SPECIAL SECTION

Effects of a Psychophysiological System for Adaptive Automation on Performance, Workload, and the Event-Related Potential P300 Component

Lawrence J. Prinzel III, NASA Langley Research Center, Hampton, Virginia, **Frederick G. Freeman, Mark W. Scerbo,** and **Peter J. Mikulka,** Old Dominion University, Norfolk, Virginia, and **Alan T. Pope,** NASA Langley Research Center, Hampton, Virginia

The present study examined the effects of an electroencephalographic- (EEG-) based system for adaptive automation on tracking performance and workload. In addition, event-related potentials (ERPs) to a secondary task were derived to determine whether they would provide an additional degree of workload specificity. Participants were run in an adaptive automation condition, in which the system switched between manual and automatic task modes based on the value of each individual's own EEG engagement index; a yoked control condition; or another control group, in which task mode switches followed a random pattern. Adaptive automation improved performance and resulted in lower levels of workload. Further, the P300 component of the ERP paralleled the sensitivity to task demands of the performance and subjective measures across conditions. These results indicate that it is possible to improve performance with a psychophysiological adaptive automation system and that ERPs may provide an alternative means for distinguishing among levels of cognitive task demand in such systems. Actual or potential applications of this research include improved methods for assessing operator workload and performance.

INTRODUCTION

Recently, there has been much interest in adaptive automation or systems that can adapt to a user's changing requirements in real time. Further, changes in the state of the system can be initiated by either the operator or the system itself (Hancock & Chignell, 1987; Rouse, 1977; Scerbo, 2001; Scerbo, Freeman, & Mikulka, 2000). It has been argued that systems using adaptive automation can provide a better match between the level of operator workload at any moment and the level of automation (Parasuraman, Bahri, Deaton, Morrison, & Barnes, 1992).

Parasuraman et al. (1992) described several techniques that could be used to trigger changes among modes of automation in an adaptive system. For instance, changes among modes could be tied to the presence of specific tactical events that occur in the task environment. Alternatively, models of operator performance or workload could be used to drive the adaptive logic. Another approach would be to assess operator performance in real time and use deviations from acceptable ranges to invoke the automation.

The last method for implementing adaptive automation described by Parasuraman et al. (1992) uses psychophysiological measures to trigger changes among the modes of automation. There are several advantages for using psychophysiological indices in an adaptive system (Byrne & Parasuraman, 1996; Gomer, 1981; Scerbo, Freeman, Mikulka, Parasuraman, Di-Nocero, et al., 2001). First, psychophysiological

Address correspondence to Lawrence J. Prinzel III, NASA Langley Research Center, Mail Stop 152, Hampton, VA 23681-0001; l.j.prinzel@larc.nasa.gov. **HUMAN FACTORS,** Vol. 45, No. 4, Winter 2003, pp. 601–613.

measures can be obtained continuously and do not necessarily require an overt response from the operator. This is important because in many current systems the operator is often placed in a supervisory role and is rarely required to make any overt responses (e.g., button presses), even though he or she might be engaged in considerable cognitive activity. Also, psychophysiological measures may provide additional information beyond behavioral measures when the two measures are used together. Further, psychophysiological measures have the potential to reflect the state of the operator as well as those functional areas of the brain that are differentially active. Consequently, it is this approach to adaptive automation that is the focus of the present paper.

A Biocybernetic Adaptive System

Pope, Bogart, and Bartolome (1995) described a system in which changes between modes of automation were triggered by the operator's own electroencephalogram (EEG) signals. More specifically, Pope et al. used an index of engagement based on ratios of EEG power bands (alpha, beta, theta, etc.). The engagement index is based on the idea that the beta band reflects increases in arousal and attention, whereas the alpha and theta bands reflect decreases (Abarbanel, 1999). Moreover, the index captures the simultaneous changes in beta activity along with changes in alpha and/or theta activity in a single variable. Several investigators have reported that EEG power band ratios may be better at distinguishing among different levels of attention than is any single power band in and of itself (Lubar, 1991; Lubar, Swartwood, Swartwood, & O'Donnell, 1995; Streitberg, Röhmel, Herrmann, & Kubicki, 1987).

Pope et al. (1995) studied several different engagement indices from a variety of sites. The engagement indices were computed using a 40-s moving window that was updated every 2 s. Their participants performed the Multi-Attribute Task (MAT) battery (Comstock & Arnegard, 1991), a PC-based suite of activities that includes compensatory tracking, monitoring, and resource management tasks. All of the tasks remained in automatic mode except the tracking task, which shifted between automatic and manual modes.

Their participants performed the task under both positive and negative feedback conditions. Under negative feedback, when the slope of the index function derived from two successive windows was positive (reflecting an increase in engagement), the tracking task was switched to or maintained in automatic mode. Similarly, a negative slope between two successive windows (reflecting a decrease in engagement) caused the tracking task to be switched to or maintained in manual mode. Thus under negative feedback, the system was expected to oscillate between automatic and manual modes more frequently, thereby maintaining a stable level of engagement. The system operated in the opposite manner under positive feedback and was expected to migrate toward extreme values of the index and remain in each task mode for longer periods, resulting in fewer switches between modes. Their results showed that more switches occurred between automatic and manual modes under negative as opposed to positive feedback and, further, that the index beta/(alpha + theta) best discriminated between positive and negative feedback conditions.

In a subsequent study, Freeman, Mikulka, Prinzel, and Scerbo (1999) extended the findings of Pope et al. (1995) to determine whether the negative feedback condition, which was shown to stabilize engagement, might actually improve performance. In addition, they modified the way the system responded to changes in the engagement index. Instead of comparing slope values from successive windows, they recorded the engagement index for a brief period to establish a stable baseline and then used deviations from the absolute value of the baseline to trigger the switches between task modes. When Freeman et al. examined those periods when participants manually performed the tracking task, they found that performance was better under negative feedback than under positive feedback.

In another study, Prinzel, Freeman, Scerbo, Mikulka, and Pope (2000) examined the effects of task load on MAT performance. Under low-workload conditions, participants performed only the tracking portion of the MAT task and the monitoring and resource management tasks remained in automatic mode. Under high-workload conditions, participants had to simultaneously perform the monitoring, resource management, and tracking tasks. In both workload conditions, however, only the tracking task switched between automatic and manual modes. Prinzel et al. observed that tracking performance was better in the low-workload condition and, once again, was better under negative feedback than under positive feedback. Further, the tracking performance of these participants was also better than that of a control group who performed the same tasks but without the benefit of the biocybernetic adaptive system. In addition, measures of subjective workload confirmed that the low-workload condition was indeed rated lower than the high-workload condition, and overall workload ratings were lower for those using the biocybernetic system in either feedback condition as compared with those in the control group.

Taken together, the work of Pope et al. (1995), Freeman et al. (1999), and Prinzel et al. (2000) shows that it is possible to moderate an operator's task load with a psychophysiologically based adaptive system using a negative feedback contingency. It is important to note that although Freeman et al. have had success with a system driven by EEG, a system of this nature may be of limited utility. An EEG-based system responds to gross changes in overall levels of engagement. It may not respond differentially to types and levels of cognitive task demand. Several methods exist for implementing automation that differ with respect to the type of aiding and level of autonomy of action. Consequently, a more effective adaptive automation scheme may result from assessing operator requirements along more dimensions than simply overall engagement.

Types and Levels of Automation

Consider the two-dimensional model for selecting among types and levels of automation recently proposed by Parasuraman, Sheridan, and Wickens (2000). The first dimension follows a straightforward view of information processing and includes four stages: information acquisition, information analysis, decision selection, and action implementation. The second dimension addresses autonomy – that is, within each of these stages the degree or level of automation can vary. At the lowest levels of automation, the system would merely provide information or suggestions to the operator. At the highest levels, however, the system would be more autonomous, having the responsibility to carry out functions on its own. For different systems, the degree of automation within each stage could also differ.

According to this model, the system described by Freeman et al. (1999) would be considered an action implementation system with high autonomy. The system usurps or relinquishes control of the tracking task according to the value of the engagement index. Moreover, although the system informs the user of its actions, it executes those actions without any user intervention.

The variety of automated systems just described suggests that a more appropriate matching of system changes to operator requirements might be achieved with a different psychophysiological index that is potentially more diagnostic of different stages of information processing. One promising alternative index is the event-related potential (ERP).

Event-Related Potentials

The ERP is derived from continuous EEG recording and represents the response of the brain to a discrete stimulus event. Unlike the EEG, which reflects brain activity in the frequency domain, the ERP represents an averaging of the EEG that is time-locked to an event. The resultant waveform is a sequence of separate (but often temporally overlapping) components, which reflect neuronal activity linked to specific sensory and cognitive processing. ERP components can be characterized in several ways, but they are typically described by their polarity and order of occurrence (P1, N1, P2, N2, P3, etc.) or mean latency in milliseconds from event onset (e.g., N170, P360; Coles & Rugg, 1995; Kramer, 1991).

The ERP components can vary as a function of factors such as a participant's attention, strategy, intentions, expectancies, and decisions coupled with changes in task parameters, relevance, and instructions (Coles & Rugg, 1995; Kramer, 1991). More important, several reviews have documented the efficacy of the ERP as a reliable and diagnostic measure of mental workload (Donchin, Kramer, & Wickens, 1986;

Kramer, 1991; Parasuraman, 1990; Wilson & Eggemeier, 1991). Most notably, the P300 component, which reflects the availability and distribution of processing resources, has been found to vary reliably with changes in task load. For example, Isreal, Chesney, Wickens, and Donchin (1980) examined the ERP in response to primary and secondary task demands. The use of a secondary task technique is predicated on the assumption that any processing resources not used on the primary task will be devoted to the secondary task. Therefore as primary task demands increase, fewer resources are available to be directed toward the secondary task, resulting in a smaller P300 amplitude (Kramer, 1991; Strayer & Kramer, 1990; Wickens, Kramer, Vanasse, & Donchin, 1983).

The better specificity of ERPs as compared with EEG with respect to information processing demands makes them a potential psychophysiological candidate for implementing adaptive automation. In particular, the sensitivity of the P300 to differences in workload might provide a better trigger for invoking task changes in action implementation systems of the type described by Parasuraman et al. (2000).

The Present Study

The primary objective of the present study was to reexamine performance with an EEG-based adaptive automation system and to determine whether the derived ERPs to a secondary task would provide an additional degree of workload specificity beyond that offered by the EEG-based index of engagement. Participants were asked to perform the MAT tracking task while performing a secondary auditory oddball task used to derive the ERPs. Further, all participants provided subjective ratings of workload.

Participants were run in one of three conditions. In the adaptive automation condition, the system switched between manual and automatic task modes based on the value of each individual's own EEG engagement index. Each of these participants was yoked to another participant, who received the exact same sequence of manual and automatic task mode switches as did their adaptive automation counterparts. This yoked group was included to determine whether differences in performance or workload ratings were attributable to the biocybernetic system or the resultant pattern of task mode switches. Finally, a third group performed the tracking task using a random sequence of task mode switches. This control group was included to determine whether there were any unique characteristics of the task mode switches generated by participants in the adaptive automation condition that might also account for effects observed in the yoked control group. Participants in all conditions were run under both negative and positive feedback conditions.

It was expected that if the schedule of task mode changes generated by the negative feedback, adaptive automation participants benefited those in the yoked control condition as well, then tracking performance would be better and workload lower for *both* of these groups, as compared with the control group, whose schedule of switches was determined at random. However, if the pattern of task mode switches is uniquely optimized for those individuals in the negative feedback, adaptive automation condition, then the adaptive automation group should be expected to show better tracking performance and have lower indices of workload than both of the control groups, whose outcomes should not differ from each other.

METHOD

Participants

Thirty-six undergraduate and graduate students (14 women, 22 men) served as participants for this experiment. The ages of the participants ranged from 18 to 40 years. Participants were given monetary compensation or extra course credit for their voluntary participation.

Apparatus

EEG activity was recorded with an Electro-Cap International Lycra sensor cap consisting of 22 recessed tin electrodes arranged according to the international 10–20 system (Jasper, 1958). A NeuroScan SynAmps was used to record EEG signals. The low- and high-pass filters were set at 100 and 1 Hz, respectively. Four channels of the SynAmps amplifier were connected via an analog output board to a Biopac EEG100A Analog/Digital converter, which in turn was connected to a Macintosh Quadra. A LabVIEW Virtual Instrument (VI) on the

Macintosh calculated total EEG power in three bands: theta, alpha, and beta. The VI also calculated the engagement index and commanded the task mode changes through serial port connections to the task computer.

The Macintosh VI was connected to a WIN 486 DX personal computer that was used to run the MAT, to command the task mode changes, and to control the auditory oddball tone sequencing and gating. The oddball tones were generated from Coulbourn Instrument analog oscillator modules.

An Analog Edge joystick was used for the compensatory tracking task. The gain on the joystick was set to 60% of its maximum with a response band of 0.8 Hz.

EEG Recording and Analysis

The EEG was recorded from sites Pz, Cz, P3, and P4. The ground site was located midway between Fpz and Fz, and each site was referenced to the left mastoid. Eye movements were measured using horizontal and vertical electrooculogram electrodes. Each amplified EEG channel was digitized at a rate of 200 samples/s in a circular buffer array. These samples were taken from the buffer and placed into four vectors, one input per channel, with each vector containing 512 data points, resulting in 2.56 s of data per channel. Each vector overlapped the previous one by about 0.5 s. Each vector was smoothed using a standard Hanning windowing procedure in order to attenuate the frequency characteristics that are artificially introduced into the power spectrum by the abrupt onset and offset of epochs.

The power spectrum was computed using a fast Fourier transformation. Bin powers were combined to calculate total power (0–256 Hz) in three EEG bands (theta or 4–7 Hz, alpha or 8–12 Hz, beta or 13–22 Hz). Bin powers are the estimates of the power spectrum within bins between discrete Fourier frequencies from 0 to 22 Hz. Band powers were divided by total power to produce percentage power. The three frequency bands do not cover the entire spectrum, and the residual frequencies were ignored. The array of percentage power for the four sites across the three bands was used to compute the engagement index: 20 beta/(alpha + theta). The engagement index, beta/(alpha + theta), has been shown to be the most effective of several indices employed by Freeman et al. (1999) and Pope et al. (1995). In the present study, beta is multiplied by 20 to produce engagement index values that fall between 2 and 20 (higher values reflect higher levels of engagement).

An artifact rejection subroutine examined the amplitudes of each epoch from the four channels of digitized EEG and compared them with pretrial baseline tests taken with eyes open and eyes closed. A power spectral distribution was derived from this baseline. If the power in any channel exceeded the threshold by more than 25%, the epoch was marked as bad. These epochs were then ignored when computing the engagement index. The data record resulting from an epoch containing an artifact was marked when it was written to the data file so that it could be ignored in subsequent data analyses. Less than 1% of any participant's data file was ever rejected.

The index was first computed over a 20-s period and then updated every 2 s using a sliding 20-s window. This moving window process continued for the duration of the trial. At each epoch, the index was compared with the mean value determined during the 5-min baseline period for each participant. An EEG index above baseline (see Experimental Procedure section) was taken to indicate that the participant's engagement level was high, whereas an EEG index below baseline was taken to indicate that engagement level was low.

To determine ERPs to the oddball stimuli, the continuous EEG data file was analyzed off line. Ocular artifacts were removed using the eye movement reduction algorithm supplied by the Neuroscan software. The continuous EEG data file was transformed into an EEG epoch file of 500 points per data file. The epoch file was baseline corrected using the baseline derived between –100 to 0 ms from the onset of the signal. The ERP P300 component was classified according to the largest base-peak amplitude and latency within a preset window (275–750 ms; see Kramer, Trejo, & Humphrey, 1996).

Tasks

Tracking task. Participants performed a modified version of the NASA MAT battery (Comstock & Arnegard, 1991). Only the compensatory tracking task was used in the present study. In the manual mode, the task requires participants to use a joystick to maintain a moving circle, approximately 1 cm in diameter, centered on a 0.5 × 0.5 cm cross located in the center of the screen. The movement of the circle was driven by the sum of five out-of-phase sine waves. Failure to control the circle results in its drifting away from the center cross. Performance was evaluated in terms of deviations from center of the tracking target in pixel units and calculated as root mean square error (RMSE). The RMSE was measured every 2 s. In the automatic mode the circle remained constrained within a small rectangular area centered over the cross, and the participant was required only to monitor the display.

Auditory oddball task. The auditory oddball secondary task consisted of high and low tones at 1100 Hz and 900 Hz, respectively. The probability of a high tone was .10, and they were randomly distributed among the low tones over the course of a trial; however, the order of tones was held consistent across participants. Tones were presented once per second with a 200-ms duration and a 800-ms interstimulus interval. Therefore, over a 16-min trial there were 96 high tones and 864 low tones. The tones were gated to provide a linear rise and fall time of 10 ms and were presented to both ears through KOSS stereo headphones at 60 dB SPL. Participants were required to silently count the number of high tones and report the sum at the end of the trial.

Experimental Design

Each participant was randomly assigned to the adaptive automation (n = 12), yoked (n = 12), or control (n = 12) group. The adaptive automation condition required participants to perform the compensatory tracking task and auditory oddball task under both negative and positive feedback conditions. Each feedback condition was run for one 16-min trial. Half of the participants began with a negative feedback trial and the other half with positive feedback. The data records of switches between task modes generated by participants in the adaptive automation group were used to determine the schedule of task mode switches for participants in the yoked condition. Thus each yoked participant performed the tracking task under the exact same schedule of task mode switches as his or her automation counterpart. The control group performed the tracking task under a random schedule of task mode switches based on the average number of switches for participants in the adaptive automation group. All three groups experienced the same number of task mode allocations. The yoked participants received the same sequence of automation task mode allocation as did their adaptive partners, whereas the control participants received a random sequence. All participants, however, had the same sequence of high and low tones in the auditory oddball task. Thus the present study used a mixed factorial design with three groups (adaptive automation, yoked, and random) manipulated between subjects and feedback condition (negative and positive) manipulated within subjects.

Experimental Procedure

All participants were fitted with an EEG electrode cap. Conductive gel was injected into the electrodes, and impedance levels were brought below 5 kΩ. Participants were then instructed how to perform the compensatory tracking and auditory oddball tasks and then practiced the tracking task for 15 min. Once the participant had an understanding of these tasks, the EEG electrode cap was connected to the SynAmps headbox connector. Participants were then asked to sit quietly with their eyes open and then with their eyes closed for 5 min each. EEG data were collected during this time to establish the baseline index for each participant. The mean value of the engagement index for each participant during this time represented the baseline criterion for determining task allocations during the experimental session.

After the baseline data were gathered, participants were given a 5-min break before beginning the experimental session. For the adaptive automation group, the system made task mode changes based on the mean baseline EEG engagement index for that participant. Under the negative feedback condition, the task was switched to or remained in manual mode when the index was below the baseline index and switched to or remained in automatic mode when the index was above the baseline index.

Under the positive feedback condition, the task was switched to or remained in manual mode when the index was above the baseline index and switched to or remained in automatic mode when the index was below the baseline. The experimental session lasted approximately 60 min.

After each trial, the participants were asked to provide ratings of subjective workload using the NASA-Task Load Index (TLX; Hart & Staveland, 1988). Workload scores were calculated by summing ratings across six subscales. Total workload scores can range from 0 to 120, with higher scores reflecting higher levels of workload (Byers, Bittner, & Hill, 1989). Upon completion of the experimental session, all participants were debriefed.

RESULTS

Two simple ANOVAs were performed on the data from only the adaptive automation group for the number of mode switches under the feedback conditions (positive, negative) and the amount of time spent in each task mode (automated, manual). The analysis for the number of switches indicated that significantly more switches occurred with negative feedback ($M = 68.9$) than with positive feedback ($M = 50.8$), $F(1, 11) = 6.50$, $p < .05$. There was no statistical difference between the amounts of time spent in the automatic and manual task modes.

Tracking Performance

A 3 Groups (adaptive automation, yoked, control) × 2 Feedback Conditions mixed ANOVA for tracking RMSE revealed a significant main effect for groups, $F(2, 33) = 3.31$, $p < .05$. Post hoc Newman-Keuls tests indicated that the adaptive automation participants demonstrated better tracking performance ($M = 8.6$) than did those in either the yoked ($M = 11.1$) or control group ($M = 10.4$). The latter two groups were not significantly different. The effect of feedback was also significant, with better tracking under the negative ($M = 8.9$) than under the positive ($M = 11.1$) condition, $F(1, 33) = 9.01$, $p < .02$. Also, the interaction of Groups × Feedback was significant, $F(2, 33) = 4.84$, $p < .01$. Post hoc analysis indicated that participants in the negative feedback, adaptive automation group performed better than those in any other combination of group and feedback condition, which were not significantly different from one another (see Table 1). Given this pattern of results, the main effect findings are largely attributable to the small tracking error of the negative feedback, adaptive automation condition, as evidenced by the interaction effect.

Subjective Workload

A Groups × Feedback Condition ANOVA was performed on the total workload scores obtained with the NASA-TLX. This yielded a significant effect for groups, $F(2, 33) = 13.76$, $p < .01$ (see Table 2). Specifically, the adaptive automation participants ($M = 63.7$) reported significantly lower workload than did either the yoked ($M = 88.0$) or control ($M = 88.5$) participants. The effect of feedback condition was also significant, $F(1, 33) = 39.8$, $p < .001$, in which lower subjective workload was reported under negative feedback ($M = 72.5$) than under positive feedback ($M = 87.8$). More important, the interaction of Groups × Feedback Conditions was significant, $F(2, 33) = 27.7$, $p < .001$. A post hoc analysis indicated that the level of subjective workload reported by the adaptive automation participants under negative feedback was significantly lower ($M = 43.1$) than in any other combination of group and feedback condition. The workload scores for the remaining five conditions ranged from 83.8 to 92.2, and there were no significant differences among them.

TABLE 1: Means (and Standard Deviations) for the RMSE Tracking Scores for the Six Experimental Conditions

	Group		
Feedback	Adaptive Automation	Yoked	Control
Negative	5.8 (0.98)	10.7 (1.21)	10.2 (0.97)
Positive	11.3 (1.12)	11.4 (1.08)	10.6 (1.11)

TABLE 2: Means (and Standard Deviations) for the NASA-TLX Scores for the Six Experimental Conditions

	Group		
Feedback	Adaptive Automation	Yoked	Control
Negative	43.0 (13.5)	84.0 (17.44)	90 (22.06)
Positive	85.0 (14.25)	92.0 (24.05)	87 (18.38)

Auditory Oddball Task Performance

Another Groups × Feedback Condition ANOVA was performed on the number of oddball tones detected out of the 96 tones. This analysis yielded a significant interaction only of groups and feedback condition, $F(2, 33) = 5.25$, $p < .01$. Post hoc analysis indicated that the negative feedback, adaptive automation participants correctly identified more of the high tones (M = 94.3) than did participants in any other combination of group and feedback condition. The scores for the remaining five conditions ranged from a high of 87.3 for the yoked, negative feedback condition to a low of 83.3 for the adaptive automation, positive feedback condition, with no significant differences among them.

Event-Related Potentials: P300

A preliminary analysis was performed on the mean peak voltage of P300 ERP using the four electrode sites as an independent variable. This analysis found no significant effect for sites. Therefore, data were collapsed across electrode sites in the subsequent analyses.

A Groups × Feedback Condition ANOVA was performed on the amplitude of the P300 response as measured in microvolts. This analysis yielded a significant main effect for groups, $F(2, 33) = 20.40$, $p < .001$. A post hoc analysis showed a significantly larger P300 response for the adaptive automation participants (M = 3.08) than for either the yoked (M = 2.09) or the control participants (M = 2.14). The main effect for feedback conditions was also significant, $F(1, 33) = 78.7$, $p < .001$, with a larger amplitude response for the negative feedback condition (M = 2.93) than for the positive feedback condition (M = 1.94). Again, the interaction of Groups × Feedback Conditions was significant, $F(2, 33) = 57.2$, $p < .001$. A post hoc analysis indicated that the adaptive automation, negative feedback condition had a significantly larger P300 amplitude (M = 4.40) than did any other combination of group and condition. The remaining conditions did not differ significantly; their P300 amplitudes ranged from 1.75 to 2.20 (see Table 3).

Engagement Index

A Task Mode (manual, automatic) × Feedback Condition ANOVA was performed on the engagement index for the adaptive automation participants. In this analysis the interaction term represents a statistical manipulation check on whether the adaptive automation system functioned correctly. This analysis yielded only a significant Task Mode × Feedback Condition interaction, $F(1, 44) = 201.31$, $p < .0001$. The post hoc analysis indicated that under negative feedback, a higher engagement index (M = 11.4) was found in the automatic mode than in the

TABLE 3: Means (and Standard Deviations) for the P300 Amplitudes to the Auditory Oddball Task for the Six Experimental Conditions

	Group		
Feedback	Adaptive Automation	Yoked	Control
Negative	4.40 (1.02)	2.20 (1.07)	2.18 (0.86)
Positive	1.75 (0.76)	1.99 (0.57)	2.10 (0.84)

manual mode (M = 8.2), whereas the opposite pattern occurred under positive feedback, with a higher engagement index in the manual mode (M = 11.9) than in the automatic mode (M = 8.1). Under negative feedback, the system was designed to switch to the automatic mode when the engagement level was above baseline and to switch to manual when the engagement index level was below baseline. The opposite shift was programmed under positive feedback.

DISCUSSION

The present study sought to add to the few empirical studies that have examined the impact of adaptive automation on behavioral, subjective, and psychophysiological measures of workload and task engagement. Specifically, performance with an EEG-based adaptive automation system was examined along with derived ERPs to a secondary task to determine the degree of workload specificity that might be achieved above and beyond that offered by the EEG-based index of engagement. Toward that end, participants were assigned to an adaptive automation condition, in which the system switched between manual and automatic task modes based on the value of each individual's own EEG engagement index; a yoked control condition, in which each participant was yoked to another participant in the adaptive automation condition and received the exact same sequence of manual and automatic task mode switches; or a second control group, in which the schedule of task mode switches followed a random pattern. Each group performed one trial under positive feedback and a second trial under negative feedback.

Task Performance and Subjective Workload

The major finding was that participants in the adaptive automation condition performed significantly better under negative feedback than under positive feedback. Moreover, the task performance of those in the adaptive automation, negative feedback condition was better than that of participants in all other combinations of groups and conditions. In addition, those in the adaptive automation, negative feedback condition reported the lowest levels of subjective workload. Because adaptive automation systems are designed to provide a tighter coupling between momentary task demands and changes between task modes, these results show that performance can be enhanced and workload lowered when the manual and automatic task mode allocations are optimized. Because the EEG engagement index is posited to accurately reflect task load demands, the biocybernetic system under negative feedback modulates the task environment to guard against overload and underload attentional states. Under positive feedback, however, the system made task allocations that did not serve to maintain the stable levels of engagement. Rather, if the participants were experiencing low engagement and, for example, were in automatic task mode, the closed-loop system did not revert to manual control, as was done under negative feedback.

For the control and yoked conditions, the task allocations were either random or based on another participant's pattern of physiological response and, therefore, were not responsive to dynamic changes in the participant's engagement state. Thus, not surprisingly, the participants operating under adaptive automation and negative feedback rated subjective workload much lower than did those performing the tasks under the other, less-than-optimal task allocation conditions.

The findings for performance and subjective workload support the earlier observations of Freeman et al. (1999) and Prinzel et al. (2000), who observed better performance under negative than under positive feedback conditions with an EEG-based adaptive automation system. Further, the results involving subjective workload were consistent with those of Prinzel et al., who found lower workload ratings associated with adaptive automation under negative feedback, as compared with a nonadaptive control condition. In the present study, as well as in the previous studies, no statistical differences were found in the amount of time spent in automatic and manual modes, although a significantly greater number of task allocations of shorter duration have been observed consistently under negative feedback. Although research has shown that short periods of manual reallocation can be beneficial to task performance (e.g., Hadley, Prinzel, Freeman, & Mikulka, 1999; Scallen, Hancock, & Duley, 1995), the differences in number and pattern of task allocations witnessed in these studies were unlikely to afford a performance advantage. In the present study, for example, there was a statistical difference but not a practical difference in the number of task allocations between positive (50.83) and negative (68.92) feedback, which equates to a task allocation every 18 and 14 s, respectively. Taken together, therefore, the results from the present experiment lend additional support to the notion that operator performance can be enhanced with a negative feedback, adaptive automation system that couples psychophysiology and changes among task modes.

P300 Event-Related Potential

A second objective of the present study was to examine ERPs to a secondary oddball task in order to determine whether the derived ERPs would provide a degree of workload specificity beyond that provided by the EEG-based index of engagement. To elicit the ERP, participants were asked to perform an auditory oddball task, which has been used with success in previous studies (e.g., Fowler, 1994; Isreal et al., 1980; Kramer, Sirevaag, & Braune, 1987; Ritter, Vaughan, & Costa, 1968; Wickens, Isreal, & Donchin, 1977). The use of the oddball paradigm is predicated on the assumption that the P300 amplitude elicited from the oddball task would be proportional to the attentional resources invested in the task (Johnson, 1986; Parasuraman, 1990). Therefore, if mental workload was increased in the primary task, there should be an increase in P300 amplitude for the primary task but a corresponding decrease in amplitude for the secondary oddball task (Wickens et al., 1983). Because the adaptive system was designed to modulate workload demands through task allocations based on real-time psychophysiological measures, the negative feedback, adaptive condition was predicted to lower subjective workload and free up attentional resources to perform the secondary task, as compared with the other experimental conditions.

The results for P300 amplitude provide support for the hypothesis that adaptive automation not only can optimize the allocation of attentional resources to the primary task to improve primary task performance but also appears to free attentional resources to improve secondary task attention. The amplitude of the P300 response to the secondary task stimulus was significantly larger in the negative feedback, adaptive automation condition than in both the yoked and control conditions, suggesting that more attentional resources were available, allowing these participants to have significantly better oddball stimulus response.

It should be noted, however, that the P300 amplitudes observed in the present study are somewhat smaller than those typically seen in other studies. It is possible that the amplitudes may have been diminished by the baseline correction procedure used to generate the ERPs from the raw EEG. Another possible explanation for the small P300 amplitudes may be that the P300 was elicited to a secondary task (i.e., participants were asked to prioritize the primary tracking task). Therefore, the P300 may have been smaller than it would have been had it been elicited from the primary task. It is also possible that the secondary oddball task and the closed-loop system employing this particular EEG engagement index definition may have interacted to diminish P300 amplitudes. Several studies have reported that oddball target stimuli enhance EEG theta response and that theta-related processes are associated with, and may even modulate, the P300. This effect may be a reflection of similar responses to signals engaging attention and memory processes (e.g., Basar-Eroglu, Basar, Demiralp, & Schürmann, 1992; Demiralp & Basar, 1992; Intriligator & Polich, 1994; Stampfer & Basar, 1985; Yordanova & Kolev, 1998). It is important to note, however, that despite these possibilities the methodology used in the present experiment was effective in discriminating P300 amplitude differences between conditions.

It should also be noted that the functional and/or direct association of theta activity and the P300 has implications for the present study. Recall that the EEG engagement index uses theta in the denominator; consequently, under negative feedback the system is designed to react to increases in theta, as well as increases in alpha and decreases in beta, by imposing perceptual-motor demands that may function to counter theta production. This may have served to diminish theta response to oddball targets, resulting in a smaller P300 component than would have been generated if a different index of engagement had been used. The opposite is true for the positive feedback condition. The yoked and control conditions were not tied to the participant's EEG, and thus the P300 amplitude would not have been affected by theta production. Therefore, although it is possible that theta suppression, resulting from use of this particular EEG engagement index definition, may have diminished the amplitude of the P300, the fact remains that significant differences among conditions were found, demonstrating the capability of this form of adaptive automation to enhance P300 response. However, an implication of this interaction of theta and the P300 is that consideration of hybrid approaches utilizing both EEG and the P300 would require the appropriate use of theta as a criterion variable for adaptive aiding.

The ERP results from the present study have several implications for the design of adaptive systems. As Byrne and Parasuraman (1996) have noted, physiological measures can have a dual role in adaptive automation research by providing information on the mechanisms underlying human interaction with automation as well as contributing to the operation of an adaptive automation system. In both of these roles, the use of psychophysiology is based on the assumption that there exists an optimal state for human operators to perform specific tasks (e.g., Gaillard, 1993; Hockey, Coles, & Gaillard, 1986). Whenever a mismatch occurs between the operator's current state and optimal state, psychological and physiological stress may occur and may be detected through psychophysiological measures (Hancock & Warm, 1989; Hockey, 1986). Adaptive automation can be used to support a human operator's internal state (Hockey, 1986) and modulate the task environment when the operator is unable to maintain the match between current state and effective state (Byrne & Parasuraman, 1996).

The results of the present study suggest that the P300 can be useful for this dual role in adaptive automation (Byrne & Parasuraman, 1996), both as a measure of the effectiveness of aiding and as a potential trigger for invoking adaptive task allocation. Although other physiological measures can provide an index of an operator's state, the ERP is unique in providing a physiological index tracking the information processing stages involved with complex human-system interaction. Further, numerous studies have demonstrated the efficacy of the P300 as a diagnostic and sensitive measure of mental workload. The present study builds on this foundation by showing that the P300 can be used to provide a measure of attentional resource allocation under conditions of adaptive aiding, making the ERP valuable as a criterion for development of future adaptive automated systems.

The ERP may also have utility as a criterion variable for controlling task allocations between the human and machine. Although the design of the present study did not specifically use ERPs to drive the closed-loop biocybernetic system, the results suggest that the P300 could be a candidate for real-time adaptive aiding. The P300 differences were found between negative and positive feedback conditions and coupled and uncoupled schedules of task mode switches. Therefore, a logical next step would be to employ ERPs in the design of the closed-loop, biocybernetic system (Scerbo, Freeman, & Mikulka, 2003).

Although the results from the present study are encouraging, it is important to recognize that the ERP results were obtained through a process of averaging EEG records over many stimulus presentations subsequent to data collection. Thus any effort to use ERPs in a real-time adaptive system would have to overcome the difficulty of extracting ERPs from the EEG signal corresponding to a limited number of stimulus presentations. However, several studies have shown that ERPs can be measured with small samples of data. For example, Humphrey and Kramer (1994) reported that a successful classification of workload states could be achieved with as few as 11 s of ERP data. Moreover, they reported a 78% correct classification of single-trial ERPs using a linear stepwise discriminant analysis. Therefore, the requirement to average over numerous stimulus events to extract the ERP may no longer be a huge barrier to implementing ERPs as part of an adaptive system.

In conclusion, the results reported here support previous research showing the potential of adaptive automation to significantly improve operator performance and lower workload based on real-time adaptive task allocation driven by the operator's EEG. Furthermore, the data confirm the efficacy of the P300 as a useful measure of attentional resources in adaptive environments and point the way for future research on adaptive systems that use ERPs as criterion measures for invoking changes among task modes.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to the anonymous reviewers whose comments helped improve this manuscript.

REFERENCES

Abarbanel, A. (1999). The neural underpinnings of neurofeedback training. In J. R. Evans & A. Abarbanel (Eds.), *Introduction to quantitative EEG and neurofeedback* (pp. 311–340). San Diego: Academic.

Basar-Eroglu, C., Basar, E., Demiralp, T., & Schürmann, M. (1992). P300-response: Possible psychophysiolgical correlates in delta and theta frequency channels. *International Journal of Psychophysiology, 13,* 161–179.

Byers, J. C., Bittner, A. C., & Hill, S. G. (1989). Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In A. Mital (Ed.), *Advances in industrial ergonomics and safety* (pp. 481–485). London: Taylor & Francis.

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

Coles, M. G. H., & Rugg, M. D. (1995). Event-related brain potentials: An introduction. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of the mind: Event-related brain potentials and cognition* (pp. 1–26). Oxford, UK: Oxford University Press.

Comstock, J. R., & Arnegard, R. J. (1991). *The Multi-Attribute Task battery for human operator workload and strategic behavior research* (NASA Tech. Memorandum No. 104174). Hampton, VA: Langley Research Center.

Demiralp, T., & Basar, E. (1992). Theta rhythmicities following expected visual and auditory targets. *International Journal of Psychophysiology, 13,* 147–160.

Donchin, E., Kramer, A. F., & Wickens, C. D. (1986). Applications of brain event-related potentials to problems in engineering psychology. In M. G. H. Coles, E. Donchin, & S. Porges (Eds.), *Psychophysiology: Systems, processes, and applications* (pp. 702–718). Middletown, NJ: Till & Till.

Fowler, B. (1994). P300 as a measure of workload during a simulated aircraft landing task. *Human Factors, 36,* 670–683.

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

Gaillard, A. W. K. (1993). Comparing the concepts of mental load and stress. *Ergonomics, 36,* 991–1005.

Gomer, F. (1981). Physiological monitoring and the concept of adaptive systems. In J. Moraal & K. F. Kraiss (Eds.), *Manned systems design* (pp. 257–263). New York: Plenum.

Hadley, J., Prinzel, L. J., Freeman, F. G., & Mikulka, P. J. (1999). Behavioral, subjective, and psychophysiological correlates of various schedules of short-cycle automation. In M. Scerbo & M. Mouloua (Eds.), *Automation technology and human performance: Current research and trends* (pp. 139–143). Mahwah, NJ: Erlbaum.

Hancock, P. A., & Chignell, M. H. (1987). Adaptive control in human-machine systems. In P. A. Hancock (Ed.), *Human factors psychology* (pp. 305–345). Amsterdam: Elsevier Science.

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31,* 519–537.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam: North-Holland.

Hockey, G. R. J. (1986). A state control theory of adaptation and individual differences in stress management. In G. R. J. Hockey, A. W. K. Gaillard, & M. G. H. Coles (Eds.), *Energetics and human information processing* (pp. 285–298). Dordrecht, Netherlands: Kluwer.

Hockey, G. R. J., Coles, M. G. H., & Gaillard, A. W. K. (1986). Energetical issues in research on human information processing. In G. R. J. Hockey, A. W. K. Gaillard, & M. G. H. Coles (Eds.), *Energetics and human information processing* (pp. 3–21). Dordrecht, Netherlands: Kluwer.

Humphrey, D. G., & Kramer, A. (1994). Toward a psychophysiological assessment of dynamic changes in mental workload. *Human Factors, 36,* 3–26.

Intriligator, J., & Polich, J. (1994). On the relationship between background EEG and P300 event-related potential. *Biological Psychology, 37,* 207–218.

Isreal, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. (1980). P300 and tracking difficulty: Evidence for multiple resources in dual task performance. *Psychophysiology, 17,* 259–273.

Jasper, H. (1958). Report on the committee on methods of clinical examination in electroencephalography. *Electroencephalography and Clinical Neurophysiology, 10,* 370–375.

Johnson, R. (1986). A triarchic model of P300 amplitude. *Psychophysiology, 23,* 367–384.

Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. In D. L. Damos (Ed.), *Multiple-task performance* (pp. 279–328). London: Taylor & Francis.

Kramer, A. F., Sirevaag, E., & Braune, R. (1987). A psychophysiological assessment of operator workload during simulated flight missions. *Human Factors, 29,* 145–160.

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

Lubar, J. F. (1991). Discourse on the development of EEG diagnostics and biofeedback for attention-deficit/hyperactivity disorders. *Biofeedback and Self-Regulation, 16,* 201–225.

Lubar, J. F., Swartwood, M. O., Swartwood, J. N., & O'Donnell, P. H. (1995). Evaluation of the effectiveness of EEG neurofeedback training for ADHD in a clinical setting as measured by changes in T.O.V.A. scores, behavioral ratings, and WISC-R performance. *Biofeedback and Self-Regulation, 20,* 83–99.

Parasuraman, R. (1990). Event-related brain potentials and human factors research. In J. W. Rohrbaugh, R. Parasuraman, & R. Johnson, Jr. (Eds.), *Event-related brain potentials: Basic issues and applications* (pp. 279–300). New York: Oxford University Press.

Parasuraman, R., Bahri, T., Deaton, J. E., Morrison, J. G., & Barnes, M. (1992). *Theory and design of adaptive automation in adaptive systems* (Progress Report No. NAWCADWAR-92033-60). Warminster, PA: Naval Air Warfare Center, Aircraft Division.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics, 30,* 286–297.

Pope, A. T., Bogart, E. H., & Bartolome, D. (1995). Biocybernetic system evaluates indices of operator engagement. *Biological Psychology, 40,* 187–196.

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 automation. *International Journal of Aviation Psychology, 10,* 393–410.

Ritter, W., Vaughan, H. G., & Costa, L. D. (1968). Orienting and habituation to auditory stimuli: A study of short term changes in averaged evoked responses. *Electroencephalography and Clinical Neurophysiology, 25,* 550–556.

Rouse, W. B. (1977). Adaptive allocation of decision making responsibility between supervisor and computer. In T. B. Sheridan & G. Johannsen (Eds.), *Monitoring behavior and supervisory control* (pp. 295–306). New York: Plenum.

Scallen, S. F., Hancock, P. A., & Duley, J. A. (1995). Pilot performance and preference for short cycles of automation in adaptive function allocation. *Applied Ergonomics, 26,* 397–403.
Scerbo, M. W. (2001). Adaptive automation. In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors* (pp. 1077–1079). London: Taylor & Francis.
Scerbo, M. W., Freeman, F. G., & Mikulka, P. J. (2000). A biocybernetic system for adaptive automation. In R. Backs & W. Bouscein (Eds.), *Engineering psychophysiology: Issues and applications* (241–253). 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.
Scerbo, M. W., Freeman, F. G., Mikulka, P. J., Parasuraman, R., Di Nocero, F., & Prinzel, L. J. (2001). *The efficacy of psychophysiological measures for implementing adaptive technology* (NASA TP-2001-211018). Hampton, VA: NASA Langley Research Center.
Stampfer, G. H., & Basar, E. (1985). Does frequency analysis lead to better understanding of human event-related potentials. *International Journal of Neuroscience, 26,* 181–196.
Strayer, D., & Kramer, A. (1990). Attentional requirements of automatic and controlled processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 67–82.
Streitberg, B., Röhmel, J., Herrmann, W. M., & Kubicki, S. (1987). COMSTAT rule for vigilance classification based on spontaneous EEG activity. *Neuropsychobiology, 17,* 105–117.
Wickens, C. D., Isreal, J. B., & Donchin, E. (1977). The event related cortical potential as an index of workload. In *Proceedings of the Human Factors Society 21st Annual Meeting* (pp. 282–287). Santa Monica, CA: Human Factors and Ergonomics Society.
Wickens, C. D., Kramer, A., Vanasse, L., & Donchin, E. (1983). Performance of concurrent tasks: A psychological analysis of the reciprocity of information processing resources. *Science, 221,* 1080–1082.
Wilson, G. F., & Eggemeier, F. T. (1991). Psychophysiological assessment of workload in multi-task environments. In D. L. Damos (Ed.), *Multiple-task performance* (pp. 329–360). London: Taylor & Francis.
Yordanova, J., & Kolev, V. (1998). Single-sweep analysis of theta frequency band during an auditory oddball task. *Psychophysiology, 35,* 116–126.

Lawrence J. Prinzel received his Ph.D. in industrial/organizational psychology from Old Dominion University in 1998. He is an aerospace technologist and human factors researcher at NASA Langley Research Center and an adjunct professor at Old Dominion University and Embry-Riddle Aeronautical University.

Frederick G. Freeman received his Ph.D. in experimental psychology from Kent State University in 1970. He is a professor of physiological psychology at Old Dominion University.

Mark W. Scerbo received his Ph.D. in experimental psychology from the University of Cincinnati in 1987. He is an associate professor of human factors psychology at Old Dominion University.

Peter J. Mikulka received his Ph.D. in experimental psychology from Rutgers University in 1967. He is a professor of experimental psychology at Old Dominion University.

Alan T. Pope received his Ph.D. in clinical psychology from the University of Florida in 1975. He is a research scientist at NASA Langley Research Center.

Date received: January 15, 2002
Date accepted: August 29, 2003