Navy Personnel Research and Development Center

San Diego, CA 92152-6800 TN-90-23 June 1990

Physiological Metrics of Mental Workload: A Review of Recent Progress

Arthur F. Kramer

NPRDC-TN-90-23 June 1990

Physiological Metrics of Mental Workload: A Review of Recent Progress

Arthur F. Kramer
University of Illinois
Champaign, IL 61820

Reviewed and released by
J. C. McLachlan
Director, Training Systems Department

Approved for public release;
distribution is unlimited.

Navy Personnel Research and Development Center
San Diego, California 92152-6800

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE June 1990	3. REPORT TYPE AND DATE COVERED Interim--1 Jan 89-1Oct 89

4. TITLE AND SUBTITLE
Physiological Metrics of Mental Workload: A Review of Recent Progress

5. FUNDING NUMBERS
PE 0602234N
PR RS34H21
TA 01

6. AUTHOR(S)
Arthur F. Kramer

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Illinois at Urbana-Champaign
Champaign, IL 61820

8. PERFORMING ORGANIZATION REPORT NUMBER
NPRDC-TN-90-23

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Technology
Code 222
Arlington, VA 22217-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
To be published in "Multiple Task Performance" (Editor--D. Damos), Taylor & Francis publishers. Sponsored in cooperation with the Office of Naval Research, Code 1142PS (Dr. H. Hawkins). Monitored by NPRDC (Dr. L. Trejo).

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
This report reviews research on physiological metrics of mental workload performed in the last decade. The focus of the review is on measurement techniques that have potential for fundamental explanation of mental workload and for use in operational environments. The techniques are examined within a framework of measurement criteria. These criteria include: sensitivity, diagnosticity, intrusiveness, reliability, and generality of application. Over 200 articles are covered by the review. Measures reviewed include: electroencephalograms, event-related potentials, magnetoencephalograms, positron emission tomography, electro-oculograms, cardiovascular measures, pupillometry, respiratory measures, and electrodermal measures. Keywords:

14. SUBJECT TERMS
Mental workload, physiological metrics, electroencephalogram, magnetoencephalogram, cardiovascular function, pupillometry, respiration, electrodermal responses

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

FOREWORD

This report reviews research on physiological metrics of mental workload performed in the last decade. The focus of the review is on measurement techniques that have potential for fundamental explanation of mental workload and for use in operational environments. The techniques are examined within a framework of measurement criteria. These criteria include: sensitivity, diagnosticity, intrusiveness, reliability, and generality of application. Over 200 articles are covered by the review. Measures reviewed include: electroencephalograms, event-related potentials, magnetoencephalograms, positron emission tomography, electro-oculograms, cardiovascular measures, pupillometry, respiratory measures, and electrodermal measures.

This effort was supported by a grant from the Office of Naval Technology, Program Element 0602234N, Work Unit RS34H21, with Dr. Leonard Trejo as the contracting officer's technical representative and by a grant from NASA Ames Research Center (NASA NAG 2-308) monitored by Dr. Sandra Hart.

J. C. McLACHLAN
Director, Training Systems Department

SUMMARY

The last in-depth review of physiological metrics of mental workload was published a decade ago (Wierwille, 1979). However, even the Wierwille review was limited in scope since its main focus was the evaluation of physiological measures for aircrew mental workload.

The present review has three goals. First, I will update Wierwille's review by examining studies performed in the last decade. Second, like Wierwille, my review will be selective. However, rather than concentrating on a specific area of application, I will focus on measurement techniques that have shown potential for making significant contributions to our understanding of the concept of mental workload as well as those techniques that have shown promise for making the transition from the laboratory to operational or simulated operational environments. Third, I will evaluate the degree to which each of several classes of physiological techniques meets a number of measurement criteria. These criteria include: sensitivity, diagnosticity, intrusiveness, reliability, and generality of application.

CONTENTS

INTRODUCTION

The last in-depth review of physiological metrics of mental workload was published a decade ago (Wierwille, 1979; but see Hancock, Meshkati, & Robertson, 1985; Wilson & O'Donnell, 1988, for more selective reviews). However, even the Wierwille review was limited in scope since its main focus was the evaluation of physiological measures for aircrew mental workload. The present review has three goals. First, I will update Wierwille's review by examining studies performed in the last decade. Second, like Wierwille my review will be selective. However, rather than concentrating on a specific area of application, I will focus on measurement techniques that have shown potential for making significant contributions to our understanding of the concept of mental workload as well as those techniques that have shown promise for making the transition from the laboratory to operational or simulated operational environments. Third, I will evaluate the degree to which each of several classes of physiological techniques meets a number of measurement criteria. These criteria include: sensitivity, diagnosticity, intrusiveness, reliability, and generality of application.

Prior to delving into the critical review, I will briefly outline the theoretical framework in which I will examine the measurement techniques. Although there is no universally accepted definition of mental workload, the recent consensus suggests that mental workload can be conceptualized as the interaction between the structure of systems and tasks on the one hand, and the capabilities, motivation, and state of the human operator on the other (Gopher & Donchin, 1986; Moray, 1989; Wickens & Kramer, 1985). More specifically, mental workload has been defined as the "costs" a human operator incurs as tasks are performed.

Early views of the mechanisms underlying the human side of the mental workload equation suggested that the "costs" could be conceptualized in terms of an undifferentiated capacity or resource (Kahneman, 1973; Moray, 1967). Additional capacity could be allocated as task difficulty increased or when operators were required to perform additional tasks. However, since the resource supply is limited, a point would eventually be reached at which additional resources would no longer be available. At this point, performance efficiency would decline. Within such a theoretical framework, the "residual capacity" remaining after the performance of the required tasks could be viewed as a measure of mental workload.

In addition to the resource-limited processing discussed above, Norman and Bobrow (1975) described another form of performance limit. In this case, the allocation of additional resources does not improve performance. As an example, consider a task in which you are required to detect a very dim signal on a noisy radar scope. In this situation, while you may try harder to distinguish the signal from the noise, the limits of your sensory system and the quality of the data may prevent you from improving your performance. Norman and Bobrow referred to such a situation as data-limited. The only way in which performance can be enhanced for a data-limited process is to improve the quality of the data (i.e., the signal/noise ratio) or the operator's sensory system (i.e., try the task again after eight hours of sleep).

While the undifferentiated view of resources in conjunction with the notion of data-limits accounted for a good deal of data, it soon became apparent that more than one resource was needed to explain the pattern of performance interactions observed when operators carried out several tasks simultaneously. A number of different multiple resource models have been proposed.

However, in each case, the major goal has been to account for the most variance in multi-task performance with the fewest types of resources. The most detailed multiple resource model has been proposed by Wickens (1980, 1984). The model divides information processing into three dichotomous dimensions with each level of a dimension representing a separate resource. Dimensions include: stages of processing (perceptual/central and response), codes of processing (verbal and spatial), and modalities of input and output (input: visual and auditory; output: speech and manual). Other multiple resource models have defined resources in terms of cerebral hemispheres (Freidman & Polson, 1981; Polson & Freidman, 1988), distance in functional cerebral space (Kinsbourne & Hicks, 1978), and arousal, activation, and effort (Sanders, 1981; see also Baddley & Hitch, 1974; Navon & Gopher, 1979; Sanders, 1979). Within these models, mental workload can be described as the cost of performing one task in terms of a reduction in the capacity to perform additional tasks, given that two tasks overlap in their resource demands. Of course, each of these models assumes that operators will expend the necessary effort to perform their assigned tasks.

The measurement techniques employed in the assessment of mental workload have kept pace with the theoretical developments in the field of timesharing. Thus, while the initial goal in the workload assessment field was the discovery of the "best" measure of capacity allocation (Knowles, 1963), more recent workload measurement reviews and taxonomies have emphasized the importance of designing a battery of measures that would tap different dimensions (resources) of mental workload (Gopher & Donchin, 1986; Leplat, 1978; Moray, 1989; O'Donnell & Eggemeier, 1986; Ogden, Levine, & Eisner, 1979; Wickens, 1979). The sensitivity of psychophysiological measures to different aspects of workload will be described below.

Criteria for Selection of Workload Measures

Given the multidimensional nature of mental workload, no single measurement technique can be expected to "tap" all of the important aspects of human mental workload. In fact, the range of diagnosticity of different techniques varies from specific resource types (e.g., perceptual resources in the Wickens, 1980 model) to global constructs such as operator effort. Thus, a technique that is adequate for one purpose may not provide the necessary information in other situations. In addition to differing in diagnosticity, workload metrics also vary along a number of other dimensions such as sensitivity, intrusiveness, reliability, and generality of application. These dimensions can be used as selection for different applications. In this section, I will briefly define each of the criteria and describe how they will be applied to the physiological measures.

The criterion of *sensitivity* refers to the capability of the measure to discriminate among variations in mental workload. For example, while a particular measure may provide a fine-grained assessment of changes in workload from low to moderate levels, it might be quite insensitive to variations from moderate to high levels. Yeh and Wickens (1988) suggested that such is the case for most subjective measures of mental load. Other measures seem to be more sensitive to changes from moderate to high levels than they are for changes from low to moderate levels of load. Many performance measures are relatively insensitive to changes in workload at low levels due to the operator's ability to maintain performance with little investment of effort. However, once a system becomes difficult to manage, small changes in workload often result in large changes in performance (e.g., either in terms of decrements in performance or changes in strategies).

Another question that can be posed when evaluating the sensitivity criterion is sensitivity to what? My description above refers to sensitivity to the magnitude of change in workload. However, recent concerns with rapid changes in workload (Wierwille, 1988) suggest that "temporal" sensitivity is also an important factor. Therefore, it would appear important to determine how quickly different measurement techniques respond to sudden changes in mental workload. In essence, this question concerns the amount of data that is necessary to provide a reliable estimate of different levels of workload.

The criterion of *diagnosticity* refers to the capability of a measure to discriminate among types of mental workload. Within the context of multiple resource models, a measure would be said to be diagnostic if it discriminated among different varieties of resources. Thus, while one technique may provide a global measure of resource allocation, another measure might prove sensitive to perceptual/central processing resources, while a third measure might be selectively sensitive to variations in spatial perceptual/central processing load. The choice of a workload measure on the basis of its degree of diagnosticity will depend on the measurement objective. If the goal is to determine whether workload differs from one task configuration to another, a measure with relatively low diagnosticity may be appropriate. However, if the objective is to assess whether a task should be implemented with visual or auditory displays or with verbal or spatial warning messages, a more diagnostic measure will be required.

The criterion of *intrusiveness* refers to the capability of measuring mental load without interfering with the operator's performance on the "primary" task. While the use of intrusive techniques can be justified if they provide more precise assessments of mental load than other, less intrusive techniques, the situations in which they can be utilized are clearly limited. Thus, while it may be acceptable to employ an intrusive measurement procedure in a laboratory or simulator setting, safety precautions preclude the use of this class of techniques in most operational environments. Furthermore, since intrusive techniques degrade performance on the task of interest, their use also complicates the interpretation of variations in mental workload.

While the *reliability* of workload measurement procedures is often assumed, there have been few formal evaluations of the reliability of these techniques. However, although formal reliability assessment procedures such as split-half, alternate-forms, and test-retest reliability (Guilford, 1954) have not traditionally be applied to workload measurement procedures, the reliability of these techniques can be estimated by comparing results obtained in similar experiments and with relatively homogenous populations. Both formal and informal estimates of reliability will be discussed during my description of each class of physiological measures.

Another important factor in the evaluation of workload metrics is the *generality of application*. While it is certainly the case that each of the previously described criteria constrain applications, I thought it important to include an explicit discussion of potential application domains for each class of physiological measures. In particular, my discussion of applications will include: (a) potential artifacts encountered with each of the measurement techniques, (b) an assessment of the degree to which particular techniques have been successfully employed in laboratory, simulator, and operational environments, (c) an evaluation of the feasibility of employing measurement procedures for purposes of training evaluation, system performance, and personnel selection, and (d) an examination of the potential for applying the measurement techniques in on-line and off-line contexts.

Physiological Measures: Strengths and Weaknesses

An important issue that is often overlooked in reviews of physiological measures of mental workload concerns the relative difficulty of collecting, analyzing, and interpreting physiological and non-physiological measures of mental load. Of course, the real question is whether physiological recording provides information about mental workload that cannot easily be obtained from subjective, primary, or secondary task measures. In an effort to provide a balanced view of physiological techniques, I will briefly enumerate and discuss the advantages and disadvantages of this class of measures.

I begin my discussion by describing the disadvantages of physiological techniques. First, although the cost of physiological recording systems has decreased dramatically over the past 10 to 15 years, the necessity for specialized equipment (e.g., amplifiers, transducers, A/D conversion, large data storage medium) renders physiological recording substantially more expensive than the collection of primary, secondary, or subjective measures of mental workload. Second, while standardized scoring procedures have been developed for subjective (Hart, Vidulich, & Tsang, 1986; Reid, 1985) and performance-based (Englund et al., 1987) workload assessment procedures, the interpretation of physiological data still requires an extensive amount of technical expertise (Kramer, 1985). Although a number of multivariate statistical procedures are commonly used in the analysis of physiological data (see Coles, Gratton, Kramer, & Miller, 1986), their selection and application is often guided by visual inspection of the voltage-x-time signals.

Third, while the discrimination between signal and noise is a problem that is encountered *during the implementation of* both physiological and nonphysiological measurement procedures, the magnitude of the problem is larger for physiological measures. For example, while low- and high-pass frequency filters may be used to eliminate a substantial portion of the noise that affects physiological measures, other varieties of noise occur within the same frequency and time domain as the signals and therefore cannot be easily filtered (e.g., alpha contamination of ERP components). Furthermore, a number of physiological signals are influenced by factors other than mental workload (e.g., physical exertion, emotional state, ambient lighting) and therefore require that experiments are conducted in well controlled settings. While careful experimental control can alleviate or at least reduce the influence of these potentially confounding factors, it also serves to complicate the use of physiological techniques in operational environments. Finally, while physiological measures provide insights into the changes in bodily functions that accompany variations in mental workload, they are further removed from operator and system performance than primary and secondary task measures of mental load. Thus, since the ultimate goal of mental workload assessment is the prediction and understanding of variations in human performance in response to changes in system demands, it is necessary to provide a strong conceptual link from the physiological measures to performance.

Given the number of potential problems associated with the use of physiological measures, why would anyone choose to use this class of techniques to assess mental workload? Obviously this chapter would not have been written if I did not believe that the strengths of physiological measures outweighed their weaknesses for at least a subset of possible applications. In the remainder of this section, I will describe some of the advantages of physiological measures of mental workload.

First, unlike secondary task measures, physiological measurement procedures are relatively unobtrusive. While most physiological measures do require the placement of recording electrodes or transducers on the body, they do not necessitate the introduction of extraneous signals into the operators task. In the past, the collection of physiological data required that the operator was tethered to an amplifier/recording system. However, the recent development of miniaturized recording and telemetry equipment has greatly enhanced the process of data collection from ambulatory operators. Thus, assuming that operators adapt to the few transducers that are affixed to their body, the collection of physiological data can be truly unobtrusive.

Second, given the recent interest in examining mental workload in semi-automated systems, it would be desirable to possess workload metrics that do not require the measurement of overt performance. Most physiological measures fulfill this criterion since they can be recorded in the absence of behavior. It is important to note, however, that, due to the multidimensional nature of mental workload, it is often advantageous to possess measures of both performance and physiology in order to infer changes in operator strategies and workload with variations in system demands.

Third, physiological measures are inherently multidimensional and therefore can be expected to provide a number of "views" of operator mental workload. For example, several mental workload measures are included within the class of central nervous system (CNS) measurement techniques. These techniques include: measures of electroencephalographic activity (EEG), event-related brain potentials (ERPs), measures of the magnetic field activity of the brain (MEG), measures of brain metabolism such as positron emission tomography (PET), and electrooculographic (EOG) activity. Each of these techniques is uniquely sensitive to different aspects of human mental workload. Furthermore, each of these techniques can be further subdivided to provide a more fine-grained analysis of processing demands. For example, ERPs are traditionally decomposed into a number of temporally and spatially definable components which differ in their sensitivity to aspects of human information processing. Moreover, different aspects of these components such as their latency and amplitude have been shown to be differentially sensitive to chronometric and energetic dimensions of human information processing (Kramer, 1987).

Fourth, since most physiological signals are recorded continuously, they offer the potential for providing measures that respond relatively quickly to phasic shifts in mental workload. However, it is important to note that, although physiological measures are often recorded continuously, the measures are differentially sensitive to the temporal dynamics of mental load. For example, changes in the amplitude and latency of ERP components often occur within several hundred milliseconds of shifts in operator strategies (Donchin, Karis, Bashore, Coles, & Gratton, 1986). Heart rate variability also responds rapidly to changes in operator workload and strategies, usually within several hundred milliseconds to several seconds (Aasam, Mulder, & Mulder, 1987; Coles & Sirevaag, 1987). On the other hand, measures of brain metabolism often require from 30 seconds to several minutes to provide an indication of changes in human information processing (Phelps & Mazziotta, 1985; Posner, Peterson, Fox, & Raichle, 1988). Thus, while some members of the class of physiological measurement techniques can index rapid and transient shifts in mental workload, other techniques are more suitable for off-line assessments of mental load.

Finally, one problem that has plagued the field of mental workload assessment has been the lack of an agreed upon method of scaling different dependent variables and tasks in terms of their resource demands (Kantowitz & Weldon, 1985). Thus, the question of how many milliseconds of reaction time (RT) are equivalent to a 1 percent change in accuracy or a 1 unit change in root-mean-square tracking error remains unanswered. A number of different transformations have been suggested to normalize these dependent measures (Colle, Amel, Ewry, & Jenkins, 1988; Mountford & North, 1980; Wickens, Mountford, & Schreiner, 1981; Wickens & Yeh, 1985). However, since different transformations differentially affect the slope of the Performance Operating Characteristic (POC: a plot of performance on one task as a function of performance on a concurrent task), which in turn has implications for the shape of the underlying resource functions, it would be preferable to possess a set of measures that could be compared across different tasks. Since physiological measures of mental workload can be recorded in a wide variety of tasks, they offer the potential for solving this scaling problem.

This section has described both the advantages and disadvantages of physiological measures of mental load in an effort to provide the reader with a framework in which to evaluate the utility of physiological measures for different applications. In the next section, I examine a number of different classes of physiological measures in terms of the selection criteria and issues described above.

PHYSIOLOGICAL MEASURES: A REVIEW AND EVALUATION

Two general classes of physiological measures will be examined in my review of measures of mental workload: central nervous system measures (CNS) and peripheral nervous system measures. Within the class of peripheral nervous system measures, I will concentrate on measures of autonomic nervous system (ANS) activity. The boundaries between the CNS and the peripheral nervous system are based on anatomical distinctions. However, it is important to note that CNS and peripheral nervous system distinction is only a shorthand for the organization of the nervous system since the two systems interact in the control of many physiological functions (see Chapters 1 through 9 in Coles, Donchin, & Porges, 1986 for an in-depth discussion of the structure and function of the nervous system).

The CNS contains all cells within the bony structures of the skull and the spinal column including the brain, the brain stem, and the spinal cord. CNS measures that will be examined in the following review include EEG activity, ERPs, MEG, measures of brain metabolism such as PET, and measures of EOG activity.

The peripheral nervous system includes all neurons outside the bony enclosures of the skull and the spinal column. One component of the peripheral nervous system is the somatic nervous system. The somatic nervous system is mainly concerned with the activation of voluntary or striated muscles. The other component of the peripheral nervous system, the ANS, controls the internal organs of the body by innervating involuntary (smooth) musculature. The ANS is subdivided into the sympathetic (SNS) and parasympathetic (PNS) nervous systems. The basic function of the SNS is the mobilization of the body to meet emergencies. This is accomplished through a complex series of responses such as the breakdown of glycogen in the liver and the decrease in blood flow near the surface of the skin so that blood flow can be increased to internal organs. The action of the SNS is diffuse and can be maintained for an extended period of time. On the other hand, the function of the PNS is to conserve and maintain bodily resources. The action of the PNS is localized and of relatively short duration compared to the SNS. It should be clear from this brief description of the

SNS and PNS that the two systems complement and counteract each other. Thus, given the reciprocal relations between these systems, it is often difficult to distinguish their influence on bodily organs. For example, heart rate may increase because of increased SNS activity or decreased activity in the PNS. In my review, I will concentrate on measures of ANS activity including: cardiovascular measures, measures of pupil diameter, respiratory measures, and electrodermal measures. It is important to note that, while I distinguish between ANS and CNS measures in my review, I do not mean to imply that the specific measures reflect the influence of only one of the nervous systems. Instead, I have classified measures on the basis of the relative influence of the CNS and ANS.

Event-related Brain Potentials (ERPs)

Overview

The ERP is a transient series of voltage oscillations in the brain that can be recorded from the scalp in response to the occurrence of a discrete event. This temporal relationship between the ERP and the eliciting stimulus or response is what differentiates the ERP from the ongoing EEG activity. Like EEG, the ERP is a multivariate measure. However, unlike EEG, the ERP is decomposed in the time, rather than the frequency, domain.

ERPs are viewed as a sequence of separate but sometimes temporally overlapping components which are influenced by some combination of the physical parameters of the stimuli and psychological constructs such as expectance, task relevance, memory processes, and resources. Figure 1 presents the series of components which are normally recorded with the presentation of an auditory stimulus. Similar diagrams can be drawn from visual and somatosensory modalities.

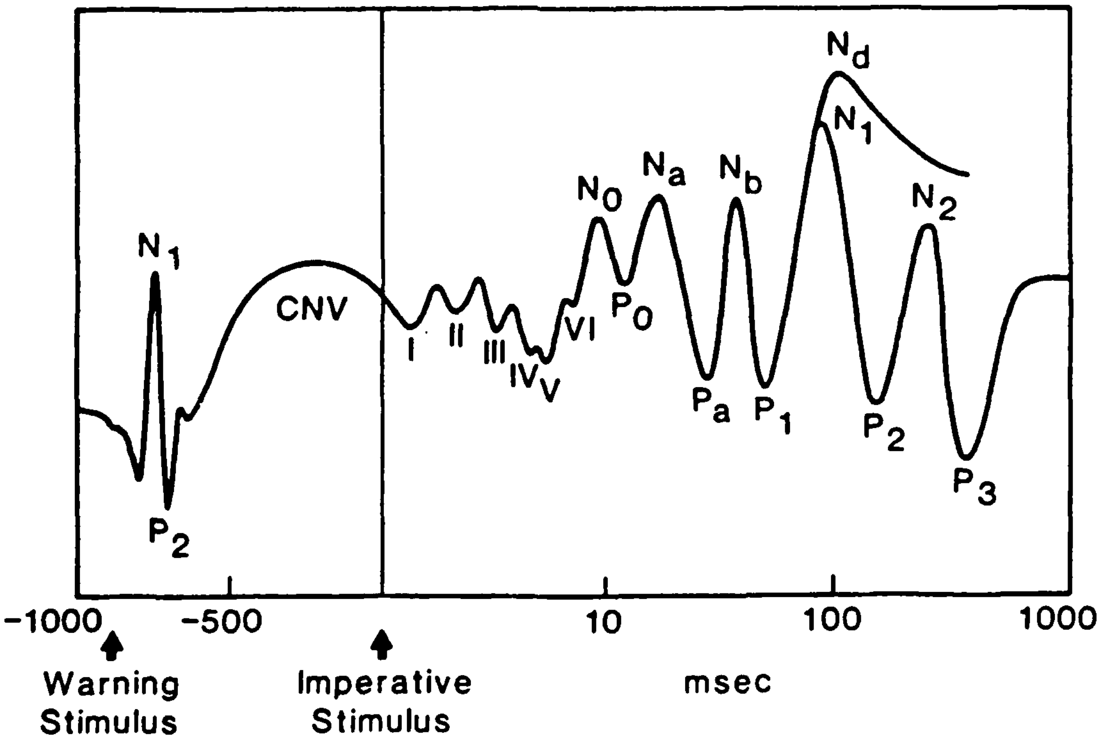

Figure 1. A graphical illustration of a prototypical auditory event-related brain potential.

Components are typically labeled with an "N" or a "P" denoting negative or positive polarity, and a number indicating their minimal latency measured from the onset of an eliciting event (e.g., N100 is a negative component which occurs at least 100 milliseconds after a stimulus). Components may be categorized along a continuum from exogenous to endogenous. The exogenous components represent an obligatory response of the brain to the presentation of a stimulus. These components are usually associated with specific sensory systems, occur within 200 milliseconds of a stimulus, and are primarily sensitive to the physical attributes of stimuli. For example, exogenous visual potentials are influenced by the intensity, frequency, hue, patterning, and location of the stimulus in the visual field. The exogenous components have been successfully used in clinical settings to monitor the functional integrity of the nervous system during surgical procedures, to assess changes in the nervous system as a result of maturation and aging, and to help diagnose various types of neuropathology including tumors, lesions, and demyelniating diseases such as multiple sclerosis (Starr, 1978; Stockard, Stockard, & Sharbrough, 1979).

The endogenous components, on the other hand, occur somewhat later than the exogenous components and are not very sensitive to changes in the physical parameters of stimuli, especially when these changes are not relevant to the task. Instead, these components are primarily influenced by the processing demands of the task imposed upon the subject. In fact, endogenous components can even be elicited by the absence of a stimulus if this "event" is relevant to the subject's task. The strategies, expectancies, intentions, and decisions of the subject as well as task parameters and instructions account for the majority of the variance in the endogenous components.

The importance of the componential nature of the ERP in the assessment of organismic state and information processing has made it imperative that components be clearly defined. The labeling of different peaks and troughs in Figure 1 suggests that some basis exists for the categorization of ERP components. The attributes of the ERP that have served as definitional criteria include: the distribution of voltage changes across the scalp, latency range, polarity, sequence, and the sensitivity of components to manipulations of instructions, task parameters and physical changes in the stimulus (Donchin, Ritter, & McCallum, 1978; Kramer, 1985).

The scalp distribution refers to the relative amplitude and polarity of the component across the scalp for a fixed temporal interval. Thus, one component may be positive at a parietal location and negative at a frontal site at time t (n), while another component might possess the opposite polarity-location relationship at time t (n). The latency range depends on the experimental manipulations as well as the specific component. For example, the components occurring within 10 milliseconds of the presentation of a stimulus, the brain-stem evoked potentials, are influenced by both organismic and stimulus variables but their latency range is only a few milliseconds. On the other hand, the latency range of the P300 component depends on the processing requirements of the task and can span several hundred milliseconds. The sensitivity of components to specific experimental manipulations is perhaps the most important of the definitional criteria. In fact, it has been suggested that components with different scalp distributions, but a similar relationship to task parameters or instructions, be defined as the same component (Ritter, Simpson, & Vaughan, 1983).

Sensitivity and Diagnosticity

Over the past decade a number of ERP components have been shown to be sensitive to variations in mental workload. The P300 component in particular has received the most extensive examination with regard to dimensions of mental load and therefore will be the starting point for my discussion of ERPs and workload. The sensitivity of the P300 component to processing demands has been extensively investigated in multi-task paradigms (Donchin, Kramer, & Wickens, 1986; Kramer, 1987). For example, Israel, Wickens, Chesney, and Donchin, (1980) required subjects to perform a simulated air traffic control (ATC) task concurrently with a visual discrimination task. Subjects were instructed to treat the ATC task as primary and the visual discrimination task as secondary. ERPs were elicited by secondary task events. The amplitude of the P300 component decreased with increases in the number of elements to be monitored in the ATC task.

Other studies have also found decreases in the amplitude of P300s elicited by secondary task events with increases in the difficulty of a primary task. These studies have employed a variety of primary tasks including pursuit and compensatory tracking, flight control and navigation, and memory/visual search as well as both visual and auditory secondary tasks (Hoffman, Houck, MacMillan, Simons, & Oatman, 1985; Kramer & Strayer, 1988; Kramer, Sirevaag, & Braune, 1987; Kramer, Wickens, & Donchin, 1983, 1985; Lindholm, Cheatman, Koriath, & Longridge, 1984; McCallum, Cooper, & Pocock, 1987; Natani & Gomer, 1981; Strayer & Kramer, in press). Capacity models predict that as the difficulty of a primary task increases, fewer resources should be available for the performance of a secondary task. The studies described above suggest that the P300s may reflect the residual resources available for secondary task performance.

Given that P300s reflect the distribution of processing resources in a dual-task situation, it would also be expected that P300s elicited by primary task events should increase in amplitude with increases in the difficulty of the primary task. Thus, capacity models predict a reciprocal relationship between the resources allocated to one task and the residual resources available to another, concurrently performed task. The question of whether P300 would reflect this reciprocity was addressed in a study conducted by Wickens, Kramer, Vanasse, & Donchin (1983). ERPs were elicited by events in both the primary and secondary tasks. In the primary task, pursuit step tracking, ERPs were elicited by changes in the spatial position of the target while in the secondary task, auditory discrimination, ERPs were elicited by the occurrence of high- and low-pitched tones. Difficulty was varied by manipulating two variables in the tracking task: the predictability of the positional changes of the target and the control dynamics. The ordering of difficulty was validated by measures of tracking performance and subjective ratings of tracking difficulty. Consistent with previous results, P300s elicited by discrete secondary task events decreased in amplitude with increases in the difficulty of the primary task. On the other hand, increasing the difficulty of the tracking task by decreasing the stability of the control dynamics and the predictability of the target resulted in a systematic increase in primary task P300 amplitude. The reciprocal relationship between P300s elicited by primary and secondary task stimuli as a function of primary task difficulty is consistent with the resource trade-offs presumed to underlie dual-task performance decrements (see also, Sirevaag, Kramer, Coles, & Donchin, 1989).

Other demonstrations of the P300 reciprocity effect have been provided in paradigms in which priority rather than difficulty was manipulated. For example, Strayer and Kramer (in press) instructed subjects to concurrently perform two tasks: recognition running memory and memory search. In different conditions, subjects were to emphasize their performance on one task or the other or treat both tasks equally. The amplitude of the P300s reflected task priority. P300s increased in amplitude with the priority of one task while simultaneously decreasing in amplitude in the other task. Thus, the demonstration of reciprocity effects with both difficulty and priority manipulations provides strong support for the argument that P300 amplitude reflects the distribution of processing resources among concurrently performed tasks. Finally, in addition to demonstrating sensitivity to processing demands in multi-task paradigms, a number of investigators have found that the P300 also reflects variations in workload within single tasks (Horst, Munson, & Ruchkin, 1984; Sirevaag, Kramer, de Jong, & Mecklinger, 1988; Ulsperger, Metz, & Gille, 1988).

With regard to the issue of diagnosticity, a number of studies have demonstrated that, while P300 is influenced by manipulations that affect perceptual/central processing resources, it is relatively insensitive to factors that influence motor processes (Israel, Chesney, Wickens, & Donchin, 1980; Kutas, McCarthy, & Donchin, 1977; McCarthy & Donchin, 1981; Ragot, 1984). On the other hand, P300 appears to be sensitive to factors that influence both verbal/spatial and visual/auditory processes. Thus, within the multiple resource framework, it appears that P300 is primarily sensitive to perceptual/central processing resources.

A second class of ERP components that are negative in polarity and occur within the first 250 milliseconds following a stimulus have also been found to be sensitive to processing demands in single and dual tasks (see Naatanen, 1988 for an in-depth review of these components). More specifically, this class of components has (a) shown a graded sensitivity to processing demands, (b) displayed a reciprocity in amplitude when recorded from two concurrently performed tasks, and (c) indicated that the limited capacity reflected by these components can be flexibly allocated among different events (Hillyard, Munte, & Neville, 1985; Kramer, Sirevaag, & Hughes, 1988; Naatanen, 1988; Parasuraman, 1985). With regard to diagnosticity, these components appear to reflect the distribution of a variety of perceptual resources.

Thus far, I have confined my discussion of ERP metrics of mental workload to two different components of the ERP: the early negativities and the P300. There is, however, some evidence to suggest that other ERP components may also be sensitive to variations in capacity in single- and dual-task conditions. For example, McCallum et al. (1987) found that a slow negative wave distinguished between levels of tracking difficulty. This negative wave was detected only with DC amplifiers and extended over most of a 20-second tracking period. In a series of simulated flight maneuvers, Lindholm et al. (1984) found that the amplitude of the N200 component discriminated between different levels of single- and dual-task demands. Horst, Ruchkin, & Munson (1987) observed an increase in negativity with increasing monitoring demands. This increased negativity occurred at both 200 to 300 milliseconds and 400 to 500 milliseconds following the presentation of a bank of gauges. Finally, Wilson and O'Donnell (1986) reported changes in the steady-state-evoked responses that were correlated with the memory search slope in a Sternberg task (1969). While the results of these studies are potentially important, additional research will be necessary to determine the sensitivity and diagnosticity of these components to varieties of processing demands.

Intrusiveness

The degree to which ERPs interfere with task performance is dependent upon the method by which the ERPs are collected. For example, in the secondary task technique, operators are required to covertly count or overtly respond to the occasional presentation of an auditory or visual probe.

Although these probes have been shown to have only a minimal effect on operators performance (Kramer et al., 1983, 1987), the imposition of additional demands is often unacceptable in operational environments.

An alternative technique is to elicit ERPs from events in the primary task. As previously described, early negativities and the P300 component show a systematic relationship to processing demands in both single- and dual-task conditions. Thus, although performance measures along are often insufficient for the measurement of mental workload in single tasks, the joint use of psychophysiological and performance measures provides an index of resource allocation.

The irrelevant probe technique has also been proposed in an effort to eliminate the additional processing demands imposed on the operator by secondary task measures (Bauer, Goldstein, & Stern, 1987; Papanicolaou & Johnstone, 1984). In this technique, irrelevant auditory or visual probes are occasionally superimposed on the subjects task. However, unlike the secondary-task-technique, subjects are not required to respond to the problems. On the other hand, the theoretical assumptions underlying the secondary task and irrelevant probe techniques are quite similar. It is assumed that the size of the ERPs elicited by the irrelevant probes will be inversely proportional to the difficulty of the subject's task. Thus, variations in the amplitude of the ERP is taken as evidence of changes in resource demands.

Although the irrelevant probe technique eliminates the problem of additional demands that is associated with the secondary task measures, it does suffer from other problems. In particular, it is necessary to assume that, as in the secondary-task-technique, residual resources that are not used in the "primary" task are devoted to the processing of the irrelevant probes. However, unlike the secondary task method, there are no performance data to corroborate this assumption. Thus, while subjects could devote additional processing capacity to the irrelevant problems, it is equally plausible that they either do not use the excess capacity or that they devote it to other functions (e.g., planning a vacation).

A technique related to the irrelevant probe technique is used in the recording of steady state potentials. Steady state responses are the result of an entrainment of the evoked response to a rapidly presented stimulus (e.g., greater than 10 flashes per second). Since the operator is not required to make overt responses to these stimuli, they do not generally interfere with performance on the primary task.

Reliability

As previously mentioned, there have been few formal assessments of the reliability of physiological measures of mental workload. Nonetheless, the repeated replication of the patterns of results described above in a variety of paradigms and with a relatively heterogenous group of subjects (e.g., pilots, students, patients) suggests that these measures do provide a reliable measure of mental load, at least in the laboratory.

In addition to this informal evidence in support of the reliability of the measures, a recent study by Fabiani, Gratton, Karis, & Donchin (1987) has formally evaluated the reliability of P300 amplitude and latency in a series of simple oddball tasks. In these tasks, subjects were asked to either covertly count or overtly respond to occasional rare probes in a train of auditory or visual stimuli (e.g., respond to a 1200 Hz tone in a train of 1300 Hz tones). The split-half reliability was .92 for P300 amplitude and .83 for P300 latency. The test-retest reliability assessed over a period of several days was .83 for P300 amplitude and .63 for P300 latency. While only 50 subjects were run in this relatively simple paradigm, the results are useful in that they provide at least a tentative benchmark for the reliability of a subset of ERP components. Additional assessments should be conducted in more complex single- and multi-task paradigms.

Generality of Application

The recording of ERPs in operational environments is complicated by a number of factors. First, ERP components possess a relatively poor signal-to-noise ratio in single trial data. For example, the single trial amplitude of relatively large ERP components such as the P300 is approximately 20 to 30 microvolts compared to 50 to 100 microvolts for the on-going EEG. Smaller components such as the N100 are usually less than 5 microvolts. While the signal-to-noise ratio problem can be overcome by averaging, this procedure requires the collection of a number of replications of relevant events and therefore limits the situations in which ERPs can be applied. However, some recent successes in the application of pattern recognition techniques to single trial data suggest that the signal-to-noise ratio problems may be overcome, at least for the larger components (Farwell & Donchin, 1988; Kramer, Humphrey, Sirevaag, & Mecklinger, 1989).

A second potential problem is the contamination of the ERP by the electrical fields produced by other physiological systems such as the heart, eyes, and muscles (ECG, EOG, and EMG, respectively). However, most of this extraneous electrical activity can be eliminated or at least reduced with suitable analog or digital filters (Nunez, 1981).

An important question is whether ERPs can be successfully recorded outside of the laboratory? Another equally important question is whether ERPs can be expected to provide information on workload in real-time. A number of recent studies suggest that ERPs can indeed be recorded in high fidelity simulators (Lindholm et al., 1984; Natani & Gomer, 1981). In one such study, Kramer et al. (1987) found that the P300 elicited by secondary task probe stimuli discriminated among flights differing in the degree of turbulence and the presence of subsystem failures. Investigations of the efficacy of ERP measures in complex operational environments still remain to be performed.

In addition to off-line assessments of mental workload, several investigators have suggested that ERPs might be useful in on-line evaluations of the moment-to-moment fluctuations in operator state and processing demands (Defayolle, Dinand, & Gentil, 1971; Gomer, 1981; Groll-Knapp, 1971; Sem-Jacobsen, 1981). While research in this area is still in its infancy, a few recent studies suggest that on-line assessment might be feasible, at least in restricted settings. For instance, Farwell and Donchin (1988) demonstrated that ERPs can be used to communicate selections from a 6 x 6 menu. In their task, subjects were instructed to attend to one item from a 6 x 6 matrix of items. The rows and columns of the matrix flashed randomly and the ERPs elicited by the flashes were used to discriminate attended from unattended items. A communication accuracy of 95 percent was achieved with 26 seconds of data. Kramer et al. (1989) found that variations in mental workload can also be discriminated with a high degree of accuracy with a relatively small amount of ERP data. While these results suggest that on-line assessment of mental workload may be feasible in the future, a good deal of additional research is required to validate and extend these initial findings to more complex scenarios.

Electroencephalographic (EEG) Activity

Overview

EEG has the longest history of any of the CNS measures that I will discuss. Berger (1929) provided the first report of changes in the frequency composition of the EEG with variations in the difficulty and type of task. Since the late 1920s, EEG has been used both clinically and experimentally to examine changes in the electrical activity of the brain in response to changes in neurological function, psychopathology, and cognitive activity.

It is perhaps not surprising that, since both EEG and ERPs are derived from the same physiological activity, they share a number of advantages and limitations. For example, they are both susceptible to the same set of artifacts which include: 60 Hz electrical "noise," eye movements (EOG), electromyographic (EMG) activity, and the electrical activity of the heart (ECG). However, since the ongoing EEG is substantially larger than ERPs, the problem of contamination is less severe for the EEG. The two aspects of the electrical activity of the brain are also similar in that they can both be recorded continuously. However, unlike the ERP, the EEG can be recorded in the absence of discrete stimuli or responses. Thus, while EEG reflects both phasic and tonic activity of the CNS, ERPs are generally employed to investigate phasic, stimulus, or response-related changes in information processing.

EEG is traditionally recorded from the scalp and is composed of a composite of waveforms with a frequency range of between 1 and 40 Hz and with a voltage range of 10 to 200 microvolts. The voltage-x-time vector is usually decomposed into a number of constituent frequency bands including: delta (up to 2 Hz), theta (4-7 Hz), alpha (8-13 Hz), and beta (14-25 Hz). In addition to differing in frequency, these components also vary in amplitude such that, while alpha and theta are relatively large, delta and beta are smaller in amplitude.

Sensitivity and Diagnosticity

The most ubiquitous changes in the EEG as a function of workload are found in the alpha band (Gale & Edwards, 1983). These changes have usually taken the form of an inverse relationship between alpha power and task difficulty (Gale, 1987; Gevins & Schaffer, 1980). For example, Natani and Gomer (1981) examined changes in EEG as pilots flew a number of missions in a fixed-base part task trainer. The most difficult missions that were characterized by pitch and roll disturbances were associated with decreased alpha power. Sterman, Schummer, Dushenko, and Smith (1987) examined EEG changes as a function of mission difficulty in a series of simulator and aircraft studies and found decreases in alpha power over the left hemisphere with decreases in flight performance. In a laboratory study, Sirevaag et al. (1988) found decreases in alpha power as subjects transitioned from a single- to a dual-disk. Finally, Pigeau, Hoffmann, Purcell, and Moffit (1987) replicated the inverse relationship between task difficulty and alpha power with a series of laboratory tasks. However, while this relationship was obtained for subjects that were classified as moderate or high alpha generators, the relationship between task difficulty and alpha power was not found for the low alpha subjects. These results suggest that the sensitivity of alpha frequencies to changes in task difficulty may be strongly influenced by individual differences among subjects. The percentage of individuals that are low, intermediate, and high alpha generators remains to be determined.

In addition to the consistent relationship between alpha power and task difficulty, the results of a number of studies suggest that activity in the theta band may also be sensitive to the level of arousal of operators. For example, Beatty and O'Hanlon (1979; see also Beatty, 1977) found that subjects who were taught to suppress theta activity performed better on vigilance tasks than control subjects and subjects who were taught to augment their theta activity. These effects were obtained for groups of college students and trained radar operators. Unfortunately, the magnitude of the performance differences was relatively small and the performance benefits were limited to situations which normally result in vigilance decrements.

More recent studies have found decreases in theta activity with transitions from single- to dual-tasks (Sirevaag et al., 1988) and with increases in multi-task difficulty (Natani & Gomer, 1981). However, in a study by Pigeau et al. (1987) theta power was found to initially increase with increments in the difficulty of an addition task and then decrease at high levels of difficulty. Although the results obtained by Sirevaag et al. and Natani and Gomer appear, at first glance, to be inconsistent with the pattern of data obtained by Pigeau et al., an examination of the task employed in the three studies may resolve this dilemma. In both the Sirevaag et al. and Natani and Gomer studies, subjects were performing in difficult multi-task settings, while in the Pigeau et al. study, subjects performed a relatively simple addition task. If we assume that subjects could perform most of the versions of the arithmetic task with little effort, it is perhaps not surprising that theta power did not decrease until the most difficult version of the task (e.g., addition of five 2-digit numbers).

With regard to diagnosticity, it appears that, while changes in the EEG spectra and particularly in the alpha and theta bands may provide an index of overall levels of arousal or alertness, they are not selectively sensitive to different varieties of processing demands. Another limitation of EEG relative to techniques such as ERPs is poor temporal resolution. While ERPs can be used to provide precise chronometric information concerning operators' strategies and workload (e.g., usually with 1-millisecond accuracy), EEG is generally used to provide average measures of alertness across time periods of several minutes. However, more diagnostic information may be available in the dynamic changes in EEG spectra across time and scalp sites than has been obtained from traditional frequency decomposition techniques (Gevins et al., 1979; Gevins, 1988).

Intrusiveness

Given the EEG can be recorded in the absence of overt behavior or the occurrence of discrete environmental events, it qualifies as a relatively unobtrusive measure of the general level of alertness of an operator. Even the constraints of bulky amplifiers and computer equipment that are employed in the laboratory may be surmounted by the use of FM recorders or telemetry devices.

Reliability

In accordance with most physiological measures, there has been a dearth of formal assessments of the reliability of EEG measures of mental workload. However, the consistent pattern of relationships between power in the alpha and theta bands and task difficulty that have been obtained in numerous studies suggests that this class of techniques provide a reliable measure of the general level of alertness of operators. It is important to note, however, that individual differences may exert a powerful influence on the reliability of the task difficulty/alpha power association (Pigeau et al., 1987)

Generality of Application

The collection of EEG data in extra-laboratory environments is susceptible to the same set of artifacts that are encountered with ERPs. These include: contamination from physiological signals such as ECG and EOG, contamination from other sources of electrical activity such as 60 Hz line noise, and contamination from changes in operator state (e.g., emotional state, physical state). While most of these potential artifacts can be minimized by the judicious selection of frequency filters and filter cutoffs (Coles, Gratton, Kramer, & Miller, 1986), the separation of mental load from emotional and physical load may be problematic in ambulatory operators who perform relatively sustained tasks. However, if it is assumed that emotional and physical load contribute to mental load (Hart et al., 1986; Reid, 1985), then the ability to separate these aspects of operator load is less important.

The question of whether EEG can be recorded in simulators and operational environments has been affirmatively answered by a number of recent studies. Systematic relationships between EEG power in the alpha and theta bands and mission difficulty have been obtained in high performance aircraft simulators (Natani & Gomer, 1981) and fixed wing military aircraft (Sterman et al., 1987). The sensitivity of these measures to variations in workload in laboratory settings has also been generalized from college students to professional radar operators (Beatty & O'Hanlon, 1979).

Magnetoencephalographic (MEG) Activity

Overview

The synchronous activation of neurons produces both electrical and magnetic fields that can be recorded from the scalp. The electrical manifestations of this neuronal activity, EEG and ERPs, have been discussed above. Magnetic fields which are much weaker than the comparable electrical activity (e.g., magnetic sensory responses are approximately 100 femtotesla as compared to urban "noise" which is approximately 100,000,000 femtotesla) may be reliably recorded with the aid of Superconducting Quantum Interference Devices (SQUIDS).

The recording of the magnetic activity of the brain during active task performance has begun relatively recently and therefore has not yet produced a wealth of information concerning human information processing (Beatty, Barth, Richer, & Johnson, 1986). However, since the MEG technique provides information that complements EEG and ERPs, it offers the potential for enhancing our understanding of the relationship between neurophysiological concepts of capacity and the psychological concept of mental workload. In particular, since MEG activity is relatively immune from "spatial smearing" that plagues the recording of electrical activity, it may be quite useful in localizing the scalp magnetic fields that are sensitive to changes in processing demands (Cuffin & Cohen, 1979; Williamson & Kaufman, 1981). However, at present the painstaking data recording techniques required to "localize" the source of the MEG activity make it an impractical tool for the analysis of complex multi-task designs. This methodological limitation should be overcome in the near future with the development of large array recording devices (Romani, 1987).

Sensitivity and Diagnosticity

Like electrical activity, the magnetic activity of the brain can be decomposed into components in both the frequency and the time domains that occur in response to perceptual, cognitive, and motor events. Thus, given that the magnetic activity includes EEG and ERP counterparts, it can be considered to be both globally sensitive to operator arousal and alertness, as is the case for EEG, and specifically sensitive to different aspects of information processing and mental workload-like components of the ERP.

While MEG can be analyzed in both the frequency and time domains, most of the empirical investigations have concentrated on uncovering the neuroanatomical loci of sensory, cognitive, and motor components of the ERPs and their magnetic counterparts. For example, a number of investigators have employed the MEG technique to examine components that are sensitive to aspects of auditory (Hari et al., 1989; Arthur & Flynn, 1987) and visual attention (Aine, George, Medvick, Oakley, & Flynn, in press). Several of these studies have found evidence for the existence of a number of neuroanatomically distinct attentional or resource sensitive components (Hari et al., 1984; Kaukoranta, Sams, Hari, Hamalainen, & Naatanen, in press; Lounasmaa, Hari, Joutsiniemi, & Hamalainen, in press; Makela, Hari, & Leinonen, 1988). While such information has not yet been applied to the study of mental workload, it may prove useful in further decomposing the processing demands that are imposed on human operators.

Intrusiveness

The intrusiveness of the MEG technique depends on whether additional signals are introduced into the operators task. For example, while event-related magnetic signals can be recorded from t -' relevant or secondary task events, MEG can also be recorded in the absence of discrete stimuli or responses. Thus, the MEG technique incorporates both the continuous recording that characterizes the EEG technique as well as the precise time locking to experimental events that is accomplished with ERPs.

Another characteristic of MEG recording, which may have a serious impact on operator state and performance strategies, is the requirement to repeat an experiment numerous times while searching for the neuroanatomical loci of scalp recorded fields. The replications are necessary to ensure sufficient spatial resolution for the derivation of topographical maps of the magnetic fields. However, this limitation is technical in nature and will be resolved with the development of large array recording systems.

Reliability

Given that the MEG technique has not yet been employed specifically in the assessment of mental workload, the reliability of the methodology is unknown. However, the reliability of recording sensory components of the MEG in relatively simple laboratory paradigms appears to be quite high for both normal as well as neurological patients (Barth, Sutherling, Engel, & Beatty, 1982, 1984; Williamson & Kaufman, 1981).

Generality of Application

The methodological constraints of the MEG technology make it impractical to record these signals outside of a well controlled laboratory environment. One such requirement is the necessity for using superconducting technology to record the magnetic fields generated by neural tissue. For instance, the sensors that are used in the SQUID are encased in a dewar filled with liquid helium which maintains the sensing apparatus near 4 degrees Kelvin. However, this limitation may be overcome in the near future with the development of high-temperature superconducting materials.

A second methodological constraint is the fact that few recording devices (from 1 to 7) are encased within a SQUID. Since the derivation of the orientation and location of the source of scalp recorded magnetic potentials requires that the signal is measured at an extensive number of scalp locations, experimental conditions must be replicated numerous times. Furthermore, since MEG components suffer from the same signal/noise ratio problems encountered with the most ERP components, averaging of several signals at each location is required. However, as indicated above, the development of large array recording devices and signal enhancement techniques should aid in the resolution of these problems.

In summary, while the recording of the magnetic activity of the brain may provide insights into operator states and performance strategies not available with other techniques, MEG will, for the foreseeable future, be limited to well-controlled laboratory settings. However, the capability of the technique to "localize" the source of scalp recorded fields may be quite useful in testing the physiological assumptions of capacity models of mental workload.

Brain Metabolism

Overview

The measurement of regional cerebral blood flow (rCBF) and the metabolic activity of the brain has recently been applied to issues of human information processing (Phelps & Mazziotta, 1985; Posner et al., 1988; Risberg & Prohovnik, 1983; Sokoloff, 1981; Ter-Pogossian, Raichle, & Soble, 1980). Although these techniques are "noninvasive" in the sense that they do not require surgical intervention, the need to employ radioisotopes necessitates that the measures be restricted to laboratory settings. Perhaps the best known of this class of techniques is PET. The PET technique involves three major components. First, glucose molecules are labeled with a radioisotope such as oxygen-15 or fluorine-18. These isotopes decay with the emission of positrons that combine with electrons to produce two gamma rays. The gamma rays are emitted 180 degrees apart from the head. The second component of the PET technique, the positron tomography, records the gamma ray activity and constructs a series of cross-sectional maps of the distribution of radioactivity in the tissue. Finally, tracer kinetic models are used to provide a mathematical description of the transport and biochemical reaction sequences of the labeled compounds.

The rCBF measurement techniques differ from PET in that blood rather than glucose molecules are tagged with a radioactive tracer such as xenon 133. Similar to PET, the electromagnetic radiation emitted from the tracer is detected by a device that surrounds the head. A computer then converts changes in the rate of flow of the tracer into a visual depiction of localized differences in cerebral blood flow.

Techniques such as PET and rCBF complement the information derived from the recording of electroencephalographic activity, since while ERPs can provide precise temporal localization of different aspects of information processing, spatial resolution is quite limited. On the other hand, while the temporal resolution of PET is limited by the decay rate of the radioisotopes (e.g., it takes at least 30 seconds to produce a PET map), spatial resolution of the metabolic activity can be quite precise. Thus, the relative strengths of electrical/magnetic and metabolic measurement techniques suggest that their joint use should provide a detailed view of the changes in brain activity that accompany variations in human information processing.

Sensitivity and Diagnosticity

A number of recent studies have found systematic relationships between measures of blood flow and task complexity in single- and dual-task settings (Gur et al., 1988; Phelps & Mazziotta, 1985). In one such study, Risberg and Prohovnik (1983) instructed subjects to view a stationary spiral, view a rotating spiral, or perform a spatial after-effects test. Increases in average cerebral blood flow in these conditions compared to a resting baseline were 5, 7, and 12 percent, respectively. Furthermore, the conditions were also distinguished on the basis of increases in blood flow in different brain regions.

A clever use of measures of cerebral blood flow and Donders' subtractive logic (1869) has been reported by Posner et al. (1988). In their study, subjects participated in a number of different conditions including: fixating a central marker, passively viewing visually presented words, repeating visually presented words, generating uses of words, and monitoring for words from specific semantic categories. Blood flow maps were obtained for each of the conditions. Assuming that each of the conditions required different forms of processing, the authors performed a number of subtractions to isolate the brain regions that were active during simple word reading. For instance, it was suggested that the processes of semantic association and attention could be isolated by subtracting the map obtained in the repeat word condition from the map obtained in the generate-word-use condition. While the Posner et al. (1988) study does not address workload issues per se, the joint use of cerebral blood flow measures and subtractive logic might prove useful in examining the type and magnitude of resources utilized during single- and dual-task performance.

With regard to diagnosticity, measures of brain metabolic activity are uniquely sensitive to changes in both the magnitude and the neuroanatomical loci of patterns of energy requirements in the brain. To the extent that models of workload (Freidman & Polson, 1981; Kinsbourne & Hicks, 1978; Wickens, 1980) specify resources or capacities that have been localized in portions of the brain, these techniques might be quite useful in decomposing the demands of tasks and task combinations. For example, Wickens's (1980) Multiple Resource model specifies that task compete for resources along three different dimensions: codes of processing (verbal and spatial), stages of processing (perceptual/central and response), and modalities of input (visual and auditory) and output (speech and manual). While the modality requirements can be observed without the use of any special measurement techniques, it is often difficult to determine whether operators process information in a verbal or spatial mode. The sensitivity of brain metabolism measures to changes in the spatial distribution of metabolic requirements may be quite useful in discriminating among these modes of processing.[1]

---

[1]Given that perceptual/central processing mechanisms appear to be widely distributed within the brain, the use of metabolic measures to discriminate among resource demands on the stages of processing dimensions is less promising.

Intrusiveness

The methodological requirements of the measurement of the metabolic activity of the brain, such as the use of radioisotopes and recording devices such as the positron tomograph, place relatively severe restrictions on the number of settings in which these techniques may be utilized. However, it appears that within the laboratory, measures of metabolic activity may be collected as subjects perform a wide variety of tasks. Thus, while this class of measures must be considered intrusive in many settings, they also have the potential to provide important information concerning the validity of the theoretical assumptions (e.g., interaction of verbal and spatial processing codes) underlying multiple resource models of multi-task processing.

Reliability

Similar to many physiological measures, there has been a lack of formal reliability assessment, especially pertaining to evaluations of mental workload. However, this lack is not particularly surprising, since the use of this class of measures in the study of human information processing is very recent. While formal reliability evaluations have not been conducted, the replicability of effects that demonstrate the sensitivity of these measures to processing demands and subject strategies provide some confidence in the reliability of these measures.

Generality of Application

The collection of brain metabolism data in extra-laboratory environments is complicated by several factors. First, depending on the decay rate of the radioisotopes, it can take anywhere from 30 seconds to several minutes to produce a measure of metabolic activity. During this imaging period, it is assumed that the subject is performing the assigned task in a uniform manner. While this assumption might be accurate for relatively simple tasks, situations in which mental workload is of interest are usually characterized by a variety of processing demands that change in relatively unpredictable ways. Thus, given the current level of temporal resolution available with this class of techniques, it may be unfeasible to assess workload in many settings.

Second, the use of radioisotopes and positron tomographs or other similar recording equipment renders the collection of metabolic activity impractical for ambulatory operators. Thus, given the limits of temporal resolution as well as the requirement for a relatively sedentary subject, these techniques are most applicable for situations in which workload is to be assessed in relatively simple tasks with nonambulatory operators (e.g., a comparison of new displays for a command, control, and communication (C3) system).

Endogenous Eye Blinks

Overview

Since a good deal of the information that is necessary to perform complex, real-world tasks is acquired through vision, it would seem reasonable to assume that measures of ocular activity might provide insights into aspects of information processing, and workload. In fact, measures of eye scanning patterns and blink characteristics have been employed for over 50 years in the investigation of mental activities (Hall & Cusack, 1972; Ponder & Kennedy, 1927). In this section, be sensitive to aspects of mental workload (see Senders, 1980; Wierwille, 1979 for reviews of the relationship between scan patterns and mental activities).

The endogenous blink has been distinguished from other blinks (e.g., reflex blinks, voluntary closures) by the absence of an identifiable eliciting stimulus (Stern, Walrath, & Goldstein, 1984). While the neurophysiology of these blinks is not well understood, it appears that they are controlled by the CNS via the VII cranial nerve. A number of techniques have been used to record blinks, including: corneal reflection methods, photographic and video scanning, and electrooculographic (EOG) procedures (Tursky, 1974; Young & Sheena, 1975). The most popular of these measures is EOG, which involves the placement of electrodes above and below an eye. The EOG measures blinks by recording changes in the potential difference between the cornea and the retina as the eyelid moves between closed and open positions.

Sensitivity and Diagnosticity

Similar to most of the other physiological techniques discussed thus far, blink activity can be decomposed into a number of different components. These components include: blink rate, blink duration, and blink latency relative to a stimulus or response. The most extensively studied characteristic of blinks has been their rate.

Blink rate has been found to decrease with the occurrence of predictable stimuli (Bauer et al., 1987) and in visual as compared to auditory tasks (Goldstein, Strock, Goldstein, Stern, & Walrath, 1985). In both of these cases, decreased blink activity is associated with the requirement to extract information from the visual environment. While this pattern of findings is consistent with the structure of the tasks that have been examined, a more confusing picture is portrayed by studies that have investigated the relationship between task demands and blink rate. For example, while Wierwille, Rahimi, & Casali (1985) found increases in blink rate when the navigational demands of a simulated flight mission increased, Stern and Skelly (1984) observed decreases in blink rate when a copilot took command of an aircraft and Sirevaag et al. (1988) found decreases in blink rate when subjects transitioned from a single to a dual task. While these discrepancies might be explained in terms of the visual requirements of the tasks (e.g., in both the Sirevaag et al. and the Stern and Skelly studies, the visual processing demands increased in the more difficult conditions, while the visual processing requirements were essentially the same in the different navigational load conditions in the Wierwille et al. study), other investigators have failed to find a significant relationship between blink rate and processing demands in a variety of visual and auditory tasks (Bauer et al., 1985; Casali & Wierwille, 1983). Thus, based on these findings, it appears that additional empirical and theoretical effort is required before blink rate could be recommended as a measure of mental workload.

In contrast to the blink rate data, other measures of blink activity appear more promising as measures of human information processing and workload. For example, the latency of blinks relative to the occurrence of task relevant information has been found to increase with increases in set size in memory-comparison tasks (Bauer et al., 1987), increase in dual- relative to single-task conditions (Sirevaag et al., 1988), and increase when responses are required in auditory discrimination tasks relative to nonresponse trials (Goldstein et al., 1985). This pattern of results is consistent with the interpretation of earlier studies which examined the relationship between blink latency and information processing (Stern et al., 1980):

> If taken at face value, these data suggest that, in the absence of a motor response, the occurrence of a blink marks the termination of the stimulus evaluation process. When a response is required, however, the blink appears to be delayed to the end of response selections, or perhaps the motor programming process. (p. 31)

Thus, it appears that blinks are inhibited until operators have had sufficient time to extract and process the critical task-relevant information.

In addition to blink latency, measures of closure duration have also been found to be systematically related to task demands. Closure duration has been found to decrease when copilots take over flight control duties from pilots (Stern & Skelly, 1984), decrease when operators are required to perform several tasks simultaneously relative to single-task control conditions (Sirevaag et al., 1988), and increase with time on task (Bauer et al., 1985; Oster & Stern, 1980), presumably due to increases in fatigue. Thus, similar to blink latency, operators appear to maintain fixation for longer periods of time when visual processing demands are high.

With regard to diagnosticity, the data obtained thus far suggest that measures of blink activity, particularly blink latency and duration, are sensitive to global aspects of information processing rather than specific components of mental workload. Additionally, it appears that blink rate and duration are sensitive to operator fatigue.

Intrusiveness

The intrusiveness of blink measurement depends on the techniques employed. For example, while the corneal reflection techniques usually require that the operator is relatively motionless, EOG can be recorded from ambulatory operators through the use of portable amplifiers and telemetry devices. Video techniques have also been developed that permit the operator a full range of motion during recording (e.g., helmet mounted video cameras). Thus, in general, the measurement of blink activity can be accomplished in a relatively unobtrusive manner.

Reliability

Given the consistent relationship obtained between task demands and blink latency/duration over a diversity of subject populations and tasks, it would appear that some characteristics of the endogenous eye blink provide a reliable measure of global aspects of task difficulty and workload. However, the fact that these measures are also sensitive to operator fatigue suggests caution when the objective is to decompose the effects of system variables on operator state and information processing strategies. Finally, the inconsistent patterns of data obtained for blink rate indicates that this aspect of the endogenous eye blink is not yet ready for application.

Generality of Application

While most of the investigations of the sensitivity of the endogenous eye blink to information processing activities have been conducted in laboratory settings, some studies have been performed in high fidelity simulators and operational systems. For example, Stern and Skelly (1984) explored the utility of a number of blink characteristics as indices of mental workload of pilots and copilots in an A7 simulator. The pilot in charge of the aircraft produced fewer and shorter duration blinks than did the pilot who was second in command. When the pilot and copilot reversed roles, the blink pattern also reversed. In a similar series of studies, Wilson, Purvis, Skelly, Fullenkamp, and Davis (1987; see also Skelly, Purvis, & Wilson, 1987) found that, for pilots flying A7 aircraft and simulators, the most difficult flight segments were associated with the lowest blink rates. Thus, based upon these studies, it appears that a number of characteristics of the endogenous blink can be reliably recorded in extra-laboratory environments.

A potential problem for the measurement of blinks in operational settings is their sensitivity to factors other than processing demands, such as air quality, defensive reactions, and fatigue. However, these potential confounds can be minimized by ensuring that these factors do not vary in the contexts which are to be compared (e.g., use short missions to reduce fatigue, record blinks in climate controlled environments, etc.).

Another important question is whether the endogenous eye blink can be used in an on-line context to measure transient changes in mental workload and information processing strategies. A potential bottleneck in the application of this technique in an on-line context is the fact that, while blink latency and closure duration have proven reliable in laboratory settings, endogenous blinks do not occur in response to every task relevant stimulus or response. Therefore, relatively rapid and short-lived changes in processing demands may not be indicated in the blink data. However, systematic evaluations of the temporal resolution of the endogenous eye blink remain to be performed.

Pupil Diameter

Overview

The observation of changes in the diameter of the pupil as a function of attention and information processing can be traced back hundreds of years to stories about merchants who claimed to be able to determine a customer's interest in a product by watching changes in their pupils (Hess, 1975; Janisse, 1977). While these anecdotal reports of the utility of pupillary changes have appeared in both eastern and western literature for centuries, empirical investigations of the association between pupillary changes and mental activities first appeared in the mid 1960s (Hess, 1965). At that time, changes in pupil diameter were related to the level of interest in an object, place, or person.

The pupil, which can vary in size from .2 to .8 mm, is controlled by a set of antagonistic muscles in the iris. One muscle group, the dilator pupillae, is innervated by fibers from the SNS. Stimulation of this muscle causes a retraction of the iris, thereby increasing the size of the pupil. The second muscle group, the sphincter pupillae, is innervated by fibers from the PNS. Stimulation of this muscle expands the iris, thereby decreasing the size of the pupil. While the relationship between the branch of the ANS (e.g., the SNS and PNS) and the muscles controlling the pupil is clear, the relative contribution of the SNS and the PNS to changes in the size of the pupil can vary. For example, pupil dilation can be accomplished by either an increase in SNS activity or a decrease in PNS activity.

It is important to note that, while our interest is in the relationship between pupil diameter and mental activities, the largest changes in the pupil occur in response to other factors (Tryon, 1975).

For example, the main function of the pupil is to protect the retina by controlling the amount of illumination that enters the eye. This light reflex is accomplished by a relatively rapid response to transient changes in illumination. A second function of the pupillary system, the near reflex, concerns the constriction of the pupil in response to a shift in fixation from a far to a near object. The constriction of the pupil, which accompanies a change in the vergence and accommodation of the eyes, presumably increases the depth of field of the visual system. The changes in the pupil that appear to reflect variations in mental activities are quite small relative to the pupillary changes observed during the light and near reflexes.

Sensitivity and Diagnosticity

The use of pupillary changes as an index of mental workload can be traced to Kahneman's (1973) seminal book on attention and effort. Kahneman reports a number of studies in which pupil diameter varied with the processing demands of the task. In his capacity model of human information processing, he employs a measure of pupil diameter as the link between the hypothetical construct of capacity and the arousal system.

More recent research has focused on explicating the sensitivity of the pupillary response to a number of task parameters (Beatty, 1982a, 1986). Pupillary changes have been found to be sensitive to perceptual (Beatty, 1988; Qiyuan, Richer, Wagoner, & Beatty, 1985), cognitive (Ahern & Beatty, 1981; Beatty, 1982a; Casali & Wierwille, 1983), and response related processing demands (Richer & Beatty, 1985, 1987; Richer, Silverman, & Beatty, 1983) in a variety of tasks. This pattern of findings suggests that, while the pupillary response is sensitive to a wide range of processing activities, it is not very diagnostic. Thus, variations in pupil diameter might best serve as an index of global changes in information processing. The sensitivity of the pupillary response to a variety of processing demands is consistent with its presumed neurophysiological role. Beatty (1982a) has suggested that "the task evoked pupillary dilations very likely reflect the cortical modulation of the reticular core during cognitive processing" (p. 290). Given that the reticular activating system receives inputs from a variety of cortical and sub-cortical structures, it is not surprising that the pupillary response is sensitive to a wide range of processing demands.

It is interesting to note that, while the pupillary response is not diagnostic with respect to the types of processing resources required for task performance, it does appear to distinguish between resource and data-limited processing. Evidence for this claim is suggested by the results of a signal detection study in which pupil diameter was insensitive to changes in the discriminability of weak auditory stimuli. However, performance measures did distinguish among experimental conditions. Beatty (1982a) interpreted these results to suggest that the pupillary response is insensitive to processes that cannot benefit from the allocation of additional resources. The auditory discrimination task employed in the study does in fact possess the attributes of a data-limited process suggested by Norman and Bobrow (1975) in which processing is limited by the quality of the data rather than the effort invested in the task.

While most investigators have found that the pupillary response provides a sensitive and reliable measure of processing demands, a few studies have obtained negative results. For example, Wierwille et al. (1985; see also Wierwille & Conner, 1983) conducted an experiment in which pilots were required to maintain a fixed airspeed, altitude, and heading in a flight simulator. In addition to straight and level flight control, the pilots were also required to perform navigational problems of varying difficulty. Measures of performance and subjective difficulty were found to discriminate among the levels of navigational complexity. However, measures of pupil diameter were insensitive to the experimental manipulations.

An examination of the pupil diameter recording methodology provides a potential explanation for these findings. In an effort to ensure that the subject's eyes and head were stationary during the measurement of pupil diameter, Wierwille et al. recorded pupil size approximately 3 seconds after the first glance at the navigational display. Given that the pupillary response is relatively rapid, usually occurring with 600 milliseconds of an eliciting stimulus, it is not surprising that measurements of pupil diameter taken 3 seconds after display did not discriminate among experimental conditions. It was also the case that only 12 pupillary responses were available for each level of navigational load. Given the fact that the magnitude of the pupillary response related to information processing is small relative to that produced in response to changes in illumination and object distance, 12 trials may be an insufficient amount of data to obtain an acceptable signal/noise ratio. Both the timing and the signal/noise ratio issues suggest caution in the application of the pupillary response to extra-laboratory environments.

Intrusiveness

The intrusiveness of the pupillary measure depends on the methodological requirements of the techniques employed during recording. Two optical techniques, photographic pupillometry and electronic video-based pupillometry, have been used in recent years. Photographic pupillometry, the simpler and less expensive of the two techniques, involves photographing changes in the pupil *during task performance*. The pupil is usually photographed every .5 to 1 second and the changes are quantified by measuring the diameter of the image of the pupil with an ordinary ruler. As might be expected, such a technique is quite time consuming when large numbers of subjects and experimental conditions are involved. This technique also requires that the head remain relatively stable during the data collection (e.g., a chin rest and a bite bar are usually employed).

The second technique, electronic video-based pupillometry involves the use of high-resolution linear infrared video cameras to obtain an image of the iris and the pupil. This technique, while more expensive than photographic pupillometry, offers more flexibility in that data can be recorded continuously without the need for stability of the operator's head.

Reliability

As described above, substantial literature suggests that the pupillary response is a sensitive and reliable index of processing demands in a wide variety of tasks. However, there have been reports of failures to find a systematic relationship between pupil diameter and task difficulty. While these data suggest the need for careful experimental control, they do not indicate a lack of reliability of the pupillary measure (see Sensitivity and Diagnosticity above). It is also important to note that the pupillary response is sensitive to factors other than processing demands including changes in illumination and in the position of fixated objects, fatigue, and emotional state.

Generality of Application

Given the requirement for precise experimental control in order to ensure that pupillary changes are not due to factors such as the light and near reflexes, it would appear that the use of the pupillary response as a measure of mental workload should be confined to laboratory settings. However, even within the laboratory, several factors must be considered prior to employing the pupillary measure. For instance, since the pupillary changes elicited by mental activities are small relative to those obtained in response to other factors, signal averaging is necessary to enhance the signal-to-noise ratio. The requirement to repeat stimulus presentations several times constrains the number of situations in which the pupillary response might serve as a workload measure.

Second, a number of investigators have distinguished between phasic and tonic changes in pupil diameter. It is generally found that tonic or baseline measures of pupil diameter are insensitive to variations in processing demands while phasic measures are responsive to changes in mental activities (Beatty, 1982b). Given that phasic pupillary responses occur in close temporal proximity to eliciting stimuli or responses, it is important to implement data recording procedures that take advantage of this relationship. However, while these procedures may increase the investigator's ability to detect processing changes, they also limit the number of situations in which pupillary response may be used to index variations in mental workload.

Cardiac Activity

Overview

Over the past 25 years, measures of cardiac activity have been the most popular physiological techniques employed in the assessment of mental workload. The sensitivity of a number of different cardiac measures to variations in workload have been examined. These techniques include: the electrocardiogram (ECG), blood pressure measures, and measures of blood volume. While each of these techniques has been used in the evaluation of workload, measures of electrocardiographic activity have shown the most promise and therefore will be the focus of this review (see Larsen, Schneiderman, & Decarlo-Pasin, 1986 for a description of the blood pressure and blood volume techniques).

Structurally, the heart is divided into four interconnected chambers: two ventricles and two atria. Oxygen depleted, venous blood returns to the heart through the right atrium. Contraction of the atrium pumps this blood into the right ventricle. The second contraction pumps the blood out of the right ventricle through the pulmonary artery to the lungs. The oxygenated blood reenters the heart through the left atrium. The next contraction pumps this supply of blood to the left ventricle where the final contraction forces the blood through the aorta to the rest of the body.

Similar to most systems influenced by the ANS, the heart is innervated by fibers from both the SNS and PNS. The SNS serves to increase the firing rate of the pacemaker cells thereby increasing heart rate. The SNS also influences the distribution of blood throughout the body by constricting and dilating the blood vessels. The PNS affects the heart through the influence of the vagal nerve. Thus, changes in heart rate can occur on the basis of SNS, PNS, or both SNS and PNS activity. While it is often difficult to discern the contribution of the SNS and PNS to changes in heart rate, this may be accomplished in at least two ways. First, drugs may be used to selectively inhibit SNS or PNS activity (Linden, 1985). Second, it has been argued that certain aspects of cardiac activity are selectively influenced by either the SNS or PNS (Furedy, 1987; Furedy & Heslegrave, 1983; Porges, 1984).

The mechanical contractions of the heart are produced by electrical impulses generated by the pacemaker cells in the sinoatrial and artioventricular nodes of the heart. This electrical activity can be measured in the form of the ECG. Figure 2 presents a prototypical ECG recording. Each of the perturbations in the voltage-x-time function can be associated with different electrical events within the heart muscles. The P wave is produced by the depolarization of the atrial muscles, the QRS complex is the result of a depolarization of the ventricles, and the T wave is produced by a repolarization of the ventricles.

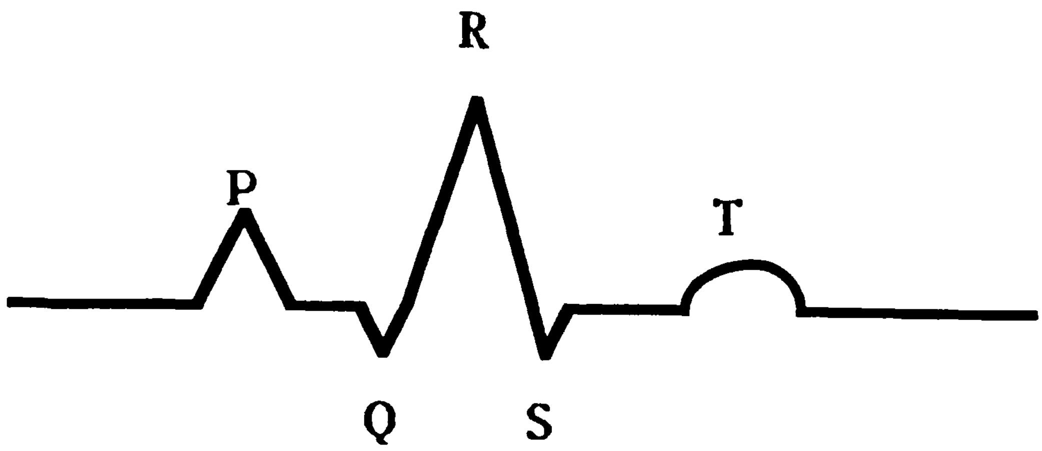

Figure 2. A graphical illustration of a normal ECG.

Given the magnitude of the signal (e.g., the QRS spike is approximately 1 millivolt), the recording of the ECG can be accomplished by the placement of two physically separated electrodes almost anywhere on the body. However, a number of standardized placements have been proposed in an effort to accentuate different aspects of the waveform (Larsen et al., 1986). Several problems can be encountered during recording. These include: low frequency artifacts caused by changes in the conductive characteristics of the skin, high frequency artifacts due to muscle activity and movement, and high frequency artifacts due to 60 Hz line noise. However, these problems can be corrected by the judicious selection of high- and low-frequency filter cutoffs.

The ECG signal is analyzed in both the time and frequency domains. The R wave is usually detected by a threshold detection device such as a Schmitt trigger and fed into a computer which is programmed to measure the number of spikes per unit time (heart rate--HR) or the inter-beat interval (IBI) between the R waves. At the level of a single observation, HR and IBI are reciprocally related. However, as soon as distributional parameters are computed, the measures are no longer linearly related. Thus, care should be taken when comparing HR and IBI averages and other distributional characteristics. Another concern is whether the data should be expressed in clock or cardiac time. Graham (1978a, 1978b) has argued that, to obtain unbiased measures, HR should be estimated in clock time, while IBI should be estimated in biological time. Frequency measures are usually estimated from R-R IBI data. This method of analysis will be discussed in detail in the following section.

Sensitivity and Diagnosticity

Of all of the measures that are derivable from the ECG, heart rate is the easiest to obtain. Simplicity of recording and analysis is an important reason why measures of heart rate have been so popular in the examination of human information processing and mental workload. Numerous studies have found systematic relations between measures of HR and a variety of information processing activities in both laboratory and field environments. For instance, several investigators have reported increases in HR during difficult mission segments in simulated (Harris, Bonadies, & Comstock, 1989; Lindhom & Cheatham, 1983; Wierwille & Conner, 1983) and actual flight in fixed wing aircraft (Roscoe, 1984; Speyer, Fort, Fouillot, & Blomberg, 1987). Unfortunately, there have also been a number of reports of failures to find systematic relationships between workload and HR (Casali & Wierwille, 1983; Hicks & Wierwille, 1979; Kalsbeek & Ettema, 1963; Salvendy & Humphreys, 1979; Wierwille et al., 1985).

One possible explanation for this seemingly inconsistent pattern of findings was offered by the Laceys in their intake-rejection hypothesis (Lacey, 1967; Lacey & Lacey, 1978). This hypothesis suggests that the direction of HR change is related to the types of task demands imposed upon an individual. HR is proposed to slow during the intake of environmental information (e.g., visual detection and discrimination, scanning, listening), while the rejection of environmental information increases HR (e.g., mental arithmetic, memory retrieval, problem solving). Thus, the inconsistent pattern of results obtained in the workload studies may be interpretable in terms of the types of task demands imposed upon the subjects. While the Laceys' theoretical formulations have been extended in a number of directions (but see Obrist, 1976, 1984 for an alternative model), researchers interested in the association between cardiac activity and workload have shifted their focus to other aspects of the ECG waveform.

The impetus for this shift can be traced to the research of Kalsbeek and colleagues (Kalsbeek & Ettema, 1963; Kalsbeek, 1971). In a series of studies, Kalsbeek found decreases in heart rate variability (HRV) with increases in the difficulty of a variety of tasks and task parameters. Small and often insignificant HR changes were obtained with the same manipulations that produced large HRV changes. In these studies HRV, which is also referred to as sinus arrhythmia, was measured as the variability of the R-R interval as a function of time. Subsequent to Kalsbeek's pioneering research, a number of different HRV measures were suggested in both the time and the frequency domains (Jenkins, Mitchel, & McClure, 1982; Opmeer, 1973; Van Dellen, Aasam, Mulder, & Mulder, 1985).

While a number of these time and frequency domain measures of HRV have shown systematic relationships with mental activities, the frequency-based measures offer a unique advantage. In particular, although time-based measures provide a global index of variability, the use of spectral analysis has enabled investigators to decompose HRV into components associated with different biological control mechanisms. Three major frequency bands have been examined: The lowest, which ranges from .02 to .06 Hz, is associated with vasomotor activity involved in the regulation of body temperature. The intermediate band, which includes frequencies from .07 to .14 Hz, is related to mechanisms involved in the short-term regulation of arterial pressure. Finally, the highest band, which ranges from .15 to .50 Hz, mainly reflects the effects of respiratory activity on HRV.

Activity in the intermediate and high frequency bands has been shown to be related to task demands. The .10 Hz component, the center point of the intermediate frequency band, has been the most extensively examined of the three frequency bands. This component has been found to decrease in power with increases in the amount of effort invested in a task (Aasam, Wijers, Mulder, & Mulder, 1988; Egelund, 1982; Hitchen, Brodie, & Harness, 1980; Mulder, 1979; Mulder & Mulder, 1980, 1981b; Mulder, Meijman, O'Hanlon, & Mulder, 1982). For example, power at .10 Hz has been found to decrease with the transition from single- to dual-task performance (Sirevaag et al., 1988), with increases in the memory load of a task (Aasman et al., 1987; Mulder & Mulder, 1981a), and with increases in subjective ratings of effort in a tracking task (Vicente, Thorton, & Moray, 1987).

It is interesting to note that under some conditions task demands appear to selectively modulate the power in the .10 Hz component without influencing the power in the low- and high-frequency bands. Van Dellen et al. (1985) found that, while the .10 Hz component decreased with increases in memory load, the other two bands were unaffected. Additional evidence for the diagnosticity of the .10 Hz component was obtained in a study by Aasman et al. (1987) in which reaction time reflected changes in the amount of visual noise and the number of memory set items, while the .10 Hz component was sensitive to only the latter manipulation. These results were interpreted to suggest that the .10 Hz component is sensitive to resource-limited, but not data-limited, processes.

In addition to the .10 Hz component, two other aspects of the HR signal appear to be potentially useful as workload metrics. Porges (1984) has argued that activity in the high frequency band, which reflects the effects of respiration on the heart, may be useful because it appears to provide a measure of the vagal influence on the heart (see also Broeckl, Jones, Johnson, & Fischer, 1989). This component has since been referred to as V to reflect its sensitivity to vagal influence. Given that the vagus nerve is primarily influenced by the PNS, the use of V may permit the investigator to decompose ANS activity during the performance of complex tasks. Furedy (1987) has suggested that the amplitude of the T wave component of the ECG may serve a similar function as V, in that T appears to primarily reflect SNS activity. In a recent study, Sirevaag et al. (1988) found that V and T could be disassociated in terms of their sensitivity to different aspects of performance in a dual-task paradigm.

Intrusiveness

Given that ECG: (a) can be recorded in the absence of discrete stimuli and responses, (b) possesses a fairly large signal/noise ratio, and (c) does not require the precise placement of electrodes to successfully detect the signal (e.g., QRS spike), it qualifies as an nonintrusive measure of mental workload. In fact, if the use of electrodes is bothersome to the subject, heart rate can be recorded by other means such as photoelectric plethysmography. In this technique, an infrared light source is directed towards a piece of tissue such as an ear or finger. The amount of light that passes through or is reflected back from the tissue is recorded by a photoelectric transducer. Since the light source is scattered by blood, the output of the photoelectric transducer provides a measure of the amount of blood in the tissue. Changes in blood volume can be used to trigger a cardiotachometer for purposes of heart rate recording.

Reliability

As outlined above, there have been considerable discrepancies in the literature concerning the efficacy of HR and HRV measures as indices of processing demands. Certainly part of this confusion can be traced to the complexity of the relationships between ECG components and the structure and processing demands of tasks (Lacey & Lacey, 1978). Similarly, the selective sensitivity of components of the HRV spectra to different biological control mechanisms further underscores the complexity of the mapping between mental activities and ECG components.

Assuming the level of complexity that is suggested by the intake-rejection hypothesis and the spectral decomposition of the HRV signal, how well do cardiac measures fare in terms of their reliability? Recent literature seems to suggest that certain components of HRV exhibit systematic and reliable relationships with task demands. The .10 Hz component decreases in power with increases in task demands. However, while this relationship is generally found for relatively large differences in task difficulty, the level of resolution available with this technique remains unexplored. The two other components described above, V and T wave amplitude, also appear to be promising candidate measures of selective aspects of mental workload. However, additional studies are needed to explore the advantages and limitations of these measures in both laboratory and applied settings.

Generality of Application

HR and HRV measures have been extensively explored in both laboratory and operational environments. Applications of HR measures have been described above. Measures of HRV have been found to discriminate between levels of task demands encountered by undersea divers (Jorna, 1985), city bus drivers (Mulder et al., 1982), driving examiners (Meijman, 1985), and keypunch operators (Kamphuis & Frowein, 1985). It is important to note that, while a number of studies have reported systematic relationships between HRV and task demands, not all applications of the HRV measures have been successful (Casali & Wierwille, 1983; Hicks & Wierwille, 1979; Wierwille & Conner, 1983). However, generally studies that have failed to report reliable relationships have used global measures of HRV rather than examining changes in the three spectral bands. Given that changes in HRV as a function of processing demands are most pronounced in the .10 Hz band, the use of global measures of HRV would appear to decrease the sensitivity of the technique (see Van Dellen et al., 1985).

As with other physiological techniques, a number of potential artifacts must be examined during the recording and analysis of HR and HRV data. First, the ECG signal can be contaminated by changes in the conductive characteristics of the skin (low frequency) as well as movements and muscle activity (high frequency). The possibility of encountering these artifacts can be reduced by careful experimental design (e.g., minimize movement and changes in emotional state) and the use of high- and low-pass filters. Second, speech tends to increase blood pressure which in turn influences power in the .10 Hz frequency band. Therefore, conditions in which there are dramatic differences in the amount of speaking may produce differential .10 Hz components despite relatively similar processing demands (in other aspects of the task). Finally, a similar effect can be produced by changes in the frequency and depth of respiration. While the .10 Hz component had originally been thought to be immune to changes in the pattern of respiration, recent research has called this assumption into question (Sirevaag et al., 1988).

Electrodermal Activity

Overview

The recording of electrodermal activity (EDA) was first reported in the late 1980s. Two different measurement techniques were developed at approximately the same time. Fere (1888) measured changes in the resistance of the skin to the passage of a small current from an external source. Modifications of this technique are used today as measures of skin resistance (SR).

Early interest in electrodermal activity concerned its sensitivity to changes in emotion and arousal. Jung (1907; Peterson & Jung, 1907) viewed EDA as a window on the unconscious and particularly on the experience of emotion. Other researchers employed measures of EDA to examine dimensions of emotion such as fear, sadness, and joy (Bayley, 1928; Linde, 1928; Waller, 1918). The sensitivity of EDA to variations in emotional experience ultimately led to its use in the detection of deception, which is still a popular application of EDA today (Waid & Orne, 1982).

As briefly described above, several different measures of EDA have been developed. While measures of the change in SR during the imposition of an external current source was popular in the past, this measure has been largely replaced by measures of skin conductance (SC). Although conductance units can be mathematically transformed to resistance units (conductance <mhos> = 1/resistance <ohms>), the distributional properties of conductance data and its systematic relationship to the underlying physiological mechanisms have made it more popular than SR measures (Fowles, 1986).

Electrodermal activity can be characterized both in terms of its baseline or tonic level as well as its phasic response to an environmental event. Measures of tonic EDA are referred to in terms of their level (SPL & SCL), while measures of phasic activity are referred to as responses (SPR & SCR). In addition to phasic and tonic activity, spontaneous or nonspecific EDA is also measured. Generally, EDA is measured as a change relative to a resting baseline. It is important to note that the amplitude of a phasic response is partially dependent on the tonic level prior to the occurrence of an environmental event, particularly when SR rather than SC is recorded. Given this dependency between level and response, Lykken, Rose, Luther, and Maley (1966) have suggested that the amplitude of the phasic response should be expressed relative to the subject's minimum and maximum tonic levels. The latency of the electrodermal response to the occurrence of stimulation is usually 1.4 to 2.5 seconds.

Changes in the electrical activity in the eccrine sweat glands form the basis of EDA. The eccrine sweat glands, which are most numerous on the palms of the hands and the soles of the feet, are under the influence of the sympathetic nervous system. In essence, the eccrine sweat glands function as variable resistors. The level of sweat in a gland is proportional to the resistance of that gland (see Fowles, 1986 for a more in-depth discussion of the physiological substrates of EDA). The major function of the glands is thermoregulation. Thus, in addition to responding to cognitive and emotional factors, EDA is sensitive to temperature, humidity, age, sex, time of day, and season.

Sensitivity and Diagnosticity

Kahneman (1973) employed a number of autonomic nervous system signals as measures of cognitive effort during the development of his Undifferentiated Capacity Theory. In one such study, Kahneman, Tursky, Shapiro, & Crider (1969) found that SR, pupil diameter, and heart rate varied with the number of digits that subjects were required to silently add.

The finding of a reliable relationship between performance and the magnitude of EDA suggested that individual differences in spontaneous levels of electrodermal activity might be predictive of the quality of task performance. This hypothesis led to a program of research that attempted to characterize individuals in terms of SC levels. Generally, subjects are classified into one of two groups: labiles who exhibit relatively large and frequent nonspecific SCRs and stabiles who exhibit much smaller and less frequent SCRs. Labiles have been found to be more resistant to vigilance decrements than stabiles (Hastrup, 1979; Sostek, 1978; Vossel & Rossman, 1984), respond more quickly in simple and choice reaction time tasks (Wilson, 1987; Wilson & Graham, 1989), and detect more targets in selective attention tasks (Straube, Schlenker, Klessinger, Himer, & Boven, 1987). However, there have been other situations in which stabiles have outperformed labiles (O'Gorman & Lloyd, 1988). This pattern of results has been taken to suggest that electrodermal lability is related to the processes of activation, arousal, and alertness (Conte & Kinsbourne, 1988; Crider, 1979; Hugdahl, Fredikkson, & Ohman, 1977). Thus, according to this interpretation, labiles would be expected to outperform stabiles in relatively simple and sustained tasks in which increases in arousal would reduce the detrimental effects of boredom and fatigue. On the other hand, the level of arousal experienced by labiles might be expected to impede the performance of more complex tasks.

The research on individual differences and performance has generally used measures of nonspecific or spontaneous EDA to classify individuals. Other researchers have examined the sensitivity of SCRs to variations in single- and dual-task difficulty and concluded that while non-specific manifestations of EDA are sensitive to general levels of arousal, SCRs appear to provide a more specific index of human information processing. For instance, Packer and Siddle (1989; see also Siddle & Packer, 1987) found that deviations in a train of repeated stimuli elicited larger SCRs and increased secondary task probe RTs than repeated stimuli. Dawson, Schell, Beers, and Kelly (1982) found that reinforced classically conditioned stimuli (CS+) elicited larger SCRs and slower probe RTs than CS-stimuli and that miscued USC-CS pairs also resulted in delayed probe RTs and large SCRs. Finally, Spinks, Blowers, and Shek (1985) presented subjects with a warning stimulus that predicted the difficulty of the subsequent imperative stimulus and found that SCRs varied with the predicted processing requirements (see also Dawson & Schell, 1982; Filion, Hazlett, Dawson, & Schell, 1989; Kazumi, Tetsuo, & Yo, 1984; Kenemans, Verbaten, Sjouw; & Slangen, 1988; Verbaten & Kenemans, 1987). These results have been interpreted in terms of the sensitivity of SCRs to the allocation of processing capacity both within as well as between tasks. Thus, while spontaneous EDA appears to be sensitive to general levels of arousal, SCRs seem to index the allocation of an undifferentiated form of processing resources.

Intrusiveness

Given that EDA can be recorded either in response to environmental events (e.g., SCR or SPR) or in the absence of stimuli (e.g., SCL, SPL, or spontaneous activity), it would appear to be a relatively flexible and noninvasive measure of ANS activity. On the other hand, the need to affix electrodes on the palms of the hands or the soles of the feet does place some restrictions on the types of tasks that can be performed during the recording of EDA.

Reliability

As with most physiological techniques, there has been a lack of formal evaluations of reliability, particularly in more complex single- and dual-task settings. However, the repeated finding of a systematic relationship between the magnitude of EDA and variations in processing demands provides some confidence in the reliability of the EDA technique. Thus, while EDA measures do not provide the level of diagnosticity that is available with measures of brain metabolism and ERPs, they do appear to provide a reliable index of general levels of arousal (e.g., nonspecific EDA activity) and resource demands (e.g., SCR and SPR).

Generality of Application

All of the studies that have been discussed in this review have been conducted in controlled laboratory settings. Although measures of EDA have been successfully collected in operational environments, such as automobile driving (Helander, 1975), a number of methodological constraints complicate the recording of EDA in extra-laboratory environments. For instance, several environmental and organismic factors can influence both the tonic and the phasic aspects of EDA. These factors include: temperature, humidity, time of day, season, sex, emotional state, and irregularities in respiration. Thus, the attribution of changes in EDA to variations in the processing demands of a task necessitates the careful control of each of these factors which in turn greatly reduces the number of non-laboratory settings in which EDA can be successfully employed.

It is also important to note that, while the magnitude of EDA provides a reliable index of processing demands in laboratory tasks, the temporal sensitivity of this technique is poorer than most of the other physiological measures. However, the level of temporal resolution of the SCR (1.3 to 2.5 seconds) may be more than adequate for many situations in which mental workload is of concern.

DISCUSSION AND CONCLUSIONS

Each of the physiological signals in this review possess a number of strengths and weaknesses as measures of mental workload. For instance, while some measures are sensitive to processing demands in general (e.g., pupil diameter, EDA), these measures are not very informative about changes in the fine-grained structure of processing requirements. However, although other measures such as ERPs, brain metabolism, and the T-wave amplitude of the ECG provide a great deal of diagnostic information concerning important aspects of mental workload, these measures are sensitive to only a small subset of the components of workload. Therefore, it would appear that the choice of measures must be guided by the breadth and level of analysis required in the evaluation of workload demands. Of course, this prescription is also true for primary, secondary, and subjective measures of mental workload. Given that mental workload is multidimensional in nature, no single measurement technique will be adequate in all settings. What I have tried to accomplish in this review, however, was to provide a theoretical and empirical basis for the selection of physiological signals for the measurement of different aspects of mental workload.

For the most part, physiological measures are relatively nonintrusive. Most of these measures can be recorded without requiring operators to perform extraneous tasks. This is a definite advantage over techniques such as secondary task measures that often interfere with performance on the task of interest. However, while physiological techniques may be nonintrusive in the sense that they do not generally require the addition of extraneous stimuli, the constraints involved in recording uncontaminated signals may encourage operators to modify the manner in which they perform their tasks. For instance, the fact that speech influences power in the .10 Hz band of the HRV signal suggests that the amount of verbal communication must be controlled when this measure is employed. Although this constraint may not be problematic in some situations, it would clearly be unacceptable in many settings (e.g., in a C3 environment, during flight, etc.). Therefore, the methodological requirements must be considered when selecting physiological measures of mental workload.

The range of sensitivity of physiological measures to the magnitude and temporal aspects of mental workload make this class of techniques potentially useful in a number of settings. For example, the relatively rapid response of ERPs and pupil diameter make these measures well suited for the evaluation of transient changes in processing demands. However, while these techniques are potentially useful in on-line contexts, they produce relatively small signals buried in a large amount of noise. Thus, the implementation of these measures must await the development of pattern recognition techniques that enable the rapid discrimination of signal and noise (for the application of such techniques, see Farwell & Donchin, 1988; Kramer et al., 1989).

Although a number of physiological techniques have been employed in operational contexts, the methodological requirements of these procedures often preclude their use in situations in which an extensive amount of movement is required. While these requirements constrain somewhat the applicability of the physiological techniques, there are more than enough environments in which cognitive aspects of performance dominate the physical aspects. Thus, given the successful resolution of a few methodological issues, we can expect to see an increase in the application of these techniques in extra-laboratory settings.

REFERENCES

Aasman, J., Mulder, G., & Mulder, L. (1987). Operator effort and the measures of heart rate variability. *Human Factors, 29,* 161-170.

Aasman, J., Wijers, A., Mulder, G., & Mulder, L. (1988). Measuring mental fatigue in normal daily working routines. In P. Hancock & N. Meshkati (Eds.), *Human mental workload.* Amsterdam: Elsevier.

Aine, C., George, J., Medvick, P., Oakley, M., & Flynn, E. (in press). Source localization of components of the visual-evoked neuromagnetic responses and effects of spatial attention. *Electronecephalography and Clinical Neurophysiology.*

Ahern, S. K., & Beatty, J. (1981). Physiological evidence that demand for processing capacity varies with intelligence. In M. Frediman, J. P. Das, & N. O'Conner (Eds.), *Intelligence and learning.* New York: Plenum Press.

Arthur, D., & Flynn, E. (1987). The effect of auditory selective attention on event-related magnetic fields of the human brain. In *Volume of Abstracts, 6th International Conference on Biomagnetism.* Tokyo.

Baddeley, A., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), *Recent advances in learning and motivation.* New York: Academic Press.

Barth, D., Sutherling, W., Engel, J., & Beatty, J. (1982). Neuromagnetic localization of epileptiform spike activity in the human brain. *Science, 218,* 891-894.

Barth, D., Sutherling, W., Engel, J., & Beatty, J. (1984). Neuromagnetic evidence of spatially distributed sources underlying epileptiform spikes in the human brain. *Science, 223,* 293-296.

Bauer, L., Goldstein, R., & Stern, J. (1987). Effects of information processing demands on physiological response patterns. *Human Factors, 29,* 213-234.

Bauer, L., Strock, B., Goldstein, R., Stern, J., & Walrath, L. (1985). Auditory discrimination and the eyeblink. *Psychophysiology, 22,* 636-641.

Bayley, N. (1928). A study of fear by means of the psychogalvanic technique. *Psychological Monographs, 38,* 1-38.

Beatty, J. (1988). Pupillometric signs of selective attention. In G. Galbraith, M. Kietzman, & E. Donchin (Eds.), *Neurophysiology and psychophysiology.* Hillsdale, NJ: Erlbaum.

Beatty, J. (1986). The pupillary system. In M. Coles, E. Donchin, & S. Porges (Eds.), *Psychophysiology: Systems, processes, and applications.* New York: Guilford.

Beatty, J. (1982a). Phasic not tonic pupillary responses vary with auditory vigilance performance. *Psychophysiology, 19,* 167-172.

Beatty J. (1982b). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91,* 276-292.

Beatty, J. (1977). Learned regulation of alpha and theta frequency activity in the human electroencephalogram. In G. Schwartz & J. Beatty (Eds.), *Biofeedback: Systems, processes, and research.* New York: Academic Press.

Beatty, J., Barth, D., Richer, F., & Johnson, R. (1986). Neuromagnetometry. In M. Coles, E. Donchin, & S. Porges (Eds), *Psychophysiology: Systems, processes, and applications.* New York: Guilford Press.

Beatty, J., & O'Hanlon, J. (1979). Operant control of posterior theta rhythms and vigilance performance: Repeated treatments and transfer of training. In N. Birbaumer & K. Kimmel (Eds.), *Biofeedback and self regulation.* Hillsdale, NJ: Erlbaum.

Berger, H. (1929). On the electroencephalogram of man. *Archives of Psychiatry and Nervous Diseases, 87,* 511-570.

Broeckl, L., Jones, K., Johnson, H., & Fischer, A. (1989). Reliability of common measures of respiratory sinus arrhythmia. *Psychophysiology, 26,* S16.

Casali, J., & Wierwille, W. (1983). A comparison of rating scale, secondary-task, physiological, and primary task workload estimation techniques in a simulated flight task emphasizing communications load. *Human Factors, 25,* 623-642.

Coles, M., Donchin, E., & Porges. (1986). *Psychophysiology: Systems, processes, and applications.* New York: Guilford Press.

Coles, M., Gratton, G., Kramer, A., & Miller (1986). Principles of signal acquisition and analysis. In M. Coles, E. Donchin, & S. Porges (Eds.), *Psychophysiology: Systems, processes, and applications.* New York: Guilford Press.

Coles, M., & Sirevaag, E. (1987). Heart rate and sinus arrhythmia. In A. Gale & B. Christie (Eds.), *Psychophysiology and the electronic workplace.* Chichester, England: Wiley.

Colle, H., Amel, J., Ewry, M., & Jenkins, M. (1988). Capacity equivalence curves: A double trdeoff curve method for equating task performance. *Human Factors, 30,* 645-656.

Conte, R., & Kinsbourne, M. (1988). Electrodermal lability predicts presentation rate effects and stimulant drug effects on paired associate learning in hyperactive children. *Psychophysiology, 25,* 64-70.

Crider, A. (1979). The electrodermal response: Biofeedback and individual difference studies. *International Review of Applied Psychology, 28,* 37-48.

Cuffin, B., & Cohen, D. (1979). Comparison of the magnetoencephalogram and electroencephalogram. *Electroencephalography and Clinical Neurophysiology, 47,* 132-146.

Dawson, M., & Schell, A. (1982). Electrodermal responses to attended and nonattended significant stimuli during dichotic listening. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 315-324.

Dawson, M., Schell, A., Beers, J., & Kelly, A. (1982). Allocation of cognitive processing capacity during human autonomic classical conditioning. *Journal of Experimental Psychology: General, 111,* 273-295.

Defayolle, M., Dinand, J., & Gentil, M. (1971). Averaged evoked potentials in relation to attitude, mental load, and intelligence. In W. T. Singleton, J. G. Fox, & D. Whitfield (Eds.), *Measurement of man at work*. London: Taylor and Francis.

Donchin, E., Karis, D., Bashore, T., Coles, M., & Gratton, G. (1986). Cognitive psychophysiology and human information processing. In M. Coles, E. Donchin, & S. Porges (Eds.), *Psychophysiology: Systems, processes, and applications*. New York: Guilford Press.

Donchin, E., Kramer, A., & Wickens, C. (1986). Applications of brain event related potentials to problems in engineering psychology. In M. Coles, E. Donchin, & S. Porges (Eds.), *Psychophysiology: Systems, processes, and applications*. New York: Guilford Press.

Donchin, E., Ritter, W., & McCallum, C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting, & S. Koslow (Eds.), *Brain event-related potentials in man*. New York: Academic Press.

Donders, F. C. (1869). Over de snelheid van psychische processen. Onderzoekingen gedaan in het Psiologish Labortorium der Utrechtsche Hoogeschool. Tweede Reeks, II, 92-120. Translated by W. G. Koster in W. G. Koester (Ed.), Attention and performance II. *Acta Psychologica, 30,* 412-431.

Egelund, N. (1982). Spectral analysis of heart rate variability as an indicator of driver fatigue. *Ergonomics, 25,* 663-672.

Englund, C., Reeves, D., Shingledecker, C., Thorne, D., Wilson, K., & Hegge, F. (1987). *Unified tri-service cognitive performance assessment battery (UTC-PAB)* (Report 87-10). San Diego: Naval Health Research Center.

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (1987). Definition, identification, and reliability of measurement of the P300 component of the event related brain potential. In P. Ackles (Eds.), *Advances in psychophysiology, Vol. 2*. New York: JAI Press.

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

Fere, C. (1986). Note of changes in electrical resistance under the effect of sensory stimulation and emotion. In M. Coles, E. Donchin, & S. Porges (Eds.), *Psychophysiology*. New York: Guilford Press.

Filion, D., Hazlett, E., Dawson, M., & Schell, A. (1989). Skin conductance orienting and resource allocation: Generality of a dissociation effect. *Psychophysiology, 4*, S23.

Fowles, D. C. (1986). The eccrine system and electrodermal activity. In M. Coles, E. Donchin, & S. Porges (Eds.). *Psychophysiology: Systems, processes, and applications*. New York: Guilford Press.

Freidman, A., & Polson, M. (1981). Hemispheres as independent resource systems: Limited capacity processing and cerebral specialization. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 1030-1058.

Furedy, J. (1987). Beyond heart rate in the cardiac psychophysiological assessment of mental effort: The T-wave amplitude component of the electrocardiogram. *Human Factors, 29*, 183-194.

Furedy, J., & Heslegrave, R. (1983). A consideration of recent criticisms of T-wave amplitude index of myocardial sympathetic activity. *Psychophysiology, 20*, 204-211.

Gale, A. (1987). The electroencephalogram. In A. Gale & B. Christie (Eds.), *Psychophysiology and the electronic workplace*. London: John Wiley & Sons.

Gale, A., & Edwards, J. (1983). The EEG and human behavior. In A. Gale & J. Edwards (Eds.), *Physiological correlates of human behavior.* New York: Academic Press.

Gevins, A. (1988). Applications of neural network (NN) signal processing in brain research. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 36*, 1152-1161.

Gevins, A., & Schaffer, R. (1980). A critical review of electroencephalographic (EEG) correlates of higher cortical functions. *CRT Critical Reviews in Bioengineering, 4*, 113-164.

Gevins, A., Zeitlin, G., Yingling, J., Doyle, J., Dedon, M., Schaffer, R., Roumasset, J., & Yeager, C. (1979). EEG patterns during cognitive tasks: I. Methodology, and analysis of complex behaviors. *Electroencephalography and Clinical Neurophysiology, 47*, 693-703.

Goldstein, R., Walrath, L., Stern, J., & Strock, B. (1985). Blink activity in a discrimination task as a function of stimulus modality and schedule of presentation. *Psychophysiology, 22*, 629-635.

Gopher, D., & Donchin, E. (1986). Workload--An examination of the concept. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and performance: Cognitive processes and performance*. New York: Wiley.

Graham, F. K. (1978a). Normality of distributions and homogeneity of variance of heart rate and heart rate period samples. *Psychophysiology, 15*, 487-491.

Graham, F. K. (1978b). Constraints on measuring heart rate and period sequentially through real and cardiac time. *Psychophysiology, 15*, 492-495.

Groll-Knapp, E. (1971). Evoked potentials and behavior. In W. T. Singleton, J. G. Fox, & D. Whitfield (Eds.). *Measurement of man at work*. London: Taylor and Francis.

Guilford, J. (1954). *Psychometric methods*. New York: McGraw Hill.

Gur, R., Gur, R., Skolnick, B., Resnick, S., Silver, F., Chawluk, J., Muenz, L., Obrist, W., & Revich, M. (1988). Effects of task difficulty on regional cerebral blood flow: Relationships with anxiety and performance. *Psychophysiology, 25*, 392-399.

Hall, R., & Cusack, B. (1972). *The measurement of eye behavior: Critical and selected reviews of voluntary movement and blinking* (Technical Memorandum 18-72). Aberdeen, MD: Aberdeen Proving Ground.

Hancock, P., Meshkati, N., & Robertson, M. (1985, November). Physiological reflections of mental workload. *Aviation, Space, and Environmental Medicine*, 1110-1114.

Hari, R., Hamalainen, M., Ilmoniemi, R., Kaukoranta, E., Reinikainen, K., Salmininen, J., Alho, K., Naatanen, R., & Sams, M. (1984). Responses of the primary auditory cortex to pitch changes in a sequence of tone pips: Neuromagnetic recordings in man. *Neuroscience Letters, 50*, 127-132.

Hari, R., Hamalainen, M., Kaukoranta, E., Makela, J., Joutsiniemi, S., & Tiihonen, J. (1989). Selective listening modifies activity of the human auditory cortex. *Experimental Brain Research, 74*, 463-470.

Harris, R., Bonadies, G., & Comstock, J. R. (1989). Usefulness of heart measures in flight simulation. *Proceedings of the Third Annual Workshop on Space Operations, Automation, and Robotics*. Houston, TX: Johnson Space Center.

Hart, S., Vidulich, M., & Tsang, P. (1986). *Collecting NASA workload ratings: A paper and pencil package* (NASA Technical Report). Moffett Field, CA: Ames Research Center.

Hastrup, J. (1979). Effects of electrodermal lability and introversion of vigilance performance. *Psychophysiology, 16*, 302-310.

Helander, M. (1975). Physiological reactions of drivers as indicators of road traffic demand. In *Driver Performance Studies* (Technical Report TRB/TRR-530). Washington, DC: Transportation Research Board.

Hess, E. H. (1965). Attitude and pupil size. *Scientific American, 212*, 46-54.

Hess, E. H. (1975). *The tell-tale eye*. New York: Van Nostrand Reinhold.

Hicks, T., & Wierwille, W. (1979). Comparison of five mental workload assessment procedures in a moving base driving simulator. *Human Factors, 21*, 129-144.

Hillyard, S., Munte, T., & Neville, H. (1985). Visual-spatial attention, orienting and brain physiology. In M. Posner & O. Marin (Eds.), *Attention and performance XI*. Hillsdale, NJ: Erlbaum.

Hitchen, M., Brodie, D., & Harness, J. (1980). Cardiac responses to demanding mental load. *Ergonomics, 23,* 379-385.

Hoffman, J., Houck, M., MacMillan, 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.

Horst, R., Munson, R., & Ruchkin, D. (1984). Event-related potential indices of workload in a single task paradigm. *Proceedings of the Human Factors Society, 28th Annual Meeting.* Santa Monica, CA: Human Factors Society.

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

Hugdahl, K., Fredikkson, M., & Ohman, A. (1977). "Preparedness" and "arousability" as determinants of electrodermal conditioning. *Behavior Research and Therapy, 15,* 345-353.

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

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

Janisse, M. P. (1977). *Pupillometry.* Washington, DC: Hemisphere Publishing.

Jenkins, J. G., Mitchel, R. H., & McClure, B. G. (1982). Heart rate variability in the newborn. In O. Rompelman & R. Kitney (Eds.), *International Workshop on the Analysis of Heart Rate Variability and Blood Pressure Fluctuations.* Delft, The Netherlands: University of Technology.

Jorna, P. (1985). Heart rate parameters and the coping process under water. In J. Orlebeke, G. Mulder, & L. van Doornen (Eds.). *The psychophysiology of cardiovascular control.* New York: Plenum Press.

Jung, C. G. (1907). On psychophysical relations of the associative experiment. *Journal of Abnormal Psychology, 7,* 247-255.

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

Kahneman, D., Tursky, B., Shapiro, D., & Crider, A. (1969). Pupillary, heart rate, and skin resistance changes during a mental task. *Journal of Experimental Psychology, 79,* 164-167.

Kalsbeek, J. (1971). Sinus arrhythmia and the dual task method in measuring mental load. In J. Fox & D. Whitfield (Eds.), *Measurement of man at work.* London: Taylor & Francis.

Kalsbeek, J., & Ettema, J. (1963). Continuous recording of heart rate and the measurement of perceptual load. *Ergonomics, 6,* 306-307.

Kamphuis, A., & Frowein, H. W. (1985). Assessment of mental effort by means of heart rate spectral analysis. In J. Orlebeke, G. Mulder, & L. van Doornen (Eds.), *The psychophysiology of cardiovascular control*. New York: Plenum Press.

Kantowitz, B., & Weldon, M. (1985). On scaling performance operating characteristics: Caveat emptor. *Human Factors, 27,* 531-548.

Kaukoranta, E., Sams, M., Hari, R., Hamalainen, M., & Naatanen, R. (in press). Reactions of human auditory cortex to a change in tone duration. *Hearing Research.*

Kazumi, M., Tetsuo, M., & Yo, M. (1984). Effects of instructions on the skin conductance response. *Japanese Psychological Research., 26,* 159-167.

Kenemans, J., Verbaten, M., Sjouw, W., & Slangen, J. (1988). Effects of task relevance on habituation of visual single trial ERPs and the skin conductance orienting response. *International Journal of Psychophysiology, 6,* 51-63.

Kinsbourne, M., & Hicks, R. (1978). Functional cerebral space. In J. Requin (Eds.), *Attention and performance VII*. Hillsdale, NJ: Erlbaum.

Knowles, W. (1963). Operator loading tasks. *Human Factors, 5,* 155-161.

Kramer, A. (1985). The interpretation of the component structure of event-related brain potentials. *Psychophysiology, 22,* 334-344.

Kramer, A. F. (1987). Event-related brain potentials. In A. Gale & B. Christie (Eds.), *Psychophysiology and the electronic workplace*. London: John Wiley & Sons.

Kramer, A., Humphrey, D., Sirevaag, E., & Mecklinger, A. (1989). Real-time measurement of mental workload: A feasibility study. *Proceedings of the Third Annual Workshop on Space Operations, Automation, and Robotics*. Houston, TX: Johnson Space Center.

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

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

Kramer, A. F., & Strayer, D. (1988). Assessing the development of automatic processing: An application of dual-task and event-related brain potential methodologies. *Biological Psychology, 26,* 231-268.

Kramer, A. F., Wickens, C. D., & Donchin, E. (1983). An analysis of the processing demands of a complex perceptual-motor task. *Human Factors, 25,* 597-621.

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

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science, 197,* 792-795.

Lacey, J. I. (1967). Somatic response patterning and stress: Some revisions of activation theory. In M. Appley & R. Trumbull (Eds.), *Psychological stress: Issues in research.* New York: Appleton Century Crofts.

Lacey, J. I., & Lacey, B. C. (1978). Two-way communication between the heart and the brain: Significance of time within the cardiac cycles. In E. Meyer & J. Brady (Eds.), *Research in the psychobiology of human behavior.* Baltimore, MD: Johns Hopkins University Press.

Larsen, P. B., Schneiderman, N., & Decarlo-Pasin, R. (1986). Physiological bases of cardiovascular psychophysiology. In M. Coles, E. Donchin, & A. Porges (Eds.), *Psychophysiology: Systems, processes, and applications.* New York: Guilford Press.

Leplat, J. (1978). Factors determining workload. *Ergonomics, 21,* 143-149.

Linde, E. (1928). Zur frage vom psychischen korrelate des psychogalvanischen reflexphanomens. *Proceedings of the Eighth International Congress of Psychology, 8,* 351-352.

Linden, R. J. (1985). Sympathetic and parasympathetic control of the heart. In J. Orlebeke, G. Mulder, & L. van Doornen (Eds.), *Psychophysiology of cardiovascular control: Models, methods, and data.* New York: Plenum Press.

Lindholm, E., & Cheatham, C. (1983). Autonomic activity and workload during learning of a simulated aircraft carrier landing task. *Aviation, Space, and Environmental Medicine, 54,* 435-439.

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

Lounasmaa, O., Hari, R., Joutsiniemi, S., & Hamalainen, M. (in press). Multi SQUID recordings of human cerebral magnetic fields may give information about memory processes in the human brain. *Europhysics Letters.*

Lykken, D., Rose, R., Luther, B., & Maley, M. (1966). Correcting psychophysiological measures for individual differences in range. *Psychological Bulletin, 66,* 481-484.

Makela, J., Hari, R., & Leinonen, L. (1988). Magnetic response to the human auditory cortex to noise/square wave transitions. *Electronecephalography and Clinical Neurophysiology, 69,* 423-430.

McCallum, C., Cooper, R., & Pocock, P. (1987). Event-related and steady potential changes in the brain related to workload during tracking. In K. Jessen (Ed.), *Electric and magnetic activity of the central nervous system: Research and clinical applications in aerospace medicine.* France: NATO Advisory Group for Aerospace Research and Development.

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science, 211,* 77-80.

Meijman, T. (1985). *Onderzoek taakbelasting rij-examinatoren CBR.* (On the workload among driving examiners) (Technical Report 87101). Groningen, The Netherlands: University of Groningen.

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

Moray, N. (1989). Mental workload since 1979. In D. Oborne (Ed.), *International Reviews of Ergonomics, 2,* 123-150.

Mountford, J., & North, R. (1980). Voice entry for reducing pilot workload. *Proceedings of the 24th Annual Meeting of the Human Factors Society,* Santa Monica, CA: Human Factors Society.

Mulder, G. (1979). Mental load, mental effort and attention. In N. Moray (Ed.), *Mental workload; Its theory and measurement.* New York: Plenum Press.

Mulder, G., & Mulder, L. J. M. (1980). Coping with mental workload. In S. Levine & H. Ursine (Eds.), *Coping and health.* New York: Plenum Press.

Mulder, G., & Mulder, L. J. M. (1981a). Information processing and cardiovascular control. *Psychophysiology, 18,* 392-405.

Mulder, G., & Mulder, L. J. M. (1981b). Task related cardiovascular stress. In J. Lo[illegible]nd A. Baddeley (Eds.), *Attention and performance IX.* Hillsdale, NJ: Erlbaum.

Mulder, H., Meijman, T., O'Hanlon, J., & Mulder, G. (1982). Differential psychophysiological reactivity of city bus-drivers. *Ergonomics, 25,* 1003-1011.

Naatanen, R. (1988). Implications of ERP data for psychological theories of attention. *Biological Psychology, 26,* 117-163.

Natani, K., & Gomer, F. (1981). *Electrocortical activity and operator workload: A comparison of changes in the electroencephalogram and in event-related potentials* (Technical Report MDC E2427). St. Louis, MO: McDonnell Douglas Corporation.

Navon, D., & Gopher, D. (1979). On the economy of the human processing system. *Psychological Review, 86,* 214-255.

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

Nunez, P. (1981). *Electrical fields of the brain: The neurophysics of EEG.* New York: Oxford University Press.

Obrist, P. (1976). The cardiovascular-behavioral interaction as it appears today. *Psychophysiology, 13,* 95-107.

Obrist, P. (1984). The orienting response, attention and learning: An information processing perspective. In D. Kimmel, E. van Olst, & J. Orlebeke (Eds.), *Handbook of perception and human performance*. New York: Wiley.

Ogden, G., Levine, J., & Eisner, E. (1979). Measurement of workload by secondary tasks. *Human Factors, 21,* 529-548.

O'Gorman, J., & Lloyd, J. (1988). Electrodermal lability and dichotic listening. *Psychophysiology, 25,* 538-546.

Opmeer, C. (1973). The information content of successive R-R interval times in the ECG. Preliminary results in factor analysis and frequency analysis. *Ergonomics, 16,* 105-115.

Oster, P., & Stern, J. (1980). Measurement of eye movement. In I. Martin, & P. Venables (Eds.). *Techniques in psychophysiology.* New York: Wiley.

Packer, J., & Siddle, D. (1989). Stimulus miscuing, electrodermal activity, and the allocation of processing resources. *Psychophysiology, 26,* 192-200.

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

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

Peterson, F., & Jung, C. G. (1907). Psychophysical investigations with the galvanometer and plethysmograph in normal and insane individuals. *Brain, 30,* 153-218.

Phelps, M., & Mazziotta, J. (1985). Positron emission tomography: Human brain function and biochemistry. *Science, 228,* 799-809.

Pigeau, R., Hoffman, R., Purcell, S., & Moffitt, A. (1987). The effect of endogenous alpha on hemispheric asymmetries and the relationship of frontal theta to sustained attention. In K. Jessen (Ed.), *Electric and magnetic activity of the central nervous system: Research and clinical applications in aerospace medicine*. France: NATO Advisory Group for Aerospace Research and Development.

Polson, M., & Freidman, A. (1988). Task sharing within and between hemispheres: A multiple resource approach. *Human Factors, 30,* 633-643.

Ponder, E., & Kennedy, W. (1927). On the act of blinking. *Quarterly Journal of Experimental Psychology, 18,* 89-110.

Porges, S. (1984). Heart rate oscillation: An index of neural mediation. In M. Coles, J. Jennings, & J. Stern (Eds.), *Psychophysiological perspectives: Festschrift for Beatrice and John Lacey.* New York: Nostrand Reinhold.

Posner, M., Petersen, S., Fox, P., & Raichle, M. (1988). Localization of cognitive operations in the human brain. *Science, 240,* 1627-1631.

Qiyuan, J., Richer, F., Wagoner, B., & Beatty, J. (1985). The pupil and stimulus probability. *Psychophysiology, 22,* 530-534.

Ragot, R. (1984). Perceptual and motor space representation: An event related potential study. *Psychophysiology, 21,* 159-170.

Reid, G. (1985). Current status of the development of the subjective workload assessment technique. *Proceedings of the 29th Meeting of the Human Factors Society.* Santa Monica, CA: Human Factors Society.

Richer, F., & Beatty, J. (1987). Contrasting effects of response uncertainty on the task evoked pupillary response and reaction time. *Psychophysiology, 24,* 258-262.

Richer, F., & Beatty, J. (1985). Pupillary dilations in movement preparation and execution. *Psychophysiology, 22,* 204-207.

Richer, F., Silverman, C., & Beatty, J. (1983). Response selection and initiation in speeded reactions: A pupillometric analysis. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 360-370.

Risberg, J., & Prohovnik, I. (1983). Cortical processing of visual and tactile stimuli studied by noninvasive rCBF measurements. *Human Neurobiology, 2,* 5-10.

Ritter, W., Simpson, R., & Vaughan, H. (1983). Event related brain potentials and two sequential stages of information processing in physical and semantic discrimination. *Psychophysiology, 20,* 168-179.

Romani, G. (1987). SQUID instrumentation for neuromagnetic study of complex brain activity. In K. Jessen (Ed.), *Electric and magnetic activity of the central nervous system: Research and clinical applications in aerospace medicine.* France: NATO Advisory Group for Aerospace Research and Development.

Roscoe, A. H. (1984). Assessing pilot workload in flight. *Flight test techniques, AGARD Proceedings* (373). France: NATO Advisory Group for Aerospace Research and Development.

Salvendy, G., & Humphreys, A. (1979). Effects of personality, perceptual difficulty and pacing of a task on productivity, job satisfaction, and physiological stress. *Perceptual and Motor Skills, 49,* 219-222.

Sanders, A. (1979). Some remarks on mental load. In N. Moray (Ed.), *Mental workload: Its theory and measurement.* New York: Plenum Press.

Sanders, A. (1981). Stress and human performance: A working model and some applications. In G. Salvendy & E. Smith (Eds.), *Machine pacing and occupational stress.* London: Taylor and Francis.

Sem-Jacobsen, C. W. (1981, January). Brain/computer communication to reduce human error: A perspective. *Aviation, Space, and Environmental Medicine,* 31-38.

Senders, J. (1980). *Visual scanning processes*. The Netherlands: Tilburg University Press.

Siddle, D., & Packer, J. (1987). Stimulus omission and dishabituation of the electrodermal orienting response: The allocation of processing resources. *Psychophysiology, 24,* 181-190.

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

Sirevaag, E., Kramer, A., de Jong, R., & Mecklinger, A. (1988). A psychophysiological analysis of multi-task processing demands. *Psychophysiology, 25,* 482.

Skelly, J., Purvis, B., & Wilson, G. (1987). Fighter pilot performance during airborne and simulator missions: Physiological comparisons. In K. Jessen (Ed.), *Electric and magnetic activity of the central nervous system: Research and clinical applications in aerospace medicine*. France: NATO Advisory Group for Aerospace Research and Development.

Sokoloff, L. (1981). The relationship between function and energy metabolism: Its use in the localization of functional activity in the nervous system. *Neurosciences Research Program Bulletin, 19,* 159-210.

Sostek, A. (1978). Effects of electrodermal lability and payoff instructions on vigilance performance. *Psychophysiology, 15,* 561-568.

Speyer, J., Fort, A., Fouillot, J., & Blomberg, R. (1987). Assessing workload for minimum crew certification. In A. Roscoe (Ed.), *The practical assessment of pilot workload.* (AGARDograph 282). Washington, DC: Advisory Group for Aerospace Research and Development.

Spinks, J., Blowers, G., & Shek, D. (1985). The role of the orienting response in the anticipation of information: A skin conductance response study. *Psychophysiology, 22,* 385-394.

Starr, A. (1978). Sensory evoked potentials in clinical disorders of the nervous system. *Annual Review of Neuroscience*. Palo Alto, CA: Annual Reviews, Inc.

Sterman, B., Schummer, G., Dushenko, T., & Smith, J. (1987). Electroencephalographic correlates of pilot performance: Simulation and in-flight studies. In K. Jessen (Ed.), *Electrical and magnetic activity of the central nervous system: Research and clinical applications in aerospace medicine*. France: NATO Advisory Group for Aerospace Research and Development.

Stern, J., & Skelly, J. (1984). The eyeblink and workload considerations. *Proceedings of the Human Factors Society, 28th Annual Meeting*. San Antonio, TX: Human Factors Society.

Stern, J., Walrath, L., & Goldstein, R. (1984). The endogenous eyeblink. *Psychophysiology, 21,* 23-33.

Sternberg, S. (1969). On the discovery of processing stages: Some extensions of Donders' method. *Acta Psychologica, 30,* 276-315.

Stockard, J., Stockard, J. E., & Sharbrough, F. (1979). Brainstem auditory evoked potentials in neurology: Methodology, interpretation and clinical application. In M. Aminoff (Ed.), *Electrophysiological approaches to neurological diagnosis.* New York: Churchill Livingstone.

Straube, E., Schlenker, R., Klessinger, T., Himer, W., & Boven, K. (1987). Electrodermal orienting response and selective attention. *Psychophysiology, 24,* 615.

Strayer, D., & Kramer, A. F. (in press). Attentional requirements of automatic and controlled processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Ter-Pogossian, M., Raichle, M., & Soble, B. (1980). Positron emission tomography. *Scientific American, 243,* 171-181.

Tryon, W. W. (1975). Pupillometry: A survey of sources of variation. *Psychophysiology, 12,* 90-93.

Tursky, B. T. (1974). Recording of human eye movement. In R. F. Thompson & M. M. Patterson (Eds.), *Bioelectric recording techniques.* New York: Academic Press.

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

Van Dellen, H., Aasman, J., Mulder, L., & Mulder, G. (1985). Time domain versus frequency domain measures of heart rate variability. In J. Orlebeke, G., Mulder, & L. van Doornen (Eds.), *Psychophysiology of cardiovascular control: Models, methods, and data.* New York: Plenum Press.

Verbaten, M., & Kenemans, J. (1987). The effect of task relevance on cortical and electrodermal aspects of the orienting response. *Psychophysiology, 24,* 619-620.

Vicente, K., Thorton, D., & Moray, N. (1987). Spectral analysis of sinus arrhythmia: A measure of mental effort. *Human Factors, 29,* 171-182.

Vossel, G., & Rossman, R. (1984). Electrodermal habituation speed and visual monitoring performance. *Psychophysiology, 21,* 97-100.

Waid, W., & Orne, M. (1982). The physiological detection of deception. *American Scientist, 70,* 402-409.

Waller, A. (1918). Galvanometric observation of the emotivity of a normal subject during the German air raid of Whit-Sunday. *Lancet, 194,* 916.

Wickens, C. D. (1979). Measures of workload, stress, and secondary tasks. In N. Moray (Ed.), *Mental workload: Its theory and measurement.* New York: Plenum Press.

Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson & R. Pew (Eds.), *Attention and performance VIII*. Hillsdale, NJ: Erlbaum.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. Davies (Eds.), *Varieties of Attention*. Orlando: Academic Press.

Wickens, C. D., & Kramer, A. F. (1985). Engineering psychology. *Annual Review of Psychology*. New York: Annual Reviews, Inc.

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

Wickens, C., Mountford, J., & Schreiner, W. (1981). Multiple resources, task hemispheric integrity, and individual differences in time sharing. *Human factors, 23,* 211-229.

Wickens, C., & Yeh, Y. (1985). POCs and performance decrements: A reply to Kantowitz and Weldon. *Human Factors, 27,* 549-554.

Wierwille, W. (1979). Physiological measures of aircrew mental workload. *Human Factors, 21,* 575-594.

Wierwille, W. (1988). Important remaining issues in mental workload estimation. In P. Hancock & N. Meshkati (Eds.), *Human mental workload*. Amsterdam, The Netherlands: Elsevier.

Wierwille, W., & Conner, S. (1983). Evaluation of 20 workload measures using a psychomotor task in a moving base aircraft simulator. *Human Factors, 25,* 1-16.

Wierwille, W., Rahimi, M., & Casali, J. (1985). Evaluation of 16 measures of mental workload using a simulated flight task emphasizing mediational activity. *Human Factors, 27,* 489-502.

Williamson, S., & Kaufman, L. (1981). Biomagnetism. *Journal of Magnetism and Magnetic Materials, 22,* 129-201.

Wilson, K. (1987). Electrodermal lability and simple reaction time. *Biological Psychology, 21,* 97-100.

Wilson, K., & Graham, S. (1989). Electrodermal lability and visual information processing. *Psychophysiology, 26,* 321-328.

Wilson, G., & O'Donnell, R. (1986). Steady state evoked responses: Correlations with human cognition. *Psychophysiology, 23,* 57-61.

Wilson, G., & O'Donnell, R. (1988). Measurement of operator workload with the neuropsychological workload test battery. In P. Hancock & N. Meshkati (Eds.), *Human mental workload*. Amsterdam, The Netherlands: Elsevier.

Wilson, G., Purvis, B., Skelly, J., Fullenkamp, P., & Davis, I. (1987). Physiological data used to measure pilot workload in actual flight and simulator conditions. *Proceedings of the Human Factors Society, 31st Annual Meeting*. New York: Human Factors Society.

Yeh, Y., & Wickens, C. (1988). Dissociation of performance and subjective measures of workload. *Human Factors, 30,* 111-120.

Young, L., & Sheena, D. (1975). Survey of eye movement recording methods. *Behavior Research Methods and Instrumentation, 7,* 397-429.

DISTRIBUTION LIST

Office of Naval Technology (Code 222) (3)
Director, Office of Naval Research (OCNR-20)
Naval Aerospace Medical Research Laboratory (Code 031)
Naval Health Research Center (Code 60)
Commander, Air Force Human Resources Laboratory, Brooks Air Force Base, TX
Program Manager, Manpower Research and Advisory Services, Smithsonian Institute
Defense Technical Information Center (DTIC) (2)