Applications of SPAM in Simulated NextGen

Battiste, Henri
Cunningham, James
Silva, Hector
2-24-2014
## Contents

Executive Summary ........................................................................................................................................ 2
Introduction and Background .................................................................................................................... 3
  Situation Awareness .................................................................................................................................. 3
  Measuring SA ......................................................................................................................................... 3
Method ...................................................................................................................................................... 3
Findings and Recommendations ............................................................................................................... 4
Limitations and Future Directions .......................................................................................................... 5
References .............................................................................................................................................. 7
Appendices .............................................................................................................................................. 8
  Appendix A: Theoretical Underpinnings of Situation Awareness ......................................................... 8
  Appendix B: Significant Statistics for Probe Types and Performance Measures ............................... 9
Executive Summary

This report was compiled as a request from current president of our human factors corporation, Dr. Tom Strybel. It is intended to provide an overview of the validity, reliability, and usability of the situation present awareness method (SPAM), and to provide our recommendations for the use of this method in airspace simulations under NextGen situations. To this end, three recent studies were analyzed and synthesized to garner more information about situation awareness and the application SPAM.

We recommend that SPAM be used in the evaluation of NextGen technologies for the following reasons:

- It can assess workload in addition and independently of situation awareness.
- It can capture workload and situation awareness objectively and without operator bias.
- It can predict performance reliably above and beyond other situation awareness metrics.
- It can discriminate between novices and experts. Training programs can benefit from this because SPAM can be used to evaluate trainee performance.
- It has higher acceptance rates than other probe methods. In addition, it is readily applicable in a wide range of domains and in field settings.
- It is able to detect subtle but meaningful changes in situation awareness.

Although these conclusions and recommendations are drawn from empirical studies, it should be noted that only three studies were analyzed and that the sample sizes of those studies were also relatively small, which could limit the generalizability of our conclusions. Further, this report is not to say that SPAM could not withstand improvement. On the contrary, efforts such as calculating an overall index of situation awareness from SPAM accuracy and probe latencies or using operator eye glance information to further validate the use of SPAM would be helpful.
Introduction and Background

This report is an examination of empirical literature on the properties and qualities of the situation present assessment method (SPAM) put together as asked by Dr. Tom Strybel, current president of our human factors corporation. Over the span of three weeks, a literature review was conducted, information in studies was analyzed and compiled, and several recommendations were made. In this report we outline what studies were analyzed, what the results were and what they mean, and provide recommendations for the use of SPAM as a measure of situation awareness.

Situation Awareness

Situation awareness (SA) is the perception of environmental elements with respect to time and/or space, the comprehension of their meaning, and the projection of their status after some variable has changed. Research has shown that low levels of SA are one of the main factors in accidents attributed to human error. In systems where information flow can be of varying degrees (high/low), decisions made with a lack of understanding about the situation may lead to serious consequences, thus having high levels of SA is important. For more information on the construct of SA, please see Appendix A.

Measuring SA

Measuring situation awareness allows knowing to some degree how an operator understands information, events, and their own actions, and how they impact the goals and objectives (immediate/future) of the system. SA can be measured by physiological methods (e.g. heart-rate), subjective methods (such as the Situation Awareness Rating Technique), and using secondary task measures (offline and on-line tasks). For our evaluation we briefly considered SAGAT and SPAM. SAGAT is an off-line method and has often been criticized that it relies on the memory of the operator to measure SA. SPAM is an on-line method (meaning the displays are active during the task) and has found to be more robust than SAGAT in measuring operator SA due to changes in the system. Of the two choices of secondary task methods for SA discussed in the literature, SPAM was mentioned as the more promising of the two, and was thus chosen, by us, for examination and evaluation.

Method

A literature review was conducted in which three recent empirical studies regarding in the context of the Next Generation Air Transportation System (NextGen) were sampled. Strybel et al. (2010) evaluated the situation present assessment method for measuring pilot situation awareness (SA) in plausible NextGen environments. Strybel et al. (2011) studied the relationships between SA, workload, and performance of air traffic controllers (ATCos) using SPAM. Lastly, Bacon et al. (2011) examined SA, workload, and performance of ATCos in simulated NextGen scenarios. The results and conclusions from these three studies were analyzed and compiled in this report. Please see the reference section for full citations.
Findings and Recommendations

The use of SPAM as a situation awareness method is recommended not only for its ability to measure situation awareness, but also for its ability to measure operator workload concurrently. As workload is the primary limiting factor to the Next-Gen proposed air traffic density increases (Strybel, 2010), the need for simplistic yet powerful measures of workload is evident. Workload has also been shown to correlate significantly with performance, meaning that increases in workload, past attentional threshold, correspond to negative performance.

Through the use the SPAM methodology a ready prompt appears to the operator through which latency is calculated. This latency has been shown to be significantly correlated to workload measures (Strybel, 2010). Thus, through implementation of SPAM techniques, we are guaranteed two measurements directly correlated with performance; for the cost of only a single measurement tool.

In comparisons to other types of situation awareness measures, probe methods provide objectivity in their assessment of situation awareness. Self-report measures may have the disadvantage of being affected by perceived performance. For example, if an operator feels as though they did well in a particular trial, they may inadvertently or advertently report they had high levels of situation awareness. Conversely, if the operator feels as though s/he did not do too well, they may report lower levels of situation awareness. SPAM queries operators with questions about their environment. The questions can address past, present, or future events, and, in studies involving air traffic control, ask about a combination of relative and absolute information. This is considered to be consistent with cognitive models of ATCos (Strybel, 2010). Thus, the answers provided by the operators can be assessed as a correct or incorrect answer by referencing the scenario at the time the question was asked. Similarly, objectivity is found in the probe question latencies by starting a timer as soon as the question is presented, stopping it when the operator answers, and therefore not relying on the operators’ self-report to assess their situation awareness.

SPAM has been shown to predict performance and SA of both Pilot and ATC. In particular, SPAM predicted overall performance and ratings of “Maintaining Situation Awareness” (Bacon, 2011) reliably. A hierarchical regression model was performed and found that shorter probe latencies were an indicator of better overall performance. Similarly, shorter probe latencies also predicted higher ratings, by expert ATCos, of maintaining situation awareness. This is consistent with the theoretical notion that shorter times to answer questions means that the operator had the information readily available in memory (and thus having high levels of SA) and did not have to look at the environment to find the answer, which lead to better performance. What is more interesting is that probe latencies were still predictive even after expertise, objective performance metrics, and percentage of alerted conflicts was removed from the model.
SPAM is recommended as a tool for analyzing training levels because it is capable of discriminating expert from novice performance. Bacon et al. (2011) found that overall response latencies were related to Expertise ($p < .01$), such that novice probe latencies were longer than experts. By nature of its design, SPAM is easily integrated into systems and requires little to no new tools to perform the probe measures.

The underlying mechanics of the methodology make it an efficient and practical tool for use in a variety of domains. To measure the diagnosticity of situation awareness probe techniques, Strybel et al. (2011) developed content and task specific questions aimed at the measurement of situation awareness and workload in Next-Gen ATC Concept of Operations (CONOPS). As this process was completed with the introduction of subject matter experts (SMEs) to identify task relevant components, the creation of context-relevant questions across varying domains is both practical and efficient; assuming the breakdown of complex tasks into specific sub-task categories can be verified by SMEs. Also, in comparison to other situation awareness methodologies, SPAM can be readily applied in real world scenarios, which is not possible with freeze-based probe methods such as SAGAT. Blanking displays and handicapping the operators during a field experiment would create a dangerous and potential lethal environment (i.e. driving). By having the operator answer the “ready” prompt only when their workload allows and offering them full access to the environment while answering probes, the practicality of using SPAM in the field increases.

The SPAM methodology also encourages responses by prompting the controllers when a relevant question is ready for presentation. At this point operators can answer the query when their workload allows and are not forced to deviate from their primary task in an immediate manner, leading to a higher acceptance rate among operators (Strybel, 2011). In this respect, SPAM is recommended as a probe method that correlates with actual real world use of primary and secondary task operations.

Spam is recommended as a measure of SA because it can measure slight changes in SA due to changes in the operator’s roles and responsibilities. One reason for this is that SPAM is not a subjective measure, in that it measures actual performance latencies and accuracies of probe measures. Strybel et al. (2010) found that conflict-type probe latencies were significantly different across different levels of pilot conflict-management responsibilities. Strybel et al. (2011) found that for scenarios that switched from high to low automation, operators had significantly higher probe latencies in the later half. This result implies that SPAM probe latencies were able to measure the change in SA due to the operator’s new roles of conflict management.

Limitations and Future Directions

Although the recommendations presented here are based on empirical examinations of the situation present assessment method, it should be noted that only three, out of what could be several dozen studies, were selected for examination. Further, sample sizes of each study were also relatively low with eight (Strybel et al., 2010), seven (Strybel et al., 2011), and 16 (Bacon et
al., 2011) total participants per study, which could limit the generalizability of the results and our recommendations.

Several changes can be made towards improving SPAM. As it is presented in this report and in the studies evaluated, the measure of SPAM that showed significance was that of various different probe latencies. In none of these studies was accuracy of the respondents found to detect changes in SPAM. It was noted in Strybel 2010 that accuracy for probes was relatively high and ranged from 71% to 89% accurate and so perhaps a ceiling effect might have obstructed some effect from coming to light. Additionally, it could be that accuracies by themselves might not be able to detect changes in SA alone, but if an overall index could be computed from accuracies and probe latencies then a possible synergistic effect of both measures might be found.

In addition, other efforts can still be taken to improve SPAM. One in particular that has been suggested is to create an improved method for creating SPAM probe questions. Currently the process is very time intensive due to the nature of validating questions.
References


Appendices

Appendix A: Theoretical Underpinnings of Situation Awareness

In recent years, the advancement of SA concepts has made it easier to convey a true definition of the construct. “Situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future” (Endsley, 1995). More importantly SA is the construct on which performance of task hinges; however, SA is not performance directly and is not decision-making. Endsley refers to three levels of SA in her description. Level 1 SA is the perception of attributes, status, and dynamics of relevant information. Level 2 SA is the synthesizing of elements of Level 1 SA into a comprehensive view of all the systems. Level 3 SA is using Levels 1 and 2 to project into the near future about an objective (Endsley, 1995). In simpler terms, Level 1 SA deals mainly with perception, (i.e. Do I see?). Level 2 SA is primarily the comprehension stage, (i.e. What am I seeing?) Level 3 SA deals mainly with projection, (i.e. What will I be seeing?) Attention and perception lead to Level 1 SA, while working memory produces Level 2 and Level 3 SA. The major limiting factors to expansion of SA in any given situation are attention, working memory, stress, workload, information complexity, and automation (Endsley, et.al. 1999).

For a measurement system to convey information it must be reliable and valid. For a measurement system to have reliability and validity the system must: A) measure the construct it states to measure consistently without measuring other processes; B) be sensitive and diagnostic; and C) it must not alter the construct in question (Endsley, 1995).

Freeze-based probe methods for measuring SA have been suggested. Freeze-based SA methods freeze the simulation at randomly selected times and the operators are questioned on their awareness of the simulation at the time (the systems are blanked, the simulation is suspended and the subjects are questioned). The advantages of this system are the data is collected immediately, the data is accurate, and the system does not interfere with actual SA during measurement. The SAGAT methodology is suggested as a direct spawn of the underlying mechanics of freeze-based simulation. Contrary to this suggestion, current literature (citation) suggests that working-memory may be heavily taxed during freeze-based SA probe methods. This need to increase discrimination of SA measure from working memory tasks has led to the development and implementation of on-line SA probe methods, in particular the Situation Present Assessment Method (SPAM).

The SPAM procedure is as follows: the participant receives an auditory alert “ready” prompt asking if they are ready to receive an element question. When their workload allows, the participant accepts the prompt. Upon accepting the ready prompt, the element question appears. The participant then answers the individual element question. After the participant selects an answer, the probe resets and the participant continues with the primary task (Durso & Dattel, 2004). All of this occurs with the operator’s displays visible and active.
## Appendix B: Significant Statistics for Probe Types and Performance Measures

### TABLE 1. Significant Statistics for Probe Types and Performance Measures

<table>
<thead>
<tr>
<th>SPAM Probe Measures</th>
<th>Conflict Accuracy</th>
<th>Conflict Latency</th>
<th>Command &amp; Communication Accuracy</th>
<th>Command &amp; Communication Latency</th>
<th>Sector Status Accuracy</th>
<th>Sector Status Latency</th>
<th>Subjective Assessment Latency</th>
<th>Ready Latencies</th>
<th>Overall Probe Latencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Spacing Error</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conflict Resolution Time</td>
<td>$r = - .39^{**}$</td>
<td>$r = .25^{*}$</td>
<td>$r = .24^{**}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expertise</td>
<td>$p &lt; .01$</td>
<td>$p &lt; .01$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Handoff Accepts</td>
<td></td>
<td>$r = .29^{**}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOS time as resolution</td>
<td>$r = .39^{**}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Spacing Error</td>
<td>$r = - .22^{**}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of LOS</td>
<td>$r = - .26^{*}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall Performance</td>
<td></td>
<td></td>
<td>$r = - .04$</td>
<td>$r = - .05$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time spent with spacing</td>
<td>$r = .27^{**}$</td>
<td>$r = .19^{*}$</td>
<td>$r = .19^{*}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time worked on AC</td>
<td>$r = - .35^{***}$</td>
<td></td>
<td>$p &lt; .03$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: *$p < .05$, **$p < .01$, ***$p < .001$
1. Strybel et al. (2010)
2. Strybel et al. (2011)
3. Bacon et al. (2010)