REPORT DOCUMENTATION PAGE

Form Approved OMB No. 074-0188

1. REPORT DATE (DD-MMM-YYYY) 15 Mar 2004

2. REPORT TYPE Journal Article

3. DATES COVERED (From – To) January 2002 – June 2003

4. TITLE AND SUBTITLE

Real-Time Assessment of Mental Workload Using Psychophysiological Measures and Artificial Neural Networks

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER 62202F

5d. PROJECT NUMBER 7184

5e. TASK NUMBER 08

5f. WORKUNIT NUMBER 64

6. AUTHOR(S)

Glenn F. Wilson
Christopher A. Russell

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Interface Division
Air Force Materiel Command
Wright-Patterson AFB OH 45433-7022

10. SPONSOR / MONITOR'S ACRONYM

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

20040524 044

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The functional state of the human operator is critical to optimal system performance. Degraded states of operator functioning can lead to errors and overall suboptimal system performance. Accurate assessment of operator functional state is crucial to the successful implementation of an adaptive aiding system. One method of determining operators' functional state is by monitoring their physiology. In the present study, artificial neural networks using physiological signals were used to continuously monitor, in real time, the functional state of 7 participants while they performed the Multi-Attribute Task Battery with two levels of task difficulty. Six channels of brain electrical activity and eye, heart and respiration measures were evaluated on line. The accuracy of the classifier was determined to test its utility as an on-line measure of operator state. The mean classification accuracies were 85%, 82%, and 86% for the baseline, low task difficulty, and high task difficulty conditions, respectively.

15. SUBJECT TERMS

Electroencephalogram (EEG), Artificial Neural Networks (ANN), Electrocardiographic (ECG), Electrooculographic (EOG)

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON / 19b. TELEPHONE NUMBER (Include area code)
UNCLAS	UNCLAS	UNCLAS	UNLIMITED	9	Glenn F. Wilson / (937) 255-8748

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39-18

SPECIAL SECTION

Real-Time Assessment of Mental Workload Using Psychophysiological Measures and Artificial Neural Networks

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

The functional state of the human operator is critical to optimal system performance. Degraded states of operator functioning can lead to errors and overall suboptimal system performance. Accurate assessment of operator functional state is crucial to the successful implementation of an adaptive aiding system. One method of determining operators' functional state is by monitoring their physiology. In the present study, artificial neural networks using physiological signals were used to continuously monitor, in real time, the functional state of 7 participants while they performed the Multi-Attribute Task Battery with two levels of task difficulty. Six channels of brain electrical activity and eye, heart and respiration measures were evaluated on line. The accuracy of the classifier was determined to test its utility as an on-line measure of operator state. The mean classification accuracies were 85%, 82%, and 86% for the baseline, low task difficulty, and high task difficulty conditions, respectively. The high levels of accuracy suggest that these procedures can be used to provide accurate estimates of operator functional state that can be used to provide adaptive aiding. The relative contribution of each of the 43 psychophysiological features was also determined. Actual or potential applications of this research include test and evaluation and adaptive aiding implementation.

INTRODUCTION

Modern complex systems can place very high cognitive demands upon their operators. The rate of information flow, the complex nature of this information, and the number and rate of required decisions can overwhelm the human operator. At the other end of the continuum, automation of tasks can lead to operator complacency and errors of inattention (Billings, 1997). However, current systems are capable of modifying themselves to meet the momentary needs of the operator. This includes assuming some task functions until the operator's mental load is reduced. In other cases, systems can adjust to improve the operator's awareness to relieve boredom or inattention. Adaptive aiding based on the current functional state of the operator can be most beneficial when supplied at the appropriate time and with the consent of the operator (Rouse, 1988). Further, accurate assessment of operator functional state is required in the test and evaluation of new and modified systems (Charlton & O'Brien, 2002). In these situations the critical factor is the accurate and reliable assessment of the operator's functional state. The functional state of an operator is defined as his or her ability to carry out the job at that moment in time.

One method of monitoring operator functional state is by examining the operator's physiology. The various physiological measures provide unique information about several aspects of operator state. Eye blink rate contains valuable information with regard to the visual demands of tasks. Heart rate is useful to determine the operator's global response to task demands (Wilson & Eggemeier, 1991). The electroencephalogram (EEG) provides useful information about both

Address correspondence to Glenn F. Wilson, AFRL/HECP, 2255 H St., Wright-Patterson Air Force Base, OH 45433-7022; glenn.wilson@wpafb.af.mil. **HUMAN FACTORS,** Vol. 45, No. 4, Winter 2003, pp. 635–643.

high workload and inattention (Gundel & Wilson, 1992; Kramer, 1991; Sterman & Mann, 1995; Wilson & Eggemeier, 1991). EEG measures have been used to classify patients with regard to types of neuropathy and psychiatric disorders using linear statistical techniques (John, Pricep, Fridman, & Easton, 1988) and artificial neural networks (ANNs; Klöppel, 1994). EEG has also been used to classify drug effects and to detect alcohol intoxification and fatigue (Gevins & Smith, 1999; Herrmann, 1982). Physiological signals are always present and can be unobtrusively collected and, thereby, are able to provide uninterrupted information about operator state (Wilson, 2001, 2002).

Several studies have used psychophysiological measures to classify operator state with regard to mental workload. Most of these studies have employed EEG, cardiac, and eye data. Several of these studies used either simple, single-task paradigms (Gevins et al., 1998; Gevins & Smith, 1999; Nikolaev, Ivanitskii, & Ivanitskii, 1998; Wilson & Fisher, 1995) or relatively few peripheral nervous system variables in the context of complex task performance (Wilson & Fisher, 1991). Others have used complex tasks with skilled operators (Russell & Wilson, 1998; Russell, Wilson, & Monett, 1996; Wilson & Russell, 2003). These papers report overall successful task classification in the 80% to 90% correct range. The success rate of correctly classifying high mental workload or altered operator state is very encouraging. This suggests that these methods could be used to provide accurate and reliable operator functional state assessment during test and evaluation and to implement adaptive aiding systems. Hilburn, Jorna, Byrne, and Parasuraman (1997) used psychophysiological measures to show that adaptive aiding controlled by the task demands of their air traffic control task reduced mental workload. This demonstrates that psychophysiological measures of operator functional state change to show reduced mental workload when adaptive aiding is applied.

Psychophysiological measures have also been used to implement adaptive aiding in laboratory situations designed to detect lowered operator engagement in the task being performed (Freeman, Mikulka, Prinzel, & Scerbo, 1999; Freeman, Mikulka, Scerbo, Prinzel & Clouatre, 2000; Pope, Bogart, & Bartolome, 1995; Prinzel, Scerbo, Freeman, & Mikulka, 1995). These investigations demonstrated enhanced operator performance when the EEG-based adaptive aiding system detected operator disengagement or lowered attention and modified the task to increase operator involvement. Prinzel, Freeman, Scerbo, Mikulka, and Pope (2000) studied the effects of utilizing their engagement index when participants performed either one or three tasks. They reported improved performance with adaptive aiding, even though the index values did not differ between the two difficulty conditions.

Most contemporary systems, such as civil and military aircraft, are a complex combination of multiple tasks that can easily place demands upon operators that may exceed the operator's cognitive capabilities. This can result in errors and catastrophic performance breakdowns that can lead to system failure. In the case of mental overload, it may be possible to avoid system failure by reducing the task demands on the operator. Accurate estimation of the operator's functional state is crucial to successful implementation of such an adaptive aiding system (Byrne & Parasuraman, 1996; Scerbo, 1996).

In the present investigation psychophysiological signals were continuously monitored on line in order to determine the participant's functional state in real time. Further, this information was used to adapt the task when high levels of mental workload were detected in order to see if task performance would be enhanced or harmed. The goal of the present study was to determine the level of accuracy that an ANN could achieve in real time using psychophysiological variables to determine participants' level of mental workload while they performed a complex task. Previous work in our laboratory has demonstrated that very accurate levels of operator functional state assessment are possible using ANNs when the data are analyzed off line using a different task (Wilson & Russell, 2003). Further, the relative contribution of EEG and peripheral nervous system measures was determined. In addition, saliency analysis was performed on all of the EEG and peripheral measures (Ruck, Rogers, & Kabrisky, 1990). This type of analysis permits one to interrogate the trained ANN to determine which of the input features provide the most relevant information to the classifier solution. This information can be used to provide a better understanding of the underlying dynamics in the data.

METHODS

Seven participants (4 women, 3 men) took part in the experiment. Their age range was from 19 to 26 years. They were trained to stable performance on the NASA Multi-Attribute Task Battery (MATB; Comstock & Arnegard, 1992). After initial familiarization with the task, they were taught to manipulate a joystick with their right hand, which controlled the position of the tracking cursor, and to use their left hand to move a mouse, which controlled a pointing cursor on the screen. All of the MATB subtasks were used: lights and dials monitoring, manual tracking, resource management, and the auditory communication task. Two levels of task difficulty were provided and were manipulated by varying the number of events that occurred during each of the 5-min trials. In order to avoid confounding by learning, performance scores from each task were recorded and practice was continued until each participant exhibited stable performance on all tasks. Stable performance was defined as level performance scores over successive trials. This required approximately 6 hr of practice spread over 3 days.

Physiological data were recorded during task performance on the 4th day and consisted of six EEG channels as well as electrocardiographic (ECG), electrooculographic (EOG), and respiration inputs. EEG electrodes were placed on the scalp at Fz, F7, T4, T5, Pz, and Oz sites of the 10-20 system. Electrodes placed on the mastoids served as reference and ground. Horizontal and vertical EOG signals were recorded from electrodes placed by the outer canthus of each eye and above and below the midline of the right eye, respectively. Grass P511 amplifiers were used to amplify and filter the signals with a band pass of 0.3 to 30 Hz for the EEG and EOG and a band pass of 10 to 30 Hz for the ECG. The respiration was recorded with a Respitrace system. For the ECG signals, R-wave peaks were detected on line and interbeat intervals were calculated. The EOG signals were evaluated by laboratory-developed software that detected blinks and provided interblink intervals. The respiration signal was used to derive interbreath intervals using a zero crossing algorithm.

On the day of data collection, Day 4, the participants practiced the tasks for 5 min prior to data collection. Then three 5-min long conditions were presented to the participants. One was a baseline condition, during which the participants merely looked at the static MATB screen. The second condition required them to perform the task at the low difficulty level. In the third condition the task was presented at the high difficulty level.

The psychophysiological data from these three conditions were input to a multiple-layer perceptron ANN classifier using backpropagation. The ANN contained three layers: an input layer, a hidden layer, and the output layer. The input and hidden layers consisted of 43 nodes representing the EEG features plus the peripheral features. The output layer consisted of three nodes representing baseline, low, and high. The ANN was trained to recognize these three conditions separately for each participant. The input to the ANN consisted of the log power of spectral EEG and EOG features, which were derived by the fast Fourier transform. The five bands included delta (1–3 Hz), theta (4–7 Hz), alpha (8–13 Hz), beta (14–30 Hz), and gamma (31–42 Hz). The low-pass filters used on the Grass P511 amplifiers are analog and pass frequencies are reduced in magnitude above 30 Hz, thereby passing some gamma band activity. Other features included ECG interbeat, EOG interblink, and respiration intervals. The 43 input features to the ANN consisted of six EEG channels and two EOG channels with five bands each, plus the three peripheral interval measures.

The data were segmented into 10-s windows with a 50% overlap. Of the 10-s segments from each of the three conditions, 75% were randomly selected and used as training data. The remaining 25% were used as test data to determine the accuracy of the ANN training. After the ANN training reached the sum squared error of .04, which usually required fewer than 10 000 passes through the data, the remaining 25% of the data were then used to test the accuracy of the classifier. These data were evaluated with the trained ANN coefficients to determine if the ANN would place the data segments in the correct class of baseline, low, or high. Using the trained ANN, the level of mental workload was determined on line to be one of the three conditions (baseline, low, or high). This was accomplished based entirely on psychophysiological data. The block of three conditions used for training was repeated twice. During both replications, on-line determination of participant mental workload level was performed every 5 s using the ANN weights derived from the training session. The output node with the largest value determined which of the three conditions the operator was in at that moment. The workload classifications were recorded to determine the accuracy of the trained ANN.

The number of correctly classified 5-s epochs during the 5-min task performance was used to determine classifier accuracy. In the other replication, adaptive aiding was applied such that when the high workload condition was detected by the trained ANN, the MATB task was adapted by "turning off" two of the subtasks. During adaptive aiding, the lights and dials monitoring and the communication tasks were turned off, and their areas on the screen were highlighted in blue to indicate that an aiding period was in progress. The participants were instructed to ignore these tasks and concentrate their efforts on the tracking and resource management tasks. They were given practice with the aiding by ignoring these tasks. The order of presentation of the classification and aiding runs was alternated across participants.

In order to assess the relative contributions of the EEG and peripheral measures, off-line analyses were performed separately on these data. The EEG and the heart, eye, and respiration rates were separated into two data sets and were individually used to train the separate ANNs. The same procedures used for training and testing the ANN with the full data set were used with these data. An additional saliency analysis was carried out to determine which of the features contributed the most information or were the most salient to the ANN. The Ruck et al. (1990) saliency measure was used to determine the relative importance of each feature to the overall solution. A partial derivative analysis was performed on the fully trained ANN, and the importance of each feature was rank ordered in the final solution. The saliency values for each participant were normalized so that their sum equaled 1.0. The results of the saliency analysis were examined via visual inspection to determine the "break point" for each participant's data – that is, the point where the saliency values noticeably changed, by showing a marked decrease, was used as the break point. The features above this point were designated as the most salient or important features. In order to determine the effects of feature reduction, a separate set of ANN analyses was completed using only the salient features.

Tracking task root mean square (RMS) error and resource management error scores were recorded so that the effects of task difficulty and the adaptive aiding could be evaluated. After performing each condition, the participants were asked to provide subjective estimates of their mental workload using an 11-point scale (0–10), with 10 representing *very high workload.*

RESULTS

Analysis of the performance data showed that the RMS error of the two difficulty levels of the tracking task and resource management task error were significantly different: tracking task, mean low = 12.4 versus mean high = 59.9, $t(6) = 1.46$, $p < .00007$; resource management task, mean low = 42.5 versus mean high = 51.0, $t(6) = 2.63$, $p < .023$. Subjective reports of overall task difficulty for the low- and high-difficulty conditions showed that the participants perceived them as different: mean low = 2.7 versus mean high = 8.3, $t(6) = 8.18$, $p < .0002$. The accuracy of the trained ANN was first tested by having it classify the withheld 25% of the training data set. These were the 25% of the data not seen by the ANN during training. The mean ANN accuracy for the training data was 98.5% correct. We have previously found this almost-perfect accuracy of classifying the test data set (Wilson & Russell, 2003). The levels of classification accuracy were very high during the test run, in which the trained ANN was used for on-line classification of the workload while the participants performed the three task difficulty levels. The mean accuracies were 84.9% for the baseline condition, 82.0% for the low-workload condition, and 86.0% for the high-workload condition (see Table 1).

These results demonstrate that an ANN can produce high levels of correct classification while participants perform complex multiple tasks.

TABLE 1: Mean Percentage Classification Accuracy

	Base	Low	High
Base	84.9	14.1	1.0
Low	14.5	82.0	3.6
High	2.6	11.3	86.0

Note. Rows are truth, columns are test.

Note the pattern of confusion by the ANN of adjacent workload conditions when misclassifications did occur. During the baseline condition, the majority of errors (14.1%) were assigned to the low condition, with only 1% misclassified as high. The errors during the low condition were primarily confusion with the baseline condition (14.5%), and most of the errors for the high condition (11.3%) were misses categorized as belonging to the low condition, with only 2.6% misclassified as baseline.

Classification accuracies for each participant showed a mean correct classification range from 69.0% to 97.8% (see Table 2). The highest accuracy for any one condition was 100%, which occurred in 4 of the 21 comparisons, and 12 of the comparisons were in the 90% range. All of the observed accuracies were well above the expected chance level of 33%. Exceptions to the very high classification levels were conditions for Participants 3, 4, and 7; each had one condition that was classified with a low percentage correct of 28.7%, 41.3%, and 41.0%, respectively. The next-lowest accuracy was 76.7%, with most of the estimates being in the 80% to 100% correct range.

The results of the off-line analysis using only the EEG data are shown in Table 3. The mean correct classification accuracy was 87.2% for the three conditions, with a mean of 85.0% for the baseline condition, 87.4% for the low condition, and 89.2% for the high condition. As was the case with the entire data set ANN, the misclassification results showed the closest neighbor receiving the highest percentage of incorrect classifications.

Using only the three peripheral measures, the overall accuracy dropped to 55.9% (see Table 4). The correct classifications for the baseline, low-workload, and high-workload conditions were 59.1%, 64.9%, and 43.8%, respectively. Because separate analyses of the EEG and peripheral features were performed off line, the entire data set was also used to train an ANN in order to provide comparison with the original on-line analysis. These results are shown in Table 5. The results of the off-line analysis are very similar to the on-line results. The mean correct classifications for the baseline, low-workload, and high-workload conditions were 86.2%, 89.6%, and 86.5%, respectively. The mean correct classification was 87.4%, compared with 84.3% found with the on-line analysis.

Table 6 shows the salient input features for each participant and rank ordered across the 7 participants based on the saliency analysis. This table can be used to show which features were the most important for each participant as well as across the 7 participants. For example, theta band EEG activity from the Fz electrode contributed the most information for the ANNs of all of the features. This was followed by F7 theta, Pz theta, and vertical electro-oculographic (VEOG) theta band activity. The mean number of salient features was 14.9. The number of salient features for Participants 1 through 7 was 17, 17, 12, 17, 28, 8, and 5, respectively.

The final analysis used only the salient features to train ANNs for each participant using combined EEG and peripheral features. This was accomplished to determine the accuracy of the off-line-trained ANN using only the salient features, compared with using all of the available

TABLE 2: Percentage Correct Classification Accuracy for the Three Conditions

	P1	P2	P3	P4	P5	P6	P7	Mean
Base	100.0	85.0	85.0	93.3	100.0	90.0	41.0	84.9
Low	95.0	86.7	28.7	76.7	91.7	100.0	95.0	82.0
High	98.3	81.7	93.3	41.3	93.3	100.0	94.3	86.0
Mean	97.8	84.4	69.0	70.4	95.0	96.7	76.8	84.3

Note. Rows are truth, columns are test.

TABLE 3: Mean Percentage Classification Scores Using Only the EEG Data

	Base	Low	High
Base	85.0	13.6	1.5
Low	8.7	87.4	3.9
High	1.0	9.9	89.2

Note. Rows are truth, columns are test.

TABLE 4: Mean Percentage Classification Scores Using Only the Peripheral Measures

	Base	Low	High
Base	59.1	37.3	3.6
Low	24.7	64.9	1.4
High	13.8	42.4	43.8

Note. Rows are truth, columns are test.

TABLE 5: Mean Percentage Classification Scores Using the Combined EEG and Peripheral Measures

	Base	Low	High
Base	86.2	11.6	2.2
Low	8.5	89.6	1.9
High	3.2	1.3	86.5

Note. Rows are truth, columns are test.

features for the on-line analysis. Only the salient features for each participant were used to train ANNs using the procedures outlined earlier. The results of this analysis showed an overall correct classification accuracy of 88.0%. The mean accuracy was 91.0% for baseline, 85.2% for low, and 88.7% for high (Table 7).

During adaptive aiding, the participants' performance on the tracking and resource management tasks was monitored. By removing the monitoring and communication tasks when the classifier determined high workload, the participants were free to focus their efforts on the remaining two tasks. Adaptive aiding resulted in a 44% reduction in RMS tracking error, $t(6) = -6.134$, $p < .0008$, compared with the nonadaptive condition. Performance on the resource management task improved with a 33% reduction in the error score that was marginally significant, $t(6) = -1.822$, $p < .06$.

DISCUSSION

These results demonstrate that an ANN using central and peripheral nervous system features can be trained to very accurately determine, on line, the functional state of an operator. This is especially significant in light of the complex multiple task that was performed by the participants. All four subtasks of the MATB were performed during both the low and high difficulty levels of task demand. Only the density of stimulus and response events was changed. The mean correct classification accuracy across participants for the three task conditions during the on-line classification ranged from 82.0% to 86.0%. These results are consistent with previous reports and demonstrate the high levels of accuracy that are possible using ANNs (Gevins & Smith, 1999; Russell et al., 1996; Wilson & Russell, 2003). These results were derived from ANNs using both central and peripheral nervous system features. The EEG-only analysis produced classification accuracies that were essentially identical to the overall accuracy of the on-line results. The analysis using the peripheral measures alone did not show very high levels of correct classification. The off-line analysis, which included both the EEG and peripheral measures, showed the same accuracies as the EEG-only analysis.

Because there are a greater number of EEG features, they probably contain more information relevant to the functional state of the participants than do the three peripheral features. This is not surprising, given that six electrodes placed over widespread scalp sites were used and their electrical activity was divided into five different frequency bands. In our analysis, only interval information was used from the three peripheral measures. Further, spectral analysis of the VEOG and horizontal electro-oculographic (HEOG) channels were included with the EEG features. One of the VEOG bands was among the five most salient features. HEOG and the interbeat and interbreath intervals were in the top third of the salient features. In order to determine the contribution of the peripheral interval data, they should be tested in other task situations. This may be especially true if the classifier results are to be used in the test and evaluation of systems and to implement adaptive aiding.

TABLE 6: Ranked Saliency Results from Highest to Lowest across All Participants

Feature	P1	P2	P3	P4	P5	P6	P7	Totals
FZ theta	.00	.1329827	.1120869	.00	.0285860	.1378697	.2620487	.6735741
F7 theta	.00	.00	.0857638	.00	.0420032	.2306649	.1873129	.5457449
PZ theta	.1049727	.1705432	.00	.0488417	.0553200	.00	.00	.3796776
VEOG theta	.0790230	.00	.0813453	.00	.0382658	.1599230	.00	.3585571
T5 theta	.0776711	.00	.1313170	.0474553	.0469925	.00	.00	.3034359
HEOG beta	.00	.0853052	.00	.00	.00	.00	.2166377	.3019429
Interbeat	.0746387	.1297913	.00	.0580571	.0366611	.00	.00	.2991481
FZ alpha	.0692260	.00	.0825001	.1070838	.0320736	.00	.00	.2908835
O2 theta	.0756671	.00	.0944248	.0621385	.0556574	.00	.00	.2878878
O2 delta	.00	.1673464	.1040000	.00	.0144115	.00	.00	.2857578
PZ delta	.00	.00	.00	.0192943	.00	.00	.2470484	.2663427
HEOG theta	.00	.00	.00	.0996785	.0476519	.0932338	.00	.2405642
FZ delta	.00	.00	.0559547	.00	.0380030	.1390996	.00	.2330574
Interbreath	.0710244	.1313430	.00	.00	.0235684	.00	.00	.2259357
F7 delta	.1500198	.00	.00	.00	.0450658	.00	.00	.1950857
O2 alpha	.00	.00	.00	.1023400	.00	.00	.0869522	.1892922
VEOG delta	.0896979	.00	.00	.0489865	.0436826	.00	.00	.1823670
T5 delta	.00	.00	.00	.0492989	.0318804	.0794221	.00	.1606014
HEOG delta	.0582251	.00	.00	.00	.0214459	.0753290	.00	.1550001
HEOG alpha	.00	.00	.0747617	.00	.0518247	.00	.00	.1265863
O2 beta	.00	.00	.0651665	.00	.0579911	.00	.00	.1231576
T5 gamma	.00	.00	.00	.00	.0284830	.0844578	.00	.1129408
T4 beta	.00	.0871344	.00	.00	.0189423	.00	.00	.1060767
F7 gamma	.00	.0955537	.00	.00	.00	.00	.00	.0955537
F7 beta	.0897480	.00	.00	.00	.00	.00	.00	.0897480
VEOG gamma	.00	.00	.0375791	.0484944	.00	.00	.00	.0860735
F7 alpha	.00	.00	.0751001	.00	.00	.00	.00	.0751001
VEOG alpha	.00	.00	.00	.0701445	.00	.00	.00	.0701445
T4 gamma	.00	.00	.00	.0401841	.0221903	.00	.00	.0623744
FZ gamma	.00	.00	.00	.0602198	.00	.00	.00	.0602198
FZ beta	.0600862	.00	.00	.00	.00	.00	.00	.0600862
T4 theta	.00	.00	.00	.0545696	.00	.00	.00	.0545696
T5 alpha	.00	.00	.00	.0432652	.00	.00	.00	.0432652
HEOG gamma	.00	.00	.00	.00	.0420444	.00	.00	.0420444
PZ alpha	.00	.00	.00	.0399481	.00	.00	.00	.0399481
O2 gamma	.00	.00	.00	.00	.0345670	.00	.00	.0345670
T4 alpha	.00	.00	.00	.00	.0321225	.00	.00	.0321225
PZ gamma	.00	.00	.00	.00	.0320942	.00	.00	.0320942
VEOG beta	.00	.00	.00	.00	.0302448	.00	.00	.0302448
Interblink	.00	.00	.00	.00	.0289569	.00	.00	.0289569
T5 beta	.00	.00	.00	.00	.0192694	.00	.00	.0192694
T4 delta	.00	.00	.00	.00	.00	.00	.00	.00
PZ beta	.00	.00	.00	.00	.00	.00	.00	.00

Very high accuracies will be required to assure acceptance by users, and the peripheral data may improve the overall accuracies in some situations. The operator functional state assessment must be very accurate and reliable in order to gain the confidence of the operators who will depend on the classifiers.

The results of the analysis using only the salient features did show an approximately 4% benefit over the on-line analysis, as has been previously reported (Wilson & Russell, 2003). The overall accuracy scores were essentially the same as those when all of the data were used off line to train the ANN. This is interesting because fewer features were used in this investigation than in the Wilson and Russell (2003) study, in which 17 EEG channels with five frequency bands and three peripheral features provided a

TABLE 7: Mean Percentage Classification Accuracy Using Only the Salient Features

	Base	Low	High
Base	90.1	9.0	1.0
Low	11.4	85.2	3.4
High	3.0	8.4	88.7

Note. Rows are truth, columns are test.

total of 88 features. The EEG electrode positions used in the present study were based on the saliency analysis of the earlier work.

ANN classification accuracy improvement has been achieved, during off-line analysis, with the addition of performance features to the psychophysiological data (Wilson & Russell, 1999). However, many modern systems do not provide adequate performance data to augment psychophysiological data because few operator responses are required (Kramer, Trejo, & Humphrey, 1996). The value of performance data to on-line ANN operator functional state assessment is yet to be determined. However, the present results demonstrate that very high levels of correct classification can be achieved using only the psychophysiological features.

The utilization of operator state information to govern the application of adaptive aiding is also interesting. Lowering task demands based on operator state resulted in large improvements in performance. The tracking task error was reduced by 44%, and the resource management error was reduced by 33%. Reduction of overall task demands by temporarily removing the burden of the monitoring and communication tasks, based on the physiologically determined operator state, freed the participants to concentrate on the two remaining tasks and greatly improve their performance.

The effect of removing the two subtasks without regard to the participant's state remains to be determined. Randomly removing these subtasks could also have improved performance. However, it is possible that there would have been no change or even degraded performance from the random removal of the two subtasks. Random removal might interfere with the participant's strategy and lead to deteriorated performance. This question will have to be determined by further research.

Acceptance of psychophysiologically determined operator functional state assessment in the workplace will be based to a large extent on the accuracy of the classification and acceptability of the data collection methods. This requires that the operator functional assessment methods must be highly accurate. If the assessment is not highly accurate and reliable, then it will not be used. The accuracy may have to approach 95% to be acceptable (Rouse, 1991). Even if the psychophysiological assessment does not meet the 95% criteria, it still may be a useful component of a procedure that incorporates other aspects of system and operator variables.

Another issue has to do with the day-to-day reliability of the measures and the effects of other factors such as illness, drugs, fatigue, and circadian shifts. Other considerations include whether or not it is necessary to establish an ANN or other type of classifier for each operator or if a generic solution can be found that would accommodate all operators. By developing an ANN for each operator, one takes advantage of the unique physiological response patterns of each person. This capability may outweigh advantages for a "one-size-fits-all" solution, which would have the benefit of rapid training or fine-tuning of the ANN.

The results of this study show that ANNs using psychophysiological measures can produce very high levels of correct classification in real time. These procedures show promise for use in applied settings where real-time operator functional state assessment is needed. This includes test and evaluation and adaptive aiding. Miniaturization of physiological recording equipment and computer hardware will make possible the development of small wearable assessment systems. Dry sensors with small telemetry units will eliminate the need for the operator to wear all of the recording and computing equipment.

REFERENCES

Billings, C. E. (1997). Some questions about advanced automation. In M. Moulaoua & J. M. Koonce (Eds.), *Human-automation interaction* (pp. 314–320). Mahwah, NJ: Erlbaum.

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

Charlton, S. G., & O'Brien, T. G. (Eds.). (2002). *Handbook of human factors testing and evaluation.* Mahwah, NJ: Erlbaum.

Comstock, J. R., & Arnegard, R. J. (1992). *The Multi-Attribute Task Battery for human operator workload and strategic behavior research* (NASA Tech. Memorandum No. 104174). Hampton, VA: NASA.

Freeman, F. G., Mikulka, P. J., Prinzel, L. J., & Clouatre, K. (2000). Evaluation of a psychophysiologically controlled adaptive automation system, using performance on a tracking task. *Applied Psychophysiology and Biofeedback, 25,* 103–115.

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. (1999). Detecting transient cognitive impairment with EEG pattern recognition methods. *Aviation, Space, and Environmental Medicine, 70,* 1018–1024.

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

Gundel, A., & Wilson, G. F. (1992). Topographical changes in the ongoing EEG related to the difficulty of mental task. *Brain Topography, 5,* 17–25.

Herrmann, W. M. (1982). Development and critical evaluation of an objective procedure for the electroencephalographic classification of psychotropic drugs. In W. M. Herrmann (Ed.), *EEG in drug research* (pp. 249–351). New York: Gustav Fischer.

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

John, E. R., Pricep, L. S., Fridman, J., & Easton, P. (1988). Neurometrics: Computer-assisted differential diagnosis of brain dysfunctions. *Science, 239,* 162–169.

Klöppel, B. (1994). Application of neural networks for EEG analysis: Considerations and first results. *Neuropsychobiology, 29,* 39–46.

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.

Kramer, A. F., Trejo, L. J., & Humphrey, D. G. (1996). In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 137–162). Mahwah, NJ: Erlbaum.

Nikolaev, A. R., Ivanitskii, G. A., & Ivanitskii, A. M. (1998). Reproducible EEG alpha-patterns in psychological task solving. *Human Physiology, 24,* 261–268.

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., Scerbo, M. W., Freeman, F. G., & Mikulka, P. J. (1995). A bio-cybernetic system for adaptive automation. In *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting* (pp. 1365-1369). Santa Monica, CA: Human Factors and Ergonomics Society.

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

Rouse, W. B. (1991). *Design for success: A human centered approach to designing successful products and systems.* New York: Wiley.

Ruck, D., Rogers, S. & Kabrisky, M. (1990). Feature selection using a multilayer perceptron. *Journal of Neural Network Computing, 2,* 40–48.

Russell, C. A., & Wilson, G. F. (1998). Air traffic controller functional state classification using neural networks. In *Proceedings of the Artificial Neural Networks in Engineering (ANNIE'98) Conference* (Vol. 8, pp. 649–654). New York: American Society of Mechanical Engineers.

Russell, C. A., Wilson G. F., & Monett, C. T. (1996). Mental workload classification using a backpropagation neural network. In C. H. Dagli, M. Akay, C. L. P. Chen, B. R. Fernandez, & J. Ghosh (Eds.), *Intelligent engineering systems through artificial neural networks* (Vol. 6, pp. 685–690). New York: American Society of Mechanical Engineers.

Scerbo, M. W. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 37–63). Mahwah, NJ: Erlbaum.

Sterman, M. B., & Mann, C. A. (1995). Concepts and applications of EEG analysis in aviation performance evaluation. *Biological Psychology, 40,* 115–130.

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

Wilson, G. F. (2002). Psychophysiological test methods and procedures. In S. G. Charlton & T. G. O'Brien (Eds.), *Handbook of human factors testing and evaluation* (pp. 157–180). 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 43rd Annual Meeting* (pp. 1099–1102). Santa Monica, CA: Human Factors and Ergonomics Society.

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

Glenn F. Wilson is a senior research psychologist at the U.S. Air Force Research Laboratory. He received his Ph.D. in physiological psychology from the University of Arizona in 1968.

Christopher A. Russell is an electrical engineer with the Human Effectiveness Directorate at the U.S. Air Force Research Laboratory, where he is pursuing his Ph.D. in electrical engineering at the Air Force Institute of Technology. He received his M.S. in electrical engineering in 1997 from Wright State University.

Date received: January 18, 2002
Date accepted: August 20, 2003