Dorneich, M.C., Ververs, P.M., Mathan, S., Whitlow, S.D Ververs, P.M., Carciofini, J., and Reusser, T. (2006). "Neuro-Physiologically-Driven Adaptive Automation to Improve Decision Making Under Stress," Proceedings of the Human Factors and Ergonomics Society Conference 2006, San Francisco, CA.

NEURO-PHYSIOLOGICALLY-DRIVEN ADAPTIVE AUTOMATION TO IMPROVE DECISION MAKING UNDER STRESS

Michael C. Dorneich, Patricia May Ververs, Stephen D. Whitlow, Santosh Mathan,
James Carciofini, Trent Reusser
Honeywell Laboratories
Minneapolis, MN

The advent of netted communications and a wide array of battlefield sensors is enabling real-time information streaming and asset management. However, the burden of information management is placed solely on the receiver of the information. Honeywell Laboratories developed a Communications Scheduler (CoS), an adaptive information management system for the dismounted Soldier, driven by an assessment of the individual's current cognitive capacity to process incoming information, in order to improve decision making under high task load conditions. An evaluation was conducted to demonstrate whether cognitive capacity to perform under differing task loads could be detected using neuro-physiological sensors and then if the adaptive automation would appropriately regulate information flow. Results revealed an improvement in primary task performance, no degradation in concurrent secondary tasks, and lower subjective workload ratings while performing cognitively challenging working memory tasks with the CoS, although a slight loss in situation awareness of lower priority message was found. The appropriate allocation of cognitive resources is key to managing multiple tasks, focusing on the most important ones, and maintaining overall situation awareness.

INTRODUCTION

One of the core capabilities of the Army Transformation is the unparalleled connectivity via netted communications enabling information sharing (Parmentola, 2004). Real-time collaboration enhances the kind of situational understanding that drives decisive actions. The inundation of information can be expected to grow between Soldier and ground and air sources. The potential data overload coupled with the efficiency of information flow required in executing Army doctrine, places on over-reliance of critical information throughput on a single point of contact, the individual Soldier. A means to help manage information overload of an individual Soldier's mental and cognitive state is needed beyond that provided by external means.

To ensure that the Soldiers are supported appropriately there needs to be intelligent information management to ensure that the system can support greater situation awareness for key information handlers on the battlefield. One way to mitigate workload is to allocate responsibility of a portion of the assigned tasks to automation. Automation is an effective means to allow users to save cognitive resources to allocate to other higher priority tasks (Dixon & Wickens, 2004; Rovira, et al., 2004). Using an assessment of the cognitive state of the user to base decisions on when to apply automation is one method of adaptive automation. Currently, adaptive systems derive their inferences about the cognitive state of the operator from mental models, performance on the task, or from external factors related directly to the task environment (Wickens & Hollands, 2000). The work described here focuses on real-time assessment of a human's capacity to understand and use information while under high task load conditions, where cognitive capacity can fluctuate greatly. In task management, mitigation strategies might include intelligent interruption to improve limited working memory, attention management to improve focus during complex tasks, or cued memory retrieval to improve situational awareness and context recovery. Ultimately, the goals of adaptive automation are similar to those of automation in general; improve overall performance while avoiding ìoperator out of the loopî conflicts or mistrust in the automation.

The Communications Scheduler was developed to mitigate the bottleneck in cognitive resources via task-based management and modality-appropriate information presentation strategies. Such technologies not only have the potential to significantly reduce the strain on the Soldiers' cognitive resources, but they also provide the opportunity to improve overall decision making by better managing information flow (Schmorrow, Raley, & Ververs, 2004). The overall result is a benefit by making smarter decisions about what information gets presented and when.

SYSTEM DESCRIPTION

The adaptive system evaluated in this paper has two principle components: 1) the Cognitive State Assessor, and 2) the Communication Scheduler.

Cognitive State Assessor

Inferring cognitive state from non invasive neuro-physiological sensors is a challenging task even in pristine laboratory environments. However, our goal was to accomplish this evaluation outside the laboratory in an outdoor environment with the participants untethered to a fixed base station of computers. In addition the information gathered on the cognitive state of the individual needed to be determined in real-time. Therefore all the artifacts known to influence the signal such as eye blinks, muscle strain, and electrical line noise needed to tagged and/or removed before being submitted for classification. See system depiction in Figure 1.

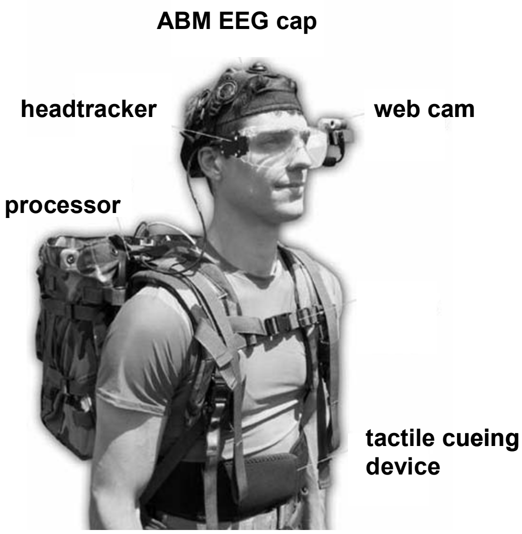

Figure 1. Fully Mobile Ensemble

.

As part of the experiment, data were collected using two EEG systems: (1) Advanced Brain Monitoringís (ABM) 6 channel differential EEG headset sampled from bipolar channels CzPOz, FzPOz, F3Cz, F3F4, FzC3, C3C4 at 256 samples per second and (2) a 32 channel BioSemi system, though only 7 channels (CZ, P3, P4, PZ, O2, P04, F7) were used for the cognitive state classification. These sites were selected based on a saliency analysis on EEG collected from various participants performing cognitive test battery tasks (Russell & Gustafson, 2001).

The EEG signals are pre-processed to remove eye blinks using an adaptive linear filter based on the Widrow-Hoff training rule (Widrow & Hoff, 1960). Information from the VEOGLB ocular reference channel was used as the noise reference source for the adaptive ocular filter. DC drifts were removed using high pass filters (0.5Hz cut-off). A band pass filter (between 2Hz and 50Hz) is also employed, as this interval is generally associated with cognitive activity.

The experimental setup supported real-time cognitive state classification by including training periods that emulated subsequent low and high workload conditions. After collecting between five and ten minutes of EEG spectra data for both low and high workload training conditions, the data were submitted to the composite classification system to identify patterns to distinguish the workload conditions. The classification system used three distinct classification approaches: K nearest neighbor (KNN), Parzen windows, and the Gaussian Mixture Models. See Mathan, Mazaeva, Whitlow, Adami, Erdogmus, Lan, and Pavel (2005) for a detailed description of the classification technique.

Communication Scheduler

The Communication Scheduler adapts the information environment via task scheduling and modality management of incoming communications. The system is tasked with determining when and how information is displayed to the Soldier. The Communications Scheduler schedules and presents messages to the soldier based on the cognitive state profile, the message characteristics, and the current context (tasks). Based on these inputs, the Communications Scheduler would pass through messages immediately, divert attention to incoming higher-priority messages, and change the modality of message presentation of lower priority messages,.

All messages had a priority associated with them, depending on their criticality. High priority items were mission-critical and time-critical and had to be heard and understood as soon as they arrived. Low priority messages were not critical (although they may still be important). When the mitigation was in effect and high workload cognitive state conditions were detected, the low priority messages were presented in text format in the message window of the participantís Personal Digital Assistant (PDA).

Poorly designed automation can be dangerous. Research shows that unless users are able to predict clearly how an automated system is likely to perform, automation may introduce more problems than it solves (Sarter, Woods, & Billings, 1997). The mitigations strategy described here has very clear rules to eliminate uncertainty and unpredictability. The Communications Scheduler benefits users by allowing responses to low messages to be deferred under conditions when attention has to be split between competing tasks, thus allowing users to focus on higher priority tasks first. However, this kind of automated system behavior has negative side effects: Loss of momentary situational awareness and lags in responses that could break coordination among teams and introduce inefficiencies in the mission. Thus it was important to invoke the Communications Scheduler only when the benefits of its use outweigh its costs. For that reason the Communications Scheduler was not used continuously, but rather only in times of high cognitive stress of the user, when faced with competing tasks that overload his or her ability to comprehend and process all incoming information.

Since the Communications Scheduler was not used continuously, the issue of automation etiquette became important. The Communications Scheduler needed to be invoked (and withdrawn) in such a manner that it would not cause confusion or induce unwanted oscillations in workload due to unpredictability. The Communications Scheduler mitigation was invoked when workload was high ñ for instance low priority messages were deferred to the PDA. However, when workload lowered below the threshold used to trigger the message deferral, the Communications Scheduler continued to defer messages. This design decision was made due to the fact that deferring communications on the basis of moment to moment fluctuations in gauge values can be confusing. Messages could be misinterpreted without surrounding context if they were to be played in audio modality after its predecessor messages have been deferred to the PDA (and remain unread for a period of time). If expected messages were not heard, it may have been hard to disambiguate whether this is because of the communications scheduler or some mission related cause. To avoid confusion, once communications scheduling was activated, all low priority messages were deferred to PDA until user was caught up on all messages and clicked a 'messages read' button (Mathan, Dorneich & Whitlow, 2005).

The Honeywell effort was concerned with mitigating high workload demands in the dismounted soldier environment, especially with regard to information overload due to communications. This particular effort demonstrated a prototype that incorporated the cognitive state of the user with an adaptive system designed to maintain high levels of performance under increasing task loads. We evaluated the effectiveness of the classification algorithms to detect the user's cognitive state by correlating classification output to performance in various task load conditions. We investigated the effectiveness of the Communications Scheduler to modify information flow based on cognitive state and thereby influence overall performance. We expected scheduling information would enhance performance on the primary tasks in high task load conditions, while not degrading performance on the remaining tasks.

METHOD

Objectives

We designed the evaluation to answer the following questions: Will the integrated sensor-driven classification of cognitive state detect a change in the participant's cognitive state between low task load and high task load conditions? Will the Communications Scheduler mitigation strategy effectively alter the participant's cognitive/attentional state in order to focus attention and improve comprehension of the highest priority items? Will there be any cost to the use of the Communications Scheduler such as a loss of situation awareness of lower priority message content?

Participants

There were eight male volunteers who participated in this evaluation. They ranged in age from 21 to 42 years of age with an average age of 29.5 years.

Hypothesis

Scheduling of information would enhance the participantís performance on the counting task and mission monitoring tasks in high task load conditions, while not degrading performance on the remaining tasks.

Tasks

Each participant played the role of a platoon leader (PL). They were responsible for managing three squads while reporting to his or her company commander (CO). Participants navigated a known and secure route while monitoring a bounded overwatch mission, maintaining radio counts, and performing a periodic interruption task presented on the handheld Personal Digital Assistant (PDA) task. The tasks are described below.

Navigation task. The participants navigated along a familiar and marked route. The main goal of this task was to keep the participants mobile during the mission.

Maintain radio counts secondary task. A simulated CO relayed messages about entities encountered by his three platoon leaders over the radio, of which the participant was one. The messages contained reports of civilians, enemies, or friendlies spotted. The participant maintained a running total of civilians, enemies, and friendlies reported to him, while ignoring the counts reported to the other two platoon leaders. This task relied heavily on the participants' ability to keep the three counts in working memory until he was asked to report the counts.

Mission monitoring secondary task. The participant organized the execution of a series of bounded overwatch maneuvers by three squads under his or her command. In bounded overwatch, one squad moved while the other two squads protected the moving squad. Participants kept track of the status of all three squads ó either "ready to move" or "ready for overwatch." Once all three squads reported that they were in position (two squads ready for overwatch and one squad ready to move), participants ordered the appropriate squad to move forward. This task required the participants to keep the track of the three squads, their locations, and readiness to advance in the mission in working memory until the final team was in position.

Math interruption tertiary task. A simple math problem was periodically presented to the participants as an interruption task during the scenarios. This task was representative of any type of unanticipated interrupt that requires significant cognitive resources and an immediate response from the platoon leader. Participants were interrupted twice per minute in both high and low task load periods. This interruption task had the potential of disrupting any of the tasks that required continual rehearsal, such as the working memory tasks of mission monitoring and maintaining counts.

Maintain situation awareness. Participants were required to maintain a situation awareness of their current location, the status of all teams and personnel reporting to them, the overall situation as relayed through radio communications, and their surroundings. Participants were also asked about the content of all messages that they received, including the delayed low priority messages.

Experimental Design.

This was a 2 (mitigation: mitigated, unmitigated) x 2 (task load block: high, low) within subjects design.

Procedure

Scenarios were run in a large grassy field surrounded by light forest situated behind Honeywell Labs in Northeast Minneapolis, Minnesota. After a training period, the participants completed two scenarios of four task load blocks each in a fixed order: High, Low, High, Low. The blocks of high and low task load conditions lasted approximately five minutes and two minutes, respectively. All four tasks: (1) navigation, (2) maintain radio counts, (3) mission monitoring, and (4) mathematical interruption task were performed simultaneously. Participants primarily interacted with a handheld radio and a PDA. Input for the mission monitoring and counts tasks came over the radio and they responded over the radio as well. The primary difference between high and low task load periods was the pace of radio communications. The math interruption task occurred at equal frequency under both task load conditions. In the mitigated condition, during detected high workload states, the participant had the Communications Scheduler to defer the low priority radio counts messages to a period of low task load by changing the modality of the messages to text on the participant's PDA.

RESULTS

There were four primary areas of interest in the data analysis. First was the accuracy of the cognitive state classification. How well could an individualís cognitive state be classified on a moment to moment basis? Second was the effectiveness of the workload manipulation. The goal was to create high and low workload periods for the participants. Subjective measures were used to determine if task load correlated with workload. Third was the participantsí confidence in the system to mitigate that workload. Subjective measures considered a participant's confidence in completing the tasks under the various experimental conditions. And finally objective task performance under the mitigated and unmitigated conditions. Task performance was assessed on the four tasks in each experimental condition.

Cognitive State Classification

A crucial component of classification in field settings was a systematic procedure for selecting a subset of EEG features that is robust to potential artifacts and provides a basis to discriminate between workload classes. With an appropriate selection of channels we were able to classify cognitive state with an accuracy that exceeded 70% for all participants, observing classification accuracies as high as 95%. Performance with both the BioSemi (2 participants) and ABM (6 participants) system was close to identical in the field environment.

Subjective Workload via NASA TLX

The Communications Scheduler mitigation significantly lowered the participants' subjective workload during high task load blocks of the scenario. The participantsí subjective assessment of mental demand ($F_{1,7}$=28.9, p=.001), temporal demand ($F_{1,7}$=15.9. p<.05) performance ($F_{1,7}$=8.8, p<.05), effort ($F_{1,7}$=35.5, p<.05), and frustration ($F_{1,7}$=10.1, p< .05) was significantly improved by the mitigation over the unmitigated trials, under high task load. Physical demand remained unchanged ($F_{1,7}$=3.7, p=0.10), as was expected given the cognitive nature of the task load manipulation.

Confidence

The participants rated their *confidence* in their ability to perform well, where a higher rating equates to an increased confidence in performance. The baseline task load condition saw no statistically significant ($F_{1,7}$=0.77, p=.41) difference in means. The participants had more confidence in their performance in both task load conditions when the mitigation was present. Confidence increased significantly ($F_{1,7}$=7.00, p<.05) from 4.3 (unmitigated) to 5.2 (mitigated) in the low task load condition and also increased significantly ($F_{1,7}$=6.45, p<.05) from 3.6 (unmitigated) to 5.1 (mitigated) in the high task load condition.

Objective Measures Performance

Maintain counts task. Participants showed a statistically significant increase in accuracy of maintaining counts in high task load condition when the Communications Scheduler mitigation was available. Participants in high task load performed at a 67.4% accuracy when unmitigated, but their performance jumped to 95.7% accuracy when the tasks were mitigated. The effect was statistically significant ($F_{1,7}$=16.8, p<.05). In the low task load participants saw performed equally in both mitigation conditions ($F_{1,7}$=0.68, p=.44); 83.3% (unmitigated) to 89.2% (mitigated). This is consistent with the hypothesis that the benefits of mitigation are realized in high task load times.

Mission monitoring task. Participants showed a statistically significant increase in accuracy of mission monitoring in high task load when the mitigation was available. Participants in high task load conditions performed at 68.2% accuracy when unmitigated, but their performance jumped to 95.8% accuracy when mitigated. The effect was statistically significant ($F_{1,7}$=18.9, p<.05). In the low task load condition, participants saw a slight increase in mean performance (92.2% to 100%) with the mitigation, although this difference was not statistically significant ($F_{1,7}$=3.72, p=.09). Again, this is consistent with the hypothesis that the benefits of mitigation are realized in high task load and the resulting high workload times.

Low priority message situation awareness. It was hypothesized that mitigation, while providing benefit, may have costs associated with it that make it inappropriate to leave the mitigation on all the time. In order to assess the possible costs of the Communication Scheduler, participants were asked situation awareness questions at the end of each high task load block that pertained to low priority messages that were deferred to the PDA for later review. Three questions were asked at the end of the two high task load blocks. Though participants did a good job filtering the low priority messages on their own so shown by the average of 30% (mean of 0.9 out of 3, standard error = .187) correct on the questions in the unmitigated scenarios. Participants in the mitigated scenario were unable to answer any of the questions, since they had not had time to review low priority messages. This effect was statistically significant ($F_{1,4}$=23.1, p<.05).

Math interruption task. Participants responded to the interruption alert much more quickly in the low task load condition than the high task load condition, indicating the availability of additional resources to attend to the interrupt in the low task load condition. In the low task load condition, mitigation slightly decreased the response time from 4.9 sec (unmitigated) to 3.7 sec (mitigated), although not significantly ($F_{1,3}$=1.85, p=0.27). In the high task load condition, where we expect to see the benefits of mitigation, the response time was faster under mitigation by almost five seconds going from 8.6 sec (unmitigated) to 3.8 sec (mitigated). The difference approached significance ($F_{1,6}$=4.8, p=.07). (Note: Due to data logging issues, only four of eight participants' data were recorded for the low task load condition. Seven of eight participants' data were used in the analysis of the high task load condition.)

DISCUSSION

This evaluation focused on the real-time cognitive state assessment in an unconventional, outside the laboratory, mobile environment. The results indicated that the task load manipulations effectively varied the participantsí workload as evidenced by the subjective workload scores and objective task performance. The classifier correctly determined the cognitive state of the participants on a moment-to-moment basis over 70% of the time based on the EEG inputs alone. The Communications Scheduler effectively mitigated performance when high workload states were detected by rescheduling lower priority messages to the PDA device. This was an effective means to improve performance as evidenced by the rebounded scores on the working memory tasks: radio counts and mission monitoring. However, the performance advantages were not realized without a cost. There was a temporary loss of situation awareness of low priority messages. We believe that it may be an acceptable cost when performance on high priority tasks is failing due to cognitive overload. The Communications Scheduler was designed to activate when the Cognitive State Assessor detected a cognitive overload and demonstrated significantly improved performance on the high priority tasks of maintain counts and mission monitoring. This is may be a case of the benefits outweighing the costs.

ACKNOWLEDGMENTS

This paper was supported by a contract with DARPA and the U.S. Army Natick Soldier Center, DAAD16-03-C-0054, for which CDR Dylan Schmorrow served as the Program Manager of the DARPA Improving Warfighter Information Intake Under Stress/AugCog program and Mr. Henry Girolamo was the DARPA Agent. The opinions expressed herein are those of the authors and do not necessarily reflect the views of DARPA or the U.S. Army Natick Soldier Center.

REFERENCES

Dixon, S.R. & Wickens, C.D. (2004). Automation reliability in unmanned aerial vehicle flight control. In D. Vincenzi (Ed.) *Proceedings of Human Performance and Situation Awareness and Automation (HPSAA) II*. Daytona, FL.

Mathan, S., Mazaeva, N., Whitlow, S., Adami, A., Erdogmus, D., Lan, T., & Pavel, M. (2005). Sensor-based cognitive state assessment in a mobile environment. In G. Salvendy (Ed.) *Proceedings of the 11th International Conference on Human-Computer Interaction* (1st Annual Augmented Cognition International conference). Mahwah, NJ: Lawrence Erlbaum.

Mathan, S., Dorneich, M., & Whitlow, S. (2005). Automation Etiquette in the Augmented Cognition Context. In G. Salvendy (Ed.), *Proceedings of the 11th International Conference on Human-Computer Interaction (1st AnnualAugmented Cognition International)*. Mahwah, NJ: Lawrence Erlbaum.

Parmentola, J. A. (2004). Army transformation: Paradigm-shifting capabilities through biotechnology, *The Bridge, 34*(3), The National Academy of Engineering of the National Academies. Retrieved January 28, 2005 from http://www.nae.edu/NAE/ bridgecom.nsf.

Rovira, E., Zinni, M., & Parasuraman, R. (2002). Effects of information and decision automation on multi-task performance. In *Proceedings of the 26th Annual Meeting of the Human Factors and Ergonomics Society*. (pp. 327-331). Santa Monica, CA: HFES.

Russell, C.A. & Gustafson, S.G. (2001). *Selecting Salient Features of Psychophysiological Measures*. Air Force Research Laboratory Technical Report (AFRL-HE-WP-TR-2001-0136).

Sarter, N. B., Woods, D. D., & Billings, C. E., (1997). Automation surprises. In G. Salvendy (Ed), *Handbook of Human Factors and Ergonomics*, (2nd Ed.). New York: Wiley.

Schmorrow, D., Raley, C., & Ververs, P. (2004). Toward effective warfighting in stressful environments. Poster presented at *Human Performance and Situation Awareness and Automation (HPSAA) II*. Daytona, FL.

Wickens, C.D. & Hollands, J. (2000). *Engineering Psychology and Human Performance*. Prentice Hall, 3rd edition.

Widrow B., & Hoff, M. E. (1960). Adaptive switching circuits. In *IRE WESCON Convention Record*, pp. 96--104.