Kramer, A. F., Trejo, L. J., & Humphrey, D. (1996). Psychophysiological measures of human information processing activities: Potential applications to automated systems. In R. Parasuraman & J. Mouloua (Eds.), Automation and Human Performance: Theory and Applications. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 137-162.

7 Psychophysiological Measures of Workload: Potential Applications to Adaptively Automated Systems

Arthur F. Kramer
Leonard J. Trejo
University of Illinois

Darryl G. Humphrey
Wichita State University

INTRODUCTION

The main goal of this chapter is to provide a brief synopsis of recent research in the field of applied psychophysiology. More specifically, we describe several studies that together begin to define the techniques and situations in which psychophysiological measures may provide important insights into human information processing activities of relevance to automated systems. Our review will focus on one class of psychophysiological measures—event-related brain potentials (ERPs)—and, for the most part, the measurement of a single but multidimensional psychological construct—mental workload. We do not mean to imply from our restricted treatment of the literature that we believe that other psychophysiological techniques have limited utility in the assessment of operator state in automated systems. In fact, psychophysiological measures such as heart rate and eye movements show great promise for the assessment of human information processing activities in simulated and operational contexts. However, we have chosen the path of providing a somewhat in-depth discussion of one particular psychophysiological technique rather than a broader but more superficial treatment of the applied psychophysiology literature (for more comprehensive treatments of this literature, see Kramer, 1991; Kramer & Spinks, 1991; Wilson & Eggemeier, 1991).

As argued by a number of researchers (Sheridan, 1987; Wickens, 1992), automation has, in many cases, changed the nature of rather than diminished the processing demands imposed on human operators. Instead of manually controlling the inner loop components of systems such as aircraft, manufacturing, and chemical processes, humans are now involved in monitoring system parameters and occasionally intervening in the operation of the system to detect, diagnose, and correct system malfunctions. One important by-product of this shift in roles for human operators is that it has become more difficult to infer the operators' information processing activities and strategies. In manual and semi-automated systems, human operators were constantly engaged in making analog and discrete inputs in an effort to maintain the state of the system within an acceptable range. These control inputs and adjustments could, in turn, be used as a yardstick against which to measure the operators' performance and to infer whether the operator was under- or overloaded or had missed critical system information.

However, overt performance is often quite sparse in automated systems, because the operators only occassionally intervene to adjust system parameters or conduct tests. In such cases, it is often difficult to determine the extent of operators engagement in automated systems. In fact, there is now a sufficient body of data to suggest that operators are often slower and less accurate in detecting and diagnosing system malfunctions when they serve as system supervisors than when they are involved in actively controlling a system (Bortolussi & Vidulich, 1989; Ephrath & Young, 1981; Kessel & Wickens, 1982). Such studies clearly suggest a need for human information processing assessment techniques that do not rely solely on the occurrence of overt control actions.

One solution that has been proposed to reduce the operator's workload while still keeping him or her "in the loop" is adaptive aiding. The concept of adaptive aiding involves the use of automation only when the operator requires assistance to meet task demands. Otherwise, the operator maintains control of the system functions, often by manually controlling system parameters, and therefore remains in the loop (Rouse, 1988; Wickens, 1992). Commercial and military piloting is one environment in which adaptive aiding has been employed for many years. In this setting the pilot can offload manual control responsibilities by engaging the autopilot.

Although the use of adaptive aiding allows for more flexible distribution of tasks between computers and human operators, and therefore can potentially enhance overall system effectiveness, the concept of adaptive aiding has also generated a number of interesting and important questions. Although many of these questions are beyond the scope of this chapter (but see Scerbo, chap. 3, this volume) one important issue that could potentially benefit from use of psychophysiological measures concerns the basis for deciding whether aiding is needed. Two different approaches have been examined in previous studies. One approach involves the use of human performance models to predict how well an operator is likely to perform a task given changing task demands and human resources (Govindaraj &

Rouse, 1986; Greenstein & Ravesman, 1986). The other approach has involved online assessment of performance, which, in turn, has been used to infer operators' intentions and capabilities to successfully complete system-relevant tasks (Geddes, 1986). Given the sparsity of overt human actions in many modern-day systems as well as the imperfect mapping of performance to intentions and mental processes, it appears reasonable to ask whether psychophysiological measures can be used to improve the assessment and prediction of human performance in complex systems. In the following sections of this chapter, we discuss the potential of psychophysiological techniques for the assessment of information processing activities of human operators in adaptively automated systems.

PSYCHOPHYSIOLOGICAL MEASURES: ADVANTAGES AND DISADVANTAGES

In an effort to adhere to the truth-in-advertising dictum, we would be remiss if we did not describe both the advantages as well as the disadvantages in the use of psychophysiological measures, and more specifically ERPs, in the assessment of psychological processes of relevance to automated systems. We begin with the disadvantages. ERPs, and psychophysiological measures in general, are relatively expensive and time consuming to acquire, analyze, and interpret, at least in comparison to most performance and subjective measures that have been obtained in extra-laboratory settings. However, over the past decade the cost for the specialized equipment necessary to record ERPs (e.g., amplifiers, transducers, a/d conversion boards, large data storage media) has decreased quite substantially such that the hardware and software can now be purchased for somewhere in the neighborhood of $30,000. The interpretation of the data is another matter. The complexity of the ERP waveform as well as the substantial theoretical and empirical literature that relates ERP components to different psychological processes precludes a cookbook approach to data interpretation. Thus, it is necessary to have a knowledgeable psychophysiologist involved in any research or assessment project.

A related point concerns the complexity of signal extraction and analysis and the detection of potential artifacts. Although artifacts are certainly a concern even with the recording of reaction time and accuracy measures, the magnitude of the problem is often larger for physiological measures. For example, many ERP components can be contaminated by other electrical activity, such as that generated by eye, neck, and body movements. Artifacts also arise from inadequate electrode placement and saturation of the a/d channels. Although knowledge of signal characteristics and analytic procedures along with careful data recording protocols can eliminate or reduce the impact of many of these potentially confounding factors, a good deal of technical expertise is necessary to ensure successful data collection and signal extraction.

Another concern with many ERP recording procedures is the potential intrusiveness of the methodology. For example, although ERPs, and the P300 component of the ERP in particular, have been found to be a sensitive index of perceptual/cognitive processing demands, many of the laboratory studies that have demonstrated this relationship have done so by using a secondary task methodology. With this method subjects are asked to perform a primary task to the best of their ability and devote any spare capacity to the performance of the secondary task. ERPs are elicited by the secondary task stimuli. The underlying assumption adopted with the use of this methodology is that any processing resources that remain after the performance of the primary task will be devoted to secondary task performance. The ERPs are assumed to tap these spare resources. In fact, a number of studies have found that the amplitude of the P300 component of the ERP elicited by the secondary task stimuli systematically decreases with increases in the difficulty of the primary task (Isreal, Chesney, Wickens, & Donchin, 1980; Isreal, Wickens, Chesney, & Donchin, 1980; Kramer, Sirevaag, & Hughes, 1988; Kramer, Wickens & Donchin, 1983, 1985) and with increases in the priority of the primary task (Strayer & Kramer, 1990). Although the ERP-based secondary task technique has been quite useful in exploring theoretical issues concerning attention and resource allocation and the development of automatic processing, the requirement to perform an extraneous task renders it difficult to apply in operational contexts in which operators may already be overburdened by task demands.

Two solutions to the intrusiveness problem have been pursued. In one procedure, hereafter to be referred to as the *primary task technique*, ERPs are elicited by discrete events within the task of interest. In this context ERP components, and in particular the P300, have been found to increase in amplitude with increases in the difficulty or priority of the task, presumably reflecting the allocation of additional processing resources or attention for more difficult or high-priority tasks (Mangun & Hillyard, 1990; Sirevaag, Kramer, Coles, & Donchin, 1989; Ullsperger, Metz, & GIlle, 1988; Wickens, Kramer,Vanasse, & Donchin, 1983). The primary task method can be quite useful in settings in which it is possible to trigger ERPs on the basis of discrete task-relevant events. Such events might include the occurrence of new aircraft on an air traffic controller's console, the presentation of updated system status information on a automated manufacturing control screen, or the presentation of new navigational fixes on an aircraft pilot's CRT. Unfortunately, however, there are situations or time periods in which few such discrete events occur but an assessment of operator state is still desired (e.g., monitoring of a sonar display or the status of a process control plant). Furthermore, it is often difficult, particularly in operational settings, to modify the system hardware and software to accommodate the acquisition of ERPs. Finally, the comparison of primary task ERPs across dissimilar tasks and systems requires the tenuous assumption that all primary task events require the same variety of processing resources or attention.

One alternative to primary and secondary task methods of ERP recording has been referred to as the *irrelevant probe technique* (Papanicolaou & Johnstone, 1984). This technique involves the recording of ERPs to auditory or visual probes that accompany the task of interest. However, unlike the secondary task method, which requires that subjects actively respond or count the probes, the probes are ignored in the irrelevant probe technique. Thus, this technique has the same advantages associated with the secondary task method while minimizing the potential for disturbing the performance of the task of interest. The theoretical rationale is essentially the same for the irrelevant probe technique as it is for the secondary task method. That is, that increases in the difficulty of the primary task will result in increased resource allocation to the primary task with a concomitant decrease in the resources available for the processing of the probes. We illustrate how this method can be used to examine mental workload in simulated real-world tasks in a later part of this chapter.

One additional concern about the use of psychophysiological measures as indices of human information processing activities is the amount of data that is necessary to reliably identify changes in mental workload, alertness, or whether an operator has failed to attend to a critical signal. In laboratory situations, ERPs are elicited by a number of presentations of a stimulus, and then these single-trial ERPs are averaged in an effort to enhance the signal-to-noise ratio of the critical ERP components. Although such a procedure is reasonable in the laboratory, it may not suffice in situations in which moment-to-moment variations in operator state is of concern. Later in this chapter we describe a program of research in which we have begun to examine the degree to which ERPs can be expected to tap dynamic changes in operator state.

Thus far, we have focused on the problems, as well as some potential solutions, in the use of psychophysiological measures for the assessment of aspects of human information processing in extra-laboratory situations. However, psychophysiological measures also possess a number of strengths that make them well suited for the assessment of aspects of human cognition of relevance to adaptively automated systems. For example, mental workload has been defined as the interaction between the structure of systems and tasks on the one hand, and the capabilities, motivation, and state of the human operator on the other (Gopher & Donchin, 1986;

Kramer, 1991). More specifically, mental workload has been defined as the information processing costs a human operator incurs as tasks are performed. In recent years, processing costs have been conceptualized in terms of multiple resources with performance decrements resulting when two or more tasks exhaust the supply of a particular variety of processing resource (Wickens, 1992).

Given the multidimensional nature of mental workload and other psychological constructs (e.g., memory, attention, language processes), it is fortunate that ERP components, which are defined with respect to their polarity, scalp distribution, and latency range, have been found to be sensitive to a variety of different information processing activities. For example, the P100 component, a positive going voltage deflection that occurs within 100 milliseconds following a stimulus, is specifically sensitive to the allocation of attention to a particular region of the visual field. The mismatch negativity (MMN), which is a negative going difference wave that occurs approximately 150 to 250 milliseconds poststimulus, provides an index of the extent to which a particular stimulus matches a predefined template (e.g., Is that the musical note I heard a few seconds ago?). The P300 component appears to reflect stimulus evaluation processing, whereas the N400 component reflects the detection of semantic mismatch. Thus, one advantage of psychophysiological measures, and ERPs in particular, is that they are inherently multidimensional in nature. That is, the components that can be found in a single one-second waveform reflect a multitude of information processing activities.

A second advantage of psychophysiological measures is that they can be recorded in the absence of overt behavior. Thus, a manual or vocal action is not required for the elicitation of many ERP components. Given that control inputs are often sparse in automated systems, psychophysiological measures may be used to provide insights into human information processing activities that would otherwise be unavailable with traditional performance measures. Finally, psychophysiological measures are recorded relatively continuously and therefore offer the potential to provide a rapid assessment of changes in operator state. However, as discussed previously, an important question concerns the amount of psychophysiological data that is required to unambiguously discriminate among different operator states. In an effort to provide a partial answer to this question, we now describe a study that was designed to examine the feasibility of employing ERPs to measure dynamic changes in mental workload.

REAL-TIME ASSESSMENT OF MENTAL WORKLOAD: A FEASIBILITY STUDY

The main goal of the study that we now briefly describe was to determine the amount of ERP data that would be necessary to reliably discriminate among several different levels of single and dual-task processing load (see Humphrey & Kramer, 1994, for a detailed description of this study). To that end, 12 young adults performed two complex tasks, monitoring six constantly changing gauges and performing mental arithmetic problems, both separately and together. ERPs were recorded from discrete events in both of the tasks, the presentation of the cursors in the monitoring task and the presentation of the operators and operands in the mental arithmetic task, and a monte carlo approach was used to relate different amounts of ERP data to different levels of accuracy in discriminating among variations in mental workload.

The two tasks that were performed by the subjects are presented in Fig. 7.1. Each of the gauges in the monitoring task were divided into 12 regions. The different regions were coded numerically (i.e., the numbers 1 to 12) and with color (i.e., 1 to 4 were green, 5 through 8 were yellow, 9 through 12 were red). The subject's task was to reset each gauge as quickly as possible once its cursor had reached the critical zone which was redundantly defined by number (> 9) and color (red). Subjects reset the gauges by depressing one of the six editing keys from a standard IBM AT keyboard with their right hand. Each key corresponded to a specific gauge. The gauge-to-key mapping was spatially compatible.

In order to encourage subjects to learn the relationships among the movement of the cursors in the different gauges, and in an effort to simulate the sampling strategies required with physically displaced gauges in operational systems, subjects could only view the position of the cursors one at a time. Thus, although the gauges were always present on the screen, the cursors were not continuously visible. To sample a gauge, (i.e., to see where the cursor was located), the subjects pressed one of a set of six keys with their left hand. Once sampled the cursor remained visible for 1,200 milliseconds. Simultaneous sampling was not possible.

The difficulty of the monitoring task was manipulated by varying the degree to which the position of the cursor on one gauge could be predicted from the position of the cursor on another gauge. In the high-predictability (HP) condition, the gauge monitoring functions were equivalent for the three gauges within a row. The only difference between these gauges was a phase offset (i.e., the cursors began at different positions on the gauges). In the low-predictability (LP) condition, each of the gauges was driven by a separate forcing function.

The center of each gauge served as a display area for the operators and operands for the arithmetic task (see Fig. 7.1). All of the operators and operands were presented simultaneously, with one operand in each gauge and one operator for every two gauges. Arithmetic problems were presented every 4 to 15 seconds following the completion of a previous problem. Subjects were instructed to complete the problems as quickly and as accurately as possible. The difficulty of the mental arithmetic task was

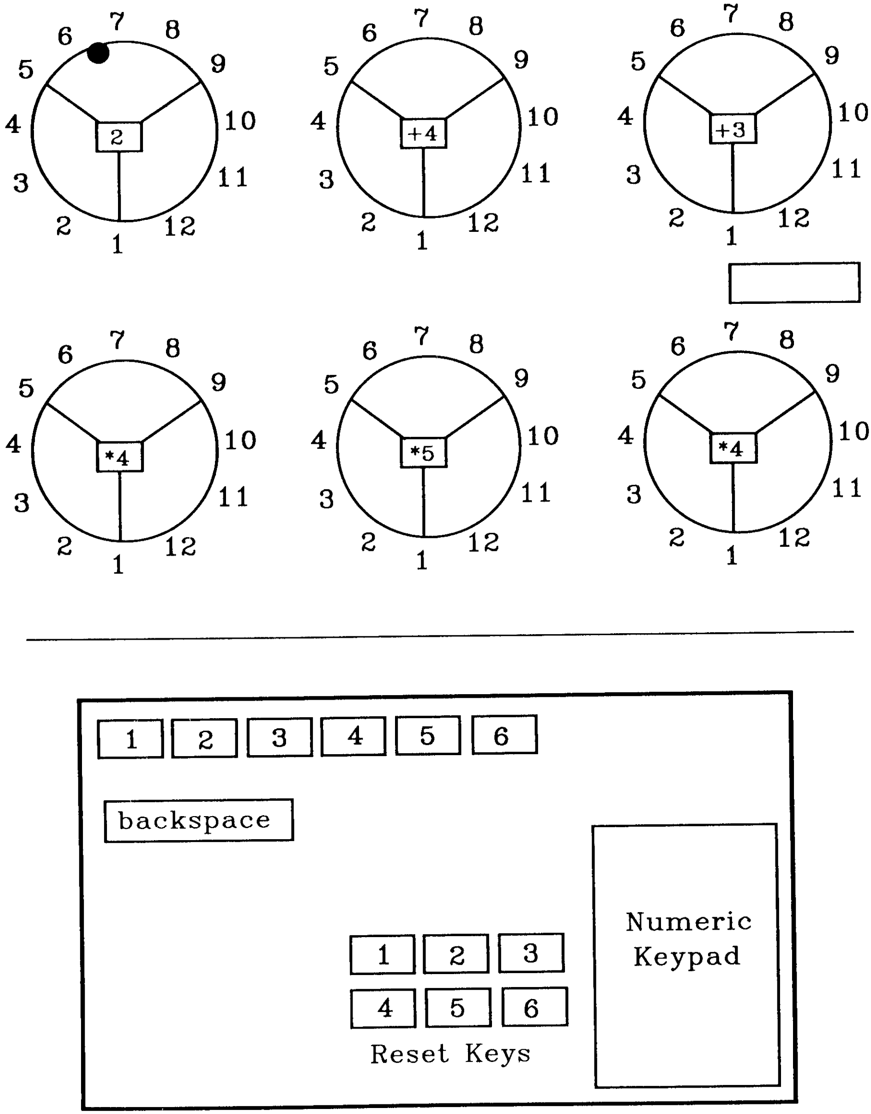

FIG. 7.1. A graphic illustration of the monitoring and mental arithmetic tasks along with the response input devices.

manipulated by varying the number of column operations necessary to complete the problem. The A2 version of the task required operations on two columns of numbers, whereas the A3 version of the task required operations on three columns of numbers. Operations included addition and multiplication. Answers were entered via the numeric keypad on the IBM AT keyboard and appeared in the window as they were typed.

Subjects performed the mental arithmetic and monitoring tasks, both separately and together, during five two-hour sessions. Electroencephalographic (EEG) activity was recorded from three midline sites—Fz, Cz, and Pz—according to the International 10–20 system (Jasper, 1958). Vertical electrooculographic (EOG) activity was recorded from electrodes placed above and below the right eye. Horizontal EOG was recorded from electrodes located lateral to each eye. The EOG and EEG were digitized every 5 miliseconds and were filtered offline (-3dB at 6.89 Hz, 0dB at 22.2 Hz). Trials that were contaminated by excessive EOG artifacts were not included in subsequent analyses of the ERP components. Fewer than 5% of the trials were rejected for excessive EOG artifacts.

RESULTS AND CONCLUSIONS

The performance data, reaction time (RT) and accuracy, were analyzed to ensure that we had successfully varied both single- and dual-task processing demands. RTs decreased from the HP to the LP condition in the monitoring task and from the two to the three column problems in the mental arithmetic task. Error rates behaved in a similar manner. Furthermore, performance was significantly poorer in the single-task than in the dual-task conditions.

After establishing that we had, in fact, successfully manipulated task difficulty as indexed by the performance measures, our next step was to determine if average ERPs differed among different task conditions. This was necessary because it would not make sense to ask how much ERP data was necessary to discriminate between different levels of mental workload if we could not show reliable ERP differences when large numbers of single trials were averaged. The grand average ERPs elicited in a subset of the task conditions for both the monitoring and mental arithmetic tasks are presented in Fig. 7.2. As can be seen from the figure, single- and dual-task ERPs are visually dissimilar for both the monitoring and mental arithmetic tasks. These differences, which were corroborated in ANOVAs, are most obvious in the region of the P300 component (i.e., 300 to 500 milliseconds poststimulus) and the later slow wave (i.e., from 750 to 1200 milliseconds poststimulus).

Given that we had now established that the different task conditions could be distinguished on the basis of averages of large numbers of single-trial ERPs, we were now able to proceed in addressing our original research question: How much ERP data is necessary to discriminate between different levels of mental workload? Of course, the answer to this question depends on the specification of a level of accuracy with which workload conditions can be discriminated. Given that the level of acceptable discrimination accuracy might vary in different situations and for

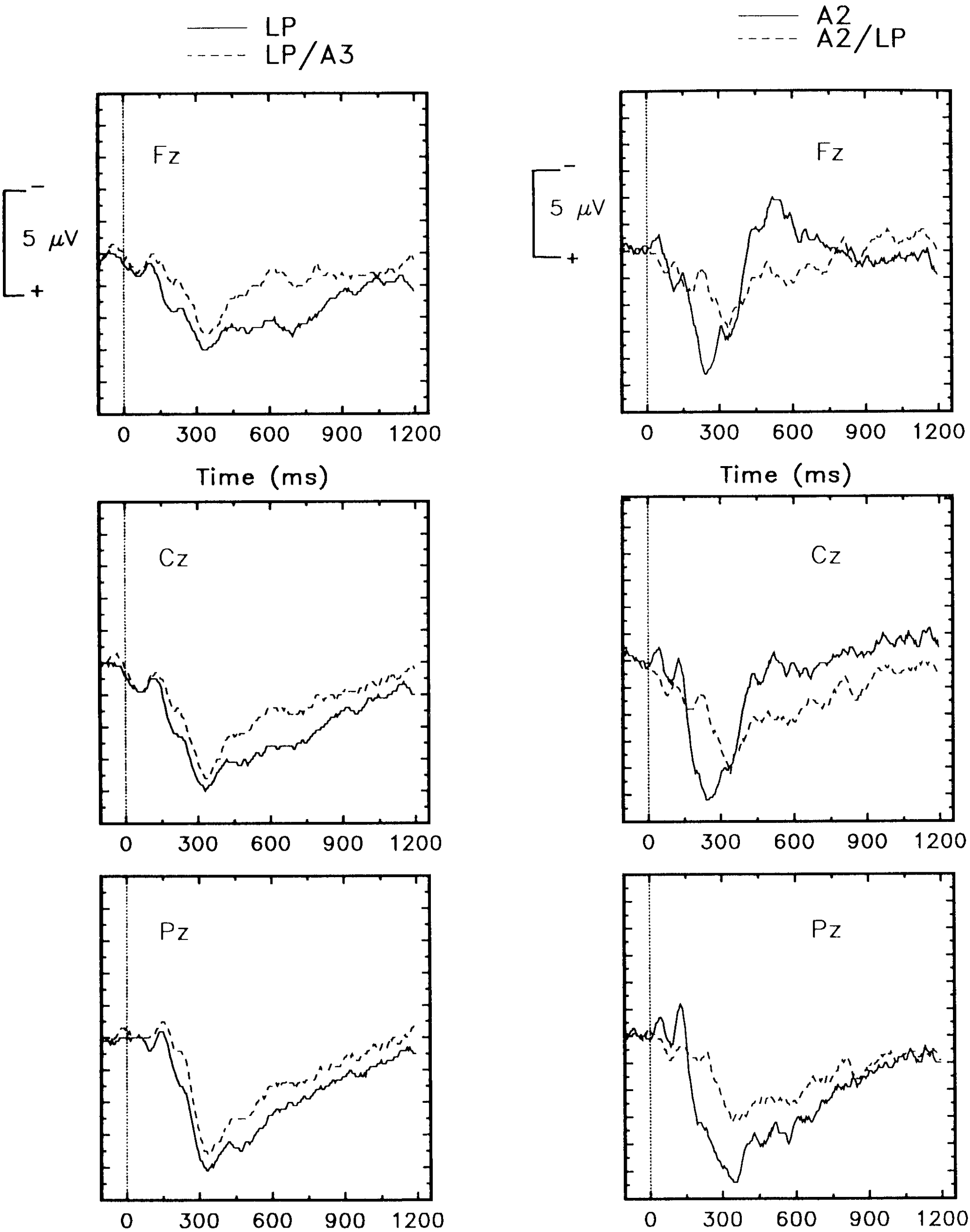

FIG. 7.2. Grand average waveforms across the 12 subjects for single and dual-task conditions in the monitoring and mental arithmetic tasks at the three electrode sites. LP refers to the single-task low predictability monitoring condition. LP/A3 refers to the ERPs elicited by the cursor updates in the monitoring task low predictability condition when performed concurrently with the three column mental arithmetic task. A2 refers to the single-task two column mental arithmetic condition. A2/LP refers to the ERPs elicited by the presentation of the operator and operands in the two column mental arithmetic task when performed concurrently with the low predictability monitoring task.

different systems, we adopted a monte carlo approach in which we systematically incremented the amount of ERP data that were averaged prior to discriminating between workload conditions. This approach has enabled us to relate, in a relatively continuous manner, different levels of accuracy of discrimination to different amounts of ERP data.

Our approach to this issue included the following steps. First, we chose two sets of experimental conditions that differed in perceived workload, performance, and average ERP measures. In an effort to evaluate the reliability of ERP measures in discriminating between levels of workload in different tasks, we also constrained our choice of conditions such that one set was chosen from the monitoring task whereas the other set of conditions was chosen from the mental arithmetic task. These conditions included the LP and LP/A3 conditions in the monitoring task and the A2 and A2/LP conditions in the mental arithmetic task.

The second step of our procedure involved the derivation of each of the ERP measures described in Table 7.1 for each of the single trials at each electrode in the selected conditions. Thus, for each experimental trial a total of 24 ERP measures were derived (i.e., the eight measures in Table 7.1 at each of the three electrode sites). The vectors of ERP measures, with one vector for each single trial, were then divided in half such that all of the even trials were placed in one pool and all of the odd trials were placed in different pool. There were approximately 75 trials in each of the two pools for each subject for the monitoring task and 35 trials in each pool for each subject for the mental arithmetic task.

The third step in our procedure involved the random selection of 1,000

TABLE 7.1
Measures Obtained From the Single-Trial Event-Related Brain Potentials in the Monitoring and Mental Arithmetic Tasks

Measure	*Description*
Base-peak amplitude (BPamp)	Largest positive voltage between 275 and 750 milliseconds poststimulus-
Base-peak mean voltage	Mean voltage in a 100-miliseconds window centered on the point picked as Bpamp
Base-peak root mean square	Root mean square amplitude computed in a 100-milliseconds window centered around the point picked as Bpamp
Cross-correlation mean voltage	Mean voltage in a 100-milliseconds window centered on the point of maximum cross-correlation between the ERP and a 300-milliseconds cosine template
Slow wave 1	Mean voltage between 750 and 1,250 milliseconds poststimulus
Slow wave 2	Mean voltage between 900 and 1,100 milliseconds poststimulus
Slow wave 1 rms	Root mean square amplitude computed between 750 and 1,250 milliseconds poststimulus
Slow wave 2 rms	Root mean square amplitude computed between 900 and 1,100 milliseconds poststimulus

samples, with replacement, of from 1 to n single-trial vectors of ERP measures from each of the even and odd pools of measures in each of the four conditions (i.e., the two conditions from the monitoring task and the two conditions from the mental arithmetic task). The vectors were averaged after each selection of 1,000 samples. Thus, for example, in the 1,000 samples of four trials the four vectors of measures selected in each sample were averaged to produce a single vector of measures. This averaging procedure was undertaken to increase the signal/noise ratio of the ERP measures.

The fourth step in our analysis procedure involved the classification of the sample vectors as representing one of the two workload levels for each of the two tasks. The classification algorithm that we applied was a linear stepwise discriminant analysis (LSDA). The discriminant functions were developed for each task and subject on one half of the data and cross-validated on the other half of the data set. In an effort to evaluate the utility of spatial information (e.g., the distribution of the ERPs across different scalp sites) in the discrimination between workload levels, separate discriminant functions were computed for the ERP data vectors at the Fz, Cz, and Pz electrode sites as well as for a combined vector of the measures across scalp electrodes. Thus, 8 ERP measures were submitted to the LSDA procedure for each of the individual electrode functions and 24 measures were submitted for each sample vector for the combined electrode function.

Figs. 7.3 and 7.4 provide a graphic representation of the efficiency of discriminating between two workload levels for the mental arithmetic and monitoring tasks, respectively. Plots are included for both the validation data, on which the discriminant functions were derived (left side), and the cross-validation data (right side), which was classified using the discriminant functions developed with the validation data. Separate panels are provided for the single electrode functions as well as for the functions, which included measures from the three different scalp locations (combined). The 12 functions in each of the panels represent the 12 subjects who participated in the study.

There are several noteworthy aspects of the figures. First, classification accuracy is monotonically related to the amount of ERP data. This is not particularly surprising, because the signal-to-noise ratio of the ERP components increases with the square root of the number of trials averaged to produce each sample. Second, combining information across different spatial locations (in the present case the Fz, Cz, and Pz recording sites) clearly improves the classification efficiency. This improved classification efficiency is observed in (a) reduced variability among subjects, (b) increases in the average asymptotic level of classification accuracy across subjects, and (c) a reduction in the amount of ERP data necessary for correctly discriminating between workload levels.

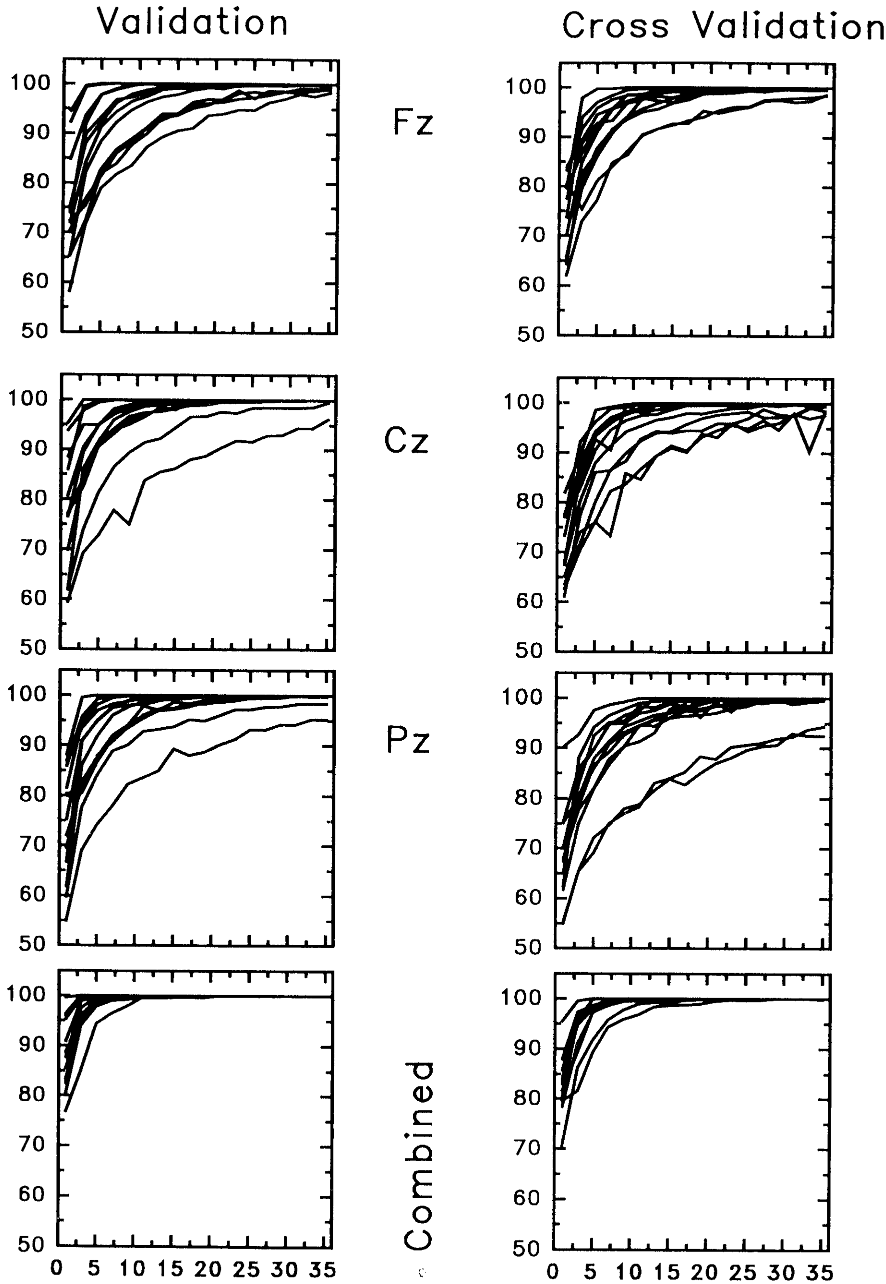

FIG. 7.3. Classification efficiency functions derived from the LSDA procedure for the mental arithmetic task for Fz, Cz, and Pz and combined electrode ERP measures. The plots on the left represent the data that was used to derive the discriminant functions while the plots on the right represent the other half of the data which was fit with the derived discriminant functions. Each of the functions in each of the panels represents a single subject. The conditions that were discriminated in these analyses were A2 and A2/LP. The y axis represents the accuracy of discriminating between the two workload conditions while the x axis represents the number of ERP samples (trials).

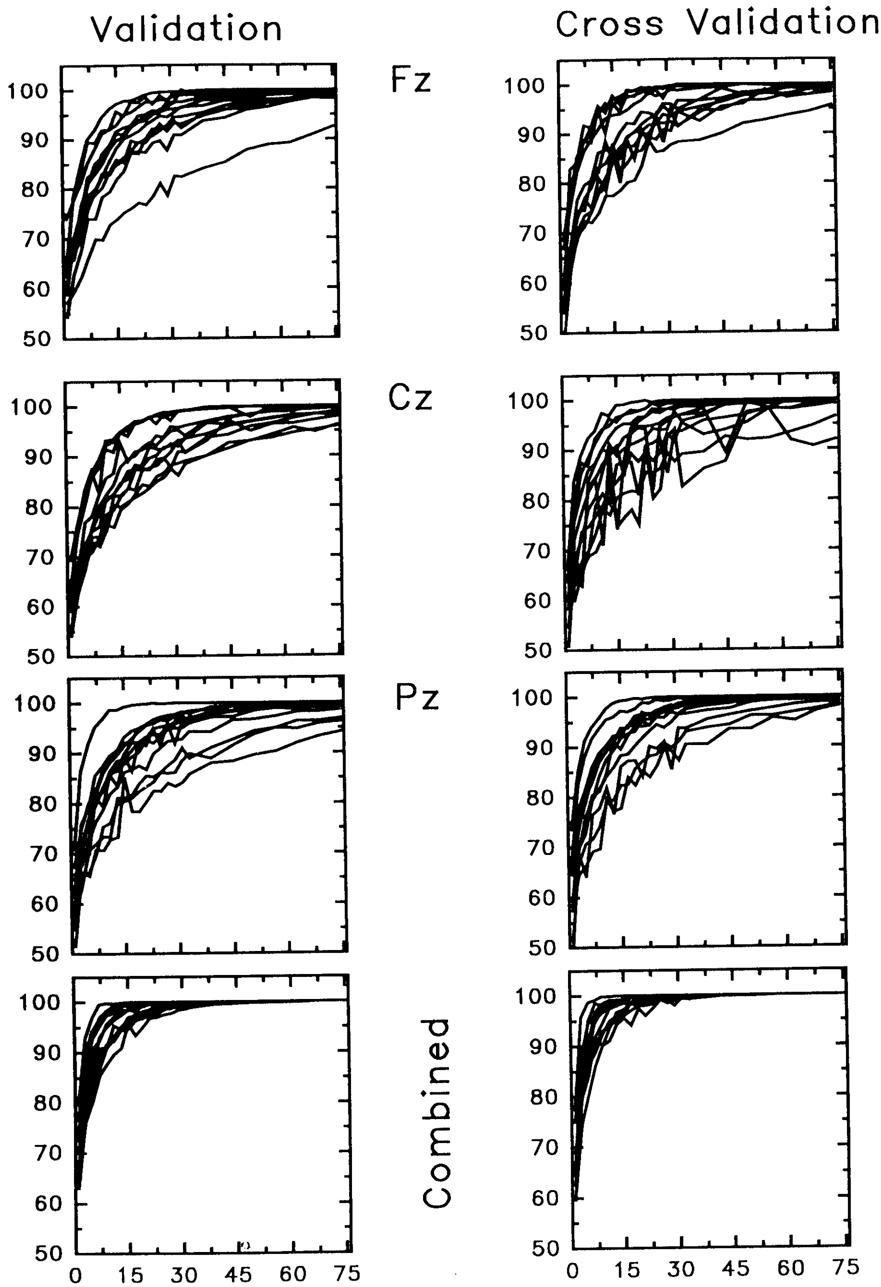

FIG. 7.4. Classification efficiency functions derived from the LSDA procedure for the monitoring task for Fz, Cz, Pz and combined electrode ERP measures. The plots on the left represent the data that was used to derive the discriminant functions while the plots on the right represent the other half of the data which was fit with the derived discriminant functions. Each of the twelve functions in each of the panels represents a single subject. The conditions that were discriminated in these analyses were LP and LP/A3. The *y* axis represents the accuracy of discriminating between the two workload conditions while the *x* axis represents the number of ERP samples (trials).

Another important aspect of the figures is the relatively small loss of classification accuracy when the discriminant functions derived on one data set (validation panels) are applied to a different set of ERP data (cross-validation panels). In fact, in the combined cross-validation sample, 90% correct classification is achieved for all of the 12 subjects within 11 and 5 trials for the monitoring and mental arithmetic tasks, respectively. These results suggest that the within-subject discriminant functions are quite reliable across similar data sets.

Another important output of the discriminant analyses is provided in Table 7.2. The table provides a summary of the ERP measures, at each electrode site and for the two tasks, that had the highest weight in the discriminant equations. Thus, the frequency of entries in different cells of the table provides an indication of the relative importance of different ERP measures in discriminating among the workload levels. There are a number of interesting aspects of the data. First, the pattern of frequencies suggests that no single measure was superior to the other measures across tasks and electrode sites. There were, however, some interesting trends. For instance, at the Pz electrode site the majority of the measures with the highest weights (i.e. the first five measures) pertained to the P300 component. On the other hand, at the Fz and Cz recording sites the measures with the highest weights were more evenly divided among measures that pertained to the P300 component and measures of different aspects of late slow wave components (i.e., the last four measures in the table). Second, the pattern of frequencies suggest that no single measure of a particular ERP component was the best discriminator for all of the subjects. Thus, in the case of the P300, both the base-peak amplitude and the cross-correlation measures proved to be good discriminators for different subjects in the sample. For the slow wave component, the best univariate measures appeared to be the mean voltage

TABLE 7.2
Number of Subjects With the Highest Weight in the Discriminant Equations for Each of the ERP Measures at Each of the Three Electrode Sites for the Monitoring and Mental Arithmetic Tasks

	Monitoring Task			*Arithmetic Task*		
Measure	*Fz*	*Cz*	*Pz*	*Fz*	*Cz*	*Pz*
Base-peak amplitude	2	2	3	2	1	6
Base-peak mean amplitude	1	1	1	0	2	2
Base-peak rms	0	1	1	2	1	0
Cross-correlation mean voltage	3	3	3	3	1	2
Slow wave 1	3	1	2	2	2	2
Slow wave 2	1	1	1	0	2	0
Slow wave 1 rms	2	3	1	1	1	0
Slow wave 2 rms	0	0	0	2	2	0

and root mean square error for the longer measurement interval (e.g., slow wave 1, slow wave 1 rms measures)

In summary, these data suggest that psychophysiological measures, and ERPs in particular, might have some utility as measures of momentary fluctuations in mental workload and, therefore, might serve as a trigger for adaptive aiding. Of course, this conclusion must remain tentative until our findings are validated with a larger variety of tasks, processing demands, and subjects. The issue of the generalizability of these findings to noisier settings such as high-fidelity simulators and operational systems also remains an open question (although for some promising results see Kramer, Sirevaag, & Braune, 1987; Sirevaag et al., 1993).

MENTAL WORKLOAD ASSESSMENT WITH IRRELEVANT PROBES

Our examination of the feasibility of employing ERPs to discriminate among momentary variations in processing demands was performed within the context of single- and dual-tasks in which ERPs were elicited by task-relevant discrete events. Although this ERP-eliciting procedure is advantageous because it does not require the addition of any extraneous tasks or stimuli, as does the secondary task procedure, it is often difficult, particularly in operational settings, to modify the system hardware and software to incorporate ERP recording. Additionally, it may be difficult to identify task-relevant discrete events that occur with sufficient frequency in many automated systems. Thus, it would be useful to possess other recording procedures for use in situations in which the primary task method is impractical or inappropriate. As we briefly described earlier, the irrelevant probe procedure may prove useful in situations that preclude the use of the primary task method.

In the study we describe next we assessed the utility of the irrelevant probe technique in a high-fidelity radar simulator with 10 highly experienced Navy radar operators. The radar operators performed a standard training exercise that contained periods of low- and high-processing demands. The ERPs were elicited by three different tones that differed in frequency of occurrence. One of the three tones occurred on 80% of the probe trials, whereas each of the other two tones occurred on 10% of the probe trials. Prior to the radar monitoring task, in which the tones were to be ignored, the radar operators performed a baseline oddball condition in which they pushed a response button every time one of the two low-probability tones was presented. The other low-probability tone and the high-probability tone did not require an overt response. The baseline condition was used to establish a record of each individual's ERP components in the absence of the demands of the radar monitoring task. The dual-deviant/single-standard (i.e., two different low-probability tones and one high-probability tone) tone presentation was used for two reasons. First, low-probability stimuli often elicit larger amplitude ERP components than do high-probability stimuli. Thus, the use of two low-probability tones would presumably enable us to increase the frequency of detecting changes in the amplitude of ERP components in response to variations in the processing demands in the radar monitoring task. Second, we were interested in examining ERP components such as the N100, N200 and P300 as well as the mismatch negativity (MMN; Naatanen, 1990). Although the N100, N200, and P300 components are often easiest to observe when subjects actively attend or respond to the ERP eliciting events, MMNs are difference waveform components that can most easily be dissociated from other ERP components when elicited by high- and low-probability events that do not require any overt action. Thus, the use of two deviant probe events, one that required a response and one that did not, would enable us to discern the MMN as well as the other ERP components in the baseline condition.

Our expectations with regard to the sensitivity of the different ERP components to the introduction of the radar monitoring task and an increase in its difficulty were as follows. The N100 component has long been interpreted to reflect early processes of selective attention or resource allocation (Hackley, Woldoroff,& Hillyard, 1990; Hillyard, Hink, Schwent, & Picton, 1973). The N100 has also shown a graded sensitivity to processing demands across both tasks and input locations (Parasuraman, 1985). Thus, it is conceivable that the irrelevant probe-evoked N100s will show a systematic decrease in amplitude with the introduction of the radar monitoring tasks as well as with an increase in its difficulty.

Previous evidence suggests that N200s may also be sensitive to changes in processing demands within and across tasks, at least when the N200s are elicited by task-relevant events. For example, Lindholm, Cheatham, Koriath, and Longridge (1984) reported increases in the amplitude of N200s with increases in the difficulty of a simulated flight mission (see also Horst, Ruchkin, & Munson, 1987).

There are now several reports suggesting that the MMN, which was once thought to be insensitive to attention, may be susceptible to attentional demands under some circumstances (Trejo, Ryan-Jones, & Kramer, in press; Woldoroff, Hackley, & Hillyard, 1991). Thus MMNs, derived by subtracting the standard from the ignored-deviant waveforms, may reflect changes in processing demands that are associated with the performance of the radar monitoring tasks as well as an increase in its difficulty.

The P300 component is well documented to be sensitive to changes in processing demands. However, most of these demonstrations have taken place within the context of primary or secondary task methods in which the eliciting stimulus is to be actively attended. There is little evidence suggesting that P300s will reliably reflect graded changes in processing demands when the P300s are elicited by task-irrelevant stimuli (but for exceptions see Sirevaag et al., 1993; Wilson & McCloskey, 1988). In fact, there is strong evidence dating back to the early research of Sutton and colleagues that P300s are not elicited by task irrelevant stimuli (see Sutton & Ruchkin, 1984). Thus, given the challenging nature of the radar monitoring task it is our expectation that irrelevant-probe P300s will not be observed in either the low- or high-load conditions of the radar monitoring task.

The radar monitoring task required the radar operators to detect and respond appropriately to the appearance of a variety of different targets. The targets included commercial aircraft and ships as well as friendly and hostile military aircraft and ships. Appropriate responses included the activation of countermeasures, notification of commanders, and the logging of the area, time, location, and bearing of the targets. The task lasted approximately 45 minutes and was subdivided, for analysis purposes, into periods of low- and high-processing demands. Low- and high-demand periods were operationally defined by two different radar instructors and corresponded, for the most part, to periods of low- and high-target density.

ERP-eliciting tones were presented in a baseline block of 500 trials and during the performance of the radar monitoring task. One of the rare tones was responded to in the baseline condition. The radar operators ignored the tones during the performance of the radar monitoring scenario. The tones varied in frequency and probability of occurrence as follows: 1500 Hz/0.80 (standards), 1000 Hz/0.10 (low deviants), and 2000 Hz/0.10 (high deviants). The interstimulus interval between tones was 700 milliseconds. The tones were 111-milliseconds bursts presented at 80 dB SPL.

Electroencephalographic (EEG) activity was recorded from three midline sites—Fz, Cz, and Pz—and the right mastoid. These electrodes were referenced to the left mastoid. Vertical EOG activity was recorded from electrodes placed above and below the right eye. Horizontal EOG was recorded from electrodes lateral to each eye. Offline, epochs of 1,000 milliseconds, including a 200-milliseconds prestimulus baseline, were extracted and the vertical and horizontal EOG records were then used to reduce EOG contamination.

ERP averages for each stimulus and experimental condition were created separately for each subject, arithmetically re-referenced to average mastoids, digitally low-pass filtered (0 to 12 Hz), and adjusted for zero-median prestimulus baseline voltage. A difference wave was computed by subtracting the average ERP for standards from the average ERP for nontarget deviants. A number of ERP components were measured on the average waveforms in each of the experimental conditions for each of the subjects. These components included N100 (latency range of 75 to 175 milliseconds poststimulus), N200 (latency range of 200 to 400 milliseconds poststimulus), and the P300 (latency range of 300 to 600 milliseconds poststimulus). The N100 and N200 amplitudes were defined as the average of 30 points centered on their respective maximal amplitudes. P300 was defined as the mean amplitude from 300 to 600 milliseconds poststimulus. Finally, two mismatch negativity components (MMN) were measured on the deviant-standard subtraction waveforms. MMN1 was defined as the mean amplitude from 150 to 250 milliseconds poststimulus. MMN2 was defined as the mean amplitude from 250 to 350 milliseconds poststimulus.

RESULTS AND CONCLUSIONS

Several important findings were obtained in our study. First, and perhaps most important, a number of ERP components were found to be sensitive to both the introduction of the radar monitoring task as well as an increase in its difficulty. The N100s and N200s elicited by the deviant tones and recorded at the Fz site systematically decreased in amplitude from the baseline condition to the low-load radar monitoring condition and from this condition to the high-load radar monitoring condition. The MMNs, illustrated in the difference waveforms presented in Fig. 7.5, behaved in a manner similar to the N100 and N200s with respect to their sensitivity to processing demands. That is, the MMN recorded at the Fz site decreased in amplitude with the introduction of the radar monitoring task as well as with an increase in its difficulty.

An interesting and unanswered question is why attention would be allocated to task-irrelevant probes. After all, the voluntary allocation of attention to irrelevant aspects of the environment would be harmless at best and costly (in terms of a reduced processing of relevant information) at worst. Thus, why do the radar operators adopt this seemingly suboptimal strategy? The tentative answer, provided by recent research on visual attention, is that subjects do not adopt such a strategy but instead attention is captured by some features of the environment.

Research by Yantis and colleagues (Yantis & Jonides, 1984, 1990) and others (Theeuwes, 1992; Todd & Van Gelder, 1979) has demonstrated that stimuli that suddenly appear in the environment capture attention regardless, for the most part, of subject's intentions. Many of these experiments are run in the following way. Subjects are asked to search displays for a particular target. On a small proportion of the search trials the target is a sudden-onset item, whereas on most of the trials the target is revealed by removing portions of one of many premasks that appear on the display.

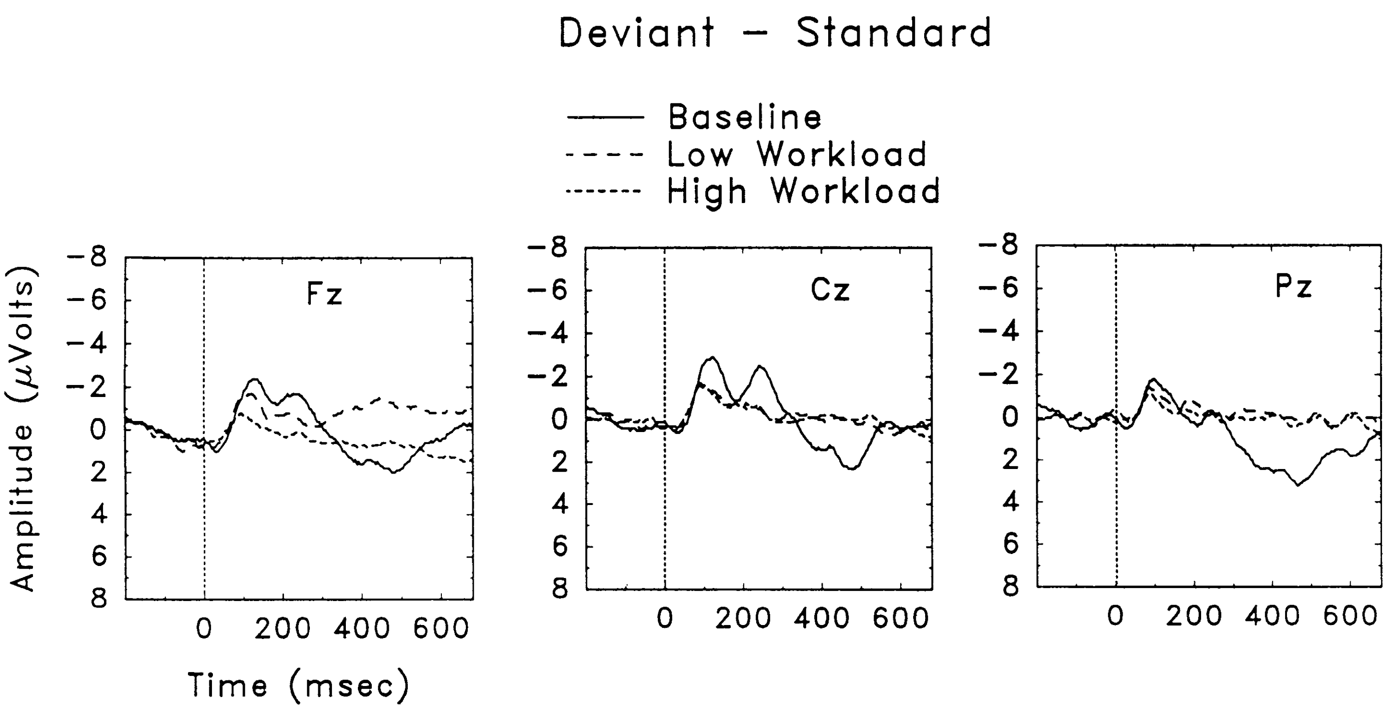

FIG. 7.5. Grand average waveforms derived by subtracting the ERPs elicited by the standard tones (e.g. the tones which occurred on 80% of the tone trials) from the ERPs elicited by the deviant tones.

These latter items are referred to as *offset-stimuli*. It is important to note that the individual onset styles of the letters within a display are independent of the target's presence or absence, and, if present, its location: The target letter is equally likely to be a sudden-onset object as any one of the remaining offset objects. Thus, there is no reason for subjects to attend to the onset object first, because it is not any more likely to be the target than any of the other (offset) objects. However, subjects do appear to attend to the sudden onset items first. This results in a search slope that is independent of the number of items on the display when the target is an onset item. Yantis (1993) suggested that the sudden appearance (sudden-onset) of a new object has a special status in that it automatically captures attention, in a bottom-up or stimulus-driven fashion, as long as attention is not tightly focused on another object in the visual field. Thus, although new objects can be said to automatically capture attention because individuals do not need to intend to process the object for the capture to occur, capture is not strongly automatic because it will fail to occur if attention is already focused elsewhere in the visual field.

We speculate that the notion of attention capture may provide an answer to why it appears that subjects attend the irrelevant probes in our study. That is, the deviant irrelevant probes capture attention because they constitute the occurrence of a new object in the environment. Attention that is not focused elsewhere (e.g., on the radar monitoring task) is captured by the deviant probes and is reflected in the amplitude of the N100, N200, and MMN components. Within this type of theoretical framework subjects do not voluntarily allocate attention to the probes but instead attentional resources that are not being used in the service of the primary task are redirected, in a stimulus-driven or bottom-up fashion, to the processing of the probes.

The P300 results can also be accommodated within this framework. It has been argued previously that P300 reflects the operation of a selection process that capitalizes on higher-order or semantic properties of stimuli. For example, whereas N100s are elicited by any stimulus that appears in an attended portion of the visual field, P300s are only elicited by target objects that occur in the attended field (Mangun & Hillyard, 1990). The fact that P300s were not elicited by the irrelevant probes in the radar monitoring task suggests that capture of attention by the probes was quite transitory. Thus, although the graded workload effect indicated by the N100, N200, and MMN components does suggest that the probes were processed to some degree during the performance of the radar monitoring task, the lack of a P300 effects suggests that this processing was aborted prior to a full evaluation of the stimulus.

We believe that our results are encouraging in that they suggest that ERPs elicited by task irrelevant probes can provide a nonintrusive method for the assessment of variations in mental workload (see also Ullsperger, Freude, & Erdmann, 1994). It is important to note, however, that the N100, N200, and MMN components are rather small relative to the ongoing electroencephalographic activity and, therefore, it is unlikely that these measures will be able to provide a real-time assessment of mental workload. On the other hand, there are many situations in which an offline assessment of mental workload is important (e.g., prototype and system evaluation, operator training assessment, fitness for duty evaluation), and it is these situations in which the ERP-based irrelevant probe technique might be effectively employed.

A BRIEF POSTSCRIPT

Our brief review of the applied psychophysiological literature suggests that psychophysiological measures have the potential to provide useful information about human information processing activities in automated systems. Our literature review was confined to one psychological construct, mental workload, and one class of psychophysiological measures, ERPs. However, other ongoing programs of research are actively exploring the utility of a variety of other psychophysiological measures for the assessment of mental workload as well as other psychological constructs of relevance to human performance in automated systems. For example, a group of Navy researchers led by Scott Makeig (Makeig & Inlow, 1993; Mullane, Makeig, & Trejo, in press) has been developing a system to predict lapses in vigilance, and the resulting decrements in performance, on the basis of an ongoing frequency domain analysis of electroencephalographic (EEG) activity. Thus far, relatively high success rates have been achieved in laboratory simulations of a sonar monitoring task using only a single recording site and a single psychophysiological measure, EEG. Potential improvements in such a system may be achieved by increasing the spatial resolution of the EEG recording (i.e., recording from more than a single electrode site) as well as by incorporating other psychophysiological measures that vary with levels of alertness such as respiration, heart rate, and eye movements.

Farwell and Donchin (1988) reported the development of an ERP-based communication device in which the P300 component of the ERP is used to index operators attention to particular objects in a 6 x 6 matrix of letters and numbers. One interesting aspect of this system is that communication can take place in the absence of eye movements to the attended objects. Thus, such a system could potentially detect the allocation of attention to a location in the periphery of the visual field while an operator is fixated, and possibly also attending, to a location in the central visual field. Gehring et al. (1993) recently reported the discovery of an ERP component, which they dubbed the *error related negativity* (ERN), and which appears to provide a manifestation of a neural system associated with the detection of and compensation for errors of responding. Such a component has the potential to provide an index of whether operators were aware of errors that they may have made in responding to system events.

In addition to the potential applicability of ERP and EEG measures described previously, a number of other psychophysiological measures (for a detailed review see Kramer, 1991) have the potential to contribute to the assessment of operator state in adaptively automated systems. One of the most promising classes of measures is heart rate and heart rate variability. Measures of heart rate have been successfully used to assess changes in mental workload in simulated and operational flight tasks in both commercial and military environments (Jorna, 1993; Roscoe, 1992; Veltman & Gaillard, 1993). One advantage of heart rate measures, compared to ERP measures, is that they can be recorded in the absence of probe stimuli. Thus, heart rate measures can be recorded in a variety of simulator and operational settings without the need to introduce extraneous and potentially disruptive stimuli. Given the continuous nature of heart rate measures it is also relatively easy to collect these data in a timely fashion. However, one concern with heart rate measures concerns the relative sensitivity of this class of measures to physical as compared to mental demands. Thus, heart rate is much more responsive to changes in metabolic demands engendered by physical than by mental demands. Fortunately, there are techniques that can be used to deconfound physical and mental influences on the heart rate measure (Jorna, 1992; Mulder, Veldman, Ruddel, Robbe, & Mulder, 1991). Heart rate measures are also less diagnostic than are ERPs with respect to changes in particular aspects of human information processing. Thus, although ERP components are sensitive to a narrow set of mental functions (e.g. N100/P100—spatial attention, N200—physical mismatch, ERN—error monitoring), heart rate measures are sensitive to changes in a wider variety of emotional and cognitive phenomena. However, although this lack of diagnosticity might be problematic in some circumstances (e.g., when trying to decide whether the perceptual or motor demands imposed on the pilot by a prototype cockpit for an advanced helicopter are excessive), there are other situations in which it is sufficient to know that the operator is overloaded and that computer-based aiding is needed. It is with this latter case that heart rate measures might be most profitably employed as an index of operator workload.

In summary, a number of laboratories are currently developing and validating psychophysiological measures of human information processing activities that have the potential to provide important insights into human cognition and performance in adaptively automated systems. Whether such potential will be realized is dependent on the demonstration of the efficacy of such psychophysiological techniques in extra-laboratory environments as well as the acceptance of these techniques by the user communities.

REFERENCES

Bortolussi, M., & Vidulich, M. (1989). The benefits and costs of automation in advanced helicopters: An empirical study. In R. Jensen (Ed.), *Proceedings of the 5th International Symposium on Aviation Psychology* (pp. 594-599). Columbus, OH: Ohio State University, Department of Aviation.

Ephrath, A., & Young, L. (1981). Monitoring versus man in the loop detection of aircraft control failures. In J. Rasmussen & W. Rouse (Eds.), *Human detection and diagnosis of system failures* (pp. 143-154). New York: Plenum.

Farwell, L., & Donchin, E. (1988). Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology, 70*, 510-523.

Geddes, N. (1986). *Opportunities for intelligent aiding in naval air-sea warfare: An A-18 war at sea study* (Technical report 8502-1). Norcross, GA: Search Technology, Inc.

Gopher, D., & Donchin, E. (1986). Workload—An examination of the concept. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and Human Performance, Vol. II.* (pp. 41.-41.49). New York: Wiley.

Govindaraj, T., & Rouse, W. (1986). Modeling the human controller in environments that include continuous and discrete tasks. *IEEE Transactions on Systems, Man and Cybernetics, SMC-11*, 410-417.

Greenstein, J., & Ravesman, M. (1986). Development and validation of a mathematical model of human decision making for human-computer communication. *IEEE Transactions on Systems, Man and Cybernetics, SMC-16*, 148-154.

Hackley, S., Woldoroff, M., & Hillyard, S. (1990). Cross-modal selective attention effects on retinal, myogenic, brainstem and cerebral evoked potentials. *Psychophysiology, 27*, 195-208.

Hillyard, S., Hink, R., Schwent, V., & Picton, T. (1973). Electrical signs of selective attention in the human brain. *Science, 182*, 177-180.

Horst, R., Ruchkin, D., & Munson, R. (1987). Event-related potential processing negativities related to workload. In R. Johnson, J. Rohrbaugh, & R. Parasuraman (Eds.), *Current trends in event-related potential research* (pp. 186-197). Amsterdam: Elsevier.

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

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

Isreal, J., Wickens, C., Chesney, G., & Donchin, E. (1980). The event-related brain potential as an index of display monitoring workload. *Human Factors, 22*, 211-224.

Jasper, H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology, 10*, 371-375.

Jorna, P. (1992). Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. *Biological Psychology, 34*, 237-257.

Jorna, P. (1993). Heart rate and workload variations in actual and simulated flight. *Ergonomics, 36*, 1043-1054.

Kessel, C., & Wickens, C. (1982). The transfer of failure detection skills between monitoring and controlling dynamic systems. *Human Factors, 24*, 49-60.

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

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

Kramer, A. F., Sirevaag, E., & Hughes, P. (1988). Effects of foveal task load on visual spatial attention: Event-related brain potentials and performance. *Psychophysiology, 25,* 512-531.

Kramer, A. F., & Spinks, J. (1991). Capacity views of information processing. In R. Jennings & M. Coles (Eds.), *Psychophysiology of human information processing: An integration of central and autonomic nervous system approaches* (pp. 179–250). New York: Wiley.

Kramer, A. F., Wickens, C. D., & Donchin, E. (1983). An analysis of the processing demands of a complex perceptual-motor task. *Human Factors, 25,* 597–622.

Kramer, A. F., Wickens, C. D., & Donchin, E. (1985). The processing of stimulus attributes: Evidence for dual-task integrality. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 393–408.

Lindholm, E., Cheatham, C., Koriath, J., & Longridge, T. (1984). *Physiological assessment of aircraft pilot workload in simulated landing and hostile threat environments* (Tech. Report AFHRL-TR-83-49). Williams Air Force Base, AZ: Air Force Systems Command.

Makeig, S., & Inlow, M. (1993). Lapses in alertness: Coherence of fluctuations in performance and in the EEG spectrum. *Electroencephalography and Clinical Neurophysiology, 86,* 23–35.

Mangun, R., & Hillyard, S. (1990). Allocation of visual attention to spatial locations: Tradeoff functions for event-related brain potentials and detection performance. *Perception and Psychophysics, 47,* 532–550.

Mulder, L., Veldman, J., Ruddel, H., Robbe, W., & Mulder, G. (1991). On the usefulness of finger blood pressure measurements for studies on mental workload. *Homeostatis in Health and Disease, 33,* 47–60.

Mullane, M., Makeig, S., & Trejo, L. (in press). *Electrophysiological monitoring and management of operator alertness I: Experimental design* (NPRDC Technical Note). San Diego: Navy Personnel Research and Development Center.

Naatanen, R. (1990). The role of attention in auditory information processing as revealed by event-related brain potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences, 13,* 201–288.

Papanicolaou, A., & Johnstone, J. (1984). Probe evoked potentials: Theory, method and applications, *International Journal of Neuroscience, 24,* 107–131.

Parasuraman, R. (1985). Event-related brain potentials and intermodal divided attention. *Proceedings of the Human Factors Society, 29th Annual Meeting.* Santa Monica, CA: Human Factors Society.

Roscoe, A. (1992). Assessing pilot workload: Why measure heart rate, HRV and respiration? *Biological Psychology, 34,* 259–287.

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

Sheridan, T. (1987). Supervisory control. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 1243–1268). New York: Wiley.

Sirevaag, E., Kramer, A. F., Coles, M., & Donchin, E. (1989). Resource reciprocity: An event related brain potentials analysis. *Acta Psychologica, 70,* 77–97.

Sirevaag, E., Kramer, A., Wickens, C., Reisweber, M., Strayer, D., & Grenell, J. (1993). Assessment of pilot performance and workload in rotary wing helicopters. *Ergonomics, 9,* 1121–1140.

Strayer, D. L., & Kramer, A. F. (1990). Attentional requirements of automatic and controlled processing. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16,* 67–82.

Sutton, S., & Ruchkin, D. (1984). The late positive complex: Advances and new problems. In R. Karrer, J. Cohen, & P. Tueting (Eds.), *Brain and information: Event-related potentials* (pp. 1–23) New York: Annals of the New York Academy of Science.

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception and Psychophysics, 51*, 599-606.

Todd, J., & Van Gelder, P. (1979). Implications of a sustained-transient dichotomy for the measurement of human performance. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 625-638.

Trejo, L. J., Kramer, A. F., & Arnold, J. A. (1995). Event-related potentials as indices as display monitoring performance. *Biological Psychology, 40*, 33-72.

Trejo, L., Ryan-Jones, D., & Kramer, A. (1995). Attentional modulation of the mismatch negativity elicited by frequency differences between binaurally presented tone bursts. *Psychophysiology, 32*, 319-328.

Ullsperger, P., Freude, G., & Erdmann, U. (1994, October). *Novelty P3 as an index of resource allocation during mental workload.* Paper presented at Thirty-Fourth Annual Meeting of the Society for Psychophysiological Research, Atlanta, GA.

Ullsperger, P., Metz, A., & Gille, H. (1988). The P300 component of the event-related brain potential and mental effort. *Ergonomics, 31*, 1127-1137.

Veltman, J., & Gaillard, A. (1993). Indices of mental workload in a complex task environment. *Neuropsychobiology, 28*, 72-75.

Wickens, C. D. (1992). *Engineering psychology and human performance* (2nd ed.). New York: HarperCollins.

Wickens, C. D., Kramer, A. F., Vanasse, L., & Donchin, E. (1983). The performance of concurrent tasks: A psychophysiological analysis of the reciprocity of information processing resources. *Science, 221*, 1080-1082.

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

Wilson, G., & McCloskey, K. (1988). Using probe evoked potentials to determine information processing demands. *Proceedings of the Human Factors Society* (pp. 1400-1403). Santa Monica, CA: Human Factors Society.

Woldoroff, M., Hackley, S., & Hillyard, S. (1991). The effects of channel selective attention on the mismatch negativity wave elicited by deviant tones. *Psychophysiology, 28*, 30-42.

Yantis, S. (1993). Stimulus-driven attention capture. *Current Directions in Psychological Science, 2*, 156-161.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 601-620.

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 121-134.