Ververs, P. M., Whitlow, S.D., Dorneich, M.C., & Mathan, S. (2005). "Building Honeywell's Adaptive System for the Augmented Cognition Program", Foundations of Augmented Cognition (D.D. Schmorrow, Ed.). Mahwah, NJ: Lawrence Erlbaum Associates. pp. 460-468.

Building Honeywell's Adaptive System for the Augmented Cognition Program

Patricia May Ververs, Stephen D. Whitlow, Michael C. Dorneich, Santosh Mathan

Honeywell Laboratories
Minneapolis, MN
trish.ververs@honeywell.com, stephen.whitlow@honeywell.com,
michael.dorneich@honeywell.com, santosh.mathan@honeywell.com

Abstract

The Honeywell AugCog team has developed a closed loop integrated prototype to address the performance advantages of a neurophysiologically driven system. The team has run a series of evaluations to validate the effectiveness of using cognitive state assessment to trigger performance mitigation strategies in an effort to improve overall system performance. This paper reviews the process for creating the closed loop integrated prototype, defining the military-relevant tasks used to test the system, and assessing the system in four separate evaluations. In parallel with the evaluations, the system was evaluated for feasibility on a mobile, dismounted soldier. The demonstration of this feasibility test is discussed.

1 Introduction

The Honeywell team is focused on reducing the 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 (Anderson, 1995). 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.

The Honeywell team is working with the U.S. Army's Future Force Warrior (FFW) program to build a system with mission relevance to the dismounted soldier. The appropriate allocation of attention is key to managing multiple tasks, focusing on the most important ones, and maintaining situation awareness. Appropriate allocation of attention is important to FFW because it directly affects two cornerstone technology thrusts: netted communications and collaborative situation awareness. See Ververs, Whitlow, Dorneich, Mathan, and Sampson (2005) for a perspective on how AugCog technology could be applied to the next-generation warfighter.

This paper outlines the development of a closed loop integrated prototype system driven by neuro-physiological and physiological state information. Described herein are the military tasks used to evaluate the effectiveness of this system, both the cognitive state classification techniques and the mitigation strategies for improving the overall system performance. The findings from four evaluations, two conducted in a desktop environment, one in an immersive virtual environment, and the most recent evaluation conducted both on a desktop and in a field test. A feasibility demonstration conducted in an Army relevant mission, known as a movement to contact, is also discussed.

2 Hardware Sensor System

The sensor suite evolved in response to the effectiveness of cognitive status assessment techniques and to meet the requirements of the various evaluations. The laboratory system consisted of a BioSemi 40 channel Active Two™ to measure EEG and EOG (electro-oculogram), a Cardiax PC-ECG (electrocardiogram) device to measure heart rate and interbeat interval (IBI), an IScan pupilometry device to measure pupil dilation, and a ProComp multi-physiologic measure device used to acquire electrodermal response (EDR) and electromyogram (EMG) for muscle activity. The laboratory navigational cueing task used a Tactile Situation Awareness System (TSAS) belt to provide the vibrotactile cues. See Dorneich, et al. (2004) for a full description of the system used in the laboratory. In the fourth evaluation and demonstration, the system consisted of the BioSemi EEG system, UFI's EZ-IBI unit. The hardware ensemble included a Fujitsu TabletPC and Anthrotronix' vibrotactile belt for tactile alerting cues.

3 Cognitive State Assessment

Researchers at the Institute for Human and Machine Cognition (IHMC) developed an agent-based architecture to provide architecture component developers with a simple interface to the system. See Figure 1. The modular architecture enables the components of cognitive state assessment such as hardware sensors (e.g., EEG, ECG) and software algorithms to be integrated and tested. Mission context is built from event recordings within the virtual environments and most recently from field sensor devices. Information regarding context is made available to the decision making elements (i.e., Augmentation Manager) through the agent architecture.

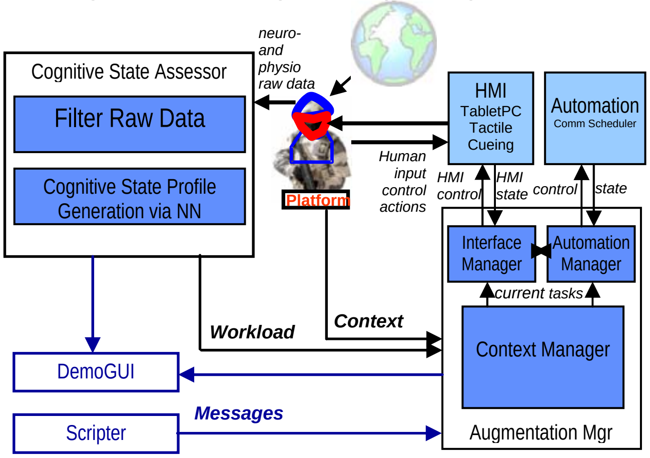

Figure 1. Current System Architecture

Several Honeywell AugCog team members have focused on the algorithms to determine cognitive state. One method used to access an individual's overall cognitive state has been to develop "gauges" to measure different aspects such as stress level, amount of executive function processing, degree of attentional resources engaged, etc. The gauges used in the first three of the evaluations are defined below.

3.1 Engagement Index

Engagement Index is an indicator of alertness. It uses a ratio of EEG power bands, beta/(alpha + theta). Freeman, Mikulka, Prinzel, and Scerbo (1999) have shown the Engagement Index to be a valid measure of an operator's task engagement in vigilance type tasks.

3.2 Arousal Meter

Arousal Meter (AM) is a real-time cardiac-based measure derived from interbeat intervals (IBIs) to status the activity of the autonomic nervous system (ANS). The heart responds to changes in the parasympathetic nervous system (PNS), which is responsible for returning an individual to a resting state. The AM, as developed by Clemson University researchers, uses the PNS subcomponent of ANS to status an individual's level of arousal (Hoover & Muth, 2004).

3.3 Stress Gauge

The Stress Gauge was developed by IHMC researchers and is a composite gauge that has used various inputs including heart rate, pupil diameter and microvolt cardiac QRS waveform root mean square (RMS) amplitude (HFQRS) to determine stress during individual trials (Raj, et al., 2004).

3.4 P300 Gauge

The P300 gauge measures the strength of EEG evoked responses following an alert tone. Once calibrated the detector is optimized to differentiate the EEG response evoked by the alert tones from that activity evoked by a frequent auditory stimulus of no significance. The idea behind the P300 is to equate the strength of the evoked potentials with the availability of attentional resources to process a message following an auditory tone. This gauge was developed by researchers at Columbia University and City College of New York (Parra, Alvino, Tang, Pearlmutter, Yeung, Osman, & Sajda, 2002; Sajda, Gerson, & Parra, 2003)

3.5 Executive Load Index

The Executive Load Index (XLI) Gauge is a measure of executive load or comprehension, where positive values indicate increasing load and negative values indicate decreasing load. It operates by measuring power in the EEG at frontal (FCZ) and central midline (CPZ) sites and was developed by Human Bionics (DuRousseau, 2004, 2004b).

For the fourth evaluation, the Honeywell team used 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 classifier used seven EEG channels (CZ, P3, P4, PZ, O2, PO4, F7) and five frequency bands: 4-8Hz (theta), 8-12Hz (alpha), 12-16Hz (low beta), 16-30Hz (high beta), and 30-44Hz (gamma) to form the features for the cognitive state assessment. 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 Erdogmus, Adami, Pavel, Lan, Mathan, Whitlow, Dorneich, 2005 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. In the event that there was no majority agreement the KDE estimation was used.

4 Task Definition

All four evaluations focused on Army relevant tasks that included a combination of communications, hostile engagements, and navigation tasks. The importance of these tasks was determined from interviews of subject matter experts situated in the FFW program. Of primary importance is the fact that the tasks are Army relevant to immediately enable the transfer of performance benefits to well-known challenges. The tasks also inherently stress the real challenges that must be overcome in order to deploy this technology. Evaluating the effectiveness of the system requires that the warfighter is mobile and frequently communicating to maintain situation awareness from the information being sent via the netted communications.

5 Building the Mitigation Strategies

A four-stage process was used for building effective mitigation strategies. First, the cognitive state assessment technique(s) was defined and built. Two “gauge” approaches to classification of cognitive state either applied preset calculations on the parameters of the signals (e.g., Engagement Index) or used a more dynamic, trained neural network classification. Next, the cognitive state assessment techniques were validated against the tasks to be used in the evaluation. Third, the mitigation strategies for alleviating the cognitive bottlenecks that negatively affect performance were created. The final stage of building the mitigation strategy was to create a ruleset for triggering the mitigation strategies based on the gauge responses. This ruleset included the automation etiquette for turning it on and maybe more importantly turning it off. The Honeywell team developed four mitigation strategies that are described below.

5.1 Communications Scheduler

The Communications Scheduler planned and presented messages to the soldier based on the cognitive state profile as defined by the gauges, the message characteristics (principally message priority), and the current context (tasks). Messaging techniques included drawing attention to higher priority items with the additional alerting tones or text messages or deferring lower priority messages until a later time or to a TabletPC device for later review.

5.2 Tactile Navigation Cueing System

Tactile Navigation Cueing System guided the soldier via tactile cues in the intended correct direction of a target location. This transformed a normally cognitively intense navigation task in an unfamiliar environment into a task that is reactionary in nature to the stimuli.

5.3 Task Offloading Negotiation Agent

Task Offloading Negotiation Agent (e.g., Medevac Agent) reduced lengthy verbal communications exchanges by automatically preloading known information from netted communications into the forms.

5.4 Mixed Initiative Target Identification Agent

The Mixed Initiative Target Identification Agent provided assistance in locating potential targets in a visual search space.

6 System Validation

Four different evaluations were conducted to determine the effectiveness of the Honeywell AugCog system. The first was conducted in a desktop environment where the participants navigated through a virtual Military Operations in Urban Terrain (MOUT) environment. See Figure 2. Additional tasks included identifying foes and shooting them and monitoring and responding to communications. Gauges measured the cognitive responses to task load and availability of attentional resources to comprehend a message, triggering the communications scheduler appropriately (e.g., repeat message, defer message). A noteworthy finding came from the incidences in which the participant failed to acknowledge the message, and the gauges indicated that the lack of attentional resources to comprehend the messages. The scheduler correctly repeated the message 72% of the time, with correctness being verified by lack of an acknowledgement. The availability of the scheduler resulted in significant improvements in situation awareness of message content in the high workload scenarios as compared the same condition when no mitigation strategy was used. In addition, approximately 24% of the time the scheduler repeated a message even if the participant acknowledged its receipt. Though this could be the result of an improper state classification, it also suggests the possibility that gauges were indicating the failure to truly comprehend the message even if the participant said, “Acknowledge.” That is, saying “acknowledge” to a message as an automatic response even though it wasn’t fully processed and understood would rightfully result in the repetition of the message. See Dorneich et al. (2004) for more details on the experimental conditions and findings.

The findings from the first evaluation indicated that the greatest benefit of the mitigation occurred at the extreme ends of the task load space. Therefore, this was capitalized on for the second evaluation in the desktop environment where long duration and clear task load differences were built into the scenarios. The tasks were similar to the previous evaluation. Several more mitigation strategies were developed. In addition to the communications scheduler, other mitigations tested were the tactile navigation cueing system, the negotiation agent for offloading components of a highly proceduralized task (i.e., Calling for a Medevac), and the 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. Findings revealed significant improvements in performance in the relevant applicable task with the availability of the automation provided by the mitigation agents. For instance, the navigation cueing reduced the number of enemy encounters by over 300% thereby increasing survivability. Message comprehension and overall situation awareness was improved by over 100% in the conditions when the communications scheduler was available. See Dorneich, et al. (2005) for more details.

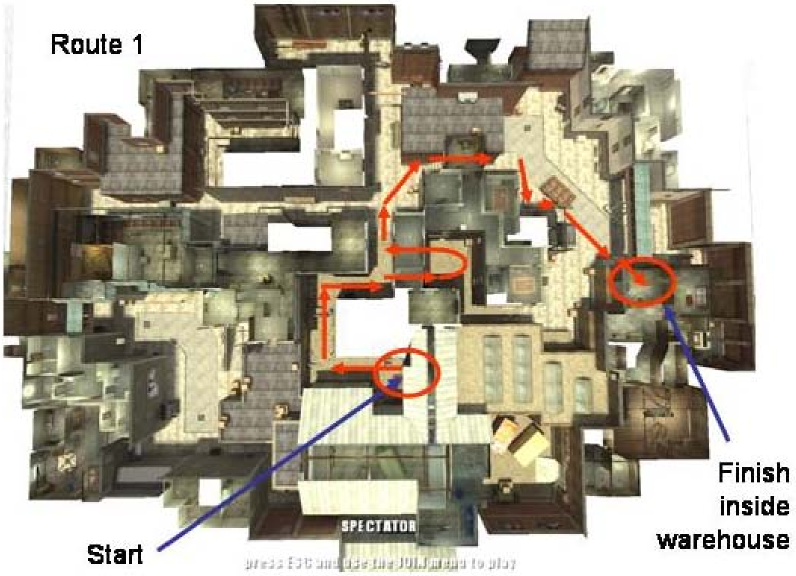

Figure 2. Top-down view of MOUT environment

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 participants stood in an 8 x 12 foot space with a motion-tracked M16 rifle prop and were tasked with visually scanning their environment for friends and foes, and shooting the foes when they appeared in the windows of buildings. See Figure 3. Participants were responsible for keeping track of the number of friends, foes and ammunition expended and to report out at various times during the scenario. In addition participants were responsible for listening for messages from team members verbally communicating their own enemy and friendly encounters among distractor messages and keeping a running count of the totals. Gauges were used to detect workload periods when the participant was maintaining the counts in working memory. When a high workload state was detected during these periods, the gauge-driven message 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 Whitlow et al. (2005) for more details.

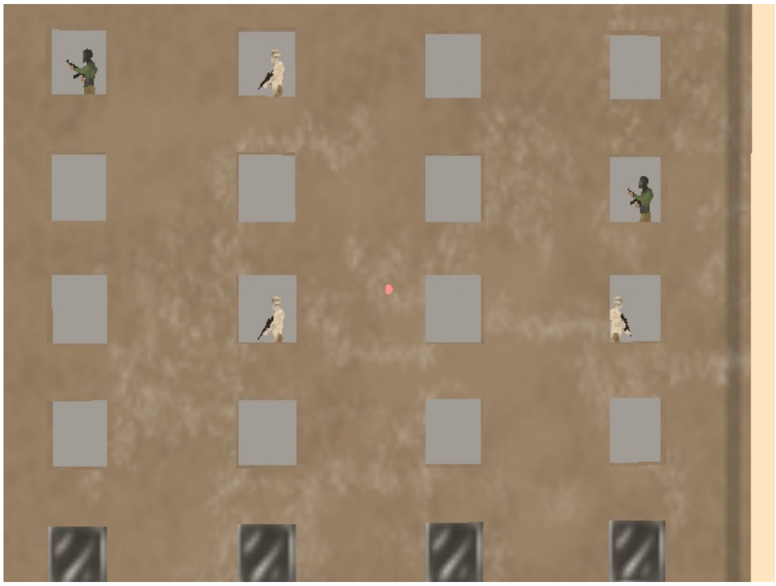

Figure 3. Face of building with 2 enemy targets (green) and 3 friendlies (tan)

In the fourth evaluation, we employed a neural network classification approach to cognitive state assessment and tested the system in a combination desktop/field study. This evaluation tested the communications scheduler and the mixed initiative target identification as mitigation strategies. For the first half of the evaluation, the six participants were seated conducting three tasks on a TabletPC. 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. There were two task load conditions, a low task load in which the participant received a limited number of communications and visual images, and a high task load condition where the rate of communications increased. The actual rate was predetermined during the training session by taxing the participants to the point where performance began to drop off. Actual rates varied from 6/min to 34/min. In the second half of the evaluation, the participants performed the first two communication tasks while in an upright and mobile position while scanning the woods in the outdoors for targets, that is, concealed "snipers."

Figure 4. Visual search task of MOUT on TabletPC. Highlighted box shows target id agent

The evaluation consisted of two training sessions. The first allowed the participants to gain familiarity with the tasks and system and the second determined the highest event rate to present the verbal messages before performance began to breakdown. This high event rate was developed to cognitively challenge the participant and was later used as the event rate for the high task load condition. Five training blocks were used to calibrate when the high workload conditions were met and to train the neural net classifier. The subsequent experimental and mobility trial scripts were generated using the lowest event rate and the highest event rate just before performance breakdown to generate the low and high task load scenarios. The participants completed five experimental sessions: three seated (low task load/unmitigated, high task load/unmitigated, and high task load/mitigated) and two mobile outdoors (low task load/unmitigated and high task load/unmitigated).

Findings from this 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. Overall there was a 94% improvement in the mission monitoring task and 36% improvement in the radio count recall task when the communications scheduler was available in the high workload trial as compared to the no mitigation high task load condition. Subjective workload levels in the mitigated conditions mirrored those of the low task load conditions and were statistically lower than in the unmitigated condition.

For the visual search task on the TabletPC (see Figure 4), 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 on the visual search task.

The mobility trials were conducted as a means to assess the feasibility to collect and clean the EEG data in real-time in the field. Hence, not all experimental manipulations needed to be evaluated. Findings indicated that performance in both the low and high task load conditions were lower in the mobile trials indicating that the addition of mobility in an actual environment made performing the mission monitoring and radio count recall tasks more cognitively challenging. Performance declined more in the high task load condition (composite metric: 58% accuracy than the low task load condition (composite metric: 66% accuracy) indicating an effective task load manipulation. An additional test of the classification method was also conducted in the field immediately prior to this evaluation. Findings indicate the ability to correctly classify cognitive state data in a mobile environment. Participants' cognitive states were classified during the execution of three tasks: navigation, visual search, and a combined navigation and communication task. By using the first third of the data to train the classifier and the second third for testing, the neural net classification accuracy was consistently above 90%. See Table 1. For more information on the real-time signal processing and classification methods see Mathan, et al. (2005).

Table 1. Probability of correctly classifying tasks. The diagonal represents the correct classification accuracy.

		Classified as…		
		Navigate	Search	Nav & Comm
True class	Navigate	0.959	0.019	0.000
	Search	0.003	0.981	0.047
	Nav & Comm	0.038	0.000	0.953

The Honeywell AugCog team also demonstrated the system feasibility in a full mission scenario lead by an "augcogified" platoon leader. This mission included a 2-person team traversing a field undercover while communicating via radios with squad leaders during a movement to objective exercise. Cognitive state was assessed during the mission, which included ambushes from opposing forces armed with laser guns. The team also had laser guns to engage the enemy. The simulation verified the feasibility of collecting EEG and IBI data in the field as well as the AugCog system being integrated with a battle dress uniform. The fully mobile system was contained in a backpack worn by the platoon leader and wirelessly transmitted data about his real-time cognitive state to the base station.

7 Conclusions and Summary

The initial findings from the evaluations described above indicate that neurophysiological and physiological data can used be to assess the cognitive state of an individual and drive mitigations to enhance performance. Studies conducted in desktop environments as well as in the field enabled the sensor suite to determine when workload was high and the performance could be enhanced by scheduling the presentation of communication messages, redirecting information to the underutilized tactile modality, offloading procedures and tasks to the automation, and utilize mixed-initiative automation assistance.

The first evaluation validated the ability to create a closed loop system and trigger a mitigation strategy solely based on the input from physiological and neurophysiological sensors. The system assessed the participant's level of comprehension of verbal messages. Findings indicated a 72% effectiveness in detecting when the participant failed to process the incoming message, resulting in a repeat of the unattended messagse. This was found even in the cases where the participant's provided an automatic acknowledgement of the truly unattended message.

In the second evaluation, during high workload periods the Communications Scheduler changed how messages were delivered, by escalating and highlighting high priority messages, and deferring low priority messages to the tablet PC. By doing this at the appropriate time as determined by gauge outputs, this resulted in 100% improvement in message comprehension and 125% improvement in overall situation awareness. These findings were enabled without a subsequent increase in subjective workload to process the communications in multiple modalities as delivered from different sources.

The third evaluation further investigated potential performance improvements of a communications scheduler due to the U.S. Army's concern over the increased information processing requirements resulting from netted communications delivered to dismounted soldiers. In a motion capture laboratory environment, performance on a working memory task was enhanced by over 150% when the scheduler was used to moderate the peaks in workload periods by delaying the arrival of new tasks if a high workload state was detected.

The fourth evaluation set out to evaluate a different approach to cognitive state classification as well as expanding the types of mitigation strategies employed. The primary goal was to move to investigate the system effectiveness with more realistic Army tasks completed in both a seated command center environment as well as a mobile environment outside the laboratory. The evaluation determined that the system could detect a high workload state and trigger the mitigation strategies, including the communication scheduler and target identification agent. These mitigations were able to offload the cognitive demands of the task to show an improvement in performance. The findings indicated a 94% improvement on the mission monitoring task, 36% improvement on the radio count recall task, and 40% improvement on the visual search task. The findings from this evaluation and demonstration determined that the cognitive state assessment from EEG could be collected in real-time on a truly mobile individual outside of the laboratory.

The mitigation strategies and tools developed and implemented for these evaluations are promising solutions to known problems as identified by the Future Force Warrior program and we will continue to explore their utility as we move to the field. However, challenges remain to fully implement a real-time neurophysiologically driven cognitive state assessment system. Real-time cognitive state classifications are difficult due to the complex nature of human physiology. Using the gauge approach limited the ruleset to a static threshold based solution that was applied universally to all participants. Using the neural net approach required specific training for the detection of the desired state. Both approaches are limited by the nonstationary nature of the EEG signals, thereby requiring the retraining of the neural net or recalibration of the ruleset thresholds. We continue to explore developments in sensor technology, such as functional imaging that uses near infrared light to detect changes in blood oxygenation during brain activity, as well as other approaches. Further challenges include the real-time detection of artifacts in data collected on a mobile dismounted soldier. Once reliable, artifact-free data can be collected and characterized, additional work will need to be done to improve the processing efficiency of the algorithms and to miniaturize the solutions in order to deploy these systems with the Future Force Warrior.

Acknowledgments

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

References

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

Dorneich, M. C., Mathan, S., Creaser, J., Whitlow, S., & Ververs, P. M. (2005). Enabling Improved Performance through a Closed-Loop Adaptive System Driven by Real-Time Assessment of Cognitive State. To appear in the *Proceedings of the 1st International Conference on Augmented Cognition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Dorneich, M., Whitlow, S., Ververs, P. M., Mathan, S., Raj, A., Muth, E., Hoover, A. DuRousseau, D., Parra, L., & Sajda, P. (2004). *DARPA improving warfighter information intake under stress – Augmented Cognition: Concept validation experiment analysis report for the Honeywell Team*. DARPA technical report (contract no. DAAD16-03-C-0054).

DuRousseau, D.R., (2004). Spatial-Frequency Patterns of Cognition, *The AUGCOG Quarterly*, 1(3), p. 10.

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

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: IEEE.

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.

Hoover, A. & Muth, E. (2004). A Real-Time Index of Vagal Activity. *International Journal of Human-Computer Interaction*, *17*(2), 197-210.

Mathan, S., Mazaeva, N., Whitlow, S., Adami, A., Erdogmus, Lan, T., & Pavel, M. (2005). Sensor-based cognitive state assessment in a mobile environment. To appear in the *Proceedings of the 1st International Conference on Augmented Cognition.* Mahwah, NJ: Lawrence Erlbaum Associates.

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).

Raj A. K., Bradshaw, J., Carff, R. W., Johnson, M., and Kulkarni, S. (2004, January). 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.

Sajda, P. Gerson, A. & Parra, L. (2003). Spatial Signatures of Visual Object Recognition Events Learned from Single-trial Analysis of EEG, Engineering in Medicine and Biology Society, 2003. *Proceedings of the 25th Annual International Conference of the IEEE*, 3, pp. 2087 – 2090.

Ververs, P. M., Whitlow, S. D., Dorneich, M. C., Mathan, S., & Sampson, J. B. (2005). Augcogifying the Army's Future Warfighter. To appear in the *Proceedings of the 1st International Conference on Augmented Cognition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Whitlow, S., Ververs, P. M., Buchwald, B., & Pratt, J. (2005). Scheduling communications with an adaptive system driven by real-time assessment of cognitive state. To appear in the *Proceedings of the 1st International Conference on Augmented Cognition*. Mahwah, NJ: Lawrence Erlbaum Associates.