Dorneich, M.C., Ververs, P.M., Mathan, S., and Whitlow, S.D (2005). "A Joint Human-Automation Cognitive System to Support Rapid Decision-Making in Hostile Environments" *Proceedings of the International Conference on Systems, Man and Cybernetics*, October 10-12; Waikoloa, HI.

A Joint Human-Automation Cognitive System to Support Rapid Decision-Making in Hostile Environments

Michael C. Dorneich, Patricia May Ververs, Santosh Mathan, Stephen D. Whitlow
Human-Centered Systems
Honeywell Laboratories
Minneapolis, MN, USA
michael.dorneich@honeywell.com

Abstract – *Honeywell has designed a joint human-computer cognitive system to support rapid decision making demands of dismounted soldiers. In highly networked environments the sheer magnitude of communication amid multiple tasks could overwhelm individual soldiers. Key cognitive bottlenecks constrain information flow and the performance of decision-making, especially under stress. The adaptive decision-support system mitigates non-optimal human performance via automation when the system detects a breakdown in the human's cognitive state. The human's cognitive state is assessed in real-time via a suite of neuro-physiological and physiological sensors. Adaptive mitigation strategies can include task management, optimizing information presentation via modality management, task sharing, and task loading. Mitigations are designed with consideration for both the costs and benefits of intermittent augmentation The paper describes the system development and evolution, explorations of usable cognitive mitigation strategies, and four evaluations that show adaptive automaton can effectively, mitigate human decision-making performance at extremes (overload and underload) of workload.*

Keywords: Augmented Cognition, decision making, decisions support system, adaptive system.

1 Introduction

Honeywell has built a joint human-computer cognitive system to support rapid decision making demands placed on the dismounted soldier. The U.S. Army is currently defining the roles of the 2010-era Future Force Warrior (FFW). The FFW program seeks to push decision making requirements to the lowest levels and posits that with enhanced netted communications capabilities a squad can cover the battlefield in the same way that a platoon now does. Among other capabilities, the application of a full range of netted communications and collaborative situational awareness will afford the FFW unparalleled knowledge as the basis of increased decision making responsibility and independence.

An approach was adopted that considers the joint human-computer cognitive system when identifying bottlenecks to improve system performance [1]. Key cognitive bottlenecks constrain information flow and decision-making, especially under stress. From an information-processing perspective, there is a limited amount of resources that can be applied to processing incoming information due to cognitive bottlenecks [2][3][4][5]. The Honeywell team is focused on reducing the cognitive limitations imposed on cognitive resources used for information processing in a highly dynamic, information rich environment. Of particular interest to Honeywell is attention, or attention as a *bottleneck* in processing. Attention can be broadly defined as a mechanism for allocating cognitive and perceptual resources across controlled processes [6]. In order to perform tasks effectively, one must have the capacity to direct attention to task relevant events and maintain a level of alertness. One must also be able to narrow (focus) or broaden (divide) one's field of attention appropriately depending on the demands of a task. Attention can be stimulated by external events (e.g., responding to an aural warning) as well as being thought of as a state where the level of awareness can also be maintained consciously, as a controlled top-down process. Breakdowns in attention lead to multiple problems: failure to notice an event in the environment, failure to distribute attention across a space, failure to switch attention to highest priority information, or failure to monitor events over a sustained period of time. The appropriate allocation of attention is important to FFW because it directly affects two cornerstone technology thrusts within the FFW program: netted communications and collaborative situation awareness.

This paper will describe the system development and evolution of the the Joint Human-Automation Augmented Cognition System (JHAACS), explorations of usable cognitive mitigation strategies, and experiments assessing their effectiveness.. JHAACS is an adaptive system that mitigated non-optimal human performance via automation when the system detects a breakdown in the human's cognitive state. The human's cognitive state is assessed in real-time via a suite of neuro-physiological and physiological sensors. Adaptive mitigation strategies can include task management, optimizing information presentation via modality management, task sharing, and task loading. For instance, 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. For example, an air traffic controller might be presented with decision aids for conflict detection and resolution by the decision-support system when it detects a rapid increase in traffic density or complexity [7]. Modality management mitigation strategies might include utilizing available resources (i.e., audio, visual) to increase information throughput. Task offloading and task sharing to automation are also mitigation strategies to reduce workload. Ultimately, the goals of adaptive decision support systems are similar to those of automation in general, such as avoiding "operator out of the loop" conflicts or mistrust in the automation.

Adaptive decision-support systems assistance triggered by real time classification of cognitive state offers many advantages over traditional approaches to automation. These systems offer the promise of leveraging the strengths of humans and automation - augmenting human performance with automation specifically when human abilities fall short of the demands imposed by task environments. However, by delegating critical aspects of complex tasks to autonomous automation components, these systems run the risk of introducing many of the problems observed in many traditional human-automation interaction contexts. The pros and cons of automating complex systems have been widely discussed in the literature (e.g. [8][9]) However, as widely noted, poorly designed automation can have serious negative effects. Automation can relegate the operator to the status of a passive observer – serving to limit situational awareness, and induce cognitive overload when a user may be forced to inherit control from an automated system. Thus the design of JHAACS has explicitly considered the costs as well as the benefits of mitigation when deciding when and how to intervene in the decision making process.

This paper describes the system architecture, design, and evaluation of JHAACS. Section 2 will briefly overview the system architecture, cognitive state assessement, and the mitigation strategies employed by JHAACS to improve joint human-automation performance while taking into account issues of automation etiquette mentioned above. Section 3 will describe various evaluations of the system, followed by a short discussion in Section 4.

2 System Description

The system architecture of JHAACS (see Figure 1) has the following components:

- *Cognitive State Assessor* (CSA) combines measures of cognitive state to produces the cognitive state profile (CSP).
- *Human-Machine Interface (HMI)*, where human interacts with the system via a TabletPC, tactor belt, and radio.
- *Automation*, where tasks can be partially or wholly automated.
- *Augmentation Manager* (AM), adapts the work environment to optimize joint human-automation cognitive abilities for specific domain tasks. The AM is comprised of three components:
  - *Interface Manager*, responsible for realizing a dynamic interaction design in the HMI
  - *Automation Manager*, responsible for the level and type of automation
  - *Context Manager*, responsible for tracking tasks, goals, and performance
- *Virtual Environment* (not shown) is a simulated approximation of the real world.
- *Experimenter's Console* gives the experimenter both insight and control over events within the system.

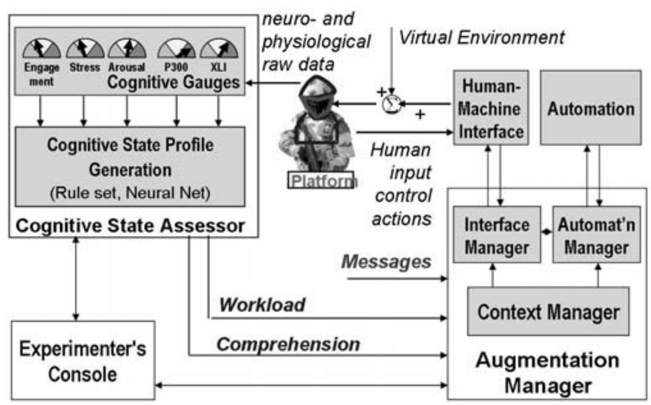

Figure 1. AugCog system architecture.

2.1 Assessing Cognitive State

Several methods of real-time assessment of cognitive state have been explored. One approach has been to develop "gauges" to measure different aspects of cognitive states. The gauges used in the first three of the evaluations were defined as:

- *Engagement Index* uses a ratio of EEG power bands, beta/(alpha + theta), and has been shown to be a valid measure of alertness in vigilance type tasks [10].

- *Arousal Meter* is a real-time measure of cognitive arousal derived from cardiac inter-beat-intervals (IBIs) [11].

- *Stress Gauge* is a composite gauge that uses heart rate, pupil diameter and microvolt cardiac QRS waveform root mean square (RMS) amplitude (HFQRS) to determine stress during individual trials.[12].

- *P300 Novelty Detector Gauge* equates the strength of the EEG evoked response following an alert tone with the availability of attentional resources to process the message following the auditory tone. [13][14].

- *XLI Gauge* measures of executive load or comprehension, by measuring power in the EEG at frontal (FCZ) and central midline (CPZ) sites [15].

The CSA outputs a Cognitive State Profile (CSP) comprised of two decision state variables: *workload* and *comprehension*. The CSP drives the mitigations of the Augmentation Manager. Currently a simple set of rules is used to derive the assessment of *workload* and *comprehension*, although work is underway to define this step with neural networks, in order to better account for individual subject differences [16]. *Workload* was considered high if any of the three gauges, Engagement, Arousal, or Stress, registered high. This was done since, for any given subject, only a subset of the gauges may be able to discriminate differing levels of workload, based on individual differences. Likewise, in order to bias *comprehension* towards false positives, both the Novelty Detector and XLI gauges had to be high (i.e. reporting a yes that the subject could comprehend the message) for the *comprehension* variable to be set to "True".

The Honeywell team has also been exploring the use of a real-time neural net classification approach to status the cognitive state of the participants. This approach used EEG data as the sole classification input for cognitive state. The neural network classification used three different techniques to estimate cognitive state: Gaussian-Mixture-Model (GMM), K-Nearest-Neighbor (KNN), and a nonparametric Kernel Density Estimate (KDE) (see [17] for more information). Cognitive state was classified into high and low cognitive workload states and was based on the agreement from two of the three models. When there was no majority agreement the KDE estimation was used.

2.2 Mitigation Strategies of the Augmentation Manager

It is a three-stage process for building effective mitigation strategies. First, the cognitive state assessment techniques needed to be developed and validated Second, the mitigation strategies are assessed to ensure that they would enhance performance if properly driven by the CSA. The final stage was to build the ruleset for applying the mitigation strategies. This includes the automation etiquette for turning it on and, maybe more importantly, turning it off. Honeywell has developed four mitigation strategies.

Communications Scheduler schedules and presents messages to the soldier based on the cognitive state profile (derived from the gauges), the message characteristics (principally priority), and the current context (tasks). Messaging techniques include drawing attention to higher priority items with the additional alerting tones or visual text messages, or deferring lower priority messages to a tablet PC device for later review. The Communications Scheduler determined the initial message presentation based on a user's current *workload*. After the first presentation of a message to the user (in audio modality), the Communications Scheduler determined whether to take further action on a message depending on the CSA's assessment of *comprehension*. Augmentation was turned off only when *workload* fell to manageable levels *and* the soldier had read ("caught up") all the messages on the Tablet PC. This was to avoid disorientation and lack of context should the system start presenting low priority messages auditorily if previous, perhaps relevant, messages,had been deferred and possibly not yet read.

Tactile Navigation Cueing System guides the soldier via tactile cues in the intended correct direction. This transforms the normally cognitively intense navigation/orientation task to a task that is reactionary in nature. The system was invoked when *workload* was high and the subject needed to navigate through an unfamiliar route. However, turning the system off as soon as *workload* fell below some threshold would leave users disoriented in an unfamiliar area. Thus once the system is turned on, the mitigation persists until users get to the safe destination.

Task Offloading Negotiation Agent (e.g., Medevac Agent) reduces lengthy verbal communications exchanges by automatically preloading information from netted communications into forms. The medevac agent only contains the most critical information needed for a Medevac. The unaugmented, more detailed information exchange might allow for safer and more efficient Medevac operations. Additionally, engaging in Medevac transactions may contribute to better situational awareness of a team's status. It is for these reasons that the Medevac Negotiation Agent is only invoked when the subject's *workload* is so high and the subject's performance so inadequate, that the differences associated with automated negotiation are acceptable in terms of overall performance.

Mixed Initiative Target Identification Agent provides assistance in locating potential targets in a visual search space. Research suggests that performance on these tasks deteriorates considerably over time. Automated systems trained to detect target stimuli in a field may not perform as well as an alert human. Consequently, they may not be able to completely replace the human operator in operational contexts. However, they could be helpful role if triggered when gauges detect a vigilance decrement.

3 Evaluation and Results

Four evaluations have been conducted to date to assess the effectiveness of JHAACS, each successive evaluation building on the findings of the previous. The first evaluation demonstrated that JHAAS mitigations were most effective at the extreme (high) end of workload. The second evaluation investigated workload at both overload and underload extremes. The third evaluation looked at cognitive state gauge assessments in a mobile environment (the previous two experiment were in a stationary desktop environment). Finally, the forth evaluation looked at a neural net approach to cognitive state assessment, as well as investigating the differences in system performance between mobile and stationary scenarios. Experimental design and results for each evaluation are briefly described here. For a full write-up of each evaluation see [18][19] [20].

The first evaluation was conducted in a desktop virtual environment with the participants tasked with navigating through a Military Operations in Urban Terrain (MOUT) (see Figure 2).

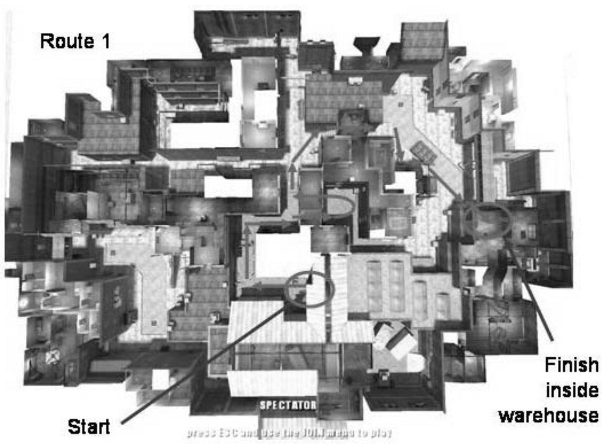

Figure 2. Top-down view of a MOUT environment.

Additional tasks included identifying friend from foes and monitoring and responding to communications. Gauges measured the cognitive responses to task load and availability of attentional resources to comprehend a message and triggered the Communications Scheduler appropriately (e.g., repeat message, defer message). The experiment was 2x2 within subjects design, with 12 subjects. Independent variables were task load (high/low) and mitigation (on/off), where subjects performed four trials in each condition (16 total). Dependent variables included measures of situation awareness via message content probes, message responses, and post-trial questions. A noteworthy finding for the cases in which the participant failed to acknowledge the message, the gauges indicated that the Communications Scheduler should repeat the message 72% of the time. This result led to a significant ($p < .02$) improvement in the situation awareness metric in the high workload scenarios for the mitigated condition. In addition, approximately 24% of the time the scheduler repeated a message even if the participant acknowledged its receipt. This is the upper bound on the false positive rate. However, anecdotal evidence suggest the possibility that at times the gauges were indicating the failure to truly comprehend the message even if the participant responded (indicating an automatic response even though they did not fully process or understand the message). See [18] for more details.

The findings from the first evaluation indicated that the greatest benefit of the mitigation occurred at the extreme end of task load space. Therefore, the second evaluation investigated both extremes of workload by including long duration and clear task load differences built into the scenarios. Several more mitigation strategies were developed. In addition to the Communications Scheduler , other mitigations included a tactile navigation cueing system, a negotiation agent for offloading components of a highly proceduralized task (i.e., calling for a Medevac), a mixed initiative target identification agent for enhancing the search in a vigilance task. The mitigations were triggered when the gauges indicated a suboptimal cognitive state. The experiment was a 2 (Mitigation) x 3 (Scenario) design. Scenarios 1 and 2 were focused on divided attention and multi-tasking, and were a within subjects design. Tasks were similar to the previous evaluation. Scenario 3 was a vigilance paradigm and between subjects. Thus subjects conducted 7 trials. Findings with 14 subjects revealed significant improvements in performance in the relevant applicable task with the availability of the automation provided by the mitigation agents. The Communications Scheduler mitigation resulted in 100% improvement in dependent metrics of message comprehension and 125% improvement in situation awareness (as measured by subject responses to messages), with no negative affect on ability to engage foes, and negative affect on workload (assessed via NASA TLX survey). The Tactile Navigation Cueing mitigation resulted in a 20% decrease in evacuation time, and reduced the number of enemy encounters by over 350% (increasing survivability). The Medevac Negotiation Assistance mitigation resulted in a 96% improvement in communication of critical information, and an over 300% improvement in time to complete the negotiation. For all three mitigations there was no negative effect on ability to engage foe, nor workload. In addition, most subjects felt the mitigated tasks were easier with augmentation. Performance for identifying targets in Scenario 3 was significantly higher for the mitigated condition, resulting in a 30% improvement over the unmitigated condition. See [19] for more details.

The third evaluation provided the opportunity to understand the cognitive state gauge assessments in a mobile environment. The scenarios took place in a motion capture laboratory at Carnegie Mellon University. The experiment was a Dual Task Design with each block consisting of 4 primary-secondary task pairs, for a total of 8 blocks (or trials) per subject. The first part (primary task) consisted of monitoring the building, shooting enemies, and maintaining 3-5 counts. In the second part (secondary task) 12 subjects noted friendly or enemy soldiers walking past any windows, monitoring radio messages, and maintaining reported counts of enemies and friendlies. The dependent variable was percentage of counts correct within Primary task. The participants stood in an 8x12 foot space with a motion-tracked M16 rifle prop. Gauges were used to detect workload periods when the participant was maintaining the counts in working memory. During these periods, the gauge-driven Communciations Scheduler delayed incoming messages, reducing the task load on working memory. This condition was compared to a random scheduler of messages. Findings indicated over a 150% improvement on the working memory tasks as well as a reduction in the subjective mental workload measures in the mitigated condition over a random message scheduler. See [20] for more details.

In a fourth evaluation, JHAACS employed a neural network classification approach to cognitive state assessment and tested the system in a combination desktop/mobile study. The Communications Scheduler and the Mixed Initiative Target Identification were mitigation strategies. The experiment was an incomplete 2 (workload high/low) x 2 (mitigation on/off) x 2 (stationary/mobile). For the first half of the evaluation, the six participants were seated conducting three tasks on a TabletPC (stationary-unmitigated-low, stationary-unmitigated-high, stationary-mitigated-high). The tasks consisted of 1) monitoring radio communication reports on the number of enemies, friendlies, and civilians encountered and maintaining the running total, 2) monitoring radio communications on the movement of three squads and directing their movements in a bounded overwatch, and 3) monitoring visual presentations of static images for signs of enemy targets and report the position of the target. Dependent variables were similar to the previous evaluations. Workl,oad was varied by varying the rate at which communications and visual images were presented The actual high workload rate was predetermined during the training session by taxing the participants to the point where performance began to drop off. Resulting rates varied from 2 to 12 messages per minute and 2 to 24 images per minute. In the second half of the evaluation, the participants performed the first two communications tasks (unmitigated-low, unmitigated-high) while in an upright and mobile position while scanning the woods in an outside environment for targets, that is, concealed "snipers." Thus subjects performed 5 trials.

Findings from this fourth evaluation indicated that performance on the radio count accuracy and mission monitoring queries during the *mitigated* high task load condition was equivalent to the performance in the unmitigated low task load condition. The Communications Scheduler (mitigation) offloaded the radio count messages to the visual modality by sending them to the TabletPC thereby allowing the participant to monitor messages during a later lower workload period. Percentage improvement in a mitigated high workload as compared to the same condition unmitigated characterizes a participant's performance improvement. Overall there was a 94% improvement in the mission monitoring task ($p < .01$) and 36% improvement ($p < .01$) in the radio count recall task when the Communications Scheduler was available in the high workload trial as compared to the no mitigation condition. Workload levels (measured via NASA TLX) in the mitigated conditions mirrored those of the low task load conditions and were statistically improved in the unmitigated condition for ratings of perceived frustration ($p < 0.05$), and approached significance for temporal demand ($p = 0.10$) and effort ($p = .06$).

For the visual search task on the TabletPC (see Figure 3), participants had an average of 87% correct identification of targets in the low task load condition. Their performance dropped to 46% in the high workload but rebounded to 61% in the mitigated condition with the assistance of the target identification agent to identify potential targets. The availability of the agent resulted in a 40% average improvement in performance.

Figure 3. Visual search task on TabletPC. Highlighted box shows automated target identification.

4 Conclusions

Experimental results support the overall efficacy of the automation support for decision making in high stress environments where multiple concurrent tasks and information overload can significantly degrade human performance. The Honeywell Joint Human-Automation Augmented Cognition System helped subjects perform significantly better on communication, fighting, navigation, and vigilance tasks with neurophysiologically triggered automation. Most subjects felt that the automation components described here actually made task execution easier. Specific strategies to minimize the negative impact of intermittent automation support were also discussed. It is clear from the findings that we were able to build a adaptive decision-support system that enhanced performance with Army relevant tasks. The mitigation strategies employed in the augmentations of the adaptive system were able to significantly improve performance on the mitigated tasks, while not decrementing performance on concurrent tasks.

Acknowledgments

This research was supported by a contract (No. DAAD-16-03-C-0054)with DARPA Augmented Cognition program (CDR (Sel.) Dylan Schmorrow, Program Manager) and funded through the U.S. Army Natick Soldier Center (Mr. Henry Girolamo, DARPA Agent). Any opinions, recommendations, findings, or conclusions expressed herein are those of the authors and do not necessarily reflect the views of DARPA or the U.S. Army.

References

[1] Dorneich, Michael C., Whitlow, Stephen D., Ververs, Patricia May, Rogers, William H. (2003). "Mitigating Cognitive Bottlenecks via an Augmented Cognition Adaptive System," *Proc. 2003 IEEE Conf. on Systems, Man, and Cybernetics*, Washington DC, October 5-8, 2003

[2] Broadbent, D. E. (1958). *Perception and Communication*. New York: Pergamon

[3] Treisman, A.M. (1964). Verbal cues, language, and meaning in selective attention. *American Journal of Psychology, 77*, 206-219

[4] Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.

[5] Pashler, H (1994) Dual-task interference in simple tasks: data and theory. *Psychological Bulletin*, *116*, 220-244.

[6] Anderson, J.R. (1995). *Cognitive Psychology and its Implications* (2nd Ed.). New York: Freeman.

[7] Hilburn, B., Jorna, P.G., Byrne, E.A., & Parasuraman, R. (1997). The effect of adaptive air traffic control (ATC) decision aiding on controller mental workload. In M. Mouloua & J. Koonce (Eds.), Human-automation interaction: Research and practice (pp. 84-91). Mahwah, NJ: Erlbaum.

[8] Parasuraman, R., & Miller, C. (2004). Trust and etiquette in high-criticality automated systems. Communications of the Association for Computing Machinery, 47(4), 51-55.

[9] Sarter, N. B., Woods, D. D., Billings, C. E., (1997) "Automation surprises," Handbook of Human Factors and Ergonomics, 2nd Edition, Salvendy, Gavriel (Ed)., Wiley and Sons, 1997.

[10] Freeman, F.G., Mikulka, P.J., Prinzel, L.J., & Scerbo, M.W. (1999). Evaluation of an adaptive automation system using three EEG indices with a visual tracking system. Biological Psychology, 50, 61-76.

[11] Hoover, A. & Muth, E.(2004). A Real-Time Index of Vagal Activity. *Int. J. of Human-Computer Interaction*.

[12] Raj A. K., Bradshaw, J., Carff, R. W., Johnson, M., and Kulkarni, S. (2004). An agent based approach for Aug Cog integration and interaction. In *Proceedings of the Augmented Cognition- Improving Warfighter Information Intake Under Stress*, Scientific Investigators Meeting (January 6-8, 2004), Arlington, VA: DARPA.

[13] Parra, L., Alvino, C., Tang, A., Pearlmutter, B., Yeung, N., Osman, A., & Sajda, P. (2002). Linear Spatial Integration for Single-Trial Detection in Encephalography, *NeuroImage*, 7(1).

[14] Sajda, P., Gerson, A., & Parra, L. (2003). Spatial Signatures of Visual Object Recognition Events Learned from Single-trial Analysis of EEG, *IEEE Engineering in Medicine and Biology Annual Meeting*, Cancun, Mexico.

[15] DuRousseau, D.R., (2004b) *Multimodal Cognitive Assessment System*. Final Technical Report, DARPA, DAAH01-03-C-R232.

[16] Mathan, S., Mazaeva, N., & Whitlow, S. (2005). Sensor-based cognitive state assessment in a mobile environment. *Proceedings of the 11th International Conference on Human-Computer Interaction (Augmented Cognition International)*, Las Vegas, Jul 22-27, 2005.

[17] Erdogmus, D., Adami, A., Pavel, M., Lan, T., Mathan, S., Whitlow, S., & Dorneich, M. (2005). Cognitive state estimation based on EEG for Augmented Cognition. To appear in the Proceedings of the *2nd International IEEE EMBS Conference on Neural Engineering*. Piscataway, NJ.

[18] Dorneich, M., Whitlow, S., Ververs, P.M., Carciofini, J., and Creaser, J. (2004) "Closing the Loop of an Adaptive System with Cognitive State," *Proceedings of the Human Factors and Ergonomics Society Conference 2004*, New Orleans, September.

[19] Dorneich, M.C., Mathan, S., Creaser, J., Whitlow, S.D., and Ververs, P.M. (2005). "Enabling Improved Performance though a Closed-Loop Adaptive System Driven by Real-Time Assessment of Cognitive State." *Proceedings of the 11th International Conference on Human-Computer Interaction (Augmented Cognition International)*, Las Vegas, Jul 22-27, 2005.

[20] Whitlow, S., Ververs, P. M., Buchwald, B., & Pratt, J. (2005). "Scheduling communications with an adaptive system driven by real-time assessment of cognitive state.", *Proceedings of the 11th International Conference on Human-Computer Interaction (Augmented Cognition International)*, Las Vegas, Jul 22-27, 2005.