistent with the performance findings. The SA measures support and complement each other from different perspectives, providing a clearer and more integrated picture of the dynamic, adaptive in-the-loop interactions between controllers and the hydropower system. The proposed SA measurement framework appears to be sensitive, useful and well adapted to this complex hydropower control environment. Therefore, to assess operators' SA in a full control room or command center environment, the core issue is to capture the dynamic relationship between the "in the head" understanding of controllers and the content of the situations with which they are interacting (Flach, 1995). Thus, the desired SA measures should be derived from multiple perspectives and should be highly context-specific.

The measures outline here were developed to handle practical constraints arising from our limited time with participants, the limited power of our simulator and shortcomings in our data capture equipment. The measures still need to be further validated in a future study.

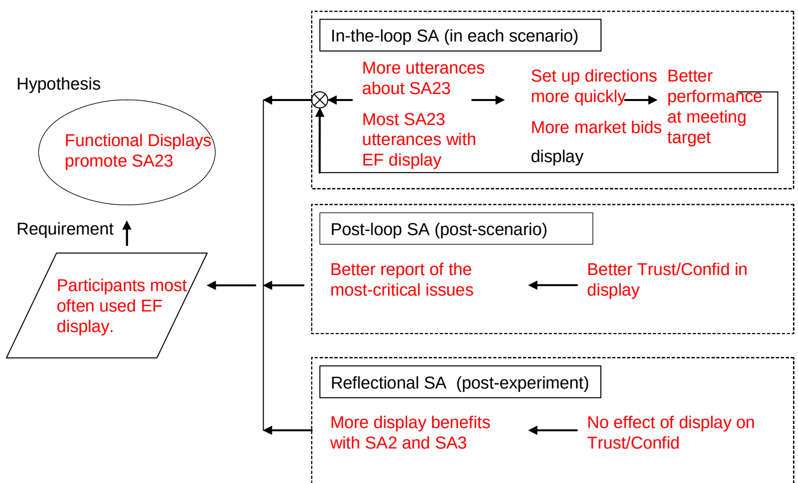

Figure 4. Mapping the experimental results on the Functional Displays into the proposed SA measurement framework

REFERENCES

Endsley, M. R. (2000). Direct measurement of situation awareness: Validity and use of SAGAT. In M. R. Endsley & D. J. Garland (Eds.), *Situation Awareness Analysis and Measurement* (pp. 147-173): Mahwah: Lawrence Erlbaum Assoc.

Flach, J. M. (1995). *Maintaining situation awareness when stalking cognition in the wild.* Paper presented at the international conference of experimental analysis and measurement of situation awareness, FL, USA.

Li, X. (in preparation). *Visualization and adaptation in a complex system: hydropower plant case study.* Unpublished Ph.D thesis, University of Queensland, St.Lucia.

Li, X., Sanderson, P., Wong, W. B. L., Memisevic, R., & Choudhury, S. (2006). Evaluating functional displays for hydropower system: Model-based guidance of scenario *Cognition, Technology, and Work, 8*, 269-282

Memisevic, R., Sanderson, P., Choudhury, S., & Wong, W. (2005). Work domain analysis and ecological interface design for hydropower system monitoring and control, *IEEE Conference on Systems, Man, & Cybernetics (IEEE-SMC2005)*. Hawaii.

Vidulich, M. A. (2000). Testing the sensitivity of Situation Awareness Metrics in Interface Evaluation. In D. J. G. Mica R. Endsley (Ed.), *Situation Awareness Analysis and Measurement*: Laurence Erlbaum Associates.

Memisevic, R., Sanderson, P., Wong, W., Choudhury, S., & Li, X. (2007). Investigating human-system interaction with an integrated hydropower and market system simulator. *IEEE Trans on Power Systems*, 22(2), 762-769

Omodei, M. M., Wearing, A. J., & Mclennan, J. (1997). Head-mounted video recording:A methodology for studying naturalistic decision making. In R. Flin, M. Strub, E. Salas & L. Martin (Eds.), *Decision making under stress:Emerging themes and applications* (pp. 137-146): Aldershot:Ashgate.

Sanderson, P., Wong, W. B. L., & Memisevic, R. (2004). *Analysing cognitive work of hydroelectricity generation in a dynamic deregulated market.* Paper presented at the Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society(HFES). Santa Monica, CA.

Smith, K., & Hancock, P. A. (1995). Situation awareness is adaptive, externally directed consciousness. *Human Factors, 37*(37), 137-148.

Taylor, R., M. (1990). *Situation awareness rating technique (SART): The development of a tool for aircrew system design.* Paper presented at the Situation awareness in aerospace operations (AGARD-CP-478), Neuilly Sur Seine, France.