Augmented Cognition for Tactical Tomahawk
Weapon Control System Operators

Patrice Tremoulet, Joyce Barton, Patrick Craven, Adam Gifford, Nick Morizio, Nadya Belov, Kathleen Stibler, and Susan Harkness Regli

Lockheed Martin Advanced Technology Laboratories
Cherry Hill, NJ 08002
ptremoul@atl.mcej.com

Abstract

Twelve individuals took part in an experiment which investigated the feasibility of using physiological sensors to apply a pacing mitigation strategy to help prevent cognitive-overload-induced performance declines. The participants performed tasks using a simulated Tactical Tomahawk Weapons Control System (TTWCS) while wearing electroencephalograph (EEG), electrocardiograph (EKG), and galvanic skin response (GSR) sensors. The tasks, retargeting missiles in response to emergent targets and responding to alerts in the form of questions about the ongoing strike, were performed in the context of low, medium, and high workload scenarios. In the control condition, data from the sensors were collected, but not used to influence task presentation. In the experimental condition, the collected data were used to apply the pacing mitigation—that is, to determine the appropriate times to present the interrupting alert tasks. Analysis revealed that using sensor data to influence task presentation led to a 66.8% reduction in the number of erroneous responses to alerts, a 34.6% decrease in decision making time for low workload scenarios and an overall reduction in decision making time of 20.4%, and a 25.5% increase in the number of missiles that participants could handle simultaneously. These results indicate that human performance may be improved by monitoring physiological responses and intervening before cognitive overload can cause performance degradation. This research effort is relevant to future TTWCS development because the cognitive workload demands placed on TTWCS operators are expected to increase as new capabilities of Tomahawk missiles are introduced and because military reduced manning initiatives will continue to require that job functions be performed by fewer personnel. However, before the technology which was tested during this experiment could be deployed in an operational environment, physiological sensors must become more comfortable, accurate and mobile and less sensitive to inter-individual variation. Furthermore, more detailed task analyses are needed to identify the specific task components within the TTWCS environment that will yield the greatest performance gain. This effort, combined with other additional research, could yield performance improvements in the operational environment that exceed those observed during the laboratory-based experiment reported here.

1 INTRODUCTION

Working under the auspices of Defense Advanced Research Projects Agency’s (DARPA) Improving Warfighter Information Intake Under Stress (IWIUS) program, Lockheed Martin Advanced Technology Laboratories (LM ATL) designed and developed a prototype augmented cognition system. Augmented cognition systems adapt to changes in a user’s cognitive state, as indicated by neuro-physiological data, by employing active mitigation strategies that may alter the way the system’s interface interacts with the user to try to maximize the use of the user’s cognitive capacities. For example, a system may reduce or increase task presentation rates, or alter task presentation formats. Previous research suggests that intelligent strategies can be used to improve performance when operators are in a high workload condition (Franke, Daniels, & McFarlane 2002).

LM ATL’s prototype has been designed to adaptively respond to changes in context

1 Context can include the criticality of an incoming alert, the nature and volume of the task or tasks that the user is currently focusing upon, and the amounts of stress, frustration, and/or fatigue that the user is currently experiencing.

Distribution Statement “A” (Approved for Public Release, Distribution Unlimited)
designed to optimize overall human-system performance—while users interact with a simulated Tactical Tomahawk Weapons Control System (TTWCS). The prototype system was designed around Performance Augmentation through Cognitive Enhancement (PACE), a highly-reusable architecture which manages user tasks and allows the implementation of a variety of mitigation strategies. PACE contains several components which are entirely domain independent and several which are designed to be extended for a particular domain. For a detailed description of the PACE architecture, see Morizio, Thomas, & Tremoulet, 2005.

This paper reviews the experimental design and empirical results of a study assessing the effectiveness of LM ATL’s prototype augmented cognition system to apply a pacing mitigation strategy designed to improve TTWCS operator performance. Pacing involves delaying lower priority tasks when the user is overloaded. For example, if physiological data indicate that the operator’s workload is above maximum threshold, pending tasks may be queued for delivery rather than presented immediately. When physiological data indicate a break in cognitive activity, the pending tasks is delivered. If too many secondary tasks compete for delivery, the primary task can be decomposed into subtasks so that cognitive breaks (during which lower priority tasks can be delivered) occur more frequently. The user is then able to complete more of the lower priority tasks while continually performing the primary task. Pacing defers messages related to lower priority tasks to periods of lower workload. Research indicates that the timing of an interruption relative to a user’s current task load can affect the user’s ability to cope with the interruption (Czerwinski, Cutrell, & Horvitz, 2000; McFarlane, 2002; Monk, Boehm-Davis, & Trafton, 2002).

We hypothesize that using a sensor-controlled pacing mitigation strategy will enable users of a simulated TTWCS system to achieve a higher level of performance during high workload periods than would be possible without augmented cognition technology. More specifically, pacing should help to prevent high workload-induced stress from causing steep performance degradation.

2 METHOD

2.1 Participants

Twelve individuals, recruited from the engineering staff of a large military industrial contractor, participated in the study. None of the participants had any prior exposure to the TTWCS environment, and all were screened for any pre-existing condition which could impact the accuracy of the physiological sensors (e.g., head trauma, seizures, or stroke).

2.2 Apparatus

All testing occurred in the LM ATL Human Testing Laboratory, a 400 square foot space with separate rooms for testing and observation. Test stimuli (scenarios) were presented on an Intel Pentium 4, 2.4 GHz Dell desktop computer with 512 MB memory, which runs Red Hat/Fedora Linux 9.0 with 2.4 kernel as its operating system. The visual interface, presented on a 17-inch screen, was jTTWCS, which is an internally developed java-based simulation of the Tactical Tomahawk Weapons Control System. All user responses were made using a standard keyboard and a two button mouse.

Participants were required to wear a set of sensors to capture physiological data while they performed test trials. This sensor suite consisted of a wireless electroencephalogram (EEG) Sensor Headset, an EKG, and GSR. The EEG Sensor Headset was designed by Advanced Brain Monitoring (ABM) to acquire six channels of wireless EEG from seven head sensors. The EKG and GSR data was collected using the Procomp Infiniti Encoder (Model #SA7500) from Thought Technology Ltd., Montreal, Canada. The EKG reading is taken using three sensors. The GSR measures the conductance level between two co-located sensors on the skin.

2.3 Stimulus

Participants performed the role of a Launch Area Coordinator (LAC) serving as a Tactical Strike Coordinator (TSC) by interacting with a simulated TTWCS. The responsibilities of a TSC include both planning future missile strikes (e.g. by allocating missions to launch platforms and requesting that new missions be developed as needed) and monitoring and adjusting ongoing strikes. The TSC also has responsibility for responding to alerts and questions about the current strike. The experiment included two main tasks: a) responding to alerts, referred to as the Alert
Task, and b) monitoring ongoing missile strikes, including retargeting missiles based on emerging targets of higher priority, referred to as the Strike Monitoring Task. The participants were given feedback on their performance only after completing a full battery of tasks in a condition.

### 2.3.1 Alert Task

Participants were required to correctly answer questions that were representative of queries that a commanding officer, subordinate, or peer level LAC might request during a missile strike. Participants responded to questions presented in a chat window located on the bottom of the display (see Figure 1). The currently active alert was displayed on the lower left of the screen and multiple choices were presented in the lower center. Scores on this task depended upon response accuracy. The answer was processed by the system once the “Done” button was clicked or the time for the alert expired. Sample alert questions include:

- Which missile will reach its target first if all missiles go directly to their default targets?
- How many default targets are you monitoring right now?
- Can missile RM007U-DL be retargeted to target T094S-EH based on the warhead type?

![Figure 1. The Multiple choice response area for the Alert Task is shown in the bottom panel.](image)

### 2.3.2 Strike Monitoring Task

During a missile strike, previously unknown, potentially mobile, emergent targets may appear. Participants were asked to retarget missiles heading toward fixed targets and reassign them to higher priority emergent targets, just as the TSC would make adjustments to the current strike as events unfold. Participants had three minutes to service as many emergent targets as possible, while maintaining coverage on as many high-priority and medium-priority default targets as possible. There were four rules for retargeting missiles: (1) Warhead types must match target types (Penetrating, Unitary, Submunition); (2) Loiter and Retarget missiles can be retargeted but NOT Fire and Forget missiles; (3) Because emergent targets are often mobile, missiles must be able to strike the emergent target within a specified window of opportunity; and (4) Credit is only given if a target is fully serviced. The number of correct missiles heading to a target must match or exceed the number of missiles required to fully service it, which was indicated by the small missiles above the target icon.
2.4 Design

The experiment used a two treatment within-subject repeated measures design, where the treatments were augmented (test) and non-augmented (control) conditions. Participants completed two blocks of trials and the order these blocks were presented and the treatment applied to each block were fully counterbalanced (see Table 1). Each block consisted of six three-minute scenarios separated by 15 second breaks. Within each block, there were two low (L), two medium (M), and two high (H) difficulty scenarios (e.g., Block A: L1, M1, H1, L2, M2, H2, Block B: L3, M3, H3, L4, M4, H4). Scenario workload was manipulated by varying the number of missiles, number of emergent targets, number of default targets, number of alerts presented per unit time and the type of alerts (e.g., difficulty) presented. Each scenario of a particular difficulty level had a matching number of missiles and targets. Performance scores from pilot testing without sensors were used to confirm that the difficulty-matched scenarios required similar amounts of effort.

<table>
<thead>
<tr>
<th>Table 1. Four orders balancing presentation order by mitigation (marked in grey)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First Presentation</strong></td>
</tr>
<tr>
<td>Order 1</td>
</tr>
<tr>
<td>Order 2</td>
</tr>
<tr>
<td>Order 3</td>
</tr>
<tr>
<td>Order 4</td>
</tr>
</tbody>
</table>

While subjects completed test scenarios, the following objective measures were collected: response time to alerts, alert response accuracy, retargeting response/decision making time, retargeting accuracy, number of sub-optimal retargeting decisions, number of ignored/unanswered alerts, cognitive state gauge levels, and the number of targets fully serviced at the end of each trial. In addition, post-session questionnaires on preferences, acceptability, and usability provided the subjective performance measures.

2.5 Procedure

Participants were trained in advance to recognize target and missile attributes such as type and priority and to understand the rules for retargeting. They were given documentation outlining the interface components and task scenarios and were given the opportunity to ask questions before and after running each training scenario. After reviewing the documentation, the participants first ran through the strike monitoring task without any alerts, and then with alerts.

Test sessions began with the application of sensors to the participants. Once the sensors were collecting data, participants completed a final training scenario. Next, participants completed two blocks of six test trials (one block was augmented and the other was not augmented). After each condition, NASA TLX was administered to capture estimates of workload. At the end of the test session the participants received an exit interview.

3 RESULTS

A one-way ANOVA with order as independent variable and retarget and alert performance measures as dependent variables was not significant, so no order effects were observed. A paired-sample t-test, with alpha set to 0.05, was used to investigate the difference in performance between mitigated and non-mitigated conditions. The average performance for the retargeting task in the mitigated condition was 66.5% correct, and the average performance in the non-mitigated condition was 58.2%. (The trial performance scores were calculated as percentages because of the varying number of item presentations across the tasks.) The average difference between mitigated and non-mitigated performance scores was significant ($t(11) = 2.66, p < .05$) and represents a 14.4% increase (percent of percent) in performance.

Missile-retargeting decision-making-time was calculated by recording the amount of time from when an emergent target first appeared to when a missile was retargeted to that emergent. In some instances, there was insufficient time for a participant to assign a missile to an emergent. In those cases, it was estimated that a participant would have completed that retargeting in 180 seconds. The average difference between mitigated ($M = 90.8$ s) and non-
mitigated (M = 114.1 s) was 23.2 s, which represents a 20.4% reduction in decision making time. The average difference for low difficulty level was 34.6%, for medium difficulty level was 15.5%, and for high difficulty was 11.4%.

The number of missiles handled by one operator was quantified by calculating the number of retargets successfully performed by an operator. The average number of retargets per participant combined across conditions for mitigated is M = 78.4, for non-mitigated M = 62.5. The difference, 15.9, represents a 25.5% increase. The number of retargets performed in the low, medium, and high workload conditions are reported in Table 2.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mitigated</th>
<th>Non-Mitigated</th>
<th>Difference</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>22.0</td>
<td>17.6</td>
<td>4.4</td>
<td>25.1%</td>
</tr>
<tr>
<td>Medium</td>
<td>25.0</td>
<td>20.5</td>
<td>4.5</td>
<td>22.0%</td>
</tr>
<tr>
<td>High</td>
<td>31.4</td>
<td>24.4</td>
<td>7.0</td>
<td>28.7%</td>
</tr>
<tr>
<td>Average</td>
<td>78.4</td>
<td>62.5</td>
<td>15.9</td>
<td>25.5%</td>
</tr>
</tbody>
</table>

4 DISCUSSION

This study indicates that human performance may be improved by developing adaptive interfaces capable of monitoring users’ physiological responses and intervening before cognitive overload can cause performance degradation. In this particular experiment, using EEG, EKG, and GSR sensor data to influence task presentation led to a 66.8% reduction in the number of erroneous responses to alerts, a 20.4% decrease in decision making time, and a 25.5% increase in the number of missiles that participants could handle simultaneously.

However, two areas require improvement before this technology can be employed in operational environments. First, several of the current limitations of sensor technologies must be overcome in order to develop a sensor suite that is mobile, comfortable, accurate, and less sensitive to inter-individual variation. While the set up and removal times and integration of the sensors have lessened over the duration of the IWIIUS program, at this point physical integration is far more difficult than electronic or data integration. Ideally, an integrated sensor suite which enables automatic placement of sensors (e.g., in a helmet), will be developed. Desirable features for this sensor suite include wireless connectivity, automatic baselining, and adequate comfort to be worn during twelve hour shifts.

Second, there is no guarantee that the simulated TTWCS tasks used in this study will prove to be operationally valid when the functionality to retarget block four tactical Tomahawk missiles is fully developed. LM ATL is addressing this concern by integrating the PACE architecture with a prototype user interface of a future release of TTWCS instead of a simulation environment. Once the integration is complete, LM ATL will test the effectiveness of pacing mitigations while active duty Navy sailors interact with the TTWCS user interface to perform operational tasks.

While not strictly required for successful transition to operational environments, improvements in the ability to transform physiological data into robust, accurate, reliable gauges of cognitive states and the development of more sophisticated, effective mitigation strategies could improve the efficacy of augmented cognition systems. Such improvements will require building on the basic and applied research performed under the IWIIUS program to identify the most promising technologies for gauging cognitive states, explore the relationships among cognitive states and mitigation strategies and developing task-specific mitigation strategies based upon detailed task analyses identifying the specific task components within operational environments that can yield the greatest performance gains.

Several interesting questions about augmented cognition technology remain which may be addressed through additional basic and applied research. However, completing this research work could yield systems that deliver performance improvements in operational environments that exceed those observed during the laboratory-based experiment reported here.
5 ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of Dylan Schmorrow, Amy Kruse, Colby Raley, and Lee Kollmorgen. The work reported here was funded by the Defense Advanced Research Projects Agency (DARPA), under contract NBCHC030032 to Lockheed Martin Advanced Technology Laboratories.

6 REFERENCES


