EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning and Memory Tasks

Chris Berka, Daniel J. Levendowski, Michelle N. Lumicao, Alan Yau, Gene Davis, Vladimir T. Zivkovic, Richard E. Olmstead, Patrice D. Tremoulet, and Patrick L. Craven

Advanced Brain Monitoring Inc., University of California, Los Angeles, United States

Introduction: *The ability to continuously monitor levels of task engagement and mental workload in an operational environment could significantly enhance performance, productivity and safety in military and industrial settings. This study establishes feasibility of operational monitoring with electroencephalographic (EEG) indices of engagement and workload acquired unobtrusively and quantified during performance of cognitive tests.* ***Methods:*** *EEG was acquired from 80 healthy participants' subjects with a wireless sensor headset (F3-F4,C3-C4,Cz-POz,F3-Cz,Fz-C3,Fz-POz) during one or more tasks including: multi-level forward/backward-digit-span, grid-recall, trails, and mental-addition, 20-minute-vigilance and image/verbal learning and memory tests. EEG metrics for engagement and workload were calculated for each 1-second of EEG using quadratic and linear discriminant function analyses of model-selected variables derived from EEG power spectra (1-Hz bins from 1-40Hz).* ***Results:*** *Across subjects, engagement but not workload decreased over the 20-minute-vigilance test. Engagement and workload were significantly increased during the encoding period of verbal and image-learning & memory when compared to the recognition/recall period. Workload but not engagement increased linearly as level of difficulty increased in forward and backward-digit-span, grid-recall and mental-addition tests. EEG measures correlated with both subjective and objective performance metrics.* ***Discussion:*** *These data suggest that EEG-engagement reflects information-gathering, visual scanning and sustained attention. EEG-workload increases with increasing working memory load and during problem-solving, integration of information, analytical reasoning and may be more reflective of executive functions. Inspection of EEG on a second-by-second timescale revealed associations between workload and engagement levels when aligned with specific task events, providing preliminary evidence that second-by-second classifications reflect parameters of task performance.*

1. INTRODUCTION

Information overload is a fact of life in the contemporary global networked society. Potentially rich sources of data are underutilized because they cannot be sorted rapidly and organized efficiently enough to accommodate the capacity of the human information processing system. The human processor can also be seriously compromised by fatigue, stress, boredom, illness and other factors. One approach to expanding the capacity of human information processing is to radically rethink the design of human-machine system interfaces to optimize the flow and exchange of data between humans and machines. This approach has been termed "neuroergonomics," an interdisciplinary area of research and practice that integrates understanding of the neural bases of cognition and behavior with the design, development and implementation of technology (13, 21, 25, 26). The vision of neuroergonomics is to use knowledge of brain-behavior relationships to optimize the design of safer, more efficient work environments that increase motivation and productivity. A complementary result of this endeavor is to better inform neuroscience regarding real-world human performance (24).

One promising avenue of research in neuroergonomics involves developing the capability to continuously monitor an individual's level of fatigue, attention, task engagement and mental workload in operational environments using physiological parameters (1, 4, 10, 14, 23, 30-32, 37). These physio-cognitive monitoring systems have a wide range of potential applications that could significantly enhance performance, productivity and safety in military and industrial settings, including evaluation of alternative interface designs, enhancing skill acquisition and optimizing the ways humans interact with technology (26). Several pioneering investigations conducted as part of the Defense Advanced Research Projects Agency (DARPA)

Accepted for Publication, May 2007. Citation: Berka C, Levendowski DJ, Lumicao MN, Yau A, Davis G, Zivkovic VT, Olmstead RE, Tremoulet PD, Craven PL (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. Aviat. Space Environ. Med. 78, B231–B244.

Augmented Cognition program explored the feasibility of integrating real-time physiological indices of user workload into the Human Computer Interface (HCI) loop to assist in managing information flow in complex task environments. Physiological indicators identified when a user was overloaded or underloaded and triggered greater information dissemination or task re-allocation. Preliminary results suggested that performance could be enhanced in these closed-loop model systems (9, 22, 35).

Heart rate variability, oculomotor activity, pupilometry, functional near infrared imaging (fNIR) and galvanic skin response have been employed to detect cognitive state changes; however, the electroencephalogram (EEG) is the only physiological signal that reliably and accurately reflects subtle shifts in alertness, attention and workload that can be identified and quantified on a millisecond time-frame. Significant correlations between EEG indices of cognitive state changes and performance have been reported based on studies conducted in laboratory, simulation and operational environments (4, 7, 8, 11, 12, 18-20, 27, 32, 33, 38). The conventional methods employed to analyze the EEG generally involve computation of the power spectral densities within the classically defined frequency bands including alpha, beta, theta, delta and gamma or ratios between these frequency bands (10, 31, 32, 37). Alternatively, the amplitudes of the N100 and P300 components of the event-related potential have been employed in some cognitive assessment models (34). The use of the event-related potential for operational applications has several limitations including the requirement for introducing "probe" stimuli into real-world tasks to elicit the potentials and the need for averaging of single trials across scalp sites or over time.

Advanced Brain Monitoring (ABM) implemented an integrated hardware and software solution for acquisition and real-time analysis of the EEG and demonstrated feasibility of operational monitoring of EEG indices of alertness, engagement and mental workload. The system includes an easily-applied wireless EEG system designed with the goal of future operational deployment. A novel analytical approach was developed that employs linear and quadratic discriminant function analyses (DFA) to identify and quantify cognitive state changes using model-selected variables that may include combinations of the power in each of the 1-hz bins from 1-40hz, ratios of power bins, event-related power and/or wavelet transform calculations. This unique modeling technique allows simultaneous selection of multiple EEG characteristics across brain regions and spectral frequencies of the EEG, providing a highly sensitive and specific method for monitoring neural signatures of cognition in both real-time and off-line analysis.

ABM has successfully applied this method to classify 1-sec or 0.5-sec segments of EEG to identify drowsiness-alertness (15-17) mental workload (4, 5), spatial and verbal processing in simple and complex tasks, (1) to characterize alertness and memory in patients with sleep apnea (6, 36) and to identify individual differences in susceptibility to the effects of sleep deprivation (2). The ABM model system has also been successfully integrated into real-time, closed-loop automated computing systems to implement dynamic regulation and optimization of performance during a driving simulation task and in the Aegis C2 and Tactical Tomahawk Weapons simulation environments (1, 3, 5).

There are several challenges that must be overcome by developers of cognitive state monitors. First, it is necessary to define a set of relatively pure tasks that consistently elicit the targeted cognitive states to provide calibration of the physiological measures and to validate the methods for cognitive monitoring. Validation of cognitive state measures generally involves experimental manipulation of task demands to induce cognitive state changes, objective measurement of performance metrics (e.g. accuracy, reaction time) and subjective measures that allow participants to describe their perceived level of difficulty as well as the amount of effort exerted in a given task. The cognitive state measures must also be validated across subjects and adjusted to account for individual differences when required.

This paper presents evidence for the utility of two EEG-based measures of cognitive states, task engagement and mental workload. Both measures increase as a function of increasing task demands but the engagement measure tracks demands for sensory processing and attention resources while the mental workload index was developed as a measure of the level of cognitive processes generally considered more the domain of executive function.

2. METHODS

EEG was acquired from 80 participants. The data from 13 participants used for model development were collected at the Lockheed-Martin Advanced Technology Lab. The study protocol was approved in advance by the Chesapeake Research Review, Inc. The data from 67 subjects obtained by Advanced Brain Monitoring and used for cross validation were acquired using protocols for 3 studies approved in advance by the Biomed IRB. Each subject provided written informed consent before participating.

All participants wore the wireless sensor headset including the following bi-polar sensor sites: F3-F4, C3-C4, Cz-PO, F3-Cz, Fz-C3, Fz-PO. Participants completed one or more of the following tests: a 20-minute 3-choice-vigilance test, an image learning and recognition memory test followed by an interference session where previously learned items appear but are no longer in the learning set, a verbal paired associate learning and memory test and multi-level forward and backward digit-span tests, grid-recall (spatial memory) and mental-addition tests. EEG metrics (values ranging from 0.1-1.0) for "engagement" and "mental workload" were calculated for each 1-second epoch of EEG using quadratic and linear discriminant function analyses of model-selected EEG variables derived from power spectral analysis of the 1-Hz bins from 1-40Hz.

Bi-polar recordings were selected in order to reduce the potential for movement artifacts that can be problematic for applications that require ambulatory conditions in operational environments. Limiting the sensors (seven) and channels (six) ensured the sensor headset could be applied within 10 minutes. The sensor montage was selected after conducting experiments using monopolar recordings and selecting the montage and channels that provided the best mental workload discrimination across subjects, tasks and conditions.

Identification and decontamination of spikes, amplifier saturation and environmental artifacts and computation of the power spectral density were applied using procedures previously described (4). Two new wavelets procedures were also applied to these data to detect excessive muscle activity (EMG) and to identify and decontaminate eye blinks. Once the artifacts are identified in the time-domain data, the EEG signal is decomposed using a wavelets transformation. Thresholds are developed for application to the wavelet power in the 64 – 128 Hz bin to identify epochs that should be rejected for EMG. The wavelets eye blink identification routine uses a two-step discriminant function analysis applied to the absolute value of the 0-2, 2-4, 4-8, 8-16, and 16-32 Hz wavelet coefficients from the 50th, 40th, 30th, 20th and 10th data points before and after the target data point in FzPOz and CzPOz. The DFA classifies each data point as a control, eye blink or theta activity. Multiple data points that are classified as eye blinks are then linked and the eye blink detection region is established based on a fixed distance before the start (e.g., 50 data points) and after the end (e.g., 50 data points) of the blink.

Decontamination of eye blinks is accomplished by computing mean wavelet coefficients for the 0-2, 2-4 and 4-8 Hz bins from nearby non-contaminated regions and replacing the contaminated data points. The EEG signal is then reconstructed from the wavelets bins ranging from 0.5 to 64 Hz. Zero values are inserted into the reconstructed EEG signal at zero crossing before and after spikes, excursions and saturations. EEG absolute and relative power spectral density (PSD) variables for each 1-second epoch using a 50% overlapping window are then computed. The PSD values are scaled to accommodate the insertion of zero values as replacements for the artifact.

Description of the Tasks: Five tasks were used by Lockheed-Martin to acquire the model development data set.

For the **Addition task**, participants were asked to add two numbers with varying numbers of digits in addends, and report the sum by typing digits through the keyboard. This task requires subjects to employ working memory and executive function resources.

During **Grid location**, an NxN grid of squares where N ranged from 3 – 6, is presented with some grids containing missiles, and participants are prompted to identify which squares contained missiles by clicking in the squares in an empty grid. This task requires spatial working memory resources.

In the **Trail-making task**, subjects are presented with a series of labeled dots on a computer screen, and asked to 'draw a trail' by clicking on the dots in series. The labels may be numbers or letters or some other code, and subjects are required to click on them in order. The first dot in the series is always pre-selected or otherwise highlighted for the subject. This task requires subjects to employ spatial memory and executive function resources.

In **Forward digit span (FDS)**, series of single digits of increasing lengths are presented and the participant responds by entering the digits in the order presented. This task requires verbal working memory resources. Similarly, the **Backwards digit span (BDS)** presents a series of single digits of increasing lengths and requires entering digits in the reverse order from the one presented.

Subjective estimates were acquired from the 13 participants evaluated in the Lockheed-Martin study using a survey administered to the participant following each difficulty level of each task. Responses to the following questions were rated on a 100-point scale: How much mental energy did you exert on this task level? (almost none… a whole lot) Objectively, how difficult was this task level? (quite easy… extremely difficult) and How much attention did you focus on this task level? (very little… I was extremely focused).

The cross validation protocols included the FDS and BDS described above, and four additional tasks.

The **3-Choice Vigilance Task (3C-VT)** incorporates features of common neuropsychological tests of vigilance, including simple or choice reaction time tests and continuous performance tests (29)**.** The 3C-VT requires subjects to discriminate primary (70%) from two secondary (30%) geometric shapes presented for 0.2 seconds and respond over a 20-minute test period. A training period is provided prior to testing to minimize practice effects.

The **Image, Verbal and Interference Learning and Memory Tests** evaluate attention, distractibility and recognition memory for images, image-number pairs or word pairs. During the training session, a group of 20 images are presented twice. The testing session presents the 20 training images randomly interspersed with 80 additional images. Subjects must indicate whether or not the image was in the training set. Five equivalent image categories were developed including animals, food, household goods, sports and travel. In the **Standard Image Recognition**, the subject must memorize 20 images and identify the 20 training images amidst 80 previously unseen testing images. For the **Interference Image Recognition**, a set of 20 new images must be memorized and distinguished from the first set of training images and 60 images previously displayed in the Standard. The **Verbal Paired Associate test (VPA)** is identical to the Standard, substituting word pairs for images. Easy (e.g. black-white, dog-cat) and difficult (e.g. table-horse, fence-towel) word pairs are included in each test.

The 67 participants in the studies conducted at Advanced Brain Monitoring completed a subjective ratings questionnaire after each level of the Forward digit span and Backward digit span. Each participant was asked to rate the difficulty of the level they just completed as either very difficult, difficult, neither easy nor difficult, easy, or very easy.

Development and Validation of the EEG-Mental Workload Index: The investigators had previously developed and validated an EEG measure that was highly correlated with task demands including the level and complexity of stimulus processing and the requirement for allocation of attentional resources. This EEG metric termed "task engagement" was found to be directly correlated with task load in simple vigilance and memory tasks and in more complex simulation tasks including the Warship Commander, a simulated naval command and control task and in an Aegis radar operations simulation environment (4, 5). Subsequent applications of the EEG metric revealed that it did not increase as a function of increasingly difficult mental arithmetic, during increasingly complex analytical reasoning, during multiple levels of difficulty in a Sternberg memory task, during the five levels of a forward or backward digit span test or in a multi-level grid location spatial memory task.

This paper presents the development and validation of a new EEG index of mental workload, designed to provide a measure of the level of cognitive processing associated with tasks involving executive function, working memory and analytical reasoning. The engagement index and the mental workload index were evaluated across multiple tasks, subjects and conditions. The final set of EEG variables selected to provide the optimal classification of EEG-engagement and mental workload are listed by channel and frequency bin in Table 1.

Table 1: EEG Variables used for Computation of Engagement and Workload

	1 – 4 Hz	5 – 7 Hz	8 – 13 Hz	14 – 24 Hz	25 – 40 Hz
Engagement and Distraction - Absolute and Relative PSD Variables Selected (1 Hz bins)					
FzPOz	0	0	0	1	6
CzPOz	1	2	5	2	6
Workload Gauge - Absolute and Relative PSD Variables Selected					
C3C4	1	0	2	2	2
CzPOz	0	0	0	2	5
F3Cz	0	3	1	0	1
F3C4	1	0	1	1	0
FzC3	1	0	0	1	1
FzPOz	1	0	1	1	2

The Lockheed-Martin data set including EEG and performance measures from 13 subjects during five tasks were used to develop and validate the mental workload gauge. The tasks were performed in the following order: grid, forward digit span, mental arithmetic, backward digit span and trails. Each task had between three and six levels of difficulty. For example, during the backward digit span, level one required the subject to memorize 2 digits and a total of 20 digit sets were presented. Level two was 4 digits and 12 digit sets, level 3 was 6 digits and 8 digit sets, and level 4 was 8 digits and 5 digit sets. For all of the Lockheed-Martin tasks, participants were allowed to self-pace with respect to the time needed to complete each problem. This differed from the testing done at ABM where all tests were performed under consistent time constraints.

Two objective performance measures were derived, the percentage of complete correct answers for each task and level, and the percentage of partially correct answers. For example, seven of nine correct numbers in a 9-digit forward digit span resulted in a 0.78 partial correct score for that problem.

The workload gauge was developed using a linear discriminant function with two classes, low and high mental workload. Different combinations of tasks and levels were evaluated as training data to derive a two-class workload gauge (i.e., low and high) that generalized across individuals. Two models were developed and employed to accommodate individual differences in the EEG across subjects. If necessary (when the EEG data are to be used in real-time as inputs to a closed loop system), the determination of which workload model best fits an individual can be made a-priori by evaluating the distribution of the probability of high workload and performance across a baseline four-level backward digit span. In order to compute workload when only the F3F4 channel was contaminated with EMG, alternative workload models were developed using inputs from five of the six channels (i.e., excluding F3F4).

There are three primary methods used to validate physiological measures of mental workload. The first is inherent in the task design; the tasks used to validate must include incrementally increasing levels of difficulty to elicit increasing levels of mental workload required by the participant. The second method of validation is to correlate objective measures of task performance with the EEG Index and the final method is to compute the correlations with the subjective reports. All three methods have strengths and weaknesses, but arguably each can contribute to the overall validation of the EEG metrics.

3. RESULTS

Figure 1 presents the mean workload levels as classified by the EEG model for each level of difficulty for the grid, addition, forward and backward digit span and trails tasks with the associated mean subjective ratings for each level of each task. EEG workload levels accurately tracked the intended pattern of the task design and the subjective ratings for all tasks except the Trails task. Figure 2 presents the same EEG workload data with the associated mean objective performance scores. As expected, performance decreases as a function of task difficulty with the exception of the Trails task where participants achieved perfect performance on all three levels of the Trails. In the Trails task, the EEG workload was more closely aligned with the objective performance measure than the task design levels or subjective rating levels. All participants achieved a perfect score for the Trails task because it was impossible to complete Trails task until the participant responded correctly.

The mean EEG-engagement values for each level of difficulty for the grid, addition, forward and backward digit span and trails tasks are presented in Figure 3. Although EEG-engagement levels changed across the levels of the tasks, the relationships were not linearly related to the task difficulty levels as demonstrated for the EEG-workload measures (based on repeated measures Analysis of Variance (ANOVA).

Canonical correlations were employed (CANCORR macro in SPSS, Release 8.0) to calculate a single aggregate measure of association in the model development data set between the EEG-workload and subjective and objective measures. As opposed to multiple bi-variate correlations, canonical correlations are linear functions that maximize the relationship between the two sets of variables. In this case, canonical correlations were calculated for each individual across the 23 tasks/levels (Figure 1) comparing the EEG-workload variable sets (%classified high workload and probability of high workload) to the objective and subjective variable sets (%complete correct, %partial correct, assigned level of difficulty, and subjective perception of difficulty). Although as a multivariate test, sample size recommendations generally specify a larger number of observations than tested here, the canonical correlation was utilized in this circumstance less as an inferential test and more as a metric of association. As outliers can have a large

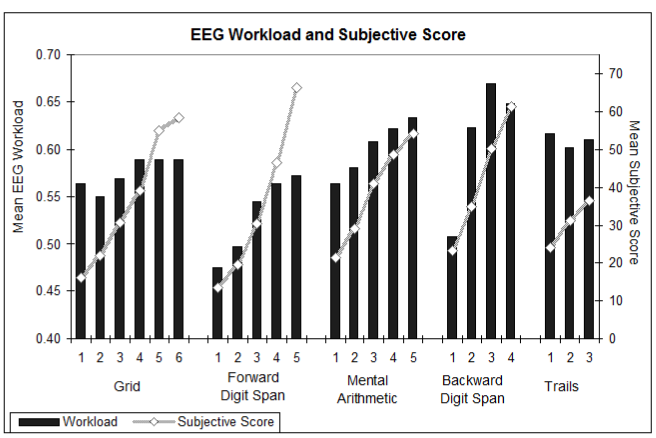

Figure 1. Mean EEG-workload and mean subjective rating scores for each difficulty level of the grid, FDS, mental arithmetic, BDS, and trails tests for the model development group (n=13).

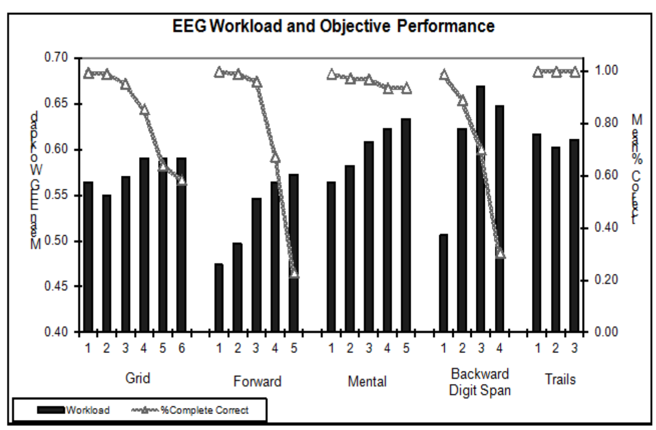

Figure 2. Mean EEG-workload and mean objective performance scores for each difficulty level of the grid, FDS, mental arithmetic, BDS, and trails tests for the model development group (n=13).

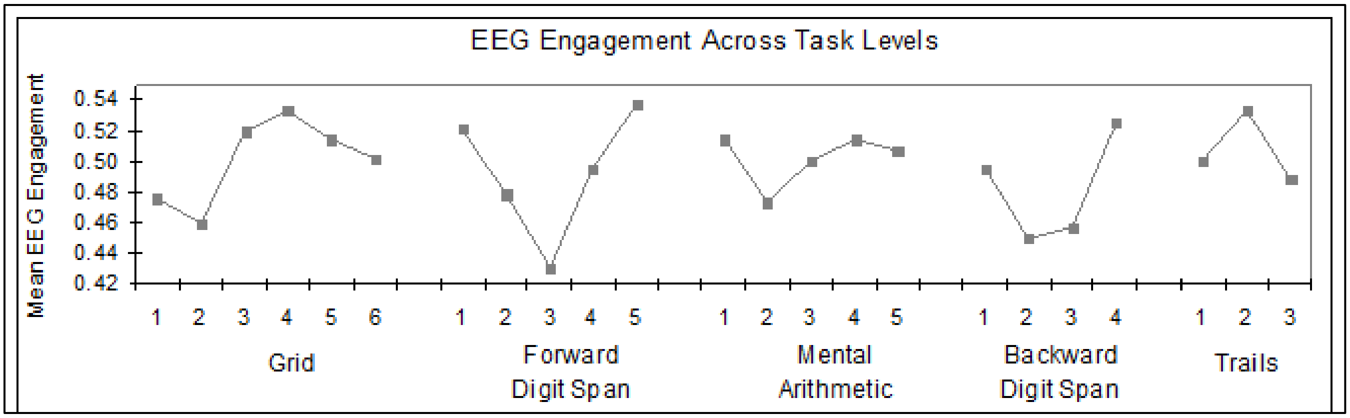

Figure 3. Mean EEG-engagement for each difficulty level of the grid, forward digit span, mental arithmetic, backward digit span and trails tests for the model development group (n=13).

impact on the calculations, all cases were examined for extreme values using Mahalanobis distance. There were no outliers noted. The box plot in Figure 4 illustrates the results of the Canonical correlations for this initial data set. Chi-squared tests were significant for the canonical correlations for 9 of the 13 subjects.

Cross-validation of the EEG-mental workload measure was completed with a new group of 17 participants evaluated at Advanced Brain Monitoring using the forward and backward digit span tests. The ABM versions of the digit span tasks were modified to limit the amount of time to respond. Two difficulty levels were added to the BDS because prior work suggested that some participants became frustrated and "gave up" at some point in the BDS and there was significant variability in when this occurred. The additional levels provided a full spectrum of EEG for all participants prior to and after giving up on the task.

Figures 5 and 6 present the mean workload levels as classified by the EEG model for each level of difficulty for the forward and backward digit span tasks with the associated mean subjective ratings and objective performance for each level of each task.

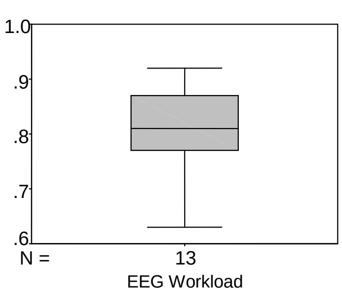

Figure 4. Results for individual subjects (n=13) from the Canonical correlations comparing EEG-workload with objective and subjective measures. The box plots display the median, with ~50% of the data represented by the shaded area, and ~99% of the data within horizontal bars (outlier cutoff points).

These data provide confirmation of the validity of the EEG measure across subjects for the forward and backward digit span tests. The mean EEG-engagement values for each level of the digit span for the cross-validation group are presented in Figure 7. These data suggest changes in engagement that are not linearly related to task difficulty and that a more complex relationship may exist between the EEG-engagement index and the task levels.

Evaluation of the EEG engagement and workload measures was also conducted during the 3C-VT (see Figure 8). Mean reaction times, EEG-engagement and EEG-workload levels were calculated for each of the 5-minute quarters of the 20-minute test. Repeated measures ANOVA was performed across the 4 quarters. Probability values reported are based on the Greenhouse-Geisser corrected degrees of freedom. As expected, repeated measures ANOVA revealed significant effects over time with reaction time increasing ($F=104.92$, $p=0.001$) and engagement decreasing ($F=216.08$, $p=0.001$). However, the workload levels did not show a significant linear increase over the 20-minute test. Effect sizes were large for reaction time (Cohen's $d=1.11$) and EEG-engagement (Cohen's $d=2.83$). These data confirm previous reports by the investigators that during the 20-minute test, subjects evidence increasing reaction time and a corresponding decrease in the EEG-engagement level. This effect becomes increasingly evident as a function of sleep deprivation or fatigue (6, 16, 36). The fact that the mental workload index did not change over time was expected in the 3C-VT, a task with minimal demands on working memory or complex cognitive processing.

For the learning and memory tests, mean EEG-engagement and workload were computed for each of the encoding and recognition periods of three types of learning and memory tests (standard, interference and verbal). For EEG-engagement, repeated

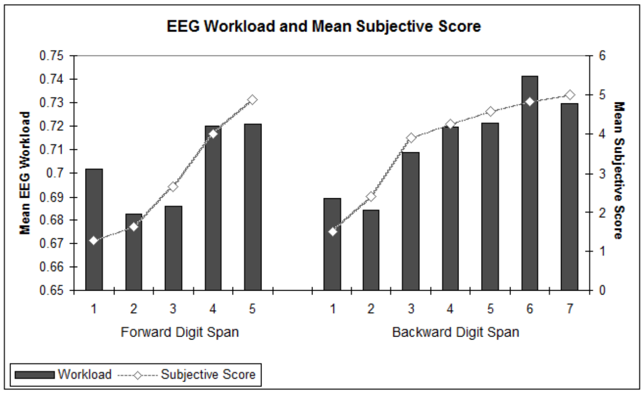

Figure 5. Mean EEG-workload and mean subjective rating scores for each difficulty level of the forward digit span and backward digit span for the cross validation group (n=17).

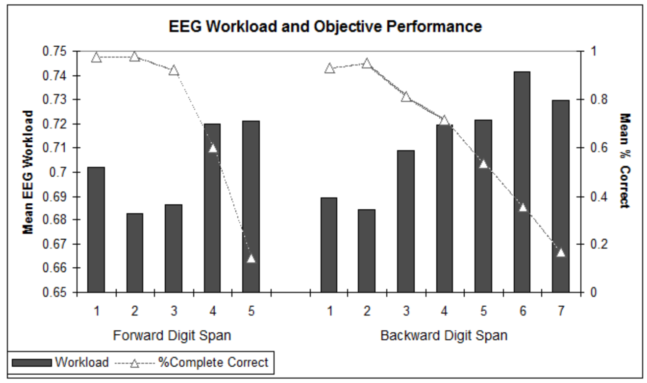

Figure 6. Mean EEG-workload and mean objective performance scores for each difficulty level of the forward digit span and backward digit span for the cross validation group (n=17).

measures ANOVA indicated a significant effect for task type (F=13.65, p=0.001) and encoding/recognition (F=15.176, p=0.001) (Figure 9). For EEG-engagement, the interference was significantly different from both standard (F=22.39, p=0.001) and the verbal (F=11.05, p=0.002), whileverbal was also significantly different from standard (F=4.73, p=0.034).

The task type effect relates to the level of task difficulty. Confirmation of the task difficulty differences was observed in the significant main effect for task type (F=44.56, p=0.001) for percentage correct performance (Figure11). This increase in task difficulty is also reflected in the increase in EEG-engagement (Figure 9).

The EEG-workload measures were also significantly increased during the encoding period of all memory tests when compared to the recognition period (see Figure 10). However, only the encoding/recognition effect for EEG-mental workload was significant (F=34.79, p=0.001). These data suggest that the EEG reflects an increased allocation of attentional resources and mental workload during the encoding period.

To summarize, EEG-workload but not engagement increased linearly across the multiple levels of FDS/BDS, grid-recall, and mental-addition tests. EEG-engagement but not EEG-workload increased as a function of time-on-task during the 20-minute 3C-VT and for the learning and memory tests, EEG-engagement and mental workload were higher during the encoding period than the recognition period and increased as a function of task difficulty. EEG measures were significantly correlated with both subjective and objective performance metrics.

4. DISCUSSION

These data confirm that the EEG can provide an unobtrusive method for monitoring dynamic fluctuations in cognitive states including task engagement and mental workload. The temporal resolution of the EEG allows for precision calculations for each one-second or half-second of data; however, the detectable states are global in nature and the validation and interpretation of changes in cognitive state on a second-by-second basis requires further investigation.

The results of the studies suggest that the EEG engagement index is related to processes involving information-gathering, visual scanning and sustained attention. The EEG-workload index increases with

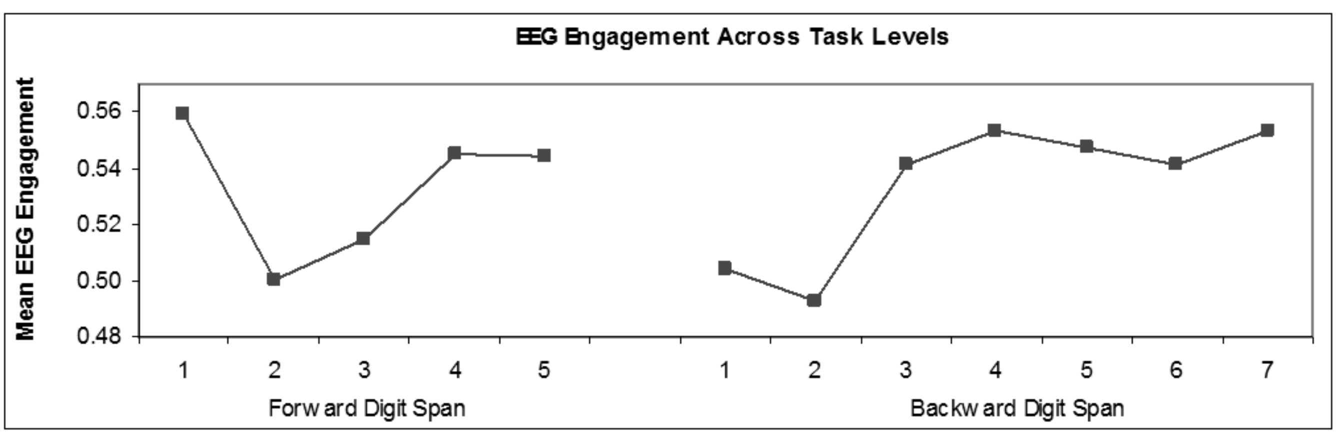

Figure 7. Mean EEG-engagement for each difficulty level of the FDS and BDS for the cross validation group (n=17).

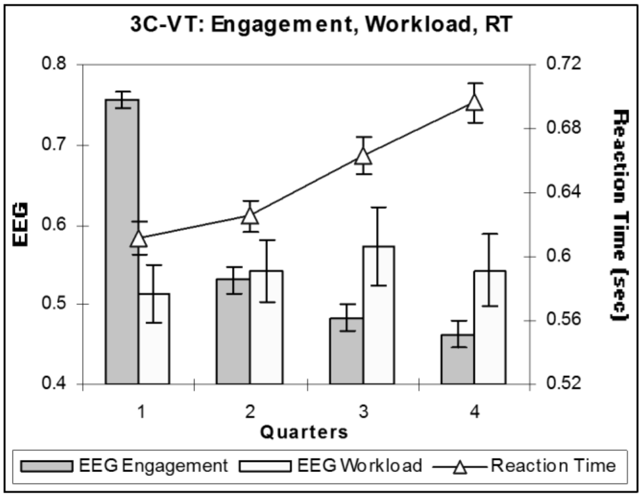

Figure 8. Mean EEG-engagement, EEG-workload and reaction time for each 5-minute quarter of the 3-CVT (n= 65 for EEG-engagement and reaction time, n=27 for EEG-workload).

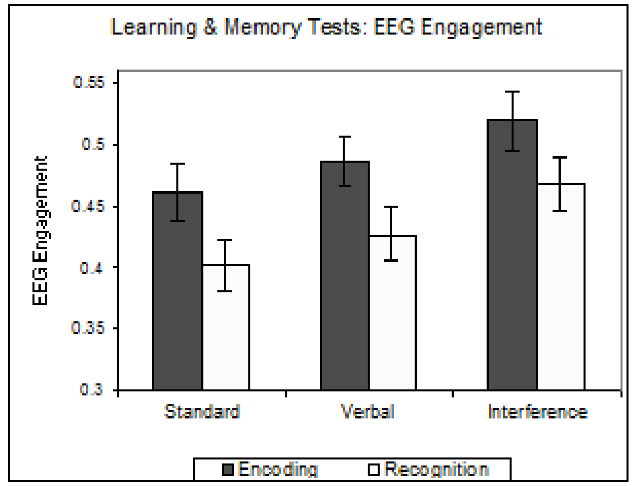

Figure 9. Mean EEG-engagement during the encoding and recognition/recall periods for the standard, verbal and interference learning and memory tests (n= 50).

working memory load and during problem-solving, integration of information, analytical reasoning and may be more reflective of executive functions. The two metrics have been shown to operate concordantly or independently, depending on the task environment, the level of task demands and the amount of effort required by the individual to complete the task. In the present study, a combination of objective performance metrics such as reaction time and percentage of correct responses and subjective ratings to assess the perceived level of effort were used to validate the EEG metrics.

Specifically, the EEG measures dissociated during a sustained vigilance task with a minimal load on working memory. Reaction time increased and EEG-engagement decreased over the 20-minute vigilance session while workload remained constant.

During multi-level mental addition, grid location and forward and backward digit span tests, the EEG-workload increased linearly as a function of increasing task difficulty. During these same multi-level tasks, EEG-engagement showed a pattern of change that was variable across tasks, levels and subjects. The pattern included a relatively high engagement during the first level of each task and a decrease in engagement for the second level of task difficulty, suggesting an initial task adaptation or novelty response. As task difficulty increased the engagement level trended upwards. In this study the first level of each task was always the easiest and the difficulty level was incrementally increased with each level. In future studies, however, a random mix of difficulty levels could be used to eliminate the novelty effects on the EEG. Because subjects were provided as much time as needed to complete the trails task, EEG engagement and workload and performance

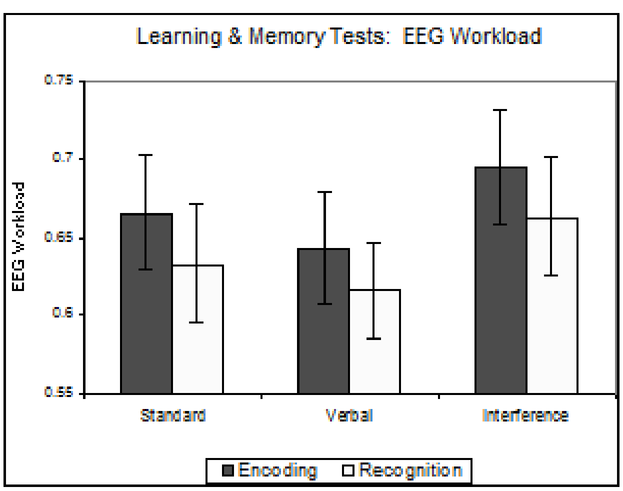

Figure 10. Mean EEG-workload during the encoding and recognition/recall periods for the standard, verbal and interference learning and memory tests (n=15).

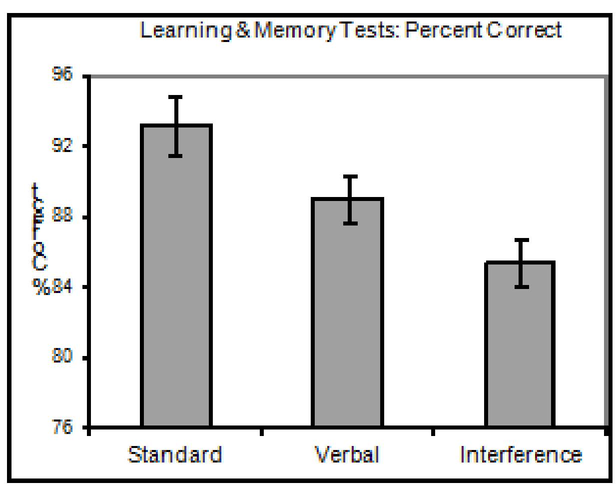

Figure 11. Mean percentage correct responses for the standard, verbal and interference learning and memory tests (n=50).

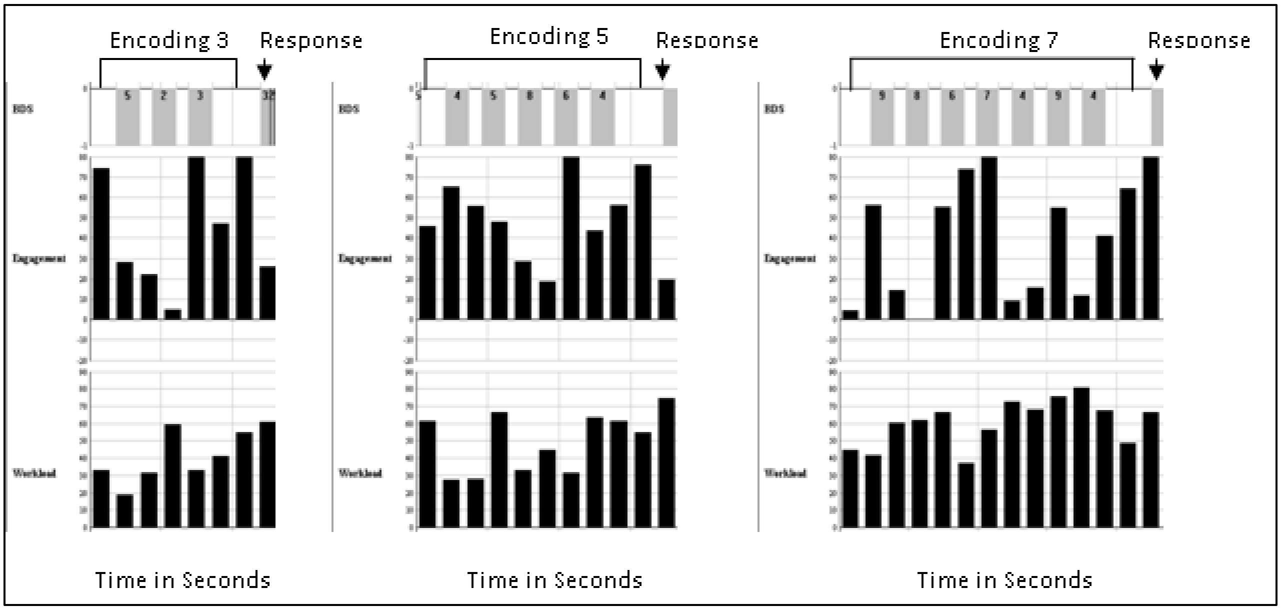

Figure 12. Second-by-second EEG-engagement and EEG-workload classifications for one subject performing the BDS. Three correctly solved problems are illustrated, representing three levels of difficulty: a) three-digit, b) five-digit and c) seven-digit.

were constant across the three difficulty levels.

In a series of image and verbal learning and memory tests designed to be increasingly difficult, both EEG-engagement and EEG-workload were higher during the encoding period than the recognition period and increased as a function of task difficulty. The level of EEG engagement and workload during encoding was positively related to the level of performance on each of the learning and memory tests. The conventional analysis of the EEG involves computation of the mean power spectral densities within the classically defined frequency bands including alpha, beta, theta, delta and gamma have been reported as the foundation for several EEG-based models of mental workload (10-12, 30-33). The modeling technique described in this paper incorporates multiple EEG variables across scalp sites and 1-Hz frequency bins (from 1 Hz – 40 Hz) to be used as inputs to quadratic and linear discriminant function analyses that provide classifications for each second of EEG. Models are constructed using stepwise multiple regression analyses to select those EEG variables that optimize the identification and classification of cognitive states within specified task environments. Simple baseline tasks are used to fit classification algorithms to the individual that can then be applied to increasingly complex task environments, providing a highly sensitive and specific technique for identifying neural signatures of cognition in both real-time and offline analysis. These methods have proven valid in EEG quantification of drowsiness-alertness during driving simulation, simple and complex cognitive tasks and in military, industrial and educational simulation environments (15-17) quantifying mental workload in military simulation environments (4, 5), distinguishing spatial and verbal processing in simple and complex tasks (1), characterizing alertness and memory deficits in patients with obstructive sleep apnea (6, 36) and identifying individual differences in susceptibility to the effects of sleep deprivation (2).

The approach described in this paper allows a complex mathematical model to be created using multiple EEG variables across frequencies and scalp locations. Table 1 lists the number, sensor locations and frequency bins of the model-selected variables used to classify workload and engagement. The engagement index relies heavily on EEG variables from the frequency bins in the beta and gamma bands, but also includes variables from delta, theta and alpha frequencies (Table 1). The workload measure also includes multiple variables from all of the frequency ranges.

In contrast, Prinzel and colleagues at NASA Langley (28) developed an EEG-engagement index based on beta activity (13 – 22 Hz) divided by alpha (8 – 12 Hz) plus theta (5 – 7 Hz) and applied it in a closed-loop system to modulate task allocation. Gevins (10-12) and Smith (30) reported frontal midline theta activity (5 – 7 Hz) increases during high task load conditions and attenuated alpha activity (8 – 12 Hz) proportional to increasing cognitive load. They suggested that these data could be combined in a multivariate approach individualized for each subject to indicate the extent to which a set of task demands activate the cortex during performance of the task. Smith and Gevins (10) recently reported a refinement of their EEG workload model that included quantification of alpha and theta band activity recorded from the frontal executive, central sensorimotor and posterior visual systems postulated to be linked to the regional cortical activation associated with decision-making, motor control and visuoperceptual demands respectively (10). All regional indices increased linearly during performance of low, medium and high load versions of a flight simulator task and correlated with subjective reports of perceived mental workload.

The value of using the established frequency band analyses is that they can be linked historically to methods applied and reported in EEG research and interpretation over the past 50 years. The risk in using simple metrics such as an increase in midline theta with a decrease in mean alpha or ratios of alpha, theta and beta is an oversimplification of cognitive state assessment. For example, Smith and Gevins repeated the flight simulator experiment after sleep depriving participants and reported that the subjective mental effort was negatively correlated with frontal activation after sleep deprivation in contrast to the positive correlation between frontal activation and subjective mental effort in the fully-rested condition (10). They interpret these data as "problematic" for the development of automated systems that use brain activation in a closed-loop system designed to identify when operators are overloaded or underloaded and trigger greater information dissemination or task re-allocation.

The model presented in this paper avoids the type of misclassifications reported in the Smith experiment (10) by combining a workload gauge, an engagement gauge and a drowsiness gauge (presented in previous work, see references 2 and 3) that are derived from independent discriminant function analyses that include variables that are sensitive to sleep deprivation. The use of three gauges derived from a complex combination of EEG variables facilitates highly sensitive and specific classifications of cognitive state changes.

Inspection of the data on a second-by-second timescale (see Figure 12) suggests that an even more significant and useful associations between EEG workload and engagement levels can be identified when aligned with specific time-locked task events. In the backward digit span example provided in Figure 12, the engagement levels are high during digit encoding and during the response period. The workload levels tended to increase during the memory rehearsal and recall period when participants were required to mentally reverse the digit sequences. These data provide preliminary evidence that the second-by-second classifications may offer valid reflections of cognitive state changes during task performance. More in-depth analyses using time series analysis are planned to further assess the relationships between EEG and stimulus characteristics (e.g. easy/difficult) and response parameters (e.g. correct/incorrect and fast/slow responses).

The workload and engagement gauges have been successfully integrated into real-time, closed-loop automated computing systems to implement dynamic regulation and optimization of performance during a driving simulation task and in the Aegis C2 and Tactical Tomahawk Weapons simulation environments (1, 3, 5). These initial findings support the possibility of integration of real-time physio-cognitive measures into the HCI loop to assist in managing information flow to increase the amount of information processed by human operators without increasing their level of stress. The envisioned outcome is a closed-loop system that utilizes physiological indices for dynamic regulation and optimization of HCI in real-time with a goal of maintaining information load within the limits of the human information processor.

Future applications of the EEG gauges presented in this paper include evaluating the effects of HCI, human factors, and ergonomic design on cognitive state to provide an objective method for guiding the development and testing of new interfaces. These EEG metrics may also be useful in assessing the effectiveness of training and simulation programs. The workload and engagement gauges are currently being used for industrial and military ergonomic assessments and as part of an interactive learning environment for high school and college science students. Additional applications include the assessment of synchronized engagement and workload acquired from multiple individuals to provide unique characterizations of group dynamics. Although the EEG provides a rich potential data source for cognitive state analysis, the addition of multiple physiological parameters such as heart rate variability, fNIR and eye-tracking may be required to extend the monitoring capabilities to include quantification of stress and emotional states.

5. ACKNOWLEDGEMENTS

Financial Support: This research was supported by the DARPA program "Improving Warfighter Information Intake Under Stress", in which Advanced Brain Monitoring is a sub-contractor to Lockheed-Martin Advanced Technology Labs and by grants NIH NHLBI HL070484 and NIH NIDA DA019357 awarded to Advanced Brain Monitoring.

Statements of financial or other relationships with potential for conflict of interest: Authors Berka, Levendowski, Lumicao, Yau, Davis, Zivkovic and Olmstead are paid salaries or consulting fees by Advanced Brain Monitoring and Berka, Levendowski, Lumicao, Davis and Olmstead are shareholders of Advanced Brain Monitoring, Inc.

Authors Tremoulet and Craven are salaried employees of Lockheed-Martin Advanced Technology Labs and served as the prime contractor for the DARPA grant that funded the project.

6. REFERENCES

1. Berka C, Levendowski D, Davis G, Lumicao M, Ramsey C, Stanney K, et al. EEG Indices Distinguish Spatial and Verbal Working Memory Processing: Implications for Real-Time Monitoring in a Closed-Loop Tactical Tomahawk Weapons Simulation. In: International Conference on Human Computer Interaction; 2005 July 2005; Las Vegas, NV; 2005.
2. Berka C, Levendowski D, P. W, Davis G, Lumicao M, Olmstead R, et al. EEG Quantification of Alertness: Methods for Early Identification of Individuals Most Susceptible to Sleep Deprivation. In: John A. Caldwell NJW, editor. SPIE Defense and Security Symposium, Biomonitoring for Physiological and Cognitive Performance during Military Operations; 2005 May 2005; Orlando, FL; 2005. p. 78-89.
3. Berka C, Levendowski D, P. W, Davis G, Lumicao M, Ramsey C, et al. Implementation of a Closed-Loop Real-Time EEG-Based Drowsiness Detection System: Effects of Feedback Alarms on Performance in a Driving Simulator. In: Human Computer Interaction Conference; 2005 July 2005; Las Vegas, NV; 2005.
4. Berka C, Levendowski DJ, Olmstead RE, Popovic MV, Cvetinovic M, Petrovic MM, et al. Real-time Analysis of EEG Indices of Alertness, Cognition, and Memory with a Wireless EEG Headset. International Journal of Human-Computer Interaction 2004;17(2):151-170.
5. Berka C, Levendowski DJ, Ramsey CK, Davis G, Lumicao MN, Stanney K, et al. Evaluation of an EEG-Workload Model in an Aegis Simulation Environment. In: SPIE Defense and Security Symposium, Biomonitoring for Physiological and Cognitive Performance during Military Operations; 2005 May 2005; Orlando, FL; 2005. p. 90-99.
6. Berka C, Westbrook P, Levendowski DJ, Lumicao MN, Ramsey CK, Zavora T, et al. Implementation Model for Identifying and Treating Obstructive Sleep Apnea in Commercial Drivers. In: International Conference on Fatigue Management in Transportation Operations; 2005 (in press) September 12-14, 2005; Seattle, WA; 2005 (in press).
7. Brookhuis KA, de Waard D. The use of psychophysiology to assess driver status. Ergonomics 1993;36(9):1099-110.
8. Brookings JB, Wilson GF, Swain CR. Psychophysiological responses to changes in workload during simulated air traffic control. Biol Psychol 1996;42(3):361-77.
9. Dickson BT. The Cognitive Cockpit - a test-bed for Augmented Cognition. In: The International Conference on Human Computer Interaction; 2005 July 2005; Las Vegas, NV; 2005.
10. Gevins A, Smith ME. Assessing Fitness-for-Duty and Predicting Performance With Cognitive Neurophysiological Measures. In: Biomonitoring for Physiological and Cognitive Performance during Military Operations; 2005 March 31-April 1; Orlando, FL, USA: SPIE - The International Society for Optical Engineering; 2005. p. 127-138.
11. Gevins A, Smith ME, Leong H, McEvoy L, Whitfield S, Du R, et al. Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. Hum Factors 1998;40(1):79-91.
12. Gevins A, Smith ME, McEvoy L, Yu D. High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. Cereb Cortex 1997;7(4):374-85.
13. Kramer A, Parasuraman R. Neuroergonomics--application of neuroscience to human factors. In: Cacciopo J, Tassinary LG, Berntson GG, editors. Handbook of Psychophysiology, 2nd ed. 2 ed. New York: Cambridge University Press; 2005.
14. Kramer HC, Trejo LJ, Humphrey DG. Psychophysiological measures of workload: Potential applications to adaptively automated systems. New Jersey: Lawrence Erlbaum Associates; 1996.
15. Levendowski DJ, Berka C, Olmstead RE, Jarvik M. Correlations between EEG Indices of Alertness Measures of Performance and Self-Reported States while Operating a Driving Simulator. In: 29th Annual Meeting, Society for Neuroscience; 1999 October 25; Miami Beach, FL; 1999.
16. Levendowski DJ, Berka C, Olmstead RE, Konstantinovic ZR, Davis G, Lumicao MN, et al. Electroencephalographic indices predict future vulnerability to fatigue induced by sleep deprivation. Sleep 2001;24(Abstract Supplement):A243-A244.
17. Levendowski DJ, Westbrook P, Berka C, Popovic MV, Ensign WY, Pineda JA, et al. Event-related potentials during a psychomotor vigilance task in sleep apnea patients and healthy subjects. Sleep 2002;25(Abstract Supplement):A462-A463.
18. Makeig S. Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. Electroencephalography and Clinical Neurophysiology 1993;86(4):283-293.
19. Makeig S, Jung TP. Changes in alertness are a principal component of variance in the EEG spectrum. Neuroreport 1995;7(1):213-6.
20. Makeig S, Jung TP. Tonic, phasic, and transient EEG correlates of auditory awareness in drowsiness. Brain Res Cogn Brain Res 1996;4(1):15-25.
21. Marek T, Pokorski J. Quo vadis, ergonomia?-25 years on. Ergonomia 2004;26:13-18.
22. Morizio N, Thomas M, Tremoulet P. Performance Augmentation through Cognitive Enhancement (PACE). In: Internation Conference on Human Computer Interaction; 2005 July 2005; Las Vegas, NV; 2005.
23. Murata A. An Attempt to Evaluate Mental Workload Using Wavelet Transform of EEG. In: Human Factors; 2005 Fall 2005: The Human Factors and Ergonomics Society; 2005. p. 498-508.
24. Mussa-Ivaldi FA, Miller LE. Brain-machine interfaces: computational demands and clinical needs meet basic neuroscience. Trends in Neuroscience 2003;26(6):329-334.
25. Parasuraman R. Neuroergonomics: Research and Practice. Theoretical Issues in Ergonomics Science 2003;4:5-20.

26. Parasuraman R, Rizzo M. Neuroergonomics: The Brain at Work. New York: Oxford University Press; 2005.
27. Pleydell-Pearce CW, Whitecross SE, Dickson BT. Multivariate analysis of EEG: Predicting cognition on the basis of frequency decomposition, inter-electrode correlation, coherence, cross phase, and cross power. In: Proceedings of the 36th Annual Hawaii International Conference on System Sciences; 2003; Hawaii; 2003. p. 131.
28. Prinzel LJ, Freeman FG, Scerbo MW, Mikulka PJ, Pope AT. A closed-loop system for examining psychophysiological measures for adaptive task allocation. Int J Aviat Psychol 2000;10(4):393-410.
29. Riccio CA, Reynolds, C.R., Lowe, P.A. Clinical Applications of Continuous Performance Tests *Measuring Attention and Impulsive Responding in Children and Adults*. New York: John Wiley & Sons, Inc.; 2001.
30. Smith ME, Gevins A. Neurophysiolgic Monitoring of Mental Workload and Fatigue and During Operation of a Flight Simulator. In: Biomonitoring for Physiological and Cognitive Performance during Military Operations; 2005 March 31-April 1; Orlando, FL, USA; 2005. p. 116-126.
31. Sterman MB. Application of quantitative EEG analysis to workload assesment in an advanced aircraft simulator. In: Proceeding of Human Factors Society; 1993; 1993.
32. Sterman MB, Mann CA. Concepts and applications of EEG analysis in aviation performance evaluation. Biol Psychol 1995;40(1-2):115-30. Review.
33. Sterman MB, Mann CA, Kaiser DA. Quantitative EEG patterns of differential in-flight workload. In: Space Operations, Applications, and Research Proceedings; 1992 June 25, 1992; Sepulveda VA Medical Center: NASA Conference Publication; 1992. p. 466-473.
34. Trejo LJ, Kochavi R, Kubitz K, Montgomery LD, Rosipal R, Matthews B. Measures and Models for Predicting Cognitive Fatigue. In: Caldwell Jr. JA, Wesensten N, editors. Biomonitoring for Physiological and COgnitive Performance during Military Operations; 2005 March 31-April 1; Orlando, FL, USA: SPIE - International Society for Optical Engineering; 2005. p. 105-115.
35. Ververs PM, Whitlow SD, Dorneich MC, Mathon S. Building Honeywell's Adaptive System for the Augmented Cognition Program. In: Internation Conference on Human Computer Interaction; 2005 July 2005; Las Vegas, NV; 2005.
36. Westbrook P, Berka C, Levendowski DJ, Lumicao MN, Davis G, Olmstead R, et al. Quantification of Alertness, Memory and Neurophysiological Changes in Sleep Apnea Patients Following Treatment with nCPAP. Sleep 2004;27:A223.
37. Wilson GF. Operator Functional State assessment for Adaptive Automation Implementation. In: Biomonitoring for Physciological and Cognitive Performance during Military Operations; 2005 March 31-April 1; Orlando, FL, USA: SPIE - The International Society for Optical Engineering; 2005. p. 100-104.
38. Wilson GF, Eggemeier FT. Physiological measures of workload in multi-task environments. London; 1991.