CHAPTER TWENTY-NINE

APPLICATIONS OF PSYCHOPHYSIOLOGY TO HUMAN FACTORS

ARTHUR F. KRAMER & TIMOTHY WEBER

The main goal of this chapter is to illustrate how psychophysiological techniques – as well as scientific theories that integrate behavioral and psychophysiological levels of description – can be used to address problems and concerns in the field of human factors. In order to accomplish this goal we will begin with a brief discussion of human factors and describe a limited but important subset of current issues. Next we will describe some of the criteria that must be met for psychophysiological measures to serve a useful function in assisting human factors researchers and practitioners in enhancing the functionality, efficiency, and safety of current and future human–machine systems. We will then briefly describe a few illustrative examples of human factors issues that are addressed with a series of converging operations, which include psychophysiological measures and models. More specifically, we will focus on the topics of the evaluation of vigilance decrements (lapses in alertness), the assessment of mental workload, and the development of adaptive automated systems. However, it is important to note that we have chosen to focus on these three application domains because psychophysiology has already made inroads – in both the laboratory and applied settings – in these research areas and not because we believe that they provide the only potential applications. There are clearly a number of additional human factors issues and concerns that are ripe for psychophysiological application. These include the assessment of operator training and skill development (Hoffman 1990; Strayer & Kramer 1990), the assessment of multimodal displays (Yagi 1997), and the examination and prediction of errors and error compensation strategies in complex systems (Gehring et al. 1993; Reason 1990; Scheffers et al. 1996).

In recent years there have been a number of reviews of applications of psychophysiology topics related to human factors, most notably mental workload assessment (Gevins et al. 1995; Kramer 1991; Kramer & Spinks 1991; Parasuraman 1990; Wickens 1990; Wilson & Eggemeier 1991) and adaptive automation (Byrne & Parasuraman 1996; Kramer, Trejo, & Humphrey 1994). Given that these authors have provided extensive historical reviews of this literature, we will mainly focus our discussion on relevant research within the past decade (but see the section entitled "A Brief History" for a review of earlier work). We will also attempt, whenever possible, to focus on empirical studies that have evaluated the major topics of interest in extralaboratory environments – that is, in simulators and operational environments. Finally, given that other chapters in this volume provide an extensive treatment of the physiological mechanisms underlying the psychophysiological measures that we discuss, our discussion will focus on the utility of these measures as indices of psychological constructs pertaining to system design, system evaluation, and operator training.

A Brief Introduction to Human Factors

Human factors has been defined as the study of human capabilities and limitations that affect the design of human–machine systems (Wickens 1992). However, the field of human factors extends beyond the theoretical and empirical study of human behavior and cognition in complex systems to the formulation of guidelines, principles, and models that can be used to design systems that accommodate human users and operators (Meister 1989). In other words, human factors is neither a domain that resides solely in the laboratory nor one that focuses solely on the engineering of new (or the retrofitting of old) human–machine systems. That is, the field of human factors endeavors moreover to provide a bridge between (i) the study of human behavior and cognition in laboratory and simulated environments and (ii) the design and evaluation of human–machine systems. These systems range from (purportedly) simple consumer products such as TVs and

John T. Cacioppo, Louis G. Tassinary, and Gary G. Berntson (Eds.), *Handbook of Psychophysiology*, 2nd ed. ISBN 62634X.

VCRs to large and complex systems such as automobiles, aircraft, process control plants, and the World Wide Web.

Over the years, the core topics of interest within the human factors community have changed with the development of technology – more specifically, technologies that have reduced the need for humans to serve as manual laborers and controllers and shifted the role of humans to system managers and supervisors. Although such developments in technology have generally been advantageous for the humans who have participated in the operation of complex systems and products, there have also been a number of costs associated with the transition of humans from the role of manual controller to that of supervisor (and occasionally more active participant) in semiautomated and automated systems.

For example, one recurrent problem has been referred to as automation-induced complacency. This occurs when human operators are expected to perform a series of manual tasks while also monitoring automated systems. Under such conditions, monitoring performance often decreases precipitously rather quickly, often within 30 minutes. However, such performance decrements occur less often when the human operator's only task is to monitor the automated systems (Parasuraman, Molloy, & Singh 1993; Parasuraman, Mouloua, & Molloy 1994). Thus, if human operators are expected to perform multiple tasks, some of which require active intervention and manual control, then the monitoring of automated systems may suffer. Of course, one solution to such a problem might be to also automate the manual tasks and thereby unburden the operator from the dual tasks of manual control and supervisory management. However, such a change is often technically impractical and may also overwhelm the human operator with excessive monitoring demands. Even when highly automated systems are a practical alternative, operators have been shown to overestimate the automated systems' reliability and succumb thereby to automation-induced complacency even in the absence of manual control demands (Lee & Moray 1992; Riley 1994).

Another problem associated with highly automated systems has been referred to as out-of-the-loop unfamiliarity or, more generally, as a lack of situation awareness (Endsley 1994; Wickens 1992). This occurs when human operators must suddenly, and often without warning, get back into the control loop and either perform manual control duties and/or detect and diagnose problems with automated systems. In such cases the human operators are both slower and more error-prone in carrying out these duties than if they had been an active participant in the operation of the system rather than a passive system monitor. Important questions with regard to this problem include how to keep the operator continuously aware of the state of important systems and how best to monitor (human) operator readiness to take over important duties should automation fail (Scerbo 1994).

There is still another set of important issues arising in the context of complex semiautomated and automated systems. How should the often overwhelming amounts of multimodal information be presented to human operators, and how should we assess whether mission-critical information has been adequately extracted and retained by the operators? These general issues include questions of both a sensory and perceptual nature. Is critical information sufficiently distinct from background information? Is critical information displayed long enough for operators to note and extract task-critical components? There are also important cognitive concerns. Is the information presented to the operator in a format that is consistent with his or her mental representation of the system? Do the working memory requirements exceed operator capacity? Given the rapid development of "virtual reality" technology for operator training and system control, the sensory, perceptual, and cognitive issues associated with information presentation and multimodal integration have become more than an academic exercise. These issues are now on the verge of becoming a serious bottleneck for the effective use of this technology.

The issues raised here (and many others) have been and continue to be addressed through the application of a variety of traditional human factors methodologies. For example, the question of when to automate system functions has been addressed with several different methods, including (1) the use of models of human performance and cognition to predict the situations in which human performance is likely to degrade and (2) the continuous assessment of human performance via measurement of overt actions and responses. Indeed, the model-based and assessment-based procedures have also been combined into a hybrid approach to enhance the efficiency of adaptively automated systems (Byrne & Parasuraman 1996). Similar techniques have been employed to evaluate new display concepts and to ensure that information is presented in formats consistent with the perceptual capabilities and cognitive representations of human operators.

Role of Psychophysiology in Human Factors

Given that there are a multitude of techniques available to address human factors problems and issues, one must ask what role psychophysiology might play in human factors research and application. Certainly, to the extent that information gained through psychophysiological measurement is redundant with that obtained from other measures and models employed by the human factors community, psychophysiological measures will be unlikely to gain wide acceptance. This is likely to be the case for the foreseeable future in light of the relatively high cost of (and substantial amount of expertise required for) collecting, analyzing, and interpreting psychophysiological measures when compared with the subjective rating and performance-based measurement techniques traditionally employed by human factors practitioners and researchers.

Thus, for psychophysiological measurement techniques to gain acceptance in the human factors community, these measures must:

1. prove to be more valid or reliable indices of relevant psychological (or, in some cases, physiological) constructs than traditional behavioral and subjective measures; or
2. enable the measurement of constructs that are difficult or impossible to measure with traditional measures; or
3. enable the measurement of relevant constructs in situations where other types of measures are unavailable.

Indeed, there is evidence (to be discussed shortly) that each of these three criteria has been or can be met within a human factors context and with measures obtained via psychophysiologically inspired models.

Another important consideration is the temporal sensitivity of psychophysiological measures. In many human factors contexts – such as the evaluation of new display concepts, the examination of the effects of different environmental conditions (e.g., differences in ambient temperature, humidity) on human performance and information processing, the evaluation of training proficiency, and the assessment of fitness for duty – data can be collected and then analyzed and interpreted off-line (i.e., at a later time that might, depending on circumstances, range from minutes to days). In such situations, enough effort can be devoted to deal adequately with potential artifacts in the psychophysiological data and enough data can usually be collected to ensure adequate signal-to-noise ratios. On the other hand, there are also a number of human factors contexts that demand almost instantaneous data processing and interpretation. For example, given the increasing trend toward adaptive automation in systems such as aircraft and process control, it has become important to develop measures that can both describe and predict changes in psychological constructs such as mental workload, alertness, and information processing strategies – in real time or near-real time. Such information could then serve – along with inferences about human information processing capacities extracted from dynamic models of the interaction between humans, tasks, and environment – as input to algorithms that determine the dynamic task allocation policy between humans and automated systems. Of course, such situations pose technical problems for psychophysiological measures that are not encountered in off-line contexts, such as rapid data collection, processing, artifact rejection, and interpretation. Additionally, the bandwidth of some systems (e.g. high-performance aircraft) may be high enough to preclude collecting sufficient amounts of (at least some types of) psychophysiological data to ensure adequate reliability or signal-to-noise ratios.

One additional issue that merits some discussion is the applicability of psychophysiological measurement techniques to extralaboratory environments. Most psychophysiological research has focused on explicating the functional significance of different measures and components with relatively simple tasks in well-controlled laboratory environments. Even so, a great deal of effort has been expended on the elimination of potential artifacts (e.g., from ambient electrical fields, contamination from other physiological signals that may mask the signal of interest, individual differences in baselines, or a measure's morphology or topography). When such artifacts are difficult or impossible to eliminate during data recording, the focus has been on adjusting the physiological measures in order to minimize the impact of artifacts on data interpretation. Given the diversity and magnitude of the artifacts encountered in such well-controlled settings, is it a reasonable expectation to collect valid and reliable psychophysiological data in less well-controlled environments such as high-fidelity (and sometimes motion-based) simulators or operational environments? Although collecting psychophysiological data in such environments clearly provides a considerable technical challenge, there have been a number of promising developments in the design of miniaturized recording equipment that can withstand the rigors of operational environments (Caldwell 1995; Miller 1995; Sterman & Mann 1995). There have also been developments in pattern recognition and signal analysis techniques that enhance the detection of some physiological signals in noise (Trejo & Shensa 1993; Westerkamp & Williams 1995), as well as development of automated artifact rejection procedures (Du, Leong, & Gevins 1994). For example, Gevins et al. (1995) reported the development of a "smart helmet" system, which incorporates a combination of 32 EEG and EOG electrodes along with miniaturized preamplifiers into a flight helmet. Barring technological roadblocks, such developments should continue to increase the potential for recording psychophysiological signals in a number of complex environments.

In summary, each of the issues discussed so far needs to be carefully considered when psychophysiological measures are to be used in addressing human factors problems and concerns. Indeed, it is likely that some types of psychophysiological measures will be appropriate for only a subset of situations in which human factors issues are examined, whereas other measures may be more widely applicable. In an effort to make some of these considerations more concrete, we now turn to a critical review of current applications of psychophysiological measurement issues in the human factors field.

Psychophysiology and Human Factors: A Brief History

Human factors developed as a unique discipline in response to human performance questions that arose around the time of World War II. For the first time, systems such as military aircraft, ships, and ground vehicles were becoming sufficiently complex that more numerous (and sometimes catastrophic) errors were observed – even though the systems were functioning as designed from a mechanical standpoint. That is, human operators either could not execute their assigned functions as expected or they did not have sufficient training to do so. As a consequence of these system problems and the well-founded suspicion that systems were not being designed to ensure adequate human performance, experimental psychologists were called upon to evaluate the human–machine interface and training regimes, to diagnose the problems (and potential problems not yet observed), and to suggest system improvements and training modifications to ensure safe and efficient system operation (Fitts & Jones 1947; Mackworth 1948).

It is interesting that psychophysiological measures, principally measures of gaze direction, played an important role in the examination of human performance in complex systems during the early years of human factors. Fitts and colleagues (Fitts, Jones, & Milton 1950; Jones, Milton, & Fitts 1950; see also Gainer & Obermayer 1964) used measures of gaze direction, gaze duration, and the sequence of eye movements to examine the information extraction strategies employed by novice and experienced aircraft pilots during instrument flight. Data acquired from these studies were used to reconfigure instrument panels to optimize the speed and accuracy with which pilots could locate and extract flight-relevant information. Eye scan measures continue to be used today, in conjunction with measures of pilot performance, to assess pilot strategies for extracting information as well as mental workload and skill acquisition in military and civilian flight (Bellenkes, Wickens, & Kramer 1997; Fox et al. 1996; Kotulak & Morse 1995).

The use of other psychophysiological measures to examine issues of interest to the human factors community soon followed the pioneering research of Fitts and co-workers. For example, Sem-Jacobsen and colleagues (Sem-Jacobsen 1959, 1960, 1961; Sem-Jacobsen et al. 1959; Sem-Jacobsen & Sem-Jacobsen 1963) recorded electroencephalographic (EEG) activity from pilots as they flew missions of varying difficulty in simulated and actual flight in an effort to examine the utility of this psychophysiological measure for assessing the deleterious effects of high-G environments and mental and emotional workload. It was also speculated that psychophysiological measures, and in particular measures of the EEG, would prove useful for the selection and evaluation of pilots for high-performance aircraft and for adaptively automated systems (see also Gomer 1981). Although Sem-Jacobsen's visions for applications have not yet been realized, our review of the current literature will indicate that at least some of these applications are soon to be realized.

Finally, measures of heart rate and heart rate variability have long been used to provide a continuous record of the cardio-respiratory function and mental workload of operators in complex simulated and real-world systems. Heart rate measures have been recorded as aircraft pilots execute a number of maneuvers in simulated and actual aircraft such as landing (Ruffel-Smith 1967), refueling during long-haul flights (Brown et al. 1969), performing steep descents (Roscoe 1975), and flying combat missions (Roman, Older, & Jones 1967). Such measures continue to be used today, often in the context of other psychophysiological measures and along with a more sophisticated appreciation for the underlying physiology.

In summary, although the application of psychophysiological techniques to issues of human factors has a relatively recent history, these measures provide useful insights into human performance and cognition in extralaboratory environments. We turn now to a discussion of recent applications of psychophysiological measures to three different topics of interest to the human factors community: assessment and prediction of vigilance decrements, assessment of mental workload, and potential psychophysiological inputs to adaptively automated systems.

Assessment and Prediction of Vigilance Decrements

Interest in the human factors community in the detection and prediction of vigilance decrements – and especially in operator performance, which is crucial to mission success – has been expressed since at least the 1950s (Broadbent 1971; Broadbent & Gregory 1965; Davies & Parasuraman 1980). Indeed, early interest in vigilance decrements focused on two aspects of the phenomenon: (i) the characterization of behavioral and information processing changes that accompany reduced vigilance, often through the utilization of signal detection theory (i.e., examining the influence of vigilance changes on the sensitivity and response criteria of human observers); and (ii) changes in physiological indices of arousal (Hockey 1984; Parasuraman 1984). However, it soon became clear that arousal could not be treated as a unitary concept but instead was multidimensional in nature (Gopher & Sanders 1984; Pribram & McGuinness 1975).

A good deal of laboratory and applied research in recent years has focused on explicating the multiple interacting mechanisms that underlie the maintenance of alertness (or, conversely, the onset of sleep – Akerstedt & Folkard 1996; Lavie & Zvulini 1992). Much of this research has concentrated on work environments, such as long-haul truck driving, train driving, and transoceanic flight. In these environments, irregular hours of sleep and activity are the norm and work often occurs during the evening or nighttime hours (Boucsein & Ottmann 1996; Kecklund & Akerstedt 1993; Miller 1995).

The measurement of EEG has been the "gold standard" against which alertness has been verified in much of this work. This measure has long served as the method of choice for the categorization of stages of sleep (Goeller & Sinton 1989; Loomis, Harvey, & Hobart 1937). Also, changes in EEG, particularly in the alpha and theta bands, have been found to be predictive of performance changes in sleep-deprived individuals (Beatty et al. 1974; Smulders et al. 1997; Townsend & Johnson 1979). Other measures – such as EOG (electro-oculogram), pupil diameter, slow eye movements, respiration, electrodermal activity, and ERPs (event-related potentials) – have also demonstrated some success in tracking the onset of sleep (Harsh et al. 1993; Miller 1995; Torsvall & Akerstedt 1987, 1988; Yamamoto & Isshiki 1992). However, within the human factors community, the interest is not in categorizing sleep stages but rather in identifying and predicting when loss of alertness or increased sleepiness will detrimentally influence performance.

In recent years, there have been a number of attempts to employ psychophysiological markers of alertness to predict vigilance decrements in laboratory and extralaboratory settings. In some cases, the identification and prediction of changes in alertness resulting in performance decrements have taken place off-line. In other studies, the focus has been on using psychophysiological measures to develop on-line, closed-loop systems to predict (and sometimes reduce) performance decrements. For example, Morris and Miller (1996) recorded EOG as ten sleep-deprived military pilots flew an extended series of instrument maneuvers in a moving-base flight simulator. The investigators were able to account for 61% of the variance in a composite measure of flight performance. The composite measure included three EOG measures, blink amplitude (i.e., the extent of eyelid movement), long closure rate (i.e., the number of closures longer than 500 msec), and blink duration. Similar relationships between eye blink measures and performance in sleep-deprived individuals have been reported in both laboratory and automobile simulator studies (Stern, Boyer, & Schroeder 1994; Wierwille 1994). Such results, when considered along with the underlying neuronal substrates, have led these researchers to suggest that endogenous blinks may be a useful index of tonic activation of the rostral central nervous system – that portion of the nervous system responsible for the maintenance of alertness.

Other researchers have examined the usefulness of ERPs and EEG in tracking performance decrements that accompany decreases in alertness. Humphrey, Kramer, and Stanny (1994) examined the changes in ERPs – more specifically, the amplitude and latency of the P300 component – as sleep-deprived subjects performed memory and visual search tasks throughout the night and into the morning. Reaction times (RTs) and lapses increased, accuracy decreased, P300 latencies increased, and P300 amplitudes decreased with increasing time on task. Indeed, the magnitude of the changes in P300 latency and RT were quite similar as a function of time on task. Given that the P300 appears to be sensitive to stimulus evaluation processes while being relatively insensitive to motor processes, it would appear that decrements in performance were due, in large part, to reduced efficiency of perceptual processes (see also Koelega et al. 1992).

Two research groups have focused on developing EEG-based closed-loop systems for the prediction of vigilance decrements. Pope, Bogart, and Bartolome (1995) had participants perform a multitask consisting of monitoring, tracking, communication, and resource management subtasks, each of which could be performed manually or automatically. The EEG was recorded continuously during performance of these tasks and was used to adjust the level of automation (i.e., the number of subtasks that were performed automatically) on the basis of changes in alertness or task engagement as inferred from a number of different EEG-based metrics. The EEG-based detection algorithms were tested in two different control modes, as both a negative and positive feedback system. In the case of the negative feedback system, the objective was to detect a decrease in task engagement and then require that a subtask be performed manually. This, in turn, would result in an EEG-based indication of increased task engagement (alertness). For the positive feedback system, the objective was to increase the discrepancy between the EEG-based indication of alertness at time n and time $n+1$ by further reducing manual control requirements when reduced alertness was detected. Thus, within this control theory approach, an effective EEG-based index of alertness or task engagement would be expected to produce relatively rapid and stable oscillatory behavior with the negative feedback system. On the other hand, a slow and less stable oscillation would be expected with the positive feedback system. Several of the EEG derivations produced such behavior. However, the derivation that was most successful was beta power divided by the sum of alpha power and theta power. This is most likely due to the fact that the other derivations included either high-frequency EEG or EMG (electromyographic) components, neither of which has been reported to be sensitive to changes in alertness.

Makeig and colleagues (1990; Makeig & Inlow 1993; Makeig & Jung 1996) have investigated the efficacy of EEG-based alertness detection systems for the prediction of missed responses during simulated sonar tasks using U.S. Navy sonar operators. In their studies, sonar operators attempted to detect 300-msec noise-burst targets embedded in a white noise background. In some of the studies, particularly those in which ERPs were of interest, brief task-irrelevant tones were also occasionally presented. Operator-specific EEG algorithms were derived by employing several different frequencies in the delta, theta, and alpha bands to predict vigilance decrements. These multiple regression–based algorithms were quite successful. The algorithms developed on data from one experimental session were capable of accounting for between 75% and 85% of the variance in error rates (i.e., response omissions or lapses) obtained in a second experimental session. Other results suggested that on-line algorithms, which detected changes in theta and gamma (> 35 Hz) activity, were capable of predicting missed target detection responses up to ten seconds in advance of the occurrence of a target. Finally, ERP components (more specifically, N200 amplitude and latency and P100/N100 amplitude difference) elicited by task-irrelevant auditory probes did a reasonable job of predicting changes in error rates within a 32-sec moving window.

The results from the Pope et al. and Makeig et al. studies suggest that it is now possible, in simulated real-world tasks, to detect and predict changes in alertness or task engagement that have important implications for system performance. One wonders, however, whether physiologically based alertness detection and prediction systems could be further improved by incorporating a number of psychophysiological measures, rather than a single measure as in the Pope et al. and Makeig et al. studies. Indeed, a number of studies suggest that this may be the case. For example, Torsvall and Akerstedt (1988) reported that changes in the alpha and theta bands of the EEG, along with slow eye movements detected in the EOG, could reliably predict that a target would be missed a full minute in the future. Varri et al. (1992) reported the development of a computerized system for predicting the onset of sleepiness that incorporated measures based on EOG, EEG, and EMG. Preliminary tests of the system showed promise when compared against the scoring of the physiological data by sleep experts. Thus, it would appear that a promising area of future research is the incorporation of a number of different physiological measures into alertness detection systems. Clearly, another important direction is the transition of psychophysiologically based alertness detection systems out of the laboratory and into operational environments. This is a particularly important step, because many of the vigilance decrements observed in laboratory research have not been reported in the field. This is likely due to the differential incentives to maintain adequate performance in these two settings (Wickens 1992).

Assessment of Mental Workload

THEORY

Although there is, at present, no commonly agreed-upon definition of the construct of mental workload, it has often been conceptualized as the processing costs incurred by a human operator in the performance of a single or multiple tasks (Kramer 1991; Wickens 1992). These costs have been associated with the effort or resources required to maintain an acceptable level of performance in the face of varying environmental and task conditions.

Early models of mental workload assumed that a single capacity or undifferentiated resource was sufficient to account for performance decrements observed with changes in task difficulty or variations in external (e.g., temperature, humidity, barometric pressure, noise, lighting) or internal (e.g., fatigue, illness) stressors (Moray 1967). In these unitary capacity models, it is assumed that a single pool of capacity is available and that the requirement to perform additional tasks (or increases in task difficulty) will require the allocation of resources. As task demands continue to increase, the supply of resources is diminished and performance declines. It is interesting to note that, within some of these early models, there was a degree of elasticity in the supply of resources available for task performance. For example, in Kahneman's (1973) model, the supply of resources could be expanded to a limited extent with increases in arousal. Thus, in a sense, humans could self-regulate the quality of their performance (at least within circumscribed limits) by varying their level of arousal.

However, this self-regulation of performance through the expansion of processing resources soon became viewed as a two-edged sword. That is, while the conceptualization of resources as a somewhat elastic commodity – driven by variations in effort, arousal, and motivation – was seen as an important method of compensating for variations in task difficulty as well as internal and external stressors, it also rendered it difficult to predict patterns of task interactions in multitask environments. Thus, an important interest in the 1970s and 1980s was the description of processing resource trade-offs between concurrently performed tasks and their implications for task performance (Norman & Bobrow 1975; Sperling & Melcher 1978). In such a context, it was important to fix the total amount of resources, so that allocating x resources to one task left $1-x$ resources available for the performance of other tasks (Navon & Gopher 1979). Without such a restriction on the notion of resources or processing capacity, it would be difficult to map resource consumption to performance quality in any meaningful way – especially in the absence of a detailed knowledge of how and when other factors (e.g., arousal, effort, and motivation) influenced the quantity of available resources.

A number of data visualization tools were developed within this fixed-resource framework as an aid to the qualitative and quantitative conceptualization of the relationship between performance and resource allocation policy. For example, a performance operating characteristic (POC) is illustrated in Figure 1: performance on two concurrently performed tasks is cross-plotted to indicate the extent to which the two tasks require resources to be adequately performed. In deriving a POC, participants are asked to vary their priority, across blocks of trials, on two tasks. For example, in one block of dual-task trials, participants might be instructed to perform as best as possible on task A and devote any spare capacity to task B. In another block of trials, participants would be instructed to treat the tasks

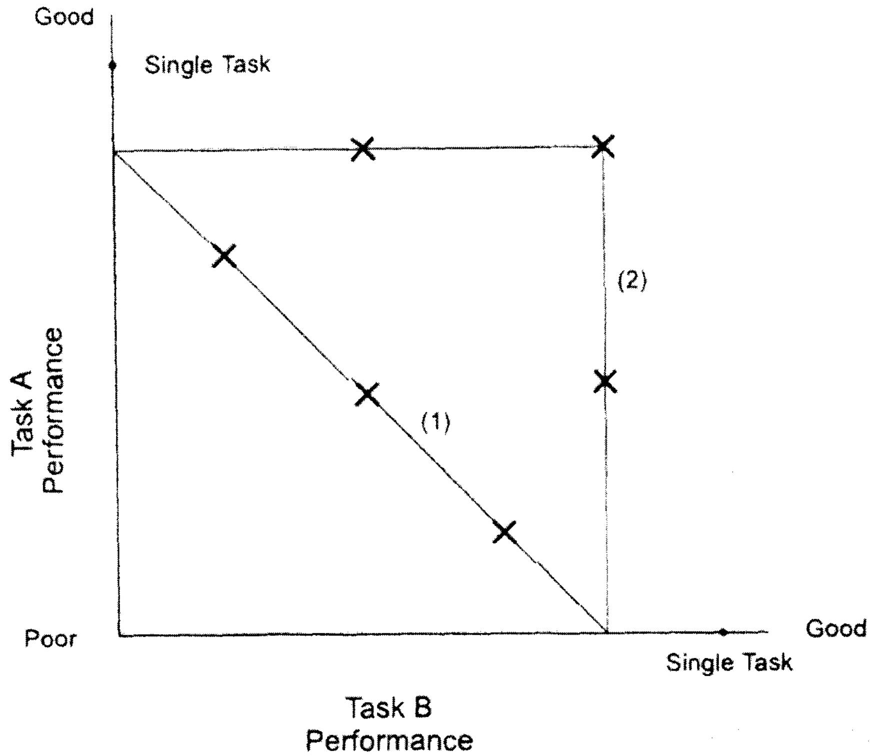

Figure 1. Illustration of a performance operating characteristic (POC). Curve 1 illustrates a situation in which the performance of one task is dependent on the performance of the other task. In such cases it is assumed that the two tasks require the same variety of processing resources or capacity. Curve 2 illustrates a situation in which the performance of one task is insensitive to the performance of the other, concurrent task. In such cases it is assumed that either the two tasks require few resources for performance (i.e., they can be performed automatically) or that the tasks require different processing resources.

with equal priority. Finally, in a third block consisting of dual-task trials, participants would be asked to favor task B. Single-task conditions, in which participants perform one of the two tasks and completely ignore the other, are also included to provide anchors for the POC.

Curve 1 illustrates a situation in which there is a 1:1 trade-off between two concurrently performed tasks; that is, performance on one task declines with increases in the performance of the other. In contrast, curve 2 illustrates a situation in which performance of one task is insensitive to the level of performance on the other, concurrent task. Although the POC illustrates only the relationship between the performance of two concurrently performed tasks, the relationship between the hypothetical construct of resources and performance can be inferred from the shape of the function in the POC. This relationship can be further illustrated in a performance resource function (PRF), which plots performance against resources.

Figure 2 illustrates the resource/performance relation that would be expected to underlie the performance functions plotted in Figure 1. Curve 1 in Figure 2 depicts the case where performance is resource-limited: performance improves in a monotonic fashion with the investment of additional resources. Such a function would be likely to underlie the performance relationship illustrated in curve 1 of Figure 1. On the other hand, curve 2 in Figure 2 indicates that task performance is data-limited; that is, the investment of additional resources beyond an initial (and minimal) allocation will have little additional influence on performance. Such a performance/resource relationship could arise for several reasons. First, the task could be very difficult, as when a human operator is required to maintain twenty unrelated pieces of information in working memory. Because such a task is obviously beyond the capabilities of most of us, investing additional resources would have little or no beneficial effect on performance. Second, a PRF function like that illustrated in curve 2 could indicate that performance on a task had been sufficiently automated that few resources are needed to achieve optimal performance – for example, the manual control of an automobile on a straight highway by an experienced driver. Curve 2 in the PRF function would likely underlie the performance relationship illustrated in curve 2 of Figure 1, where changes in the performance on one task had little influence on the performance of the other task (i.e., assuming the two tasks were data-limited).

A number of psychophysiological studies have been conducted to examine the predictions of the resource models

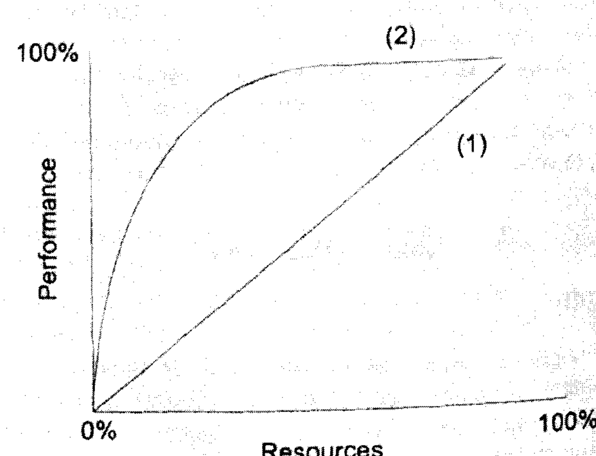

Figure 2. Illustration of a performance resource function (PRF). Curve 1 illustrates a situation in which the investment of additional processing resources results in a corresponding improvement in performance. Curve 2 represents a situation in which optimal performance is achieved after the investment of minimal resources, with the additional investment of resources having little or no effect on the quality of performance.

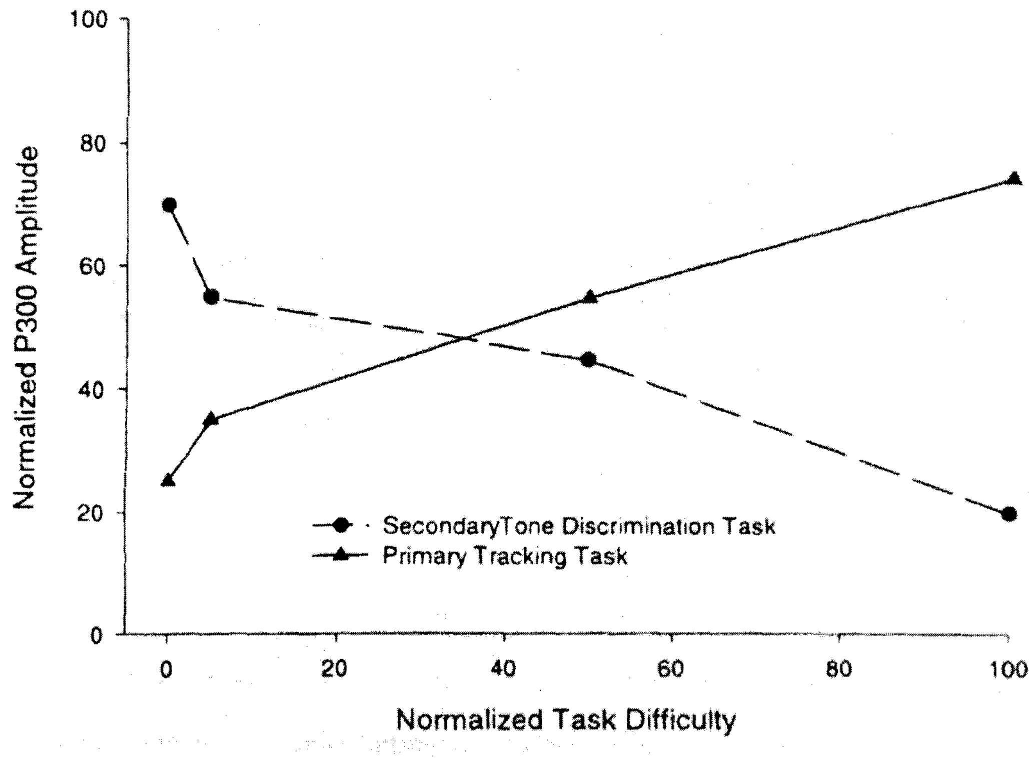

Figure 3. Illustration of a P300 reciprocity effect, where P300s elicited by events in a tracking task increase with the difficulty of the tracking task while P300s elicited by tones in a secondary tone discrimination task decrease in amplitude with increases in the difficulty of the tracking task. The performance and P300 measures were normalized (i.e., by subtracting the minimum score in each task from each condition and dividing these products by the differences between the minimum and maximum score in each task) to facilitate comparisons across the primary and secondary tasks.

of mental workload. Studies examining the P300 and N100 components of the ERP have confirmed a number of predictions of resource models of mental workload. One important prediction of resource models is that resource-limited tasks should entail a monotonic relationship between resources and performance. Kramer, Wickens, and Donchin (1985) confirmed this prediction when they found that increasing the difficulty of a tracking task systematically decreased the amplitude of the P300 component of the ERP elicited by a secondary visual discrimination task. Thus, it would appear that the P300 might provide an index of the allocation of resources: increased resource demands by the primary task would lead to diminished resources for the secondary task, which in turn would be reflected by decreases in the P300 amplitude. Indeed, this relationship between P300 and task difficulty under dual-task conditions was confirmed in a variety of studies that used tasks such as visual search, memory search, visual and auditory discrimination, and tracking (Hoffman et al. 1985; Kramer & Strayer 1988; McCallum, Cooper, & Pocock 1987; Natani & Gomer 1981; Strayer & Kramer 1990).

However, if ERPs and more specifically P300s reflect resource allocation, then we would expect a reciprocity in P300s elicited by two concurrently performed tasks. That is, fixed-capacity models (Navon & Gopher 1979; Norman & Bobrow 1975) argue that, as one task becomes more difficult or more important, additional resources will be allocated for the performance of that task, leaving fewer resources for the concurrently performed task. If P300s reflect such a process then we would expect a reciprocity in P300 amplitude, so that increases in the amplitude of P300s elicited by events in one task should be accompanied by decreases in the amplitude of P300s elicited in another, concurrent task. Indeed, such an effect has been reported. Sirevaag et al. (1987; see also Wickens et al. 1983) found that P300s elicited by changes in target position in a tracking task increased in amplitude with increases in the difficulty of the task. Correspondingly, the amplitude of P300s elicited by events in a task of lesser importance (an auditory discrimination task) decreased with increases in the difficulty of the primary (tracking) task. The results from this study are illustrated in Figure 3. Thus, it would appear that changes in P300 amplitude in dual-task studies mimic the resource reciprocity effects predicted by fixed-capacity models.

The P300 reciprocity effects are important for several reasons. First, they provide converging support for the notion of resource trade-offs among tasks, support that is independent of the performance effects traditionally used to define resource allocation policies. Second, the P300 data can be obtained in the absence of overt responses, thus enabling the assessment of resource allocation policies in situations where actual behavior occurs infrequently (e.g., a quality control inspector monitoring a visual display, a pilot monitoring the status of the aircraft while flying on autopilot). Finally, as we review next, reciprocity effects have been found for components of the ERP other than the P300, suggesting the potential of resource trade-offs for different processing operations.

Similar reciprocity effects have been reported for other components of the ERP, such as the P100, N100, and P200 (Mangun & Hillyard 1990; Parasuraman 1985). Note, however, that in the study of Mangun and Hillyard (1990), the P100 and N100 components showed reciprocity patterns like that illustrated in curve 1 of Figure 1 (or similar to that shown in Figure 3), reflecting a trade-off in the processes underlying these components between locations in a spatial attention task (e.g., attend to the left or the right to detect an infrequent target), whereas P300s and performance measures displayed a pattern more like that illustrated in curve 2 of Figure 1. The authors explained this apparent dissociation between reciprocity patterns for the earlier ERP components (P100 and N100) and the later ERP component and performance measures by suggesting that, "as attention was increasingly withdrawn from one visual field and allocated to the other, higher perceptual processes were still able to extract and analyze the information from the progressively diminishing sensory signal" (Mangun & Hillyard 1990, p. 548). In other words, more than a single variety of processing resources or capacity can underlie performance in a divided attention task.

The suggestion that a single resource or capacity is insufficient to account for the pattern of performance and processing interactions in dual-task situations has been supported by a variety of behavioral and electrophysiological studies. For example, there have been a number of reports of failures to find performance trade-offs between two resource-limited tasks (North 1977; Shaffer 1975; Wickens 1992). There have also been reports that some psychophysiological measures such as the P300 and heart rate variability (HRV) are sensitive to only a limited subset of processing demands. For example, P300 appears to be sensitive to central processing but not motor demands (Isreal et al. 1980); P100 and N100 appear to reflect early selective attention processes and more specifically the distribution of attentional resources in visual space (Mangun & Hillyard 1990). The 0.10-Hz component of HRV appears to be sensitive to working memory demands but insensitive to response or motor demands (Aasman, Mulder, & Mulder 1987; Jorna 1992). As a result of findings like these and others, a number of theorists have proposed that mental workload should be conceptualized as a multidimensional rather than a unidimensional construct, so that two concurrently performed tasks will show trade-offs only when the tasks require the same types of processing resources or capacity (Polson & Freidman 1988; Wickens 1992).

It is interesting that some psychophysiological measures appear to be diagnostic of particular varieties of processing resources whereas other measures are less diagnostic and appear instead to reflect general or undifferentiated processing demands imposed upon the human. Psychophysiological measures that fall into this latter category include respiration, heart rate (HR), eye blinks, electrodermal activity, and some components of electroencephalographic activity (Fogarty & Stern 1993; Kramer 1991; Wilson & Eggemeier 1991). The sensitivity of these measures to general or undifferentiated processing demands can have advantages and disadvantages both. On the positive side, such measures could be used in a wide variety of settings and across a number of different systems to provide a general indication of the mental workload experienced by the human operator. In many cases, such information can be extremely valuable to system designers who are interested in the overall magnitude of processing demands imposed upon the human operator. However, if more specific information concerning the type of processing demands is needed – for example, to discern whether

TABLE 1. Assessment Measures Mapped to Workload Components

	Psychological Constructs	
Physiological Measure	**Primary**	**Additional**
P300 component of the ERP	Perceptual/central processing resources	Memory updating
P100/N100 components of the ERP	Spatial attention	
EEG – alpha activity	General processing demands	Alertness
EEG – theta activity	General processing demands	Alertness
Heart rate	General processing demands	Physical and emotional workload
Heart rate variability (0.10-Hz component)	Working memory demands	Problem-solving demands
Eye scan pattern	Visual information extraction strategies	
Blink rate	Visual demands	Alertness
Blink duration	Visual demands	Alertness

Note: Listed are psychophysiological measures that have been employed in the assessment of the multidimensional construct of mental workload in applied settings, along with a hypothesized mapping between measures and mental workload components.

the response demands of a new control system or the perceptual demands of a new display configuration are responsible for increased mental workload – then more diagnostic measures will be necessary. A summary of the inferred relationship between different psychophysiological measures and different aspects or components of mental workload is presented in Table 1.

APPLICATIONS

We have provided a brief description of the theoretical framework in which psychophysiological measures of mental workload have been explored. Now we turn to a discussion of extralaboratory applications of physiological measurement of mental workload in simulator and operational environments. Aircraft flight is the context in which, by far, the most extralaboratory research has been pursued. Indeed, early applications of psychophysiological measurement of workload and other psychological constructs date back to the late 1940s. At that time, Paul Fitts and his colleagues (Fitts et al. 1950; Jones et al. 1950) employed measures of eye gaze direction to determine the manner in which pilots extracted information from instrument panels. The results of these studies led to the development of aircraft instrument panels that configured instruments on the basis of importance (defined in terms of the number and duration of eye fixations) and sequential scan strategies.

In the past decade, a number of studies have been conducted to evaluate the utility of different psychophysiological measures as metrics of mental workload in simulator and operational environments. For example, Kramer, Sirevaag, and Braune (1987) examined whether the inverse relationship between P300s elicited by secondary task events and the difficulty of a primary task, which had previously been reported in laboratory studies, would be obtained in an aircraft simulator. Seven student pilots flew an instrument flight plan with a single-engine aircraft simulator while concurrently counting one of two tones that were presented via earphones. The P300s elicited by the secondary task tones decreased in amplitude with increases in the difficulty of the flight task, which was produced by increasing turbulence and subsystem failures. Similar effects were reported for P300 amplitude and latency (Fowler 1994) and P200 amplitude (Wilson, Fullenkamp, & Davis 1994) for simulated visual flight (i.e., flight in which the pilot flies with reference only to the ground) and actual flight in military aircraft, respectively. It is important to note that, in each of these studies, ERPs were elicited by tones from an "oddball" task in which the pilots were required to either covertly count or manually respond to the relevant tones.

Even though this "relevant probe" technique appears to produce relatively consistent data in simulator and operational environments, there is an important drawback to using this method for eliciting ERPs in nonlaboratory environments. Namely, it is conceivable that the requirement to count or overtly respond to auditory (or visual) probes will increase workload in already demanding environments. Furthermore, safety concerns will likely preclude the imposition of additional task demands on operators in many real-world environments. Therefore, although the relevant probe technique would appear appropriate for the assessment for mental workload in laboratory and simulator environments, it is unlikely to see wide application in operational contexts.

Given these potential limitations of the relevant probe technique, what other psychophysiological measures might be used to examine mental workload in simulator and operational environments? There are at least three possibilities. First, an "irrelevant probe" technique has been used to elicit ERPs in aircraft simulators. In this technique, additional stimuli (e.g. tones) are presented but participants are not required to count or overtly respond to them (Papanicolaou & Johnstone 1984). Thus, at the very least, the irrelevant probe technique minimizes response or motor interference with the task of interest. Sirevaag et al. (1993) employed this technique to assess the mental workload experienced by senior helicopter pilots using a variety of different communication systems. The P300s elicited by irrelevant probes decreased in amplitude with increases in the communication load in different phases of low-level high-speed flight in a high-fidelity helicopter simulator. Thus, like the P300 effects obtained with secondary task-relevant probes, it would appear that the P300 provides an index of the residual processing resources that remain after performing the primary task (in both cases, flight control). However, there are also cases in which P300s elicited by irrelevant probes have not proven sensitive to the mental workload experienced in simulators. Kramer, Trejo, and Humphrey (1995) had ten highly trained U.S. Navy radar operators perform a simulated radar task that was varied in difficulty by manipulating the number of targets to be tracked in a limited period of time; ERPs were elicited by auditory irrelevant probes. Although the amplitude of the P300 decreased when the tone task alone was compared to the radar task with the irrelevant tones, no further decrease in P300 amplitude was observed with increases in the difficulty of the radar task.

An important question is why the irrelevant probe task was sensitive to levels of mental workload in the Sirevaag et al. (1993) study but not in the Kramer et al. (1995) experiment. Although there are a number of differences in these two studies, including the type of system being simulated, one intriguing possibility concerns the nature of the operators' tasks. In the Sirevaag et al. (1993) study, the pilots were constantly communicating with air traffic control, ground controllers, and other aircraft. That is, they were monitoring auditory messages presented via headphones. In contrast, verbal communication was quite infrequent for the radar operators in the Kramer et al.

(1995) study. Therefore, it seems conceivable that the P300 elicited by the irrelevant auditory probes reflected variations in mental workload in the Sirevaag et al. study because the auditory channel was being actively monitored and attended, whereas the auditory channel was not very important in the radar monitoring task examined in Kramer at al. (1995). In other words, it may be that P300s elicited by irrelevant probes reflect mental workload only if they occur within an attended modality or source of information. This hypothesis is consistent with the findings of Verbaten, Huyben, and Kemner (1997) and Makeig et al. (1990). These researchers found that ERPs elicited by task-irrelevant probes were sensitive to variations in mental workload in cases where the irrelevant probes were presented in the same modality as the task of interest. In any event, additional research is warranted before the irrelevant probe technique will be ready for application in simulator and operational settings.

Another alternative to the relevant probe technique for the assessment of mental workload in simulator and operational environments has been referred to as the "primary task" technique (Kramer et al. 1985). In this technique, psychophysiological measures (in particular, ERPs) are elicited by relevant events in the task of interest. For example, in the Sirevaag et al. (1993) study, P300s elicited by changes in the position of the target in the tracking task increased in amplitude with increases in the difficulty of tracking. A clear advantage of this technique is that the mental workload (as well as other processing operations) can be assessed with respect to specific events that occur within the operator's task(s). Furthermore, unlike the relevant or irrelevant probe techniques, no additional stimuli need be introduced into the task, thereby negating concerns regarding performance disruption and operator safety. Thus, within aircraft simulators and during actual flight, psychophysiological measures could be recorded based on radio communications, the presentation of navigational fixes on multifunction displays, and the occurrence of various warning indicators in the cockpit. However, there are also several drawbacks associated with the primary task technique with regard to psychophysiological measurement. First, the technique requires that relevant discrete events be found in the tasks of interest and that the simulators (or operational systems) be modified so that these events can be used as triggers for the psychophysiological measures. However, modification of complex systems – particularly operational systems – is usually quite difficult, if not impossible. Second, given the low signal-to-noise ratio of many ERP components (i.e., those psychophysiological measures that are triggered by discrete events), there must be a sufficient number of discrete events within the task of interest in order for the technique to be feasible. Given these constraints, there are limited opportunities for the use of the primary task technique in simulator and operational environments.

Nonetheless, psychophysiological measures that do not require the imposition of additional stimuli might still be employed. Indeed, the great majority of psychophysiological measures that have been used to assess mental workload in simulator and operational environments fall into this category. These measures include eye movements, blinks, heart rate, heart rate variability, respiration, electrodermal measures, hormonal measures, and EEG activity. The only potential drawback of these measures is that they are not, in general, diagnostic with regard to the varieties of mental workload that are experienced by the human operator. However, these techniques have been used successfully in a number of applied contexts to track changes in mental workload with variations in task and environmental demands.

For example, measures of HR and HRV have proven sensitive to variations in the difficulty of flight maneuvers and phases of flight (e.g., straight and level, takeoffs, landings) in both fixed-wing and rotary-wing military and commercial aviation (Boer & Veltman 1997; Roscoe 1993; Veltman & Gaillard 1996; Wilson & Fisher 1995), automobile driving (Brookhuis & de Waard 1993), air traffic control (Brookings, Wilson, & Swain 1996), and electroenergy process control (Rau 1996). In each of these cases, heart rate measures have served as a relatively continuous index of mental workload in simulated or operational environments.

It is important to note, however, that changes in heart rate and components of heart rate variability do not always produce the same pattern of effects with regard to their sensitivity to mental workload and task difficulty (Jorna 1992; Mulder 1992; Porges & Byrne 1992; Wilson 1992). This is likely due, in large part, to the fact that these measures are influenced by different physiological systems responsible for maintaining homeostasis. Thus, we wish to underscore that, although it may be relatively easy to record heart rate measures in applied situations (but see Wilson 1992 and Jorna 1992 for in-depth discussion of potential artifacts in applied settings), changes in these measures are multiply determined by different physiological systems and therefore are likely to be influenced by different physical and psychological phenomena (Berntson, Cacioppo, & Quigley 1993; Cacioppo et al. 1994). In fact, spectral analysis of heart rate is traditionally differentiated into three functionally distinct bands as follows.

1. A low-frequency band ranges from 0.02 to 0.06 Hz and appears to reflect the regulation of body temperature.
2. A mid-frequency band ranges from 0.07 to 0.14 Hz and is apparently related to the short-term regulation of blood pressure.
3. A high-frequency band ranges from 0.15 to 0.50 Hz and appears to reflect momentary respiratory influences (i.e. respiratory sinus arrhythmia, RSA) on heart rate.

The relative sensitivity of HR and HRV measures to changes in mental workload has been examined in a number of applied studies. Tattersall and Hockey (1995) analyzed HR and HRV measures recorded during a three-hour cockpit training study of eleven novice flight engineers. They reported that (i) HRV in the mid-frequency band was reduced during demanding problem-solving activities during straight and level flight and (ii) HR was higher during the take-off and landing phases of flight. The authors speculated that HR provides a sensitive index of the higher arousal or stress experienced by the flight engineers during takeoff and landing phases of flight, whereas HRV in the mid-frequency band reflects the mental effort expended during difficult problem-solving operations. Wilson (1993) recorded HR from aircraft pilots and weapon systems officers as they flew a variety of maneuvers in an F4 Phantom aircraft. For the pilots, HR increased from ground-based operations to flight as well as during a number of different flight segments (i.e., takeoff and landing, low-level flight, target acquisition). The HRs were generally lower for the weapon systems officer than for the pilot, with one important exception: weapon systems officer HRs were higher than pilot HRs during the one flight segment in which the weapon systems officers flew the aircraft.

In general, HRV was less sensitive to changes in workload than HR. For the pilot, HRV decreased from preflight to flight segments, but this pattern was reversed for the weapon systems officers. Wilson speculated that the different HRV patterns for the pilot and weapon systems officer might be attributable to different patterns of respiration. Indeed, Veltman and Gaillard (1996; see also Sirevaag et al. 1993) reported a confound between HRV in the mid-frequency band and respiration in a simulated flight task. However, when they deconfounded HRV and respiration by scaling HRV with blood pressure variability (BPV) – essentially removing the influence of respiration on the derived measure – they found that the derived HRV measure did discriminate among preflight and different flight segments. Thus, although there are important concerns with regard to respiratory confounds on mid-frequency HRV measures, there also appear to be procedures that can be used to address such concerns (see also Mulder 1992). Therefore, as suggested by our preceding discussion as well as by the mapping of psychophysiological measures to cognitive constructs (Table 1), the choice of whether HR or a component of HRV is to be employed in any particular setting should be predicated upon (i) the potential artifacts that might be encountered as well as (ii) the mental workload aspect(s) of interest to the researcher or human factors practitioner.

In addition to HR and HRV, a number of other measures have been used to assess mental workload in applied settings without requiring additional task requirements (e.g., the probe techniques necessary for elicitation of ERPs). Electro-oculographic measures – including blink rate, blink amplitude, blink duration, and saccade length and velocity – have been employed to assess mental workload in simulator and operational environments. Consistent with laboratory studies, blink rate has been found to reflect changes in mental workload in applied settings, although these changes are in visual workload rather than mental workload in general (Sirevaag et al. 1993; Wilson 1994; Wilson et al. 1994). For example, Veltman and Gaillard (1996) found that blink rate discriminated between flight conditions and landing (with lower blink rates during landing) yet was insensitive to different levels of task demands during flight. Likewise, Brookings et al. (1996) reported that blink rate discriminated among a number of different aircraft density conditions during a simulated air traffic control task (with reduced blink rates while monitoring an increasing number of aircraft) yet did not discriminate among scenarios characterized by different levels of complexity (i.e., as defined by increasing the heterogeneity of aircraft types to be controlled). These results led the researchers to conclude that – although blink rate distinguishes among different levels of *visual* load, with decreasing blink rates while extracting more information from the visual environment – blink rates are evidently insensitive to cognitive load in applied settings.

These findings and conclusions are somewhat perplexing when viewed in the context of laboratory-based research that has found blink rate to be sensitive to both visual and cognitive workload (Kramer 1991; Stern, Walrath, & Goldstein 1984). For example, Bauer, Goldstein, and Stern (1987) reported that blink rate reflected the memory processes necessary to maintain either few or many items in memory for a brief period of time, a situation in which cognitive but not visual load varied. Given such a pattern of results, why then does blink rate appear to be insensitive to cognitive load in applied settings? One possible explanation concerns the relative sensitivity of blink rate to visual and cognitive load. It is conceivable that blink rate is not sufficiently sensitive to relatively fine distinctions between high levels of cognitive load that are experienced in applied settings (e.g., the distinction between straight and level flight versus landing). That is, unlike laboratory conditions in which gradual increases in task demands are implemented, comparisons between task conditions in applied settings are often between very low-demand situations (e.g. resting baselines) and a variety of what are often high-demand task conditions. Blink measures may just not be sensitive to cognitive load differences in high-demand situations.

Blink duration measures – that is, the amount of time the eyelids are closed during a blink – have produced more variable results than have blink rate measures in applied settings. Sirevaag et al. (1993) reported decreases in blink duration with increases in communication load during rotary wing flight, regardless of whether communication was carried out visually (by reading messages from a multifunction display and responding manually) or orally. However, other researchers have either found blink duration to discriminate between the same conditions as blink rate measures (Veltman & Gaillard 1996) or to be less sensitive to visual load conditions than blink rate (Wilson 1994; Wilson et al. 1994). Thus, clearly more research is needed to discern the relative sensitivity of different blink measures to mental workload as well as to examine the range of sensitivity of electro-oculographic measures to cognitive aspects of mental workload in applied settings.

Unlike ERPs, which require the presentation of a discrete event for their elicitation, electroencephalographic measures can be recorded independently of ongoing stimulus and response activity. Indeed, EEG measures in the form of the traditional frequency bands (see Chapter 2 of this volume for an in-depth discussion of the spectral decomposition of EEG activity) have been used in a limited number of applied settings (aircraft flight and automobile driving) as indices of mental workload. Measures of EEG have more frequently been employed in the assessment of low levels of arousal in vigilance situations and as input for closed-loop adaptive systems to monitor alertness in real time or near-real time. These applications will be discussed next.

Sterman and colleagues (1994; Sterman & Mann 1995; Sterman, Mann, & Kaiser 1992) examined EEG changes in the 8–12-Hz (alpha) band during a series of simulated and operational military flights. For example, Sterman et al. (1992) reported a systematic decrease in the power of 8–12-Hz EEG activity as control responsiveness was degraded in a T4 aircraft. Furthermore, these spectral changes were sensitive to the time course of the variations in task difficulty within the flights. Sterman and Mann (1995) likewise reported graded decreases in this alpha power as U.S. Air Force pilots flew progressively more difficult in-flight refueling missions in a B2 aircraft simulator. As in their previous study, alpha suppression varied within each flight according to momentary demands of the tasks and mission. Brookings et al. (1996) reported changes in both theta (4–8-Hz) and alpha (8–12-Hz) power as a number of experienced air traffic controllers performed a series of control tasks varying in task complexity and aircraft density. Alpha power decreased with increases in the heterogeneity of the aircraft types to be controlled, whereas theta power increased with the absolute number of aircraft to be controlled. Brookhuis and de Waard (1993) reported that a derived EEG measure – (alpha + theta)/beta – reflected the difficulty of an automobile driving task, decreasing in power as driving difficulty increased.

Thus, the EEG measures, particularly in the alpha and theta bands, have proven sensitive to variations of mental workload in applied settings. Furthermore, these measures have the potential to track momentary fluctuations in mental workload that result from relatively rapid changes in task demands. However, it still remains to be determined whether changes in these components of the EEG reflect general variations in the arousal or preparatory state of the organism or instead more specific cognitive operations or processes. Topographic analyses of spectral changes across the scalp are likely to provide insight into this issue (Andrew & Pfurtscheller 1996; Gevins et al. 1995; Sterman & Mann 1995).

In summary, a number of psychophysiological recording techniques that have been developed in the laboratory have been successfully implemented in extralaboratory contexts such as simulators and operational environments. Although the psychophysiological measures discussed here have proven useful in the assessment of mental workload in applied contexts, there is clearly a need for the development of additional and more efficacious procedures for signal extraction, pattern recognition, and artifact detection and compensation if psychophysiological measures are to become more widely used in the human factors context.

In addition to the need for further development of technology and methodology, there is also a need for a reconsideration of the theoretical framework in which mental workload has been examined in applied contexts. As described earlier in this section, the modal view of mental workload has been the fixed (unitary or multidimensional) resource or capacity view. Although this conceptualization has provided a reasonable starting point for the examination of performance trade-offs between concurrently performed tasks in applied settings, it is unnecessarily restrictive when complex task performance is considered within the context of the stresses (e.g., sleep loss, fatigue, illness, variations in motivation) of everyday life. That is, the fixed-capacity view does not permit the compensatory and often strategic control of performance that is needed to cope with the common stressors encountered in most jobs and tasks. Indeed, theorists such as Hockey (1997; see also Gaillard 1993; Gopher & Sanders 1984; Pribram & McGuinness 1975) have argued the need for multiple-level compensatory control mechanisms. According to this view, performance is monitored, and if it is found to be deficient then either of two solutions is available. In one solution, additional resources are recruited to improve performance – with the assumption of concomitant expenses in subjective effort and behavioral and physiological cost (both short- and potentially long-term costs in the form of psychosomatic illnesses). In the second solution, performance goals are revised and performance strategies are modified. Of course, the examination of mental workload within such a theoretical framework, although much more inclusive in terms of the important factors such as stress effects, also requires a more detailed understanding of the interaction among physiological, behavioral, and subjective factors that influence performance in applied contexts. Yet such a level of complexity is clearly necessary to develop an adequate understanding of human performance and information processing in extralaboratory environments.

The fixed-resource notion has also come under attack in recent years by researchers who have suggested that, rather than conceptualizing multitask performance in terms of graded capacity sharing, additional consideration should be given to all-or-none performance trade-offs and strategic modifications of processing strategies under difficult dual-task conditions (Allport 1987; Meyer & Kieras 1997; Navon & Miller 1987). Indeed, Pashler (1994) provided relatively strong empirical evidence in favor of bottleneck models (cf. Broadbent 1958). According to these models, the availability of a processing operation is restricted to a single task at a time. Therefore, multitask decrements can be attributed to the delay in the availability of the processing mechanisms rather than a lack of capacity or resources.

The important question, however, is what implications these theoretical considerations have for the psychophysiological assessment of mental workload in applied settings. At the very least, such considerations suggest a broader role for psychophysiological measures. For example, the concern for compensatory regulation of behavior suggests that psychophysiological measures may serve a role in assessing and predicting short- and long-term physiological or health costs as well as in assessing momentary fluctuations in mental workload. Indeed, psychophysiological measures have played an increasingly important role in the investigation of the efficacy of different coping styles in the workplace (Gaillard & Kramer in press). Similarly, the concern with changes in information processing strategies with variations in mental workload could be addressed with psychophysiological measures that have not been traditionally used in workload assessment. For example, ERP components have been identified and characterized that would provide additional insights concerning: the monitoring and detection of errors (the error related negativity – Gehring et al. 1993; Scheffers et al. 1996); the programming and execution of overt actions (the lateralized readiness potential – Coles, Scheffers, & Fournier 1995; Osman, Moore, & Ulrich 1995); the updating of working memory (the slow-wave component – Rosler, Heil, & Roder 1997); and the detection of semantically incongruent information (the N400 – Kutas & Van Petten 1994). Thus, psychophysiology could play an important role in the examination of a broadened conceptualization of mental workload in applied settings.

Psychophysiological Inputs to Adaptively Automated Systems

In a previous section (Assessment and Prediction of Vigilance Decrements), we briefly described two alertness detection systems (Makeig & Inlow 1993; Pope et al. 1995) that, in essence, provide the possibility for the allocation of tasks between humans and machines on the basis of an assessment of the human operator's level of alertness or task engagement. In many ways, the application of psychophysiology to the on-line detection and prediction of vigilance decrements is well on its way to implementation in operational systems, at least in those systems where operator movements and physical activity are somewhat constrained (e.g., automobile driving, train driving, piloting, quality control inspection, process control). This is due in large part to the fairly well-developed concept of alertness at both a physiological and a psychological level of description. Additionally, there is a substantial body of empirical research both in the laboratory and in simulators that suggests physiological measures can indeed be successfully employed to detect and predict vigilance decrements, particularly in terms of response omissions or lapses.

However, can we expect on-line applications of psychophysiological measures to other issues of concern to human factors professionals, such as the assessment of changes in mental workload or the direction of attention? In many ways, the detection and prediction of behavioral implications of variations in mental workload and attention are harder than assessing performance implications of lapses in alertness. This follows because in many cases the major concern in alertness detection is whether behavior is present or absent – that is, whether the human operator is awake and performing or asleep. Although this is certainly somewhat of an oversimplification of the vigilance or alertness problem(s), just being able to make the binary decision that the operator will be awake or asleep during critical task periods would be a major contribution to the field. However, the assessment of mental workload and attention goes beyond simply making a binary decision (although in many cases even this would be a great improvement in our knowledge about the psychological state of the human operator). Instead, it often involves the assessment of graded changes in performance quality and oftentimes changes in information processing and performance strategies. Thus, these areas of application of psychophysiology demand much greater precision of measurement than does alertness detection.

Given these situational constraints, several important questions must be answered before psychophysiological measures can be employed in real-time adaptive systems. One question, which has been addressed (in part) by research discussed in the previous section on the assessment of mental workload, is the sensitivity of different psychophysiological measures to levels and types of processing demand or mental workload. As discussed earlier, some psychophysiological measures (e.g., ERP components) are quite diagnostic with regard to the nature of processing demands. Other measures, however, are sensitive to changes in a host of psychological and sometimes physical constructs (e.g., respiration, eye blinks, electrodermal activity) but are not diagnostic of specific types of demands. Clearly, the decision of which measure to employ depends upon the nature of the question, particularly in terms of the system performance implications for detecting and predicting general or specific changes in psychological processes. Another related question – which has received much less attention but is no less important in the context of real-world systems – is whether particular psychophysiological measures are sensitive to the entire range or only a limited range of the psychological construct of interest (e.g., mental workload, attention, alertness).

A number of studies have examined the extent to which EEG can be used to distinguish among the types of processing required to perform different tasks as a starting point for the development of psychophysiologically based communication systems. For example, Wilson and Fisher (1995) examined the extent to which EEG data could be used to classify which of 14 different tasks (e.g., simple auditory and visual RT, spatial processing, memory search, visual monitoring) a subject was performing. A principal components analysis (PCA) was used to determine the EEG frequency bands, which were then submitted to a stepwise discriminant analysis procedure. This was done to classify the EEG according to which of the different tasks was being performed during the recording (see also Mecklinger, Kramer, & Strayer 1992 for further discussion of the PCA technique applied to EEG frequency band determination). An average classification accuracy of 86% (with a range of 61%–95%) was achieved across seven participants. However, the frequencies above 30 Hz at lateral recording sites were heavily represented in the classifier. Thus, it is conceivable that the high classification accuracy might to some extent be due to muscle activity of the neck and scalp rather than to electrical activity of the brain. Kerin and Aunon (1990) performed a similar study in an effort to determine whether EEG frequency band asymmetry ratios (the ratio of power in the traditional frequency bands across homologous sites on the right and left side of the scalp) obtained from 2-sec data samples could be used to discriminate among the performance of a variety of different tasks (mental rotation, mental multiplication, mental composition of a letter, visual imagery) under ten different experimental conditions. In this case, however, the investigators ensured that the EEG effects were not contaminated by muscle artifacts. The classification accuracies (percentages) ranged from the mid-80s to 90s, which were similar to those reported by Wilson and Fisher (1995). Unfortunately, however, it is difficult to directly compare the findings of the Kerin and Aunon (1990) and Wilson and Fisher (1995) studies owing to their use of different classification procedures and tasks. Nonetheless, the data from the two studies are promising and suggest that the varieties of processing associated with different types of tasks can be distinguished by an on-line analysis of relatively short samples of EEG. Clearly, however, additional studies are needed that systematically compare the efficacy of different classification algorithms with a large corpus of perceptual, cognitive, and psychomotor tasks.

Other researchers have demonstrated that on-line analysis of EEG and ERPs can be used to communicate at least two-state information to a computer. Farwell and Donchin (1988) developed an ERP (P300-based) communication system in which participants attended to 1 of 36 cells in a 6 × 6 matrix of letters and symbols. The rows and columns of the matrix were randomly flashed and ERPs were elicited by the flashes. Discriminant analysis algorithms were developed to capitalize on differences in amplitude in the P300 and slow wave, which discriminated between attended and unattended elements in the matrix. The system was able to communicate, with 95% accuracy, approximately 2.3 letters per minute. Although this communication rate was quite slow, the system was far from optimized in that only a portion of the ERP waveform was used from a single scalp site for classification. Indeed, Humphrey and Kramer (1994) demonstrated classification accuracies above 95% between conditions of high versus low workload. They used dual tasks with average ERPs consisting of 10 sec of data (i.e., ten 1-sec ERPs) and combined the ERP data from several electrodes (see also Trejo, Kramer, & Arnold 1995).

Other researchers (Pfurtscheller et al. 1996; Wolpaw & McFarland 1994) have trained participants to utilize their EEG to move cursors around a computer screen. Wolpaw et al. (1991) trained participants, over the course of five one-hour sessions, to modify the amplitude of their 8–12-Hz mu rhythm to move a cursor into a target. The target was randomly positioned at the top or the bottom of a computer screen. Participants learned to modify their mu rhythm, which was recorded over the motor cortex, by thinking about performing either a physical activity such as lifting weights (which increased the amplitude of the mu rhythm) or relaxing (which decreased the amplitude of the mu rhythm). The investigators persuasively ruled out contamination of the mu rhythm by eye blinks or by other potential activity that might increase local EMG activity.

In terms of assessing the direction of attention, particularly in the visual domain, another promising technology is eye tracking. In recent years eye trackers have evolved from cumbersome devices that require constraining the observer with a chin rest and bite bar – which, of course, precluded speaking and many other complex actions – to relatively light, head-mounted devices or (in some cases) completely unobtrusive recording devices. Although there are certainly still constraints on the conditions under which eye movements can be reliably recorded (e.g., the observer must be relatively nonambulatory and, in the case of nonobtrusive eye trackers, facing the tracker), it is now possible to record the position of the eyes – with relatively high temporal and spatial precision – in a laboratory, a simulator, and a number of operational environments. Indeed, given the research suggesting visual attention and eye position are often closely coupled (Deubel & Schneider 1996; Kowler et al. 1995; Zelinsky & Sheinberg 1997; but see Fox et al. 1996), eye trackers can be used to dynamically track the allocation of attention to different regions of the visual field.

Such information can be extremely valuable in the context of on-line monitoring of operator information extraction strategies. This ensures that critical information has been noted and that operators are sampling information with sufficient frequency to ensure an up-to-date mental model of the task and environment.

In summary, although the application of psychophysiology to on-line assessment is still in its infancy, the laboratory and simulator research that has been conducted thus far has made a promising start toward the development of physiologically based on-line assessment of alertness, attention, and mental workload. Clearly, continued progress will depend on the integration of multiple psychophysiological measures with other measures of the constructs of interest. It depends, as well, on the continued development of signal detection and pattern recognition techniques that can hasten the extraction of the measures from the background noise experienced in simulator and operational settings.

Psychophysiological Inference in Human Factors

Human factors researchers and practitioners have used psychophysiological measures in at least two different ways to make inferences about important psychological processes in applied settings. At the most fundamental level, human factors researchers have used psychophysiological measures as an index of whether two conditions, systems, or individuals differ. In such a case, the main interest is often whether a particular display, control device, or novel design produces a general difference in brain function (e.g., via measures of EEG or ERPs) or autonomic nervous system responsivity (via measures of heart rate, heart rate variability, or respiration). The psychophysiological information obtained in these studies, often along with behavioral and subjective assessments, is then used to decide whether the modified system produces equivalent human responses as compared to a baseline system (especially when additional functions or features are added to the system and the question is whether the human operator can still adequately perform the requisite tasks) or perhaps whether the modified system has led to an enhancement in human responses.

However, the ability to discern whether two systems result in a difference in the physiological responsivity of human operators is not always sufficient. In many cases, the human factors researcher is interested in discerning the nature of the psychological difference engendered by two or more tasks or systems. Only with such information can the proposed tasks or systems be further refined or modified to best accommodate the human operator. For example, in order to reduce mental workload it is often important to know whether high levels of workload are the result of excessive perceptual, memory, or motor demands (or some combination of these different types of processing demands). Such knowledge can then enable system modifications that are targeted to the specific nature of processing demands, thereby reducing the time and cost necessary for system improvements. Indeed, as has been illustrated in previous discussions and the hypothesized mapping in Table 1, psychophysiological measures have been used to indicate whether specific types (and magnitudes) of processing demands differ across systems, settings, and individuals in applied contexts such as piloting, air traffic control, sonar and radar monitoring, and automobile and truck driving. However, although such inferences are routinely made, it is important to keep in mind that the mapping between physiological measures and psychological constructs is rarely one-to-one (Cacioppo & Tassinary 1990). That is, as illustrated in Table 1, the great majority of psychophysiological measures are related in a one-to-many fashion with psychological constructs, so there clearly is a need for the use of converging operations (and measures – including physiological, behavioral, and subjective) to isolate the influence of system changes on psychological processes.

Conclusions

In this chapter we have provided a brief synopsis of several current issues in the field of human factors that would likely benefit from the application of psychophysiological techniques and psychophysiologically inspired models and theories. In discussing each of these potential application areas, we have endeavored, whenever possible, to describe studies in which psychophysiological measures have been used to address issues of concern to the human factors community in applied settings – that is, in complex simulators and in operational environments. Indeed, if psychophysiology is to make a lasting contribution to the field of human factors, it is important that we "transition" our measurement techniques from the relatively sterile yet well-controlled environment of the laboratory to the much richer but less controlled operational settings. Clearly, as evidenced by our critical review of the literature, such transitions are beginning to take place.

In each of the research and application domains that we discussed – the assessment of mental workload, the detection and prediction of lapses in alertness, and the on-line assessment of information processing activities and strategies – there have been demonstrations of successful applications of psychophysiology. In these cases, psychophysiological measurement has either (a) provided converging support, along with performance and subjective measures, of important changes in information processing strategies, alertness, or attention, or (b) provided insights that were not available with other measures – for example, by providing information concerning physiological coping strategies with implications for short- and long-term psychological and physical health, by indicating changes in resource allocation strategies with implications for multitask performance, and by predicting when vigilance decrements will be observed. Clearly, given the continued development of semiautomated and automated systems in which human operators monitor rather than actively control system functions, there will be numerous additional opportunities for the use of psychophysiological measures to provide insights into the covert processes of the mind.

However, in each of the research domains that we have discussed, there remain a number of important challenges for psychophysiological measurement. These challenges include:

1. the development and further refinement of signal extraction, pattern recognition, and artifact rejection and compensation algorithms that can be employed in relatively noisy environments;
2. the continued development of physiological and psychological models of psychophysiological measures; and
3. the mapping of these models and measures to models developed by other research domains, such as cognitive science and neuroscience.

Indeed, there appears to be activity on each of these fronts and in particular on the integration of psychophysiological, neuroscience, cognition, and emotion in the development of macro models of human psychological function.

NOTE

The preparation of this chapter was supported by the U.S. Army Research Laboratory under a cooperative research agreement (DAAL01-96-2-0003) and by a grant from the Office of Naval Research N00014-93-1-0253.

REFERENCES

Aasman, J., Mulder, G., & Mulder, L. (1987). Operator effort and the measurement of heart-rate variability. *Human Factors, 29*, 161–70.

Akerstedt, T., & Folkard, S. (1996). Predicting sleep latency from the three-process model of alertness. *Psychophysiology, 33*, 385–9.

Allport, A. (1987). Selection for action: Some behavioral and neurophysiological considerations of attention and action. In H. Heuer & A. Sanders (Eds.), *Perspectives on Perception and Action*, pp. 395–418. Hillsdale, NJ: Erlbaum.

Andrew, C., & Pfurtscheller, G. (1996). Event-related coherence as a tool for studying dynamic interactions of brain regions. *Electroencephalography and Clinical Neurophysiology, 98*, 144–8.

Bauer, L. O., Goldstein, R., & Stern, J. A. (1987). Effects of information processing demands on physiological response patterns. *Human Factors, 29*, 213–34.

Beatty, J., Greenberg, A., Deibler, W., & O'Hanlon, J. (1974). Operant control of occipital theta rhythm affects performance in a radar monitoring task. *Science, 183*, 871–3.

Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual scanning and pilot expertise: The role of attentional flexibility and mental model development. *Aviation, Space and Environmental Medicine, 68*, 869–79.

Berntson, G. G., Cacioppo, J. T., & Quigley, K. S. (1993). Respiratory sinus arrhythmia: Autonomic origins, physiological mechanisms, and psychophysiological implications. *Psychophysiology, 30*, 183–96.

Boer, L. C., & Veltman, J. A. (1997). From workload assessment to system improvement. Paper presented at the NATO Workshop on Technologies in Human Engineering Testing and Evaluation (June 24–26, Brussels).

Boucsein, W., & Ottmann, W. (1996). Psychophysiological stress effects from the combination of night-shift work and noise. *Biological Psychology, 42*, 301–32.

Broadbent, D. (1958). *Perception and Communication.* London: Pergamon.

Broadbent, D. (1971). *Decision and Stress.* New York: Academic Press.

Broadbent, D., & Gregory, M. (1965). Effects of noise and signal rate upon vigilance as analyzed by means of decision theory. *Human Factors, 7*, 155–62.

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

Brookings, J., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology, 42*, 361–77.

Brown, W. K., Rogge, J. F., Buckley, C. J., & Brown, C. A. (1969). Aeromedical aspects of the first non-stop transatlantic helicopter flight: II. Heart rate and ECG changes. *Aerospace Medicine, 40*, 714–17.

Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology, 42*, 249–68.

Cacioppo, J. T., Berntson, G. G., Binkley, P. F., Quigley, K. S., Uchino, B. N., & Fieldstone, A. (1994). Autonomic cardiac control. II. Noninvasive indices and basal response as revealed by autonomic blockades. *Psychophysiology, 31*, 586–98.

Cacioppo, J. T., & Tassinary, L. G. (1990). Psychophysiology and psychophysiological inference. In J. T. Cacioppo & L. G. Tassinary (Eds.), *Principles of Psychophysiology: Physical, Social and Inferential Aspects*, pp. 3–33. Cambridge University Press.

Caldwell, J. (1995). Assessing the impact of stressors on performance: Observations on levels of analysis. *Biological Psychology, 40*, 197–208.

Coles, M. G. H., Scheffers, M., & Fournier, L. (1995). Where did you go wrong? Errors, partial errors, and the nature of human information processing. *Acta Psychologica, 90*, 129–44.

Davies, D. R., & Parasuraman, R. (1980). *The Psychology of Vigilance.* London: Academic Press.

Deubel, H., & Schneider, W. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research, 36*, 1827–37.

Du, W., Leong, H., & Gevins, A. (1994). Ocular artifact rejection by adaptive filtering. Paper presented at the 7th IEEE SP Workshop on Statistical Signal and Array Processing (Quebec City, CA).

Endsley, M. R. (1994). Automation and situation awareness. In M. Mouloua & R. Parasuraman (Eds.), *Human Performance in Automated Systems: Current Research and Trends*, pp. 163–81. Hillsdale, NJ: Erlbaum.

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*, 510–23.

Fitts, P., & Jones, R. (1947). Analysis of factors contributing to 460 "pilot error" experiences in operating aircraft controls. Memorandum report no. TSEA 4-694-12, Harry Armstrong Aerospace Medical Research Laboratory, Wright-Patterson AFB, Ohio.

Fitts, P., Jones, R., & Milton, J. (1950). Eye fixations of aircraft pilots III: Frequency, duration and sequence of fixations while flying Air Force Ground Controlled Approach System (GCA). Air Material Command Technical report no. USAF TR-5967.

Fogarty, C., & Stern, J. (1993). Eye movements and blinks: Their relationship to higher cognitive processes. *International Journal of Psychophysiology, 8*, 35–42.

Fowler, B. (1994). P300 as a measure of workload during a simulated aircraft landing task. *Human Factors, 36*, 670–83.

Fox, J., Merwin, D., Marsh, R., McConkie, G., & Kramer, A. (1996). Information extraction during instrument flight: An evaluation of the validity of the eye–mind hypothesis. Proceedings of the Human Factors Society, 40th Annual Meeting (Philadelphia).

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

Gaillard, A. W. K., & Kramer, A. F. (in press). Theoretical and methodological considerations in psychophysiological research. In R. Backs (Ed.), *Engineering Psychophysiology*. Hillsdale, NJ: Erlbaum.

Gainer C. A., & Obermayer, R. W. (1964). Pilot eye fixations while flying selected maneuvers using two instrument panels. *Human Factors, 6*, 485–501.

Gehring, W., Goss, B., Coles, M. G. H., Meyer, D., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science, 4*, 385–90.

Gevins, A., Leong, H., Du, R., Smith, M., Le, J., DuRousseau, D., Zhang, J., & Libove, J. (1995). Towards measurement of brain function in operational environments. *Biological Psychology, 40*, 169–86.

Goeller, C., & Sinton, C. (1989). A microcomputer-based sleep stage analyzer. *Computer Methods and Programs in Biomedicine, 29*, 31–6.

Gomer, F. E. (1981). Physiological monitoring and the concept of adaptive systems. In J. Morel & K. F. Kraiss (Eds.), *Manned Systems Design*, pp. 271–87. New York: Plenum.

Gopher, D., & Sanders, A. F. (1984). S-Oh-R: Oh stages! Oh resources! In W. Printz & A. F. Sanders (Eds.), *Cognition and Motor Processes*. Berlin: Springer-Verlag.

Harsh, J., Voss, U., Hull, J., Schrepfer, S., & Badia, P. (1993). ERP and behavioral changes during the wake/sleep transition. *Psychophysiology, 31*, 244–52.

Hockey, G. R. J. (1984). Varieties of attentional state: The effects of environment. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of Attention*, pp. 449–84. New York: Academic Press.

Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive energetical framework. *Biological Psychology, 45*, 73–94.

Hoffman, J. (1990). Event-related potentials and automatic and controlled processes. In J. Rohrbaugh, R. Parasuraman, & R. Johnson (Eds.), *Event-Related Brain Potentials: Basic Issues and Applications*, pp. 145–57. New York: Oxford University Press.

Hoffman, J., Houck, M., MacMillian, F., Simons, R., & Oatman, L. (1985). Event related potentials elicited by automatic targets: A dual-task analysis. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 50–61.

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

Humphrey, D., Kramer, A. F., & Stanny, R. (1994). Influence of extended-wakefulness on automatic and non-automatic processing. *Human Factors, 36*, 652–69.

Isreal, J., Chesney, G., Wickens, C., & Donchin, E. (1980). P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. *Psychophysiology, 17*, 259–73.

Jones, R., Milton, J., & Fitts, P. (1950). Eye fixations of aircraft pilots IV: Frequency, duration and sequence of fixations during routine instrument flight. Technical report no. 5795, Wright-Patterson AFB, Ohio.

Jorna, P. G. A. M. (1992). Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. *Biological Psychology, 34*, 237–57.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kecklund, G., & Akerstedt, T. (1993). Sleepiness in long distance truck driving: An ambulatory EEG study of night driving. *Ergonomics, 36*, 1007–17.

Kerin, Z., & Aunon, J. (1990). Man–machine communications through brain processing. *IEEE Engineering in Medicine and Biology, 90*, 55–7.

Koelega, H. S., Verbaten, M. N., van Leeuwen, T. H., Kenemans, J. L., Kemner, C., & Sjouw, W. (1992). Time effects on event-related brain potentials and vigilance performance. *Biological Psychology, 34*, 59–86.

Kotulak, J. C., & Morse, S. E. (1995). Oculomotor responses with aviator helmet-mounted displays and their relation to in-flight symptoms. *Human Factors, 37*, 699–710.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research, 35*, 1897–1916.

Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. In D. 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*, 145–60.

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., & Strayer, D. (1988). Assessing the development of automatic processing: An application of dual-task and event-related brain potential methodologies. *Biological Psychology, 26*, 231–68.

Kramer, A. F., Trejo, L., & Humphrey, D. (1994). Psychophysiological measures of workload: Potential applications to adaptively automated systems. In M. Mouloua & R. Parasuraman (Eds.), *Human Performance in Automated Systems: Current Research and Trends*, pp. 137–62. Hillsdale, NJ: Erlbaum.

Kramer, A. F., Trejo, L., & Humphrey, D. (1995). Assessment of mental workload with task-irrelevant auditory probes. *Biological Psychology, 40*, 83–100.

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

Kutas, M., & Van Petten, C. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M. Gernsbacher (Ed.), *Handbook of Psycholinguistics*, pp. 83–144. San Diego: Academic Press.

Lavie, P., & Zvulini, A. (1992). The 24 hour sleep propensity function: Experimental bases for somnotypology. *Psychophysiology, 29*, 566–75.

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human–machine systems. *Ergonomics, 35*, 1243–70.

Loomis, A. L., Harvey, E., & Hobart, G. A. (1937). Cerebral states during sleep as studied by brain potentials. *Journal of Experimental Psychology, 21*, 127–44.

Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology, 1*, 5–61.

Makeig, S., Elliot, F. S., Inlow, M., & Kobus, D. A. (1990). Predicting lapses in vigilance using brain evoked responses to irrelevant auditory probes. Unpublished manuscript.

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

Makeig, S., & Jung, T. P. (1996). Tonic, phasic and transient EEG correlates of auditory awareness in drowsiness. *Cognitive Brain Research, 4*, 15–25.

Mangun, G. R., & Hillyard, S. A. (1990). Allocation of visual attention to spatial locations: Tradeoff functions for event-related brain potentials and detection performance. *Perception and Psychophysics, 47*, 532–50.

McCallum, C., Cooper, R., & Pocock, P. (1987). Event-related and steady state changes in the brain related to workload during tracking. In K. Jensen (Ed.), *Electric and Magnetic Activity of the Central Nervous System: Research and Clinical Applications in Aerospace Medicine*. France: NATO AGARD.

Mecklinger, A., Kramer, A. F., & Strayer, D. L. (1992). Event related potentials and EEG components in a semantic memory search task. *Psychophysiology, 29*, 104–19.

Meister, D. (1989). *Conceptual Aspects of Human Factors*. Baltimore: Johns Hopkins University Press.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive control processes and multiple-task performance: Part I. Basic Mechanisms. *Psychological Review, 104*, 3–65.

Miller, J. C. (1995). Batch processing of 10,000 hours of truck driver EEG data. *Biological Psychology, 40*, 209–22.

Moray, N. (1967). Where is capacity limited? A survey and a model. *Acta Psychologica, 27*, 84–92.

Morris, T. L., & Miller, J. C. (1996). Electro-oculographic and performance indices of fatigue during simulated flight. *Biological Psychology, 42*, 343–60.

Mulder, L. J. M. (1992). Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological Psychology, 34*, 205–36.

Natani, K., & Gomer, F. (1981). Electrocortical activity and operator workload: A comparison of changes in the electroencephalogram and event-related potentials. Technical report no. MDC E2427, McDonnel Douglas Corporation, St. Louis.

Navon, D., & Gopher, D. (1979). On the economy of the human processing system. *Psychological Review, 86*, 214–55.

Navon, D., & Miller, J. (1987). The role of outcome conflict in dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 435–48.

Norman, D., & Bobrow, D. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7*, 44–64.

North, R. (1977). Task functional demands as factors in dual-task performance. In *Proceedings of the 21st Annual Meeting of the Human Factors Society*. San Francisco: Human Factors Society.

Osman, A., Moore, C., & Ulrich, R. (1995). Bisecting RT with lateralized readiness potentials: Precue effects after LRP onset. *Acta Psychologica, 90*, 111–27.

Papanicolaou, A. C., & Johnstone, J. (1984). Probe evoked potentials: Theory, method and applications. *International Journal of Neuroscience, 24*, 107–31.

Parasuraman, R. (1984). Sustained attention in detection and discrimination. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of Attention*, pp. 243–72. New York: Academic Press.

Parasuraman, R. (1985). Event related brain potentials and intermodal divided attention. *Proceedings of the Human Factors Society, 29*, 971–5.

Parasuraman, R. (1990). Event-related brain potentials and human factors research. In J. Rohrbaugh, R. Parasuraman, & R. Johnson (Eds.), *Event-Related Brain Potentials: Basic Issues and Applications*, pp. 279–300. New York: Oxford University Press.

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation induced complacency. *International Journal of Aviation Psychology, 3*, 1–23.

Parasuraman, R., Mouloua, M., & Molloy, R. (1994). Monitoring automation failures in human machine systems. In M. Mouloua & R. Parasuraman (Eds.), *Human Performance in Automated Systems: Current Research and Trends*, pp. 45–9. Hillsdale, NJ: Erlbaum.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116*, 220–44.

Pfurtscheller, G., Kalcher, J., Neuper, C., Flotzinger, D., & Pregenzer, M. (1996). On-line EEG classification during externally-paced hand movements using a neural network-based classifier. *Electroencephalography and Clinical Neurophysiology, 99*, 416–25.

Polson, M., & Freidman, A. (1988). Task sharing within and between hemispheres: A multiple resource approach. *Human Factors, 30*, 633–43.

Pope, A., Bogart, E., & Bartolome, D. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology, 40*, 187–95.

Porges, S. W., & Byrne, E. A. (1992). Research methods for the measurement of heart rate and respiration. *Biological Psychology, 34*, 93–130.

Pribram, K. H., & McGuinness, D. (1975). Arousal, activation and effort in the control of attention. *Psychological Review, 27*, 131–42.

Rau, R. (1996). Psychophysiological assessment of human reliability in a simulated complex system. *Biological Psychology, 42*, 287–300.

Reason, J. (1990). *Human Error.* Cambridge University Press.

Riley, V. (1994). A theory of operator reliance in automation. In M. Mouloua & R. Parasuraman (Eds.), *Human Performance in Automated Systems: Current Research and Trends*, pp. 8–14. Hillsdale, NJ: Erlbaum.

Roman, J., Older, H., & Jones, W. L. (1967). Flight research program. VII: Medical monitoring of Navy carrier pilots in combat. *Aerospace Medicine, 38*, 133–9.

Roscoe, A. H. (1975). Heart rate monitoring of pilots during steep gradient approaches. *Aviation, Space and Environmental Medicine, 46*, 1410–15.

Roscoe, A. H. (1993). Heart rate as a psychological measure for in-flight workload assessment. *Ergonomics, 36*, 1055-62.

Rosler, F., Heil, M., & Roder, B. (1997). Slow negative brain potentials as reflections of specific modular resources of cognition. *Biological Psychology, 45*, 109–42.

Ruffel-Smith, H. P. (1967). Heart rate of pilots flying aircraft on scheduled airline routes. *Aerospace Medicine, 38*, 1117–19.

Scerbo, M. (1994). Theoretical perspectives on adaptive automation. In M. Mouloua & R. Parasuraman (Eds.), *Human Performance in Automated Systems: Current Research and Trends*, pp. 37–64. Hillsdale, NJ: Erlbaum.

Scheffers, M., Coles, M. G. H., Bernstein, P., Gehring, W., & Donchin, E. (1996). Event-related brain potentials and error related processing: An analysis of incorrect responses to go and no-go stimuli. *Psychophysiology, 33*, 42–53.

Sem-Jacobsen, C. W. (1959). Electroencephalographic study of pilot stress. *Aerospace Medicine, 30*, 797–803.

Sem-Jacobsen, C. W. (1960). Recording of in-flight stress in jet fighter planes. *Aerospace Medicine, 31*, 320–35.

Sem-Jacobsen, C. W. (1961). Black-out and unconsciousness revealed by airborne testing of fighter pilots. *Aerospace Medicine, 32*, 247–53.

Sem-Jacobsen, C. W., Nilseng, O., Patten, C., and Eriksen, O. (1959). Electroencephalographic recording in simulated flight in a jet fighter plane. *Journal of Electroencephalography and Clinical Neurophysiology, 11*, 22–9.

Sem-Jacobsen, C. W., & Sem-Jacobsen, J. E. (1963). Selection and evaluation of pilots for high performance aircraft and spacecraft by in-flight EEG study of stress tolerance. *Aerospace Medicine, 34*, 605–9.

Shaffer, L. (1975). Multiple attention in continuous tasks. In P. Rabbitt & S. Dornic (Eds.), *Attention and Performance.* London: Academic Press.

Sirevaag, E., Kramer, A., Coles, M., & Donchin, E. (1987). Resource reciprocity: An event-related brain potentials analysis. *Acta Psychologica, 70*, 77–90.

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*, 1121–40.

Smulders, F., Kenemans, J., Jonkman, L., & Kok, A. (1997). The effects of sleep loss on task performance and the electroencephalogram in young and elderly subjects. *Biological Psychology, 45*, 217–39.

Sperling, G., & Melcher, M. (1978). The attention operating characteristic: Examples from visual search. *Science, 20*, 315–18.

Sterman, B., & Mann, C. (1995). Concepts and applications of EEG analysis in aviation performance evaluation. *Biological Psychology, 40*, 115–30.

Sterman, B., Mann, C., & Kaiser, D. (1992). Quantitative EEG patterns of in-flight workload. In *1992 Space Operations Applications and Research Proceedings*, pp. 466–73. Houston, TX: Johnson Space Center.

Sterman, B., Mann, C., Kaiser, D., & Suyenobu, B. (1994). Multiband topographic EEG analysis of a simulated visuo-motor aviation task. *International Journal of Psychophysiology, 16*, 49–56.

Stern, J. A., Boyer, D., & Schroeder, D. (1994). Blink rate: a possible measure of fatigue. *Human Factors, 36*, 285–97.

Stern, J., Walrath, L. C., & Goldstein, R. (1984). The endogenous eyeblink. *Psychophysiology, 21*, 22–33.

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

Tattersall, A. J., & Hockey, G. R. (1995). Level of operator control and changes in heart rate variability during simulated flight maintenance. *Human Factors, 37*, 682–98.

Torsvall, L., & Akerstedt, T. (1987). Sleepiness on the job: Continuously measured EEG changes in train drivers. *Electroencephalography and Clinical Neurophysiology, 66*, 502–11.

Torsvall, L., & Akerstedt, T. (1988). Extreme sleepiness: Quantification of EOG and spectral EEG parameters. *International Journal of Neuroscience, 38*, 435–41.

Townsend, R. E., & Johnson, L. C. (1979). Relation of frequency analyzed EEG to monitoring behavior. *Electroencephalography and Clinical Neurophysiology, 47*, 272–9.

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

Trejo, L. J., & Shensa, M. J. (1993). Linear and neural network models for predicting human signal detection performance from event-related potentials: A comparison of wavelet transforms with other feature extraction methods. *Proceedings of the International Simulation Conference* (November 7–10). San Diego: Society for Computer Simulation.

Varri, A., Hirvonen, K., Hasan, J., Loula, P., & Hakkinen, V. (1992). A computerized analysis system for vigilance studies. *Computer Methods and Programs in Biomedicine, 39*, 113–24.

Veltman, J. A., & Gaillard, A. W. K. (1996). Physiological indices of workload in a simulated flight task. *Biological Psychology, 42*, 323–42.

Verbaten, M. N., Huyben, M. A., & Kemner, C. (1997). Processing capacity and the frontal P3. *International Journal of Psychophysiology, 25*, 37–48.

Westerkamp, J., & Williams, R. (1995). Adaptive estimation of single response evoked potentials. *Biological Psychology, 40*, 161–8.

Wickens, C. D. (1990). Applications of event-related potential research to problems in human factors. In J. Rohrbaugh, R. Parasuraman, & R. Johnson (Eds.), *Event-Related Brain Potentials: Basic Issues and Applications*, pp. 301–9. New York: Oxford University Press.

Wickens, C. D. (1992). *Engineering Psychology and Human Performance.* New York: Harper-Collins.

Wickens, C. D., Kramer, A. F., Vanesse, L., & Donchin, E. (1983). The performance of concurrent tasks: A psychophysiological analysis of the reciprocity of information processing resources. *Science, 221*, 1080–2.

Wierwille, W. W. (1994). Overview of research on driver drowsiness definition and driver drowsiness detection. Paper presented at the 14th Annual Technical Conference on the Enhanced Safety of Vehicles (ESV) (May, Munich).

Wilson, G. F. (1992). Applied use of cardiac and respiration measures: Practical considerations and precautions. *Biological Psychology, 34,* 163–78.

Wilson, G. F. (1993). Air-to-ground training missions: A psychophysiological workload analysis. *Ergonomics, 36,* 1071–87.

Wilson, G. F. (1994). Workload assessment monitor (WAM). In *Proceedings of the Human Factors and Ergonomics Society,* pp. 944–8. Santa Monica, CA: Human Factors Press.

Wilson, G. F., & Eggemeier, F. T. (1991). Psychophysiological assessment of workload in multi-task environments. In D. Damos (Ed.), *Multiple Task Performance,* pp. 329–60. London: Taylor & Francis.

Wilson, G. F., & Fisher, F. (1995). Cognitive task classification based on topographic EEG data. *Biological Psychology, 40,* 239–50.

Wilson, G. F., Fullenkamp, P., & Davis, I. (1994). Evoked potential, cardiac, blink and respiration measures of pilot workload in air to ground missions. *Aviation, Space and Environmental Medicine, 65,* 100–5.

Wolpaw, J. R., & McFarland, D. J. (1994). Multichannel EEG-based brain computer communication. *Electroencephalography and Clinical Neurophysiology, 90,* 444–9.

Wolpaw, J. R., McFarland, D. J., Neat, G. W., & Forneris, C. A. (1991). An EEG-based brain computer interface for cursor control. *Electroencephalography and Clinical Neurophysiology, 78,* 252–9.

Yagi, A. (1997). Application of saccade evoked cortical models to display evaluation. Paper presented at the Workshop on Applied Psychophysiology (June, Helsinki).

Yamamoto, Y., & Isshiki, H. (1992). Instrument for controlling drowsiness using galvanic skin reflex. *Medical and Biological Engineering, 30,* 562–4.

Zelinsky, G. J., & Sheinberg, D. L. (1997). Eye movements during parallel-serial visual search. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 244–62.