Human Factors: The Journal of the Human Factors and Ergonomics Society

Performance Enhancement in an Uninhabited Air Vehicle Task Using Psychophysiologically Determined Adaptive Aiding

Glenn F. Wilson and Christopher A. Russell

The online version of this article can be found at:

Published by:

$SAGE

On behalf of:

Human Factors and Ergonomics Society

Additional services and information for *Human Factors: The Journal of the Human Factors and Ergonomics Society* can be found at:

Email Alerts: http://hfs.sagepub.com/cgi/alerts

Subscriptions: http://hfs.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://hfs.sagepub.com/content/49/6/1005.refs.html

>> Version of Record - Dec 1, 2007

What is This?

Performance Enhancement in an Uninhabited Air Vehicle Task Using Psychophysiologically Determined Adaptive Aiding

Glenn F. Wilson and **Christopher A. Russell,** Air Force Research Laboratory, Wright Patterson Air Force Base, Ohio

Objective: We show that psychophysiologically driven real-time adaptive aiding significantly enhances performance in a complex aviation task. A further goal was to assess the importance of individual operator capabilities when providing adaptive aiding. **Background:** Psychophysiological measures are useful for monitoring cognitive workload in laboratory and real-world settings. They can be recorded without intruding into task performance and can be analyzed in real time, making them candidates for providing operator functional state estimates. These estimates could be used to determine if and when system intervention should be provided to assist the operator to improve system performance. **Methods:** Adaptive automation was implemented while operators performed an uninhabited aerial vehicle task. Psychophysiological data were collected and an artificial neural network was used to detect periods of high and low mental workload in real time. The high-difficulty task levels used to initiate the adaptive automation were determined separately for each operator, and a group-derived mean difficulty level was also used. **Results:** Psychophysiologically determined aiding significantly improved performance when compared with the no-aiding conditions. Improvement was greater when adaptive aiding was provided based on individualized criteria rather than on group-derived criteria. The improvements were significantly greater than when the aiding was randomly provided. **Conclusion:** These results show that psychophysiologically determined operator functional state assessment in real time led to performance improvement when included in closed loop adaptive automation with a complex task. **Application:** Potential future applications of this research include enhanced workstations using adaptive aiding that would be driven by operator functional state.

INTRODUCTION

The human operator is a key component of manned systems, and optimal system performance is predicated upon unimpaired operator capabilities. Less than optimal operator functional state (OFS) can reduce the efficiency of the operator and result in degraded system performance. The functional state of the human operator, unlike other system components, is not monitored during job performance. This can lead to situations in which unrecognized operator impairment results in lowered system performance and errors. Continuous monitoring of OFS would permit the ongoing assessment of operator cognitive capability. When evaluated in the context of current and predicted system demands, OFS assessment may indicate that the operator is not capable of successfully dealing with the cognitive demands of the task. It may be possible to adjust the task demands to match the current OFS (Rouse, 1988).

Adaptive aiding is a method of providing assistance to operators by introducing automation only when it is required (Parasuraman, Mouloua, & Molloy, 1996; Scerbo, 1996). Parasuraman et al. (1996) proposed that there are five strategies used to implement adaptive aiding. Adaptive aiding may be provided based upon critical events, operator performance, operator physiology, models of operator cognition, and hybrid methods that combine the other four techniques. Each of these strategies has attributes that support its use and others that may be problematic in certain situations. It may be that the hybrid approach will be the most

useful in many situations. The focus of the current work is on using psychophysiological measures to implement adaptive aiding.

Accurate and reliable assessment of OFS is essential for adaptive aiding schemes having the goal of dynamically matching the momentary cognitive capabilities of the operator with the demands of the task. In theory, adaptive aiding should be provided only when it is required (Scerbo, 1996) and must improve performance when compared with static and fully automated systems (Hancock & Verway, 1997; Parasuraman & Riley, 1997). To reduce performance errors that would be caused by not providing aiding when required or providing it when it is not required depends upon accurate and reliable assessment of the cognitive capabilities of an operator from moment to moment during task performance (Parasuraman, 2003).

Several authors have argued that psychophysiological measures are worthy candidates for this role (Byrne & Parasuraman, 1996; Gevins et al., 1998; Gevins, Smith, McEvoy, & Yu, 1997; Scerbo, Freeman, & Mikulka, 2003; Wilson & Russell, 2003a, 2003b). Unlike performance and subjective measures, psychophysiological measures are continually available and can be collected without intruding into the operator's task (Kramer, 1991; Wilson & Eggemeier, 1991). Some tasks are highly automated and place the operator in a primarily monitoring role, which greatly reduces overt performance. The collection of subjective measures can intrude into the operator's task, and if assessment is delayed to avoid this intrusion, the responses may suffer from memory lapses and operator bias (Moroney, Biers, & Eggemeier, 1995).

Numerous studies have demonstrated that psychophysiological measures are sensitive to changes in OFS, including mental workload and fatigue (Caldwell, Caldwell, Brown, & Smith, 2004; Gevins et al., 1997; Kramer, 1991; Wilson & Eggemeier, 1991). Further, psychophysiological data have been recorded in a variety of operational environments, including flight and driving automobiles, trucks, and trains (Apparies, Riniolo, & Porges, 1998; Brookhuis & de Waard, 1993; Wilson, 2001, 2002b).

By using multiple psychophysiological measures it is possible to gain a more complete estimation of OFS in complex, multifaceted tasks such as air traffic control (Brookings, Wilson, & Swain, 1996) and piloting aircraft (Hankins & Wilson, 1998). For example, using data collected from

general aviation pilots, Wilson (2002a) reported that eye blink rate decreased when pilots flew using only the aircraft instruments, heart rate increased during takeoffs and landings, and electroencephalographic (EEG) activity was affected by navigation tasks.

When multiple measures are collected, a method of data synthesis may be desirable to provide a single index that characterizes OFS. Several methods have been used, including multivariate statistical analysis (Berka et al. 2004; Pleydell-Pearce, Whitecross, & Dickson, 2003; Smith, Gevins, Brown, Karnik, & Du, 2001; Wilson & Fisher, 1991, 1995), ratio of the power in EEG bands (Freeman, Mikulka, Prinzel, & Scerbo, 1999; Pope, Bogart, & Bartolome, 1995) and artificial neural networks (ANNs; Gevins et al., 1998; Wilson & Russell, 2003a, 2003b).

A further advantage of psychophysiological measures is that they can provide OFS information in real time (Berka et al., 2005; Wilson & Russell, 2003b, 2004). These online measures of OFS have been derived from several EEG channels (Berka et al., 2005; Freeman et al., 1999) or EEG and peripheral nervous system measures of heart and eye activity (Wilson & Russell, 2003b, 2004).

Real-time analysis of EEG activity has been used to modify task characteristics online to better match OFS in a limited number of reports. Reports from the National Aeronautics and Space Administration (NASA) and Old Dominion University used various ratios of theta, alpha, and beta EEG bands to develop an engagement index (EI; Freeman et al., 1999; Pope et al., 1995; Prinzel, Freeman, Scerbo, Mikulka, & Pope, 2000). This work was based on the assumption that the EI was related to arousal or task engagement. Unfortunately, this critical assumption was never tested using their task or operators (Scerbo et al., 2003).

If the EI indicated a low level of task engagement, a tracking task would be switched from automatic to manual mode, which required operator control. High levels of engagement would result in converting the tracking task to automatic mode, which required no operator input. Better performance was reported under negative versus positive feedback conditions. Unfortunately, it is difficult to interpret these data because the tracking task that was producing the dependent variable, tracking error, was the task that was turned off and on. This occurred rapidly, with a reported mean of switching modes every 14 to 19 s for the duration

of the 16-min trials (Prinzel, Freeman, Scerbo, Mikulka, & Pope, 2003). Further, because the tracking task reset to the zero or home position during automatic mode, the error score would be reduced for the following manual mode periods while the cursor moved from the home position. The effect of this error reduction was not reported.

The utility of the EEG ratio measures for high cognitive workload in complex tasks is uncertain. The EI is designed to monitor arousal based on changes in theta, alpha, and beta power. The assumption was that theta and alpha activity increase with low arousal or low engagement states and decrease with arousal, whereas beta band activity does the opposite. However, other reports show that high cognitive workload levels increase theta and reduce alpha over the parietal scalp areas where the EEG activity for the EI was recorded (Brookings et al., 1996; Fairclough, Venables, & Tattersall, 2005; Fournier, Wilson, & Swain, 1999; Gevins & Smith, 2003; Hankins & Wilson, 1998).

Additionally, Prinzel et al. (2000) reported no differences in the EI between their easy single task condition and their difficult multiple task condition. They reported better tracking performance between an experimental and a control group. Because only the experimental group experienced the single-task tracking condition, it was not possible to make comparisons that would have indicated whether the two groups possessed equivalent performance on the tracking task in the single-task condition. In a later report (Prinzel et al., 2003), no differences were found among the EI values when three experimental conditions were used.

Wilson and Russell (2004) employed an uninhabited air vehicle (UAV) task with two levels of difficulty. In this complex task, operators monitored the progress of four UAVs and were required to download radar images and to locate and designate targets under two levels of visual complexity and time pressure. The operators were also required to simultaneously perform a running memory task. Using an ANN that was trained to recognize low and high mental workload based upon psychophysiological data, the authors reported high levels of ANN correct discrimination between low and high task performance.

Further, when the ANN detected levels of high workload, adaptive aiding was provided to assist the operators, which resulted in increased number of targets hit, targets designated, and weapon release points achieved when compared with the nonaiding condition. By closing the loop between operator and the UAV using psychophysiological measures, they were able to dynamically match the task demands to the operator's momentary cognitive state and produce significant improvements in complex task performance.

Similar results were described by Parasuraman, Mouloua, and Hilburn (1999), who reported that adaptive aiding enhanced performance in a flight simulation task when the aiding was provided during high workload, but not during low-workload flight segments. Although they did not use real-time psychophysiological data to assess OFS, the adaptive aiding did enhance performance.

Because individual operators possess different cognitive abilities and skill levels, optimal adaptive aiding should occur when the aiding is implemented based upon the momentary capabilities of each operator (Hedge & Borman, 2006). Differences in cognitive ability have been found to have correlates in brain activity (Gevins & Smith, 2000). Although group-determined criteria for implementing adaptive aiding have been shown to improve performance, it would seem propitious if the decision to provide aiding were tailored to the capabilities of the individual operator (Wilson & Russell, 2004). In many situations it is easier to use the same task difficulty levels for all operators. However, we have observed that this produces suboptimal results because the difficult level of task demand may be too challenging for some operators and too easy for others.

In our laboratory, when a fixed level was used as the difficult task condition, the accuracy of our psychophysiologically driven classifier varied among the operators and the accuracy was correlated with proficiency. This suggests that optimal adaptive aiding will be achieved if the capabilities of each operator are determined and used to provide adaptive aiding (Gevins & Smith, 2003; Parasuraman, 2003). Recent data from Gonzalez (2005) showed that high task demands had a more detrimental effect on the performance of operators with lower cognitive ability when compared with operators with higher abilities.

An additional consideration is the method used to implement the adaptive aiding. Psychophysiological measures are subject to rapid fluctuations, which may produce the presentation and withdrawal of adaptive aiding at a too-rapid rate. One strategy that is used to smooth the transitions is to apply a data window with an overlap of tens of

seconds (Freeman et al., 1999; Gevins et al., 1998; Wilson & Russell, 2003a). Also, if successful, the adaptive aiding will provide task demands that match the current capabilities of the operator, which will change the operator's physiology and result in the removal of the aiding. If this occurs quickly, turning the aiding on and off could actually interfere with the operator's performance in complex tasks.

An alternative would be to couple the OFS with task information when determining the duration of each aiding episode. For example, if the OFS classifier determines that the operator requires assistance, the adaptive aiding system could interrogate the system and provide the aiding until the demands of the task diminish (Bonner, Taylor, & Miller, 2000). This would avoid multiple occurrences of aiding onset and offset while the task demands remain high. The dynamic nature of the adapting system must not interfere with the operator's task performance.

This paper describes a project in which the OFS of UAV operators was assessed using psychophysiological measures, online, while they performed tasks having different levels of cognitive difficulty. This information was used to modify task demands to determine if operator performance would improve. A complex, simulated, UAV attack scenario was used in which each operator was simultaneously responsible for four vehicles and was required to locate and designate targets using preestablished rules. The most difficult level of the task was individually determined for each operator. This level and a task difficulty level representing the group's mean were used to determine the importance of individual thresholds to determine adaptive aiding implementation.

The goals of this project were to demonstrate that psychophysiologically driven adaptive aiding would improve performance and that randomly presented aiding would not show the same level of performance improvement. Further, the importance of using individual operator capabilities to determine the threshold for implementing the aiding would show greater performance improvement than when a group-determined performance threshold was used. The method of terminating the aiding was also explored to determine if ending the aiding when the psychophysiological OFS assessment indicated that it was no longer needed was superior to ending the aiding when the current subtask ended.

METHODS

Ten volunteers with a mean age of 24.9 years, who had given informed consent, were practiced until they showed stable performance on a simulated UAV task. The practice took a mean of 10.6 hr over 3 to 4 days. The operators monitored the progress of four autonomous vehicles, on two abutted 20-inch (~51-cm) diagonal computer screens separated by 4 inches (~10.2 cm) of bezel, as they flew a preplanned bombing mission. When the vehicles reached designated way points, radar images of the target area were available to the operators. The operators gave commands to download and view the images and then performed a visual search of the images. They marked six targets for bombing before the vehicle reached the weapons release way point. If the targets were not selected and/or the weapons release command was not given in time, the weapons from that vehicle could not be released, thereby reducing the effectiveness of the entire mission.

Three categories of targets were used, and the operators were required to employ a predetermined set of priorities when selecting targets. Because the entire radar image could not be viewed at one time, the operators had to pan around the image and remember the types and locations of the targets that they found. Then they would designate the targets based upon the preestablished set of priorities (see Figure 1).
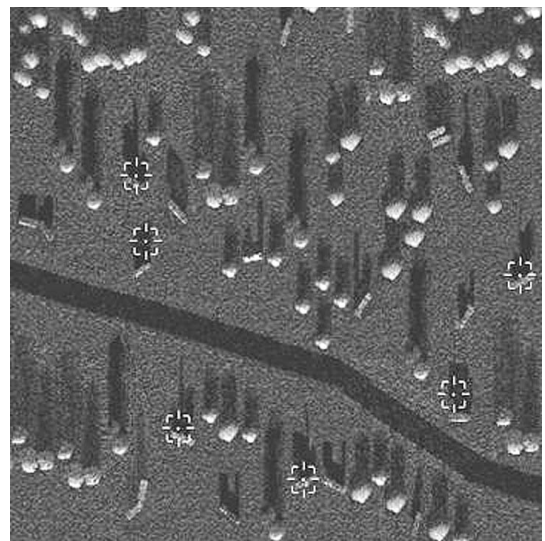

Figure 1. Radar image showing an entire difficult-level image with six targets designated.

The complexity of the images was presented at two levels. The more difficult level contained a larger number of distractors in the form of nontarget vehicles and buildings and also required more complex decisions concerning target priority. Four easy and four difficult images were presented in a balanced order for each condition. Simultaneously, the operators monitored the well-being of each vehicle (vehicle health task, VHT) by observing messages showing potential vehicle problems, such as loss of communication. These messages appeared in the upper left of the display screen and occurred throughout the easy and difficult conditions. One distractor message was also presented, and all status messages appeared for 10 s, although the command that identified the problem that had reached a critical level appeared for 15 s.

Memory was manipulated by having the operators keep two aircraft-problem combinations (excluding the distractor message) in memory until the command was given that identified the aircraft in which the problem had reached a critical level and required attention. The operators then selected the appropriate vehicle designation from a pull-down menu and, using other pull-down menus, found and selected the appropriate fix for that vehicle's problem.

The number of targets selected, the number of nontargets selected, the number of targets hit, and whether or not the command to release the weapons was executed in time (successful weapons release) were recorded. The VHT was scored by the number of correct solutions, the number of time-outs, and the reaction times for responding to a critical malfunction. Following each mission the operators gave estimates of their mental workload using the NASA Task Load Index (NASA-TLX) for the easy and difficult radar images. Each data collection run took approximately 14 min.

The difficult task level for testing was determined for each operator, after he or she had reached stable performance, by using a titration procedure. This was accomplished by increasing the speed of the UAVs during the difficult radar image conditions until the operator successfully completed only 25% to 30% of the weapon release points. This vehicle speed was then designated as that operator's individual level for the difficult radar image processing. The group mean of these titration runs was determined and used as a second vehicle speed for all of the operators as the group level of the difficult radar image processing. The perfor-

mance of 5 of the operators (the low-performance group) fell below or at the mean level, whereas the other 5 operator's titration speeds were above the mean level (the high-performance group).

Five channels of EEG, electrocardiograph (ECG), and vertical and horizontal electro-oculograph (EOG) activity were collected. The EEG data were recorded from scalp sites F7, Fz, Pz, T5, and O2 of the 10/20 electrode system using an Electrocap (Electrocap International, Eaton, OH). These sites have previously been shown in our laboratory to provide good discrimination between task levels in complex cognitive tasks (Russell & Wilson, 2005). Electrodes attached to the mastoid processes were used as reference and ground. Eye and cardiac activity were recorded using disposable Ag/AgCl electrodes. The EOG electrodes were placed above and below the midline of the right eye to record vertical movement and blink activity. Electrodes placed next to the outer canthus of each eye recorded horizontal ocular activity.

The ECG electrodes were placed on the sternum and the left clavicle. These data were amplified and filtered by a small, operator-worn telemetry device (BioRadio 110, Cleveland Biomedical Inc.). The sampling rate was 200 Hz with a bandpass from 0.5 to 52.4 Hz. The digitized data were reduced online and simultaneously stored on a computer disk by a laboratory-developed software program, NuWAM (Krizo, Wilson, & Russell, 2005). Eye artifacts in the EEG data were corrected using an adaptive filter with inputs from the vertical and horizontal EOG channels (He, Wilson, & Russell, 2004). The corrected EEG and the EOG data were submitted to a fast Fourier transformation every second. These data were separated into five bands for further statistical analysis: delta 2.0 to 4.0 Hz, theta 5.0 to 8.0 Hz, alpha 9.0 to 13.0 Hz, beta 14.0 to 32.0 Hz, and gamma 33.0 to 43.0 Hz. Interbeat intervals were calculated online from the ECG data.

These reduced data were then provided to an ANN every second. A 10-s window with a 9-s overlap was used as input to the ANN. The ANN had a total of 37 input features with a hidden layer with 37 nodes and 2 output nodes, easy and difficult. Because there were more data in the easy condition, training examples for ANN were randomly selected so that the number of examples were the same for the easy and difficult ANN training data sets. Of the 10-s segments from each of the two

ANN training conditions, 75% were randomly selected and used as training data, whereas 25% were used as validation data to determine the point at which the ANNs were trained but not overtrained. The validation data were also used to test the accuracy of the trained ANN (Wilson & Russell, 2003b).

After the operators had been practiced to stable performance and their titration levels established, they returned on a separate day for test data collection, which began with collecting data that were used for training the ANN. The ANN training data represented periods of easy and difficult task levels recorded while each operator performed the UAV task at his or her titrated vehicle speed and also at the easy condition vehicle speed. The data from two separate ANN training runs were combined. Separate ANNs were trained for each operator. During subsequent task performance the ANN provided estimates of the operator's state every second.

Four conditions were used; each comprised a single data collection run (see Table 1):

No adaptive aiding. During this condition, only operator performance and ANN accuracy were recorded. This was done for the individually determined (*no aiding-individual*) and the group (*no aiding-group*) vehicle speeds.

Adaptive aiding. When the ANN estimates indicated that the operator was in a state of high cognitive workload, the UAV task was modified to reduce the cognitive demands on the operator.

Adaptive aiding was accomplished with two manipulations: (a) decreasing the velocity of the vehicle whose targets were being evaluated by 50% and (b) displaying the current VHT status messages in a dropped tab at the VHT display area. This provided more time to complete target selection before the weapons release point was reached and provided a visual list of currently unresolved VHT messages to ease the memory load for the VHT task. This was accomplished for both the individually determined vehicle speed (*aiding-individual*) and the group determined vehicle speed (*aiding-group*). The same group mean speed was used for all operators. When the ANN indicated the easy operator state, the vehicle speed was returned to the faster level.

Random aiding. In this condition aiding was provided at randomly determined intervals for each operator. The total amount of aiding and the number of times aiding was provided were the same as for the aiding condition. The length of each aiding period was the mean for that condition (total time/number of times aided for that operator). This was accomplished for the individual (*random aiding-individual*) and group (*random aiding-group*) vehicle speeds.

Leave on aiding. Using the individually determined vehicle speed, the aiding was turned on at the first instance of ANN-determined high workload level and left on until the weapons release command was given or the release way point was crossed.

TABLE 1: List of Experimental Conditions With Brief Descriptions

Condition	Description
Training	ANN training only; individual and group speeds used separately.
No aiding-individual	Performance only, no aiding, used to test ANN accuracy and provide baseline performance. Individually determined speeds used.
No aiding-group	Performance only, no aiding, used to test ANN accuracy and provide baseline performance. Group speed used.
Aiding-individual	Aiding presented using ANN trained with individual speeds.
Aiding-group	Aiding presented using ANN trained with group speeds.
Random aiding-individual	Aiding presented randomly; total aiding time was the same as each operator's aiding-individual total times.
Random aiding-group	Aiding presented randomly; total aiding time was the same as each operator's aiding-group total times.
Leave on aiding	Aiding presented using ANN trained with individual speeds. Aiding left on until weapons were released or weapons release point met.

Note. ANN = artificial neural network.

On the day of data collection the operators practiced the tasks by completing a warm-up scenario prior to data collection. The order of presentation was blocked with the constraints that the two ANN training runs had to occur first and the aiding-individual aiding-group aiding had to occur prior to their respective random aiding conditions.

The performance, psychophysiological, and subjective data were statistically evaluated using a within-operator ANOVA. Significant ANOVAs were followed with paired comparisons, t tests, to determine significant differences using $p \leq .05$.

RESULTS

The ANN classification accuracies – that is, correctly determined easy and difficult task levels based upon task condition, for the training and the two nonaided conditions – are presented in Figure 2. The easy versus difficult workload comparison was significant, $F(1, 9) = 15.94$, $p < .0002$, with a mean correct classification for the easy condition of 89.7% and the difficult condition of 80.1% correct. There was a significant effect among the training, no aiding-individual, and no aiding-group conditions, $F(2, 18) = 23.95$, $p < .0001$. The correct classification means for the training, no aiding-individual, and no aiding-group conditions were 95.7%, 83.6% and 75.5%, respectively. Paired comparisons showed that the ANN did significantly better discriminating between the easy and difficult task levels for the training condition than for both the individual and group conditions.

The classification accuracies were significantly higher for the individual than the group conditions. The comparison between the high- and low-performance groups was significant, $F(1, 4) = 5.24$, $p = .027$, with the mean correct percentage for the high performers of 87.7% and 82.2% for the low performers. The interaction of task difficulty and the type of aiding was significant, $F(2, 8) = 17.57$, $p < .0001$. The test data for the training run are those data that were withheld from the ANN training and belonged to the same overall data set resulting in the very high classification accuracies. For the low- and high-performance groups in the training condition the ANNs did very
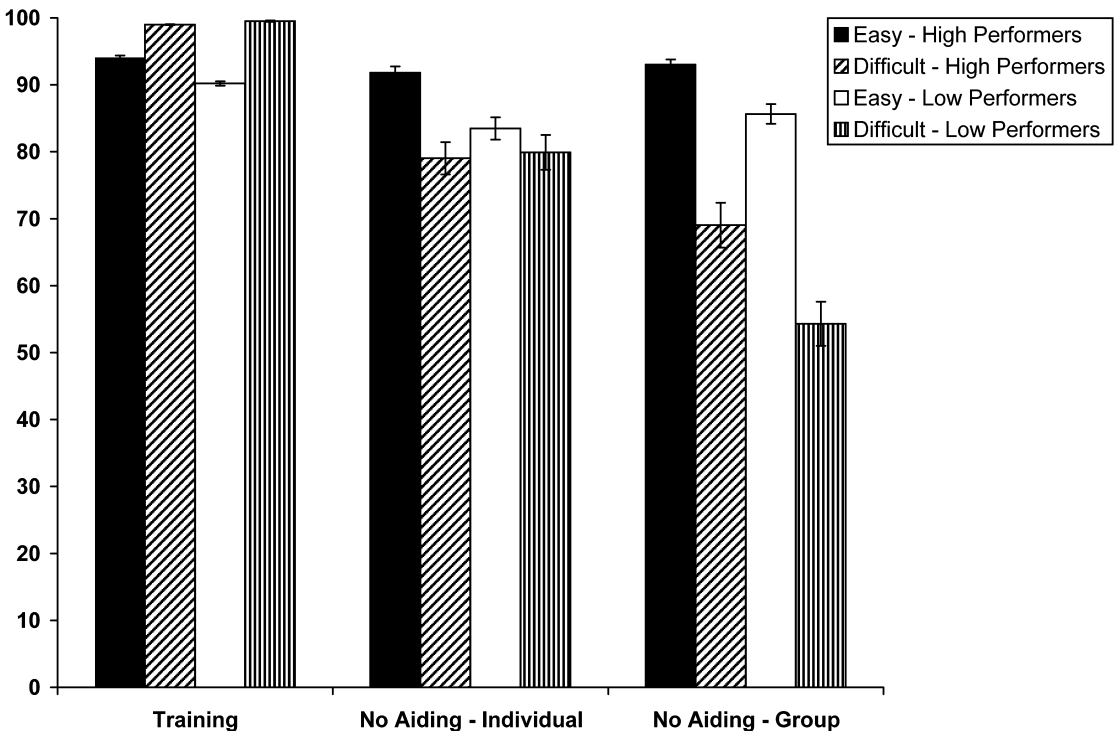

Figure 2. Mean artificial neural network (ANN) classifier accuracies for the training, no aiding-individual, and no aiding-group conditions for the high- and low-performance groups. Standard error bars are shown.

well, with a range of correct classification of the easy and difficult conditions from 89% to 100%.

For the two nonaiding runs the data were not part of the original training data set and the accuracies were lower, ranging from 54% to 93% correct. The nonaiding runs using the individually determined task difficulty resulted in a mean correct classification of 83.6% with a range from 79% to 91.8%. The ANN accuracies when the operators performed the group mean task difficulty level was 75.5%, with a range from 54.3% to 93% correct. Although the accuracy of correctly determining the easy task demand level was essentially the same as for the no aiding-individual condition, the accuracy of correctly determining the difficult task level dropped to a mean of 61.7% for the group difficulty level, compared with 79.5% for the individually determined difficult task level condition.

The number of successful weapons releases (SWRs) was greatly affected by the task difficulty, $F(1, 9) = 203.51$, $p < .0001$. The percentage of SWRs during the easy level was almost perfect, mean of 97.7%, whereas the overall difficult task performance was 51.3%. Because only the difficult task condition was affected by experimental conditions, the statistical tests on only those data will be reported.

For the difficult condition there was a significant effect of aiding type, $F(7, 63) = 4.90$, $p < .0002$. As shown in Figure 3, there were dramatic differences for the SWRs during the difficult task levels associated with the various aiding conditions for the combined low- and high-performance groups. The goal of only 25% to 30% completed SWRs in the nonaided difficult task level during the training and no aiding-individual titrated conditions was achieved, 27.5% and 30% respectively. The no aiding-group was slightly higher, 35%, because the mean difficulty level of the high and low performers was used.

The largest improvement in performance was during the aiding-individual condition, which was significantly greater than the three nonaiding conditions and the random aiding-individual condition. It was not significantly different from the aiding-group, random aiding-group, and the leave on aiding condition. The aiding-group percentage SWRs was significantly larger than all three of the nonaiding conditions. The leave on aiding condition also demonstrated significantly improved percentage SWRs as compared with the three nonaiding conditions and the random aiding-individual conditions.

Examination of the low- and high-performance groups' data separately showed that the various aiding conditions had differential effects. The low-performance group's data from the difficult task level is shown in Figure 4. The best performance
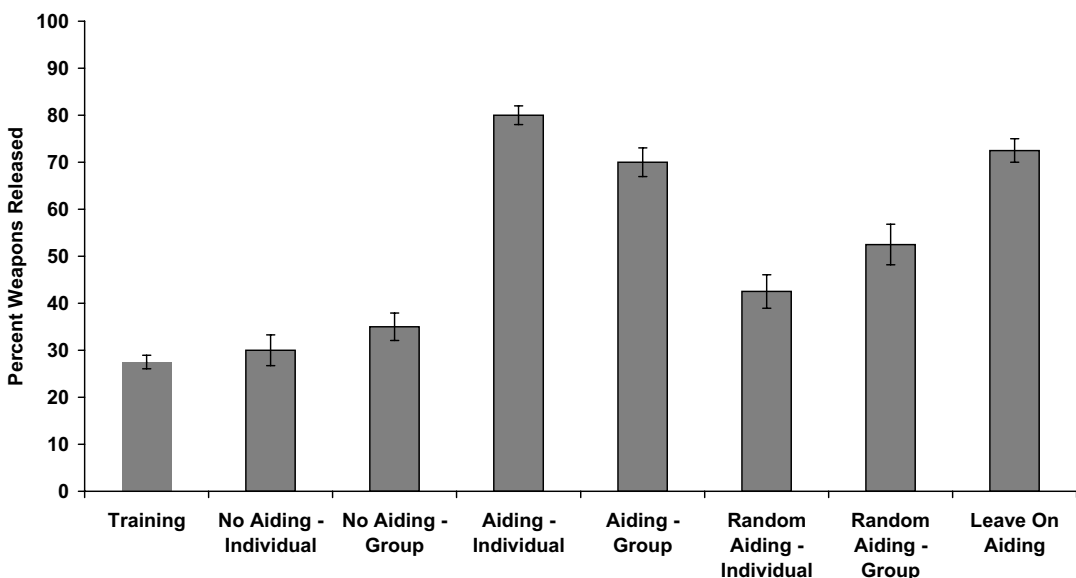

Figure 3. Mean percentage successful weapons releases (SWRs) completed for the difficult task level for each of the conditions for the 10 operators. Bars are standard errors.
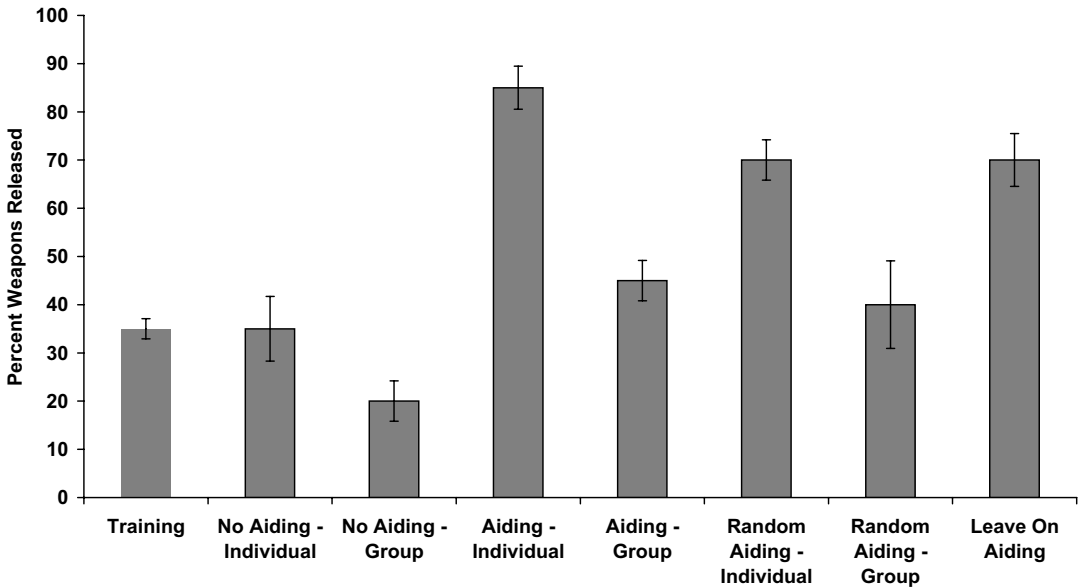
Figure 4. Mean percentage successful weapons releases (SWRs) by aiding condition for the low-performance group during the difficult task level. Error bars are standard errors.

was during the aiding-individual condition, which was significantly larger than the three nonaiding conditions and the aiding-group, random aiding-individual, and random aiding-group conditions. Only the leave on aiding condition was statistically equivalent. The percentage SWR during the random aiding-individual condition was significantly larger than all three of the nonaiding conditions. The leave on aiding condition produced better performance than the three nonaiding conditions and the aiding-group conditions.

The high performers' data showed a more complex picture of the effects of the various aiding conditions (see Figure 5). The titrated vehicle speeds were higher for this group than for the low-performance group. The highest percentage SWRs was during the aiding-group condition, which was as high as the easy task difficulty condition results, mean of 95%. This was significantly larger than the three nonaiding conditions and the aiding-individual and random aiding-individual conditions.

The next-highest percentage SWR was the same for the aiding-individual and leave on aiding conditions. They were both significantly larger that the training, no aiding-individual, and random aiding-individual conditions. The random aiding-group results were significantly larger than the training, all no-aiding, and random aiding-individual condi-

tions. The no aiding-group results were significantly larger than the training and random aiding-individual conditions results. The vehicle speed during the group condition was below the titrated speed for the entire high-performance group.

The number of targets selected was significantly affected only by task difficulty, $F(1, 9) = 64.7$, $p < .0001$. The mean percentage targets selected was very high for both the easy and difficult tasks, 99.5% and 92.8% respectively. This almost-perfect selection of targets during the easy task level and lower performance during the difficult task level was uniform across the various aiding conditions. Aiding and the type of aiding had no significant effect on target selection. Taken together with the SWR data, it appears that the operators chose accuracy (target selection) over speed (SWR completion) during the difficult task level.

The number of false alarms (incorrectly chosen distractors) was also significantly affected only by task difficulty, $F(1, 9) = 257.9$, $p < .0001$. The mean percentage of false alarms was 6.2% for the high-difficulty condition; there were no false alarms during the low-difficulty level tasks. Given the very low number of false alarms, the number of targets hit was determined by the SWRs, and the statistical results were identical to those of the SWRs and will not be discussed. None of the VHT measures were significantly affected by aiding type.
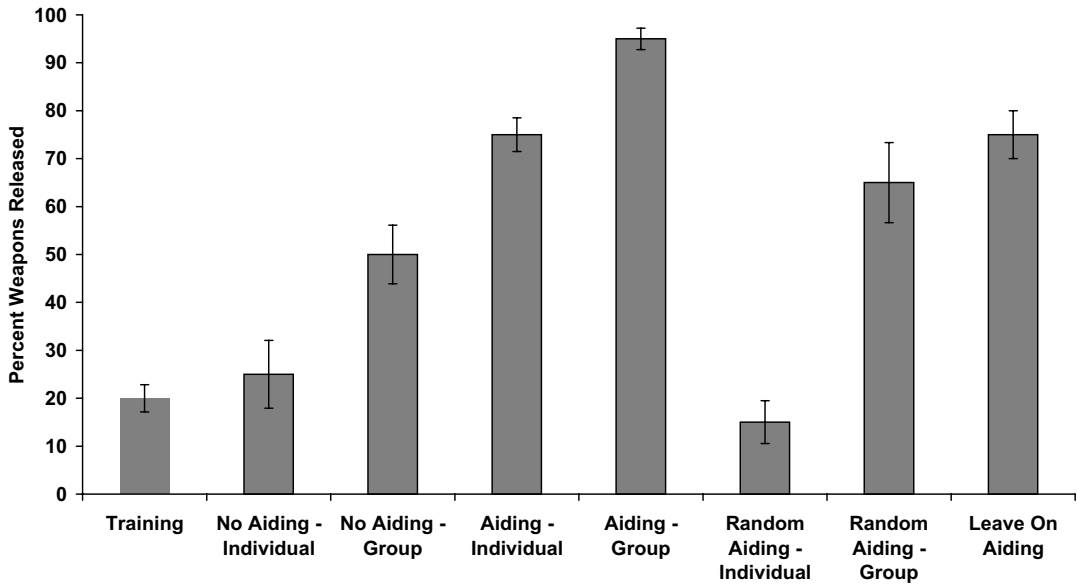
Figure 5. Mean percentage successful weapons releases (SWRs) by condition for the high-performance group during the difficult task level. Standard error bars are shown.

The subjective measure of mental workload, NASA-TLX composite, was significantly influenced by task difficulty, $F(1, 9) = 68.52, p < .0001$ (see Figure 6). The overall mean NASA-TLX composite score for the low-difficulty conditions was 15.3, whereas the mean for the difficult condition was 60.2. The interaction of task difficulty and performance group was also significant, $F(1, 4) = 5.90, p = .017$. Paired comparisons showed that the subjective workload composite score for the low-performance group during the difficult task was significantly higher than their scores during the easy condition and both the easy and difficult conditions for the high-performance group. Further, the difficult level subjective scores for the high-performance group were significantly higher than the easy task scores for both performance groups.

Separate ANOVAs were performed on the data of the high- and low-performance groups. For the low-performance group the effects only of task difficulty were significant, $F(1, 4) = 61.37, p = .0014$. However, for the high-performance group task difficulty, $F(1, 4) = 21.39, p = .0098$, and task difficulty by aiding condition, $F(7, 28) = 2.44, p = .044$, were significantly different. Paired comparisons showed that the aiding-individual condition scores were significantly lower than those from the no aiding-training, no aiding-individual, and random aiding-individual conditions. The subjective work-load estimates for the random aiding-group were significantly lower than those from the no aiding-training condition. Conversely, the no aiding-individual scores were significantly higher than those from the aiding-group, random aiding-group, and leave on aiding conditions.

DISCUSSION

Adaptive aiding based upon psychophysiological measures using an ANN classifier produced a 50% improvement in performance on the UAV task. Eighty percent of the weapons release way points were completed during the aiding-individual condition, as compared with only 30% completed without the aiding during the no aiding-individual condition. The task difficulty that was used to elicit the aiding was based upon each operator's capability as determined by the titration procedure. When the adaptive aiding was accomplished using the group-determined mean vehicle speed, the overall improvement in performance was only 35%.

This difference between the aiding-individual and aiding-group conditions represents 38.4 targets that were destroyed rather than 16.8. In operational terms this is a substantial difference. Basing the implementation of the adaptive aiding upon the capabilities of each operator would
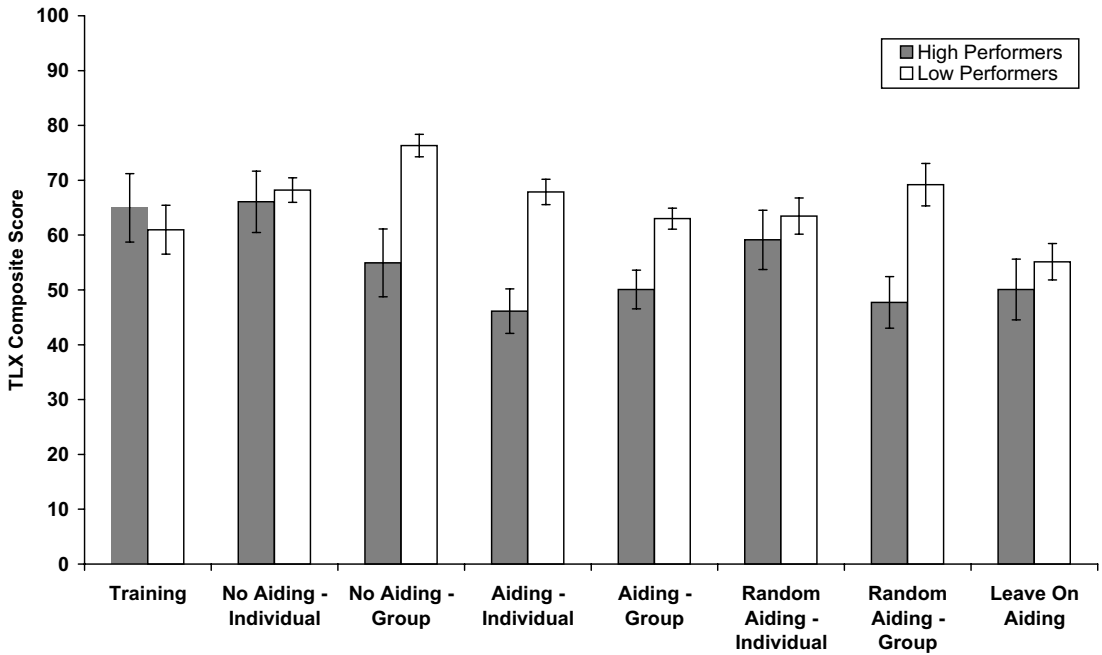
Figure 6. NASA-TLX composite scores by high- and low-performance groups for the difficult task condition by aiding type. Error bars are one standard error.

greatly improve performance and would have a tremendous impact upon operational outcome. Further, when the same amount of aiding was presented at randomly chosen times, the improvement in performance was 12.5% and 17.5% for the random aiding-individual and random aiding-group conditions, respectively. This shows that aiding has a much greater impact when it is presented based upon the psychophysiologically determined OFS rather than randomly presented during task performance.

The basis of the psychophysiologically determined adaptive aiding was dependent upon the success of the ANN classifier. The mean correct classification percentage was 83.5% for the no aiding-individual and 75.5% for the no aiding-group conditions. This was accomplished online in essentially real time and far above the 50% expected by chance. If the classifier was not able to accurately determine the functional state of the operators, then the adaptive aiding would not have been provided or would have been given at inappropriate times, when it was not needed. In either case, less performance improvement would be expected.

Even providing the same amount of aiding but at random times did not produce the high levels of performance improvement found when the aiding was given based upon the psychophysiologically determined need. Further, the ANNs were trained specifically for each operator. Using the same pool of psychophysiological features, the ANNs derived solutions that were optimized for each operator. Although not addressed in the current study, an earlier report found that the ANN classifier did not generalize very well to different manipulations of task difficulty in an air traffic control task (Wilson & Russell, 2003a). This suggests that ANN classifiers may have to be trained on the specific tasks being performed by operators in operational environments.

Examination of the high- and low-performance groups' SWRs showed differential effects of the aiding. The greatest improvement for the low-performance group was during the aiding-individual condition, when the task difficulty level was based upon their predetermined capabilities. On the other hand, the best performance for the high-performance group was during the aiding-group condition, when the group-determined difficulty level was used. This difficulty level was below the group's capability, and they were able to produce almost-perfect SWR scores matching those of the easy task level. The low-performance

group's scores during the aiding-group condition were low because the group-determined task difficulty level was above their individual capabilities.

Examination of the performance of the two groups during the random aiding conditions is very interesting. The low-performance group's scores were enhanced during the random aiding-individual condition and were only 15% below their aiding-individual scores. However, during the more difficult group difficulty task level, the randomly presented aiding resulted in only 40% SWRs. The effects of randomly providing the aiding for the high-performance group are very intriguing. The random aiding-individual condition produced the lowest percentage of way points met, 15%. This occurred even though the task difficulty was at their titrated speed. Debriefing comments by this group revealed that they all felt that the randomly presented aiding greatly interfered with their performance.

One of the differences between the high- and low-performance groups could be that the high-performance group developed strategies that they depended upon and the randomly presented aiding interfered with these strategies. The low-performance group did not depend upon well-developed strategies, and the randomly presented aircraft slowing helped them because the group condition task difficulty level was beyond their capabilities. This may add to Gonzalez's (2005) finding that high workload is more detrimental to low-ability operators in that high-ability operators may develop more successful strategies but interference with these strategies is more detrimental to their performance.

Examination of the percentage time that the aiding was on during the difficult task condition revealed that there was no significant difference between the low- and high-performance groups. The high-performance group showed higher scores during the random aiding-group condition because the task difficulty level was below their capabilities and they did not have to depend upon performance strategies. An examination of the time it took them to give permission for weapons release revealed that the times were longer in this condition than the aiding-individual condition, which suggests a strategy shift.

The subjective results support this contention because the random aiding-individual condition was rated as significantly more difficult than the aiding-individual condition. This was not the case

for the low-performance group, who showed significant differences only between the easy and difficult tasks, not between the various difficult conditions. Apparently, for the low-performance group all of the difficult task conditions were perceived as equally difficult, and the various manipulations did not affect their perceptions. This suggests not only that operators have different abilities but that the manner in which aiding is presented affects them differently.

Providing the aiding only when the classifier determined it was required (aiding-individual) provided slightly better performance versus leaving it on until the task demands changed. The difference, 7.5%, was not statistically significant. However, because 7.5% represents 2.4 more targets destroyed, this is an operationally relevant increase in targets destroyed. These results suggest that a mitigation manager based upon task context coupled with psychophysiologically driven OFS assessment may produce significant enhancements in more complex tasks.

Tasks having a richer set of cognitive demands may benefit by exactly matching specific mitigations with the current task situation. If the psychophysiological OFS assessor is capable of determining only global mental workload, a mitigation manager could provide the most appropriate mitigation in the current situation. This would represent the hybrid model of adaptive aiding suggested by Parasuraman et al. (1996), which would combine the psychophysiological and critical events techniques.

These results confirm that psychophysiologically determined OFS assessment can be used to provide adaptive aiding and result in overall system performance enhancement (Byrne & Parasuraman, 1996; Scerbo, 1996). These results show that psychophysiologically determined adaptive aiding significantly enhanced the performance of the operators and that tailoring the onset of the aiding based on the capabilities of each operator provided the most improvement.

Unlike data from earlier studies that were based upon assumptions concerning the relationship between EEG and OFS and the assumed outcome of higher rates of switching between task modes, these data provide direct evidence that psychophysiologically driven aiding produces dramatic improvement in operator performance (Freeman et al., 1999; Prinzel et al., 2000). These results also support the report by Wilson and Russell (2004),

which demonstrated performance improvement in complex task performance using psychophysiologically determined OFS to provide mitigations. Whereas Veltman and Jansen (2003) argued that adaptive aiding should be based upon both physiological and subjective effort measures, the current results demonstrate that in some circumstances adaptive aiding driven by only physiological features can provide performance enhancement.

In contemporary systems the task demands on the operator and OFS are not dynamically matched. System demands are driven by the task, and it is typically assumed that with a well-designed system the operator has sufficient cognitive capacity to perform the required tasks. However, OFS can vary from moment to moment in response to changing task demands in the context of the internal milieu of the operator. Therefore, the operator's current cognitive capabilities may not meet the momentary requirements of system operation. This leads to reduced operator effectiveness and errors.

The strong coupling between the cognitive demands placed on an operator and psychophysiological measures of OFS permits the rapid evaluation that is necessary for online assessment and adaptive aiding. These results suggest that psychophysiologically driven adaptive aiding will have application in operational environments. The task used here was complex, requiring visual search and decision making using specified rules of engagement, much like tasks in actual operational settings.

The inclusion of performance measures and task variables into the mitigation decision loop should improve the accuracy and utility of online operator functional state assessment and provide further enhancement of complex task performance (Bonner et al., 2000; Wilson & Russell, 1999). Actual operational settings will no doubt provide for several mitigations rather than only one, as used here.

Choosing the appropriate mitigation will require a much more complex mitigation manager that will be required to consider the many critical performance and scenario variables. It will also need to consider the individual needs of each operator based upon his or her unique skills and capabilities. Operators' preferences for which aspects of the task that they prefer to have mitigated will also be considered. It may be possible to use OFS and projected task demands to predict periods of future overload. This would permit the implementation of mitigations or changes to task requirements prior to the period of high task demand to ensure that the operator is capable of meeting these future challenges.

ACKNOWLEDGMENTS

REFERENCES

Apparies, R. J., Riniolo, T. C., & Porges, S. W. (1998). A psychophysiological investigation of the effects of driving longer-combination vehicles. *Ergonomics, 41,* 581–592.

Berka, C., Levendowski, D. J., Cvetinovic, M. M., Petrovic, M. M., Davis, G., Lumicao, M. N., et al. (2004). Real-time analysis of EEG indexes of alertness, cognition, and memory acquired with a wireless EEG headset. *International Journal of Human-Computer Interaction, 17,* 151–170.

Berka, C., Levendowski, D. J., Ramsey, C. K., Davis, G., Lumicao, M. N., Stanney, K., et al. (2005). Evaluation of an EEG-workload model in the Aegis simulation environment. In *Proceedings of SPIE* (Vol. 5797, pp. 90–99). Bellingham, WA: International Society for Optical Engineering (SPIE).

Bonner, M. C., Taylor, R. M., & Miller, C. A. (2000). Tasking interface manager: Affording pilot control of adaptive automation and aiding. In P. T. McCabe, M. A. Hanson, & S. A. Robertson (Eds.), *Contemporary ergonomics 2000* (pp. 70–74). London: Taylor & Francis.

Brookhuis, K. A., & de Waard, D. (1993). The use of psychophysiology to assess driver status. *Ergonomics, 36,* 1099–1110.

Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology, 42,* 361–377.

Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology, 42,* 249–268.

Caldwell, J. A., Caldwell, J. L., Brown, D. L., & Smith, J. K. (2004). The effects of 37 hours of continuous wakefulness on the physiological arousal, cognitive performance, self-reported mood, and simulator flight performance of F-117A pilots. *Military Psychology, 16,* 163–181.

Fairclough, S. H., Venables, L., & Tattersall, A. (2005). The influence of task demand and learning on the psychophysiological response. *International Journal of Psychophysiology, 56,* 171–184.

Fournier, L. R., Wilson, G. F., & Swain, C. R. (1999). Electrophysiological, behavioral, and subjective indexes of workload when performing multiple tasks: Manipulations of task difficulty and training. *Biological Psychology, 31,* 129–145.

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 task. *Biological Psychology, 50,* 61–76.

Gevins, A., & Smith, M. E. (2000). Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral Cortex, 10,* 829–839.

Gevins, A., & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science, 4,* 113–131.

Gevins, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., et al. (1998). Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human Factors, 40,* 79–91.

Gevins, A., Smith, M. E., McEvoy, L., & Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice. *Cerebral Cortex, 7,* 374–385.

Gonzalez, C. (2005). Task workload and cognitive abilities in dynamic decision making. *Human Factors, 47,* 92–101.

Hancock, P. A., & Verwey, W. B. (1997). Fatigue, workload, and adaptive automation. *Accident Analysis and Prevention, 29,* 495–506.

Hankins, T. C., & Wilson, G. F. (1998). A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. *Aviation, Space, and Environmental Medicine, 69,* 360–367.

He, P., Wilson, G. F., & Russell, C. (2004). Removal of ocular artifacts form electro-encephalogram by adaptive filtering. *Medical and Biological Engineering and Computing, 42,* 407–412.

Hedge, J. W., & Borman, W. C. (2006). Personnel selection. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (3rd ed., pp. 458–471). Hoboken, NJ: Wiley.

Kramer, A. F. (1991). Physiological measures of mental workload: A review of recent progress. In D. Damos (Ed.), *Multiple task performance* (pp. 279–238). London: Taylor & Francis.

Krizo, E. J., Wilson, G. F., & Russell, C. A. (2005). NuWAM: The development of a modular, real-time augmented cognition system. In *Proceedings of the 11th International Conference on Human-Computer Interaction: Vol. 11. Foundations of augmented cognition* [CD-ROM]. Mahwah, NJ: Erlbaum.

Moroney, W. F., Biers, D. W., & Eggemeier, F. T. (1995). Some measurement and methodological considerations in the application of subjective workload measurement techniques. *International Journal of Aviation Psychology, 5,* 87–106.

Parasuraman, R. (2003). Adaptive automation matched to human mental workload. In G. R. J. Hockey, A. W. K. Gaillard, & O. Burov (Eds.), *Operator functional state: The assessment and prediction of human performance degradation in complex tasks* (pp. 177–193). Amsterdam: IOS Press.

Parasuraman, R., Mouloua, M., & Hilburn, B. (1999). Adaptive aiding and adaptive task allocation enhance human-machine interaction. In M. W. Scerbo & M. Mouloua (Eds.), *Automation technology and human performance: Current research and trends* (pp. 119–123). Mahwah, NJ: Erlbaum.

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors, 38,* 665–679.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39,* 230–253.

Pleydell-Pearce, C. W., Whitecross, B. T., & Dickson, B. T. (2003). Multivariate analysis of EEG: Predicting cognition basis of frequency decomposition, inter-electrode correlation, coherence, cross phase and cross power. In R. H. Sprague (Ed.), *Proceedings of the 36th Annual Hawaii International Conference on Systems Sciences* [CD-ROM]. Los Alamitos, CA: IEEE Computer Society Press.

Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology, 40,* 187–195.

Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., & Pope, A. T. (2000). A closed-loop system for examining psychophysiological measures for adaptive task allocation. *International Journal of Aviation Psychology, 10,* 393–410.

Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., & Pope, A. T. (2003). Effects of a psychophysiological system for adaptive automation on performance, workload, and the event-related potential P300 component. *Human Factors, 45,* 601–613.

Rouse, W. B. (1988). Adaptive aiding for human/computer control. *Human Factors, 30,* 431–443.

Russell, C. A., & Wilson, G. F. (2005). Feature saliency analysis for operator state estimation. In *Proceedings of the 11th International Conference on Human-Computer Interaction: Vol. 11. Foundations of augmented cognition* [CD-ROM]. Mahwah, NJ: Erlbaum.

Scerbo, M. W. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* [CD-ROM]. Mahwah, NJ: Erlbaum.

Scerbo, M. W., Freeman, F. G., & Mikulka, P. J. (2003). A brain-based system for adaptive automation. *Theoretical Issues in Ergonomic Science, 4,* 200–219.

Smith, M. E., Gevins, A., Brown, H., Karnik, A., & Du, R. (2001). Monitoring task loading with multivariate EEG measures during complex forms of human-computer interaction. *Human Factors, 43,* 366–380.

Veltman, H. H. A., & Jansen, C. (2003). Differentiation of mental effort measures: Consequences for adaptive automation. In G. R. J. Hockey, A. W. K. Gaillard, & O. Burov (Eds.), *Operator functional state: The assessment and prediction of human performance degradation in complex tasks* (pp. 249–259). Amsterdam: IOS Press.

Wilson, G. F. (2001). In-flight psychophysiological monitoring. In F. Fahrenberg & M. Myrtek (Eds.), *Progress in ambulatory monitoring* (pp. 435–454). Seattle, WA: Hogrefe and Huber.

Wilson, G. F. (2002a). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *International Journal of Aviation Psychology, 12,* 3–18.

Wilson, G. F. (2002b). Psychophysiological test methods and procedures. In S. G. Charlton & T. G. O'Brien (Eds.), *Handbook of human factors testing and evaluation* (2nd ed., pp. 127–156). Mahwah, NJ: Erlbaum.

Wilson, G. F., & Eggemeier, F. T. (1991). Physiological measures of workload in multi-task environments. In D. Damos (Ed.), *Multiple-task performance* (pp. 329–360). London: Taylor & Francis.

Wilson, G. F., & Fisher, F. (1991). The use of cardiac and eye blink measures to determine flight segment in F4 crews. *Aviation, Space, and Environmental Medicine, 62,* 959–961.

Wilson, G. F., & Fisher, F. (1995). Cognitive task classification based upon topographic EEG data. *Biological Psychology, 40,* 239–250.

Wilson, G. F., & Russell, C. (1999). Operator functional state classification using neural networks with combined physiological and performance features. In *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting* (pp. 1099–1102). Santa Monica, CA: Human Factors and Ergonomics Society.

Wilson, G. F., & Russell, C. A. (2003a). Operator functional state classification using multiple psychophysiological features in an air traffic control task, *Human Factors*, *45,* 381–389.

Wilson, G. F., & Russell, C. A. (2003b). Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. *Human Factors, 45,* 635–643.

Wilson, G. F., & Russell, C. A. (2004). Psychophysiologically determined adaptive aiding in a simulated UCAV task. In D. A. Vicenzi, M. Mouloua, & P. A. Hancock (Eds.), *Human performance, situation awareness and automation: Current research and trends* (Vol. 2, pp. 200–204). Mahwah, NJ: Erlbaum.

Glenn F. Wilson is the chief scientist at Physiometrex, Springfield, OH. He received his Ph.D. in physiological psychology from the University of Arizona in 1968.

Christopher A. Russell is an electrical engineer at the Human Effectiveness Directorate, U.S. Air Force Research Laboratory at Wright-Patterson Air Force Base, OH. He received his Ph.D. in electrical engineering from the Air Force Institute of Technology in 2005.