A Psychophysiological Assessment of the Efficacy of Event-Related Potentials and Electroencephalogram for Adaptive Task Allocation

Lawrence J. Prinzel III
Old Dominion University

A PSYCHOPHYSIOLOGICAL ASSESSMENT OF THE EFFICACY

OF EVENT-RELATED POTENTIALS AND ELECTROENCEPHALOGRAM

FOR ADAPTIVE TASK ALLOCATION

by

Lawrence J. Prinzel III

M.S. August 1995, Old Dominion University
B.S. May 1993, Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

INDUSTRIAL-ORGANIZATIONAL PSYCHOLOGY

OLD DOMINION UNIVERSITY
August 1998

Approved by:

Frederick G. Freeman (Director)

Mark W. Scerbo (Member)

Peter J. Mikulka (Member)

Alan T. Pope (Member)

ABSTRACT

A PSYCHOPHYSIOLOGICAL ASSESSMENT OF THE EFFICACY OF EVENT-RELATED POTENTIALS AND ELECTROENCEPHALOGRAM FOR ADAPTIVE TASK ALLOCATION.

Lawrence J. Prinzel III
Old Dominion University, 1998
Director: Dr. Frederick G. Freeman

The present study was designed to test the efficacy of using Electroencephalogram (EEG) and Event-Related Potentials (ERPs) for making task allocations decisions. Thirty-six participants were randomly assigned to an experimental, yoked, or control group condition. Under the experimental condition, a compensatory tracking task was switched between manual and automatic task modes based upon the participant's EEG. ERPs were also gathered to an auditory, oddball task. Participants in the yoked condition performed the same tasks under the exact sequence of task allocations that participants in the experimental group experienced. The control condition consisted of a random sequence of task allocations that was representative of each participant in the experimental group condition. Therefore, the design allowed a test of whether the performance and workload benefits seen in previous studies using this biocybernetic system were due to adaptive aiding or merely to the increase in task mode allocations.

The results showed that the use of adaptive aiding improved performance and lowered subjective workload under

negative feedback as predicted. Additionally, participants in the adaptive group had significantly lower tracking errors scores and NASA-TLX ratings than participants in either the yoked or control group conditions. Furthermore, the amplitudes of the N1 and P3 ERP components were significantly larger under the experimental group condition than under either the yoked or control group conditions. These results are discussed in terms of their implications for adaptive automation design.

ACKNOWLEDGMENTS

Dr. Fred Freeman deserves a considerable amount of thanks for more than just helping me to complete this dissertation. For the past eight years, he has been so supportive and helpful that I don't think I would have completed my Ph.D. without him. Thank you very much.

Thanks also to my committee members: Dr. Mark Scerbo, Dr. Peter Mikulka, and Dr. Alan Pope. Each of you had a significant impact on my education, and I greatly appreciate the considerable amount of time that you invested in helping me reach my educational goals.

The other two people that deserve acknowledgment are Elizabeth Shirer and Kristen Koontz. I have known Elizabeth before I began college and, without her encouragement, I probably would be a 28 year old lifeguard right now. She has always been there for me, and I can't say enough to express how much I love and appreciate her.

Kristen, we have been many things to each other, but always have been the greatest of friends. Thanks for being there for me. I will always love you.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Automation refers to "...systems or methods in which many of the processes of production are automatically performed or controlled by autonomous machines or electronic devices (p.7)." Automation is a tool, or resource, that the human operator can use to perform some task that would be difficult or impossible without the help of machines (Billings, 1997). Therefore, automation can be thought of as a process of substituting some device or machine for some human activity; or it can be thought of as a state of technological development (Parsons, 1985). However, some people (e.g., Woods, 1996) have questioned whether automation should be viewed as a substitution of one agent for another. Nevertheless, the presence of automation has pervaded every aspect of modern life. We have built machines and systems that not only make work easier, more efficient and safer, but also have given us more leisure time. The advent of automation has further enabled us to achieve these ends. With automation, machines can now perform many of the activities that we once had to do. Now, automatic doors open for us. Thermostats regulate the temperature in our homes for us. Automobile transmissions

The *Publication Manual of the American Psychological Association* (4th ed.) was used in the preparation of this manuscript.

shift gears for us. We just have to turn the automation on and off. One day, however, there may not be a need for us to do even that.

Impact of Automation Technology

Advantages of Automation. Wiener (1980; 1989) noted a number of advantages to automating human-machine systems. These include increased capacity and productivity, reduction of small errors, reduction of manual workload and fatigue, relief from routine operations, more precise handling of routine operations, and economical use of machines. In an aviation context, for example, Wiener and Curry (1980) listed eight reasons for the increase in flight-deck automation: Increase in available technology, such as the Flight Management System (FMS), Ground Proximity Warning System (GPWS), Traffic Alert and Collision Avoidance System (TCAS); concern for safety; economy, maintenance, and reliability; decrease in workload for two-pilot transport aircraft certification; flight maneuvers and navigation precision; display flexibility; economy of cockpit space; and special requirements for military missions.

Disadvantages of Automation. Automation also has a number of disadvantages. Automation increases the burdens and complexities for those responsible for operating, troubleshooting, and managing systems. Woods (1996) stated that automation is "...a wrapped package -- a package that

consists of many different dimensions bundled together as a hardware/software system. When new automated systems are introduced into a field of practice, change is precipitated along multiple dimensions (p.4)." Some of these changes include: (a) adding to or changing the task, such as device setup and initialization, configuration control, and operating sequences; (b) changing cognitive demands, such as decreased situational awareness; (c) changing the role that people in the system have, often relegating people to supervisory controllers; (d) increasing coupling and integration among parts of a system often resulting in data overload and "transparency"; and (e) increasing complacency by those who use the technology. These changes can result in lower job satisfaction (automation seen as dehumanizing), lowered vigilance, fault-intolerant systems, silent failures, an increase in cognitive workload, automation-induced failures, over-reliance, increased boredom, decreased trust, manual skill erosion, false alarms, and a decrease in mode awareness (Wiener, 1989).

Adaptive Automation

The disadvantages of automation have resulted in increased interest in advanced automation concepts. One of these concepts is automation that is dynamic or adaptive in nature (Hancock & Chignell, 1987; Morrison, Gluckman, & Deaton, 1991; Rouse, 1977; 1988). In adaptive automation,

control of tasks can be passed back and forth between the operator and automated systems in response to the changing task demands. Consequently, this allows for the restructuring of the task environment based upon (a) what is automated, (b) when it should be automated, and (c) how it should be automated (Rouse, 1988; Scerbo, 1996). Rouse (1988) described the criteria for adaptive aiding systems:

> The level of aiding, as well as the ways in which human and aid interact, should change as task demands vary. More specifically, the level of aiding should increase as task demands become such that human performance will unacceptably degrade without aiding. Further, the ways in which human and aid interact should become increasingly streamlined as task demands increase. Finally, it is quite likely that variations in level of aiding and modes of interaction will have to be initiated by the aid rather than by the human whose excess task demands have created a situation requiring aiding. The term *adaptive aiding* is used to denote aiding concepts that meet [these] requirements (p.432).

Adaptive aiding attempts to optimize the allocation of tasks by creating a mechanism for determining when tasks need to be automated (Morrison & Gluckman, 1994). In adaptive

automation, the level or mode of automation can be modified in real-time. Further, unlike traditional forms of automation, both the system and the operator share control over changes in the state of automation (Scerbo, 1994; 1996). Parasuraman, Bahri, Deaton, Morrison, and Barnes (1992) have argued that adaptive automation represents the optimal coupling of the level of operator workload to the level of automation in the tasks. Thus, adaptive automation invokes automation only when task demands exceed the operator capabilities to perform the task(s) successfully. Otherwise, the operator retains manual control of the system functions. Although concerns have been raised about the dangers of adaptive automation (Billings & Woods, 1994; Wiener, 1989), it promises to regulate workload, bolster situational awareness, enhance vigilance, maintain manual skill levels, increase task involvement, and generally improve operator performance (Endsley, 1996; Parasuraman et al., 1992; Parasuraman, Mouloua, & Molloy, 1996; Scerbo, 1994, 1996; Singh, Molloy, & Parasuraman, 1993).

Perhaps, the most critical challenge facing system designers seeking to implement adaptive automation concerns how changes among modes or levels of automation will be accomplished (Parasuraman et al., 1992; Scerbo, 1996). The best approach involves the assessment of measures that index the operators' state of mental engagement (Parasuraman et

al., 1992; Rouse, 1988). The question, however, is what should be the "trigger" for the allocation of functions between the operator and the automation system. Numerous researchers have suggested that adaptive systems respond to variations in operator workload (Hancock & Chignell, 1987; 1988; Hancock, Chignell & Lowenthal, 1985; Humphrey & Kramer, 1994; Reising, 1985; Riley, 1985; Rouse, 1977), and that measures of workload be used to initiate changes in automation modes. Such measures include primary and secondary-task measures, subjective workload measures, and physiological measures. This, of course, presupposes that levels of operator workload can be specified so as to make changes in automation modes (Scerbo, 1996). Rouse (1977), for example, proposed a system for dynamic allocation of tasks based upon the operator's momentary workload level. Reising (1985) described a future cockpit in which pilot workload states are continuously monitored, and functions are automatically reallocated back to the aircraft if workload levels get too high or too low. However, neither of these researchers provided specific parameters in which to make allocation changes (Parasuraman, 1990).

Morrison and Gluckman (1994), however, did suggest a number of workload indices candidates that may be used for initiating changes among levels of automation. They suggested that adaptive automation can be invoked through a

combination of one or more real-time technological approaches. One of these proposed adaptive mechanisms is biopsychometrics. Under this method, physiological signals that reflect central nervous system activity, and perhaps changes in workload, would serve as a trigger for shifting among modes or levels of automation (Hancock, Chignell, & Lowenthal, 1985; Morrison & Gluckman, 1994; Scerbo, 1996).

Byrne and Parasuraman (1996) discussed the theoretical framework for developing adaptive automation around psychophysiological measures. The use of physiological measures in adaptive systems is based on the idea that there exists an optimal state of engagement (Gaillard, 1993; Hockey, Coles, & Gaillard, 1986). Capacity and resource theories (Kahneman, 1973; Wickens, 1984; 1992) are central to this idea. These theories posit that there exists a limited amount of resources to draw upon when performing tasks. These resources are not directly observable, but instead are hypothetical constructs. Kahneman (1973) conceptualized resources as being limited, and the limitation is a function of the level of arousal. Changes in arousal and the concomitant changes in resource capacity are thought to be controlled by feedback from other ongoing activities. An increase in the activities (i.e., task load) causes a rise in arousal and a subsequent decrease in capacity. Kahneman's model was derived from research

(Kahneman et al., 1967, 1968, 1969) on pupil diameter and task difficulty. Therefore, physiological measures have been posited to index the utilization of cognitive resources.

Several biopsychometrics have been shown to be sensitive to changes in operator workload suggesting them as potential candidates for adaptive automation. These include heart rate variability (Backs, Ryan, & Wilson, 1994; Itoh, Hayashi, Tsukui, & Saito, 1989; Lindholm & Cheatham, 1983; Lindqvist et al., 1983; Opmeer & Krol, 1973; Sayers, 1973; Sekiguchi et al., 1978), EEG (Natani & Gomer, 1981; O'Hanlon & Beatty, 1977; Sterman, Schummer, Dushenko, & Smith, 1987; Torsvall & Akerstedt, 1987), eyeblinks (Goldstein, Walrath, Stern, & Strock, 1985; Sirevaag, Kramer, deJong, & Mecklinger, 1988), pupil diameter (Beatty, 1982; 1986; 1988; Qiyuan, Richer, Wagoner, & Beatty, 1985; Richer & Beatty, 1985; 1987; Richer, Silverman, & Beatty, 1983), electrodermal activity (Straube et al., 1987; Vossel & Rossmann, 1984; Wilson, 1987; Wilson & Graham, 1989) and event-related potentials (Defayolle, Dinand, & Gentil, 1971; Gomer, 1981; Hancock, Chignell, & Lowenthal, 1985; Reising, 1985; Rouse, 1977; Sem-Jacobson, 1981).

The advantage to biopsychometrics in adaptive systems is that the measures can be obtained continuously with little intrusion (Eggemeier, 1988; Kramer, 1991; Wilson &

Eggemeier, 1991). Also, because behavior is often at a low level when humans interact with automated systems, it is difficult to measure resource capacity with performance indices. Furthermore, these measures have been found to be diagnostic of multiple levels of arousal, attention, and workload. Therefore, it seems reasonable to determine the efficacy of using psychophysiological measures to allocate functions in an adaptive automated system. However, although many proposals concerning the use of psychophysiological measures in adaptive systems have been advanced, not much research has actually been reported (Byrne & Parasuraman, 1996). Nonetheless, many researchers have suggested that perhaps the two most promising psychophysiological indices for adaptive automation are the electroencephalogram (EEG) and event-related potential (ERP) (Byrne & Parasurman, 1996; Kramer, Trejo, & Humphrey, 1996; Morrison & Gluckman, 1994; Parasuraman, 1990; Scerbo, 1996).

Mental Workload

The use of psychophysiological measures in adaptive automation requires that such measures are capable of representing mental workload. Mental workload has been defined as the amount of processing capacity that is expended during task performance (Eggemeier, 1988). The basic concept refers to the difference between the processing resources available to the operator and the

resource demands required by the task (Sanders & McCormick, 1993). Essentially, workload is invoked to describe the interaction between an operator performing the task and the task itself. In other words, the term "workload" delineates the difference between capacities of the human information processing system that are expected to satisfy performance expectations and that capacity available for actual performance (Gopher & Donchin, 1986). However, there is disagreement on the definition of the term, on the best means for measuring it, and on the most effective ways for moderating workload. Some psychologists have defined it in terms of the perceptual and cognitive demands imposed on the operator, whereas engineers tend to prefer a definition based on the scheduling of tasks in multi-task environments or on control theory models (Parasuraman, 1990). An emerging consensus is that workload is a multidimensional construct, rather than a scalar quantity, that cannot be uniquely specified by any one measurement technique (Howell, 1990). Despite this, research has shown that both the EEG and ERP are useful as a metric of mental workload (Byrne & Parasurman, 1996; Gale & Christie, 1987; Kramer, 1991; Parasuraman, 1990)

Electroencephalogram

Physiological Basis. The EEG derives from activity in neural tissue located in the cerebral cortex, but the

precise origin of the EEG, what it represents, and the functions that it serves are not presently known. Current theory suggests that the EEG originates from post synaptic potentials rather than action potentials. Thus, the EEG is postulated to result primarily from the subthreshold post-synaptic potentials that may summate and reflect stimulus intensity instead of firing in an all-or-none fashion (Gale & Edwards, 1983).

Description of the EEG. The EEG consists of a spectrum of frequencies between 0.5 Hz to 35 Hz (Surwillo, 1990). Delta waves are large amplitude, low frequency waveforms that typically range between 0.5 and 3.5 Hz in frequency, in the range of 20 to 200 *u*V (Andreassi, 1995). Theta waves are a relatively uncommon type of brain rhythm that occurs between 4 and 7 Hz at an amplitude ranging from 20 to 100 *u*V. Alpha waves occur between 8 and 13 Hz at a magnitude of 20 to 60 *u*V. Finally, beta waves are an irregular waveform at a frequency of 14 to 30 Hz at an amplitude of about 2 to 20 *u*V (Andreassi, 1995). An alert person performing a very demanding task tends to exhibit predominately low amplitude, high Hz waveforms (beta activity). An awake, but less alert person shows a higher amplitude, slower frequency of activity (alpha activity). With drowsiness, theta waves predominate and in the early cycles of deep slow wave sleep, delta waves are evident in the EEG waveform. The

generalized effect of stress, activation or attention is a shift towards the faster frequencies, lower amplitudes with an abrupt blocking of alpha activity (Horst, 1987).

Laboratory Studies. Gale (1987) found that there exists an inverse relationship between alpha power and task difficulty. Other studies have also demonstrated the sensitivity of alpha waves to variations in workload associated with task performance. Natani and Gomer (1981) found decreased alpha and theta power when high workload conditions were introduced to pilots during pitch and roll disturbances in flight. Sterman, Schummer, Dushenko, and Smith (1987) conducted a series of aircraft and flight simulation experiments in which they also demonstrated decreased alpha power and tracking performance in flight with increasing task difficulty.

Numerous studies have also demonstrated that theta may be sensitive to increases in mental workload. Subjects have been trained to produce EEG theta patterns to regulate degrees of attention (Beatty, Greenberg, Diebler, & O'Hanlon, 1974; Beatty & O'Hanlon, 1979; O'Hanlon & Beatty, 1979; O'Hanlon, Royal, & Beatty, 1977). In particular, Beatty and O'Hanlon (1979) found that both college students and trained radar operators, who had been taught to suppress theta activity performed better than controls on a vigilance task. Though theta regulation has been shown to affect

attention, the magnitude of the effect is often small (Alluisi, Coates, & Morgan, 1977). More recent research, however, has demonstrated its utility in assessing mental workload. Both Natani and Gomer (1981) and Sirevaag, Kramer, deJong, and Mecklinger (1988) found decreases in theta activity as task difficulty increased and during transitions from single to multiple tasks, respectively.

Field Research. More recent research has demonstrated the utility of EEG in assessing mental workload in the operational environment. Sterman et al. (1993) evaluated EEG data obtained from 15 Air Force pilots during air refueling and landing exercises performed in an advanced technology aircraft simulator. They found a progressive suppression of 8-12 Hz activity (alpha waves) at medial (Pz) and right parietal (P4) sites with increasing amounts of workload. Additionally, a significant decrease in the total EEG power (progressive engagement) was found at P4 during the aircraft turning condition for the air refueling task (the most difficult flight maneuver). This confirmed other research that found alpha rhythm suppression as a function of increased mental workload (e.g., Ray & Cole, 1985).

Event-Related Potential

Description. The event-related potential, or ERP, is a transient series of voltage oscillations that occurs in response to the occurrence of a discrete event. This

temporal relationship between the ERP and an event is what discriminates the ERP from the ongoing electroencephalogram (EEG) activity. The ERP, like EEG, is a multivariate measure; however, unlike EEG, the ERP is broken down into a series of time rather than frequency domains (Kramer, 1991).

ERPs can be seen as a sequence of separate but often temporally overlapping components that are affected by a combination of the physical parameters of the stimuli and psychological constructs such as motivation, expectancy, resources, task relevance, memory, and attention (Kramer, 1987). Although the ERP has been found to be dependent upon both the psychological and physical characteristics of the eliciting stimuli, in some instances the ERP has been found to be independent of specific stimuli (Andreassi, 1995). For example, ERPs have been found to occur at the same time that the stimuli were *expected* to occur but were not actually presented (Sutton, Teuting, Zubin, & John, 1967).

Classification. The ERP can be classified as either being an evoked potential or an emitted potential. The "evoked potentials" (EPs) are ERPs that occur in response to physical stimulus presentation whereas "emitted potentials" occur in the absence of any invoking stimulus. Emitted potentials may be associated with a psychological process, such as recognition that a stimulus component is missing from a regular train of stimulus presentations or with some

preparation for an upcoming perceptual or motor act (Picton, 1988).

ERP components can also be categorized along a continuum from endogenous to exogenous. The endogenous components are influenced by the processing demands imposed by the task, and are not very sensitive to changes in the physical parameters of stimuli, especially when these changes are not relevant to the task. In fact, endogenous components can be elicited by the absence of an eliciting stimulus if this "event" is relevant to the subject's task. Subject's strategies, expectancies, intentions, and decisions, in addition to task parameters and instructions, account for most of the endogenous components (Kramer, 1991).

The exogenous components, on the other hand, represent a response to the presentation of some discrete event. These components tend to occur somewhat earlier than endogenous components and they are usually associated with specific sensory systems, occur within 200 msec after the presentation of a stimulus, and are elicited by the physical characteristics of stimuli. For example, exogenous auditory potentials are influenced by the intensity, frequency, patterning, pitch, and location of the stimulus in the auditory field (Kramer, 1987; 1991).

The difference between the endogenous and exogenous components suggest the need for components to be clearly defined. ERP components are typically labeled with either a "N" or "P", for negative and positive polarity, respectively. Also, a number is assigned indicating the minimal latency measured from the onset of a discrete event. The attributes of the ERP that have served as definitional criteria have included: the arrangement of transient voltage changes across the scalp, polarity, latency range, sequence, and the sensitivity of these components to task instructions, parameters, and physical changes in the eliciting stimulus (Donchin, Ritter, & McCallum, 1978; Kramer, 1985; 1987; 1991).

The scalp arrangement concerns the amplitude and polarity of the components across various locations on the scalp. For example, research has demonstrated that the P300 component becomes increasingly smaller in amplitude from the parietal to the frontal sites, whereas the N100 is largest over the Fz, Cz, and Pz sites (see Figure 3). The latency range is influenced by both experimental manipulations and whether it is an endogenous or exogenous component. For example, brainstem evoked potentials occur within 10 ms after the presentation of a stimulus. These ERPs are influenced by both organismic and stimulus variables; however, the latency range is only 2-5 ms. This is

contrasted with the latency range of the P300 which depends on the processing requirements of the task and has been shown to span 300-900 ms (Kramer, 1991).

Physiological and Theoretical Basis. The ERP is composed of a sequence of "components" that are generated by groups of cells in different locations of the brain which become active at different times after presentation of a stimulus. Although there is little consensus as to what the different components are thought to measure, the early components have been argued to represent the delivery of sensory input from various modalities through the afferent pathways. The later components originate in the primary projection systems, the different association areas, and the non-specific parietal and frontal regions (Vaughan & Arezzo, 1988).

To complicate matters further, the later the ERP components (e.g., P300), the more the components represent "memory-driven" rather than "data-driven" processes. For example, Hillyard and Picton (1979) have argued for a two-stage process for the ERP. The primary sensory system carries out a feature analysis and evaluates characteristics of the stimulus and, if it passes some criteria for selection, it then passes the sensory input to a second system. This second system evaluates the stimulus with

comparison to memory models of expected or salient events (Gopher & Donchin, 1986).

The two-stage model of attentional processes involved in the etiology of the ERP has implication for the study of mental workload. Donchin and his colleagues (Donchin, 1981; Donchin, McCarthy, Kutas, & Ritter, 1983) argued that, because the P300 is elicited by improbable or unexpected events, the P300 represents a "context-updating" of the mental model of the environment. The mental model is continually assessed for deviations from expected sensory inputs and, when the events exceed some criterion, the mental model is updated. The frequency at which the mental model is updated is based on the surprise value and task relevance of the event. Donchin (1981) further developed a subroutine metaphor for the various activities of the ERP components. The P300 subroutine was posited to be invoked whenever there is a need to evaluate unusual, novel events in the environment (Gopher & Donchin, 1986; Kramer, 1987; Kramer; 1991).

The finding that the subroutine, characterized by the P300, is invoked only with task-relevant or surprising events has been important in the use of the ERP as a measure of mental workload. Consider a situation in which a participant must perform an oddball task while performing another task simultaneously. Now, imagine that the

difficulty of the primary task is increased. Would the P300 subroutine still be invoked? If so, would the amplitude of the P300 reflect the increased workload demands and, therefore, serve as an index of the resources demanded by these two tasks? Such questions as these served as the impetus for researchers to begin to investigate the use of the P300 in the assessment of workload (Kramer, 1987; Gopher & Donchin, 1986; Parasuraman, 1990).

Dual-Task ERPs. The earlier ERP studies of mental workload were driven by research findings connecting changes in ERP components to state variables, such as fatigue and arousal. Haider, Spong, and Lindsley (1964) first reported that shifts in the N100 visual and auditory ERP during discrimination tasks reflected both states, such as fatigue, arousal, and vigilance, as well as discrimination task performance. Thereafter, ERPs were linked to the secondary-task method, a method that was emerging as a technique for assessing primary task workload demands. The earlier dual-task ERP studies of mental workload concentrated on stimulus-evoked, exogenous, rather than task-evoked, endogenous ERP components. For example, Defayolle, Dinand, and Gentil (1971) reported that the P100 component of the ERP to flashes of red light was reduced when subjects performed a reasoning task as opposed to a control condition in which no task was performed. Furthermore, as the

difficulty of the reasoning task was increased, the amplitude of the P100 showed further reductions. Spyker, Stackhouse, Khalafall, and McLane (1971) demonstrated that the P250 component of the ERP was also affected by the difficulty of the task. They reported that the amplitude of the P250 component of the ERP to visual probe stimuli was reduced as the dynamic complexity of a tracking task was increased (Parasuraman, 1990).

In a recent review of the research, Parasuraman (1990) concluded that these early studies were plagued by lack of experimental control over the processing of the probe stimulus. The experimental tasks were either not integrated with the presentation of the probe or, as in the case of Defayolle, Dinand, and Gentil (1971), time domains of ERPs were not averaged separately for various response categories and different stimuli. More recent research, however, requires subjects to process the discrete event to some degree. A separate task is associated with the ERP stimuli making this method a more exact analog of the dual-task procedure (Parasuraman, 1990).

Many of these more recent studies have focused on the P300 component. These studies were based upon the notion that P300 amplitude in a task should be proportional to the attentional resources invested in the task (Johnson, 1986; Parasuraman, 1990). Put another way, if subjects are given

one task to perform while performing another task concurrently, the demands imposed by the secondary task would impact the "memory-driven" processes and, therefore, can be assessed by evaluating how the amplitude of the P300 changes in the primary task (Parasuraman, 1990).

One of the first such studies was performed by Wickens, Isreal, and Donchin (1977). In this study, the P300 amplitude to counted tones decreased when a visual tracking task was also performed. This finding is not much different than the earlier ERP studies, except that the effect was for a task-evoked, endogenous rather than a stimulus-evoked, exogenous ERP component. However, P300 amplitude was not found to be sensitive to increases in the difficulty of the tracking task, either when the number of tracked dimensions was increased from one to two (Wickens et al., 1977) or when the bandwidth of the tracking task was increased (Isreal, Chesney, Wickens, & Donchin, 1980). The fact that the P300 did not vary much as a function of primary task difficulty was attributed to the idea that primary and secondary tasks draw on different "resource pools." This view contends that the tracking task difficulty taps response-related resources; however, the P300 counting task taps perceptual resources.

In another study, Isreal, Wickens, Chesney, and Donchin (1980) coupled a counting task with a visual monitoring

task. Subjects were asked to monitor the visual task for changes in the intensity or direction of squares and triangles that moved over a visual display. In this study, perceptual factors were manipulated by requiring subjects to monitor either four or eight display elements. The results showed that the P300 amplitude to the stimuli in the visual task was smaller in the dual-task conditions. Moreover, P300 was decreased further in the high-load, eight display element condition; however, this effect was found only for the direction-change primary task. Similar studies (e.g., Kutas, McCarthy, & Donchin, 1977; McCarthy & Donchin, 1981; Ragot, 1984) have also found that the P300 is influenced by perceptual factors. Taken together, these studies support the view that P300 amplitude can be used as a measure of workload of a perceptual and cognitive, but not response-related nature. Further, P300 latency has been found to change with stimulus parameters, such as masking, that are known to affect encoding and central processing, but not for stimulus-response processing, such as stimulus-response compatibility (McCarthy & Donchin, 1981; Parasuraman, 1990). These results have been discussed in terms of the multiple-resource view of workload that holds that several separate resource pools exist corresponding to different modalities, perceptual versus response processes, and so on (Wickens, 1984). The fact that the P300 amplitude was not sensitive

to tracking difficulty suggests that this factor depletes resources that are not used by the P300 process (Hoffman, 1990; Parasuraman, 1990).

Primary Task ERPs. The afore-mentioned studies utilized a dual-task methodology to assess ERP as a metric to resources of a perceptual/cognitive nature and were taken as supporting the multiple-resource view of workload. The results demonstrated that, if the primary task difficulty is manipulated and yields secondary task performance decrements, in addition to secondary task P300 amplitude decrements, then the results can be taken as reflecting competition for perceptual/central processing resources over and above those placed upon the response/ output system. However, according to Sirevaag, Kramer, Coles, and Donchin (1989), the P300 associated with the primary task has been overlooked. They contended that, if P300 amplitude does indeed evince resource competition shown to occur during dual-task performance, logically then the P300s elicited by the primary task should result in an increase in amplitude as the workload of the primary task is increased. Further, in dual-task studies where ERPs can be recorded in response to both discrete primary and secondary task events, one should find a reciprocal relationship between primary and secondary task P300 amplitudes (Sirevaag et al., 1989).

The amplitude reciprocity hypothesis was tested in a study by Wickens, Kramer, Vanasse, and Donchin (1983) in which subjects were asked to track a target with a cursor. The ERPs elicited by the discrete changes of the primary task were recorded in one experimental run. ERPs for tones counted during the secondary task were also recorded in a separate trial. In this study, task demands were manipulating by changing the number of integrations between the joystick output and the movements of the cursor on the screen. They found that the P300 associated with the step changes increased in amplitude with increasing primary task difficulty; whereas secondary task P300 amplitudes decreased.

Recent studies have also found that P300s elicited to events from the primary task increase in amplitude with increases in primary task difficulty (Sirevaag et al., 1989; Strayer & Kramer, 1990; Ullsperger, Metz, & Gille, 1988) For example, Sirevaag et al. (1989) employed a method where both primary and secondary ERPs could be concurrently recorded within the same experimental condition. Measures of P300 amplitude and performance were obtained from 40 subjects within the context of a pursuit step tracking task performed alone and with a concurrent secondary auditory discrimination task. The pursuit tracking task difficulty was manipulated by varying both the velocity and

acceleration control dynamics as well as the number of dimensions, either one or two, to be tracked. ERPs were recorded for both the tracking task setup changes and for the secondary task tones. The results showed that, as the primary task difficulty was increased as reflected in increased root mean squared error (RMSE) scores, there was decreased secondary task P300 amplitudes and increased primary task P300 amplitudes. Moreover, the increases in primary task P300 amplitudes were concomitant with the amplitude decrements obtained for the secondary task. These findings were taken as supporting the amplitude reciprocity hypothesis between primary and secondary task P300 amplitudes as a function of primary task difficulty.

Simulation Research. The previously mentioned research has provided important evidence about the relationship between the P300 and mental workload. However, these studies have not addressed whether such findings can generalize to real-world environments. This is especially important if such studies are to be applied to adaptively automated systems. Fortunately, much research has been conducted that has addressed this issue. Studies have employed a number 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 et al., 1985; Humphrey & Kramer,

1994; Kramer & Strayer, 1988; Kramer, Sirevaag, & Braune, 1987; Kramer, Wickens, & Donchin, 1983; 1985; Lindholm, Cheatham, Koriath, Longridge, 1984; Natani & Gomer, 1981; Sirevaag et al., 1993; Strayer & Kramer, 1990; Theissen, Lay, & Stern, 1986). For example, Lindhom et al. (1985) elicited ERPs to auditory stimuli during simulated landings and attack scenarios. They reported a larger P300 amplitude decrease as the workload in the primary task was increased. A related study used an oddball, or rare event, secondary-task to elicit ERPs as subjects performed a flight task simulation (Natani & Gomer, 1981). This study found significant P300 amplitude decrements as well as longer P300 latencies under the high workload conditions. However, similar results were not found for a second replication of the task (Wilson & Eggemeier, 1991).

Theissen, Lay, and Stern (1986) employed a visual oddball task to elicit ERPs while electronic-warfare officers performed various tasks in a fighter aircraft simulator. Task difficulty levels were manipulated by changing task parameters, such as target characteristics (e.g., number and type) and threats to aircraft. The results demonstrated smaller P300 amplitudes in the single-task control condition than in the simulated flight conditions. Kramer, Sirevaag, and Braune (1987) evaluated workload during a flight simulation experiment that used an

auditory, rather than visual, oddball task that required subjects to discriminate infrequent from frequent tones. They found that the P300 component of the ERP consistently indexed changes in flight difficulty level with a finding of decreased P300 amplitude with increased primary-task difficulty. Further, P300 amplitude demonstrated a negative correlation with deviations from flight headings. Such a finding suggests that primary task data can be coupled with ERP data to make allocation decisions in an adaptively automated environment.

Sirevaag et al. (1993) elicited ERPs to irrelevant probes as helicopter pilots flew a series of reconnaissance missions in a motion-based, high-fidelity helicopter simulator. They reported smaller P300s amplitudes to probes as the communication load imposed on the pilots was increased. Biferno (1985) also looked at communication load and ERPs. He recorded ERPs from radio call signs as subjects performed flight simulator missions. P300 amplitude was found to be smaller as the workload increased. Furthermore, both fatigue and subjective workload estimates of workload were reported to discriminate between various levels of workload. These results suggest that ERPs are associated with other measures of taskload thereby attesting to their utility for workload estimation and adaptive automation.

Most of the research conducted with ERPs and mental workload has been focused on flight simulation. In one of the few applications of ERPs outside of aviation, Wesensten et al. (1993) recorded auditory ERPs from 10 male participants at 0900, 1600, and 1830 hours. P300s were collected while participants were at sea level and another one was collected following a rapid ascent to a simulated 4,300 meter altitude. The results of the study were a decrease in P300 amplitude, while P300 latency and reaction time increased, following the ascent. Another study (Janssen & Gaillard, 1985) used an auditory Sternberg memory task to elicit ERPs from automobile drivers as they drove on three different types of roadway: rural, city, and highway. Expressway driving was found to elicit the smallest P300 amplitudes, and this was interpreted as being the driving segment with the highest workload (Wilson & Eggemeier, 1991).

Conflicting Simulator Studies. A number of field studies have demonstrated that the ERP reliably varies with workload. However, a few studies exist that have not shown such clear-cut evidence (e.g., Fowler, 1994; Jannsen & Gaillard, 1985; Natani & Gomer, 1981). For example, Fowler (1994) elicited ERPs using auditory and visual oddball tasks as subjects flew a flight rules final approach and landing manuever under workloads varied by manipulating turbulence

and hypoxia. The oddball tasks required subjects to detect infrequent tones or flashes of an artificial horizon. Although RMSE flying performance was found to be systematically degraded by the two workload conditions, the P300 amplitude was not strongly related to performance. However, P300 amplitude was inversely related to high taskload when the visual condition was analyzed separately. The authors accounted for this result by invoking the amplitude reciprocity hypothesis. As stated previously, this hypothesis suggests that, as the primary task difficulty is increased and the P300 amplitude elicited by the secondary task decreases, P300 amplitude for task-relevant events embedded in the primary task increases. Therefore, the flashing horizontal horizon was processed as part of the primary task causing the P300 amplitude to increase as a function of task difficulty. However, this cannot account for the results reported for the auditory condition as no systematic pattern emerged in contrast to a similar study done by Kramer, Sirevaag, and Braune (1987).

Fowler (1994) also reported that P300 latency was found to covary with flight performance, increasing as a function of workload in both modalities. O'Donnell and Eggemeier (1986) suggested that the P300 amplitude indexes workload because it is sensitive to subject expectancy that is disrupted by workload. This would explain the

disassociation between latency and amplitude because the mechanisms controlling expectancy would be different than those indexing the speed of perceptual/cognitive processing. According to this view, the instrument flight rules flying task used by Kramer, Sirevaag, and Braune (1987) primarily interrupted subject expectancy whereas the visual flight rules task used by Fowler (1994) primarily slowed stimulus evaluation. The authors noted that this possibility suggests that both P300 amplitude and latency can be used as indices of mental workload, depending on the nature of the task (Fowler, 1994)

In a second study, Janssen and Gaillard (1985) were unable to replicate the finding of a smaller P300 amplitude to probes during expressway driving despite the fact that heart-rate variability was found to be significantly decreased in the more demanding expressway segment in both studies. Also, Natani and Gomer (1981) were unable to replicate the findings of their first study. Similar to Fowler (1994), however, Janssen and Gaillard reported that P300 latency was sensitive to increases in taskload.

Real-Time Assessment of Mental Workload. Although the simulator studies cited above, have yielded useful information, they have not addressed whether ERPs could measure dynamic changes in mental workload. For example, in simulator studies, 50-100 single trial ERPs may be collected

and then averaged to determine whether ERP components discriminate workload or performance levels. In an adaptively automated environment, collection of this quantity of ERP data may not be practical. A number of earlier studies, however, have suggested that ERPs can be used for on-line evaluations of moment-to-moment fluctuations in operator workload (Defayolle et al., 1971; Gomer, 1981; Sem-Jacobsen, 1981). Although research on real-time assessment of mental workload is still in its infancy, this line of research has been expanded in several recent studies that have suggested that on-line assessment may soon be feasible. For example, Farwell and Donchin (1988) asked subjects to attend to one item in a 6 x 6 matrix of items. The columns and rows flashed randomly and ERPs elicited from the flashes were used to discriminate between the attended and unattended items. A 95 percent accuracy level was found using just 26 seconds of ERP data. Kramer, Humphrey, Sirevaag, and Mecklinger (1989) also found that on-line assessment of mental workload can be performed with a small amount of ERP data (Kramer, 1991).

Humphrey and Kramer (1994) also reported a study that examined whether ERPs could measure dynamic changes in mental workload. They examined how much ERP data is necessary to discriminate between levels of mental workload in complex, real-world tasks. In order to address this

question, they employed a bootstrapping approach to investigate the accuracy of discriminating between workload levels using different amounts (e.g., 1 to 75 sec) of ERP data. Participants were asked to perform two tasks, monitoring and mental arithmetic, both separately and together. Following an analysis of the performance, subjective workload ratings, and average ERP data in the single- and dual-task conditions, two different conditions from each of the tasks were selected for further analysis. The results of the study indicated that 90% correct discrimination could be achieved with from 1 to 11 sec of ERP data. These results were discussed in terms of real-time assessment of mental workload using ERP data. Kramer, Trejo, and Humphrey (1996) discussed these results as evidence that event-related potentials can be useful in the design of adaptive systems.

Research Purpose

The EEG and ERP represent viable candidates for determining shifts between modes of automation in adaptive systems. Because real-time assessment of workload is the goal of system designers wanting to implement adaptive automation, it is likely that these measures will become the focus of research on adaptive automation. This optimism stems from a number of studies that have suggested that they might be useful for on-line evaluations of operator workload

(Defayolle et al., 1971; Farwell & Donchin, 1988; Gomer, 1981; Humphrey & Kramer, 1994; Kramer, 1991; Kramer et al., 1989; Sem-Jacobsen, 1981). Although these results suggest that on-line assessment of mental workload may be possible in the near future, a good deal of additional research is needed.

The determination of measures on which to dynamically allocate automation does not represent the only area that needs exploration. Other areas include the frequency with which task allocations are made, when automation should be invoked, and how this invocation changes the nature of the operator's task (Parasuraman et al., 1992). Specifically, it is not known how changing among automation task modes impacts the human-automation interaction.

The present study attempted to examine the impact of cycles of automation on behavioral, subjective, and psychophysiological correlates of operator performance. Furthermore, the efficacy of use of EEG and ERPs for adaptive task allocation was also examined. The study was an off-shoot of previous research by Pope, Bogart, and Bartolome (1995) who examined the use of EEG as an adaptive trigger for changing among automation task modes.

The Biocybernetic System

Electroencephalogram. Pope, Bogart, and Bartolome (1995) reported one of the few studies examining the utility

of EEG for adaptive automation technology. These researchers developed an adaptive system that uses a closed-loop method to adjust modes of automation based upon changes in the operator's EEG patterns. The closed-loop method was developed to determine optimal task allocation using an EEG-based index of engagement or arousal. The system uses a biocybernetic loop that is formed by changing levels of automation in response to changing taskload demands. These changes were made based upon an inverse relationship between the level of automation in the task set and the level of pilot workload.

The level of automation in a task set could be such that all, none, or a subset of the tasks could be automated. The task mix is modified in real time according to operator's level of engagement. The system assigns additional tasks to the operator when the EEG reflects a reduction in task set engagement. On the other hand, when the EEG indicates an increase in mental workload, a task or set of tasks may be automated, reducing the demands on the operator. Thus, the feedback system should eventually reach a steady-state condition in which neither sustained rises nor sustained declines in the EEG are observed.

One issue for the biocybernetic system concerns the nature of the EEG signal used to drive changes in task mode. Pope, Bogart, and Bartolome (1995) argued that

differences in task demand elicit different degrees of mental engagement that could be measured through the use of EEG-based engagement indices. These researchers tested several candidate indices of engagement derived from EEG power bands (alpha, beta, & theta). These indices of engagement were derived from recent research in vigilance and attention (Davidson, 1988; Davidson et al., 1990; Lubar, 1991; Offenloch & Zahner, 1990; Streitberg, Rohmel, Herrmann, & Kubicki, 1987). For example, Davidson et al. (1990) argued that alpha power and beta power are negatively correlated with each other to different levels of arousal. Therefore, these power bands can be coupled to provide an index of arousal. For example, Lubar (1991) found that the band ratio of beta/theta was able to discriminate between normal children and those with attention deficit disorder.

Pope and his colleagues (1995) reasoned that the usefulness of a task engagement index would be determined by a demonstrated functional relationship between the candidate index and task operating modes (i.e., manual versus automatic) in the closed-loop configuration. They used both positive and negative feedback controls to test candidate indices of engagement because each should impact system functioning in the opposite way, and a good index should be able to discriminate between them. For example, under negative feedback conditions, the level of automation in the

tasks was lowered (i.e., automated) when the EEG index reflected increasing engagement. On the other hand, when the EEG reflected increases in task demands, automation levels were increased. Task changes were made in the opposite direction under positive feedback conditions; that is, the level of automation in the tasks was maintained when the EEG engagement index reflected increasing task demands. If there was a functional relationship between an index and task mode, the index should demonstrate stable short-cycle oscillation under negative feedback and longer and more variable periods of oscillation under positive feedback. The strength of the relationship would be reflected in the degree of contrast between the behavior of the index under the two feedback contingencies.

Pope, Bogart, and Bartolome (1995) found that the closed-loop system was capable of regulating participants' engagement levels based upon their EEG activity. They reported that the index 20 beta/(alpha+theta) possessed the best responsiveness for discriminating between the positive and negative feedback conditions. The conclusion was based upon the increased task allocations in the negative feedback condition witnessed under this index than under either the beta/alpha or alpha/alpha indexes. These results were taken to suggest that the closed-loop system provides a means for evaluating the use of psychophysiological measures for

adapting automation. A number of subsequent studies (Prinzel, Scerbo, Freeman, & Mikulka, 1995; Prinzel, Hitt, Scerbo, Freeman, & Mikulka, 1995; Prinzel, Scerbo, Freeman, & Mikulka, 1997) have also reported that the system is capable of moderating workload on behavioral, physiological, and subjective dimensions.

Recently, an improvement had been made to the biocybernetic system. The previous system used by Pope, Bogart, and Bartolome initiated changes in automation levels based on the slope of the index taken from successive measurements. One problem with using a slope measure concerns its sensitivity to changes in operator arousal and its reflection of levels of operator engagement. The system makes task allocation decisions regardless of whether the engagement level is high or low. In other words, an operator's overall engagement level may be quite low relative to his or her normal baseline engagement level. However, the system may make a task allocation decision to automate a task merely because the arousal level is higher, when the next EEG engagement index is derived, despite the fact that the overall arousal level is still low (Hadley, et al., 1997; Prinzel, Scerbo, Freeman, & Mikulka, 1997). Therefore, the system makes task allocation decisions without a consideration of individual differences in engagement.

One strong candidate for making such decisions is the use of algorithms based on the absolute levels of the EEG engagement index. In such a system, baseline data could be obtained, such as the mean of the EEG engagement index, for an individual operator. That data could then be fed into a biocybernetic system and task allocation decisions made based upon the absolute value of the index relative to the mean data obtained during the baseline period.

Prinzel, Freeman, Scerbo, and Mikulka (1997) reported on such a biocybernetic system. They examined the effectiveness of the three indices derived from the same four cortical sites as Pope, Bogart, and Bartolome (1995). Their system used the average index derived from the participant's baseline EEG to make task allocation decisions. Participants were asked to perform a compensatory tracking task under both negative and positive feedback conditions. The results were that participants performed better under the negative feedback condition than under the positive feedback condition. Also, the index 20 beta/(alpha+theta) was found to be superior in distinguishing between negative and positive feedback in terms of behavioral, subjective, and physiological correlates. Thus, the task allocation and physiological data were found to be comparable to the previous results of studies using a slope method to drive the biocybernetic

system. However, the results demonstrated that the system was also better able to improve performance and moderate workload demands. These findings are important for the design of adaptive automation. The use of a slope approach may work well with binary types of adaptive automation. More complex systems incorporating multiple levels of automation, however, would require algorithms that can trigger task allocations based upon differences among several engagement levels. These findings suggest that a system, using absolute measures of operator engagement, may be used to allocate tasks among various task engagement levels.

Short-Cycle Automation. Clearly, the research on automation has shown that a number of deleterious effects on human performance often accompany the advantages that automation provides. As Endsley and Kiris (1994) have noted, research is needed that examines various techniques that would establish human-centered automation that minimizes the negative effects of automation while maximizing overall human-system performance. Adaptive automation has been touted as just such a remedy. However, although much speculation has been made concerning adaptive automation, it remains to be seen whether adaptive automation can deliver on its promises (Glenn et al., 1994). Woods (1996) stated that, "conventional wisdom about

automation makes technology change seem simple....However, the reality of technology change ...is that technological possibilities often are used clumsily, resulting in strong, silent, difficult-to-direct systems that are not team players (p. 15)"; this is what he calls "apparent simplicity, real complexity". What is required then is to examine whether adaptive automation really can provide anything additional not already present by other, less technological approaches.

A number of studies have demonstrated that cycling between automation modes may be beneficial (Ballas, Heitmeyer, & Perez, 1991; Hadley, Prinzel, Freeman, & Mikulka, 1998; Hilburn, Molloy, Wong, & Parasuraman, 1993; Parasuraman, Molloy, & Singh, 1993; Parasuraman, Bahri, Molloy, & Singh, 1992; Johannsen, Pfendler, & Stein, 1976; Scallen, Hancock, & Duley, 1995). These studies have shown that short-cycle automation can significantly improve performance and lower workload. For example, Scallen, Hancock, and Duley (1995) had six pilots perform tracking, fuel management, and system monitoring tasks for nine trials lasting five minutes each. The nine trials consisted of factorial combinations of three conditions of tracking difficulty (low, medium, and high) and three conditions of cycle duration (15, 30, or 60 seconds). These researchers found that tracking performance was significantly better at

the 15-sec cycle duration, but there were no differences in mental workload across the three cycling conditions ($p <$ .07).

Hadley, Prinzel, Freeman, and Mikulka (1998) expanded on the Scallen, Hancock, and Duley (1995) study. These researchers asked nine participants to perform a tracking task and an auditory, oddball task for three trials consisting of a 15-, 30-, and 60-sec cycle durations. ERPs were gathered to infrequent, high tones presented in an auditory oddball task. The results showed that tracking performance was significantly better under the 15-sec duration, but participants rated workload significantly higher under this condition. These results were interpreted in terms of a micro-tradeoff; that is, participants did better under the 15-sec condition at the expense of working harder. The conclusion was supported by the ERP results. An examination of the EEG gathered five seconds after each task allocation revealed that P300 latency was found to be considerably longer and the amplitude considerably smaller under the 15-sec cycle duration than under either the 30- or 60-sec cycle conditions. Therefore, these results suggest that short periods of manual reallocation may prove beneficial to performance and moderating workload demands. However, such benefits are tempered by increased return-to-manual deficits (Wiener & Nagel, 1988). Moreover, they

support the use of ERPs metrics of workload in the design and implementation of adaptive automation technology.

Note that the question of adaptive automation does not hinge on its conceptual underpinnings. Inherently, it makes sense to transform the operator's task at times when the operator's mental state is less than optimal. However, this is not to say that adaptive automation provides utility that supersedes the difficulties that we, as researchers, designers, and practitioners, may face with the implementation of this type of technology. Such studies, as those discussed previously, demonstrate that schedules of static automation can also have positive effects on performance and workload. Therefore, it is of theoretical and practical interest to determine what benefits, if any, that adaptive automation provides beyond that of static automation which cycles between automation modes based upon scripted automation schedules.

The present study sought to examine the impact that adaptive automation has on performance as well as subjective and psychophysiological measures of workload. To assess the research question, the biocybernetic system was used to dynamically allocate tasks between manual and automatic modes. Participants were yoked to other participants that also performed the experimental tasks. However, the task allocations that these participants experienced was based

upon the exact cycle scheduling of their yoked counterpart. Therefore, it was possible to examine the impact that adaptive and static cycling has on various correlates of workload. Further, the design allowed a comparison of these two forms of task allocation.

Event-Related Potentials. As noted, many theories, models, and platforms for implementing adaptive automation have already been proposed (Mouloua & Parasuraman, 1995), including the use of biopsychometric measures, such as ERPs, as indices of operator states in adaptive systems (Defayolle, Dinand, & Gentil, 1971; Gomer, 1981; Hancock, Chignell, & Lowenthal, 1985; Reising, 1985; Rouse, 1977; Sem-Jacobson, 1981). The use of ERPs in the design of adaptive automation systems was considered some years ago in the context of developing "biocybernetic" communication between the pilot and the aircraft (Donchin, 1980; Gomer, 1981). The idea concerned systems in which tasks or functions could be allocated flexibly to operators, using ERPs, that may allow the optimization of mental workload to be sought in a dynamic, real-time environment. For example, a method might be developed for obtaining momentary workload levels allowing an index to be derived, such as the amplitude of the P300 wave of the ERP. The workload index could then be compared in real-time to a stored profile of the ERP associated with that task(s). The profile would be

generated from initial baseline data. If the optimal physiological level for a task is exceeded, then the task(s) could be off-loaded from the operator and allocated to the system. Further, if the workload levels become too low, then the task(s) could be transferred back to the operator (Parasuraman, 1990). In recent reviews, however, Parasuraman (Byrne & Parasuraman, 1996; Parasuraman, 1990) concluded that although many proposals have been made concerning the use of ERPs in adaptive systems, little actual research has been conducted.

The proposed study attempted to further the research on the use of ERPs for adaptive automation. What is proposed is that the absolute biocybernetic system be used to make task allocation decisions between manual and automatic task modes as previously described. Participants were also asked to perform an oddball, auditory task concurrently with the compensatory tracking task. The EEG signal was fed to both the biocybernetic system and to a data acquisition system that permited the analysis of ERPs to high and low frequency tones. Such results are hoped to assess the efficacy of using ERPs in the design of adaptive automation technology.

Research Hypotheses

1. Based upon previous findings (Hadley, Mikulka, Freeman, Scerbo, & Prinzel, 1997; Prinzel, Freeman, Scerbo,

& Mikulka, 1997) with the absolute biocybernetic system, it is predicted that the system will make significantly more task allocations under the negative feedback condition than under the positive feedback condition. The hypothesis is confined to the data gathered from the adaptive automation group as the schedules of task allocations for the yoked and control groups are determined based upon the data gathered from the former group.

2. Parasuraman, Molloy, and Singh (1993) demonstrated that manual task reallocation may be a potential countermeasure to decrements in performance often observed with automation. They found a temporary return to manual control of a monitoring task from automated functioning reduced failures of ommissions for both pilots and nonpilots. Furthermore, more sustained benefits were observed with multiple or repetitive manual reallocations. Similar findings were reported by Scallen, Hancock, and Duley (1995) and Hadley, Prinzel, Freeman, and Mikulka (1998) for tracking performance. Therefore, because the negative feedback condition is predicted to produce the most task allocations, the increase in manual reallocations should result in significantly better tracking performance and lower subjective workload scores than under the positive feedback condition.

3. Another hypothesis concerns how behavioral and subjective measures are moderated by adaptive automation relative to static automation. It is predicted that participants in the adaptive automation condition will have better tracking performance and lower subjective workload ratings than the yoked participants in the static automation condition. Therefore, performance and workload metrics should evince significant differences between the adaptive and yoked group conditions with no differences observed between the yoked and control group conditions. Specifically, participants in the adaptive automation, negative feedback condition should have significantly better performance and lower workload scores than all other group, feedback conditions.

4. Because task allocations between operating modes are not contingent upon changes in workload as measured by EEG patterns for those participants in the yoked condition, there should be no differences in tracking performance or subjective workload estimates between negative and positive feedback conditions. The reason is that any task allocations made are determined based upon the schedule of their yoked counterpart and are unrelated to the mental state of the participant. Likewise, there should be no differences in performance or subjective workload metrics between the yoked and control group conditions.

Conversely, if the effects are due to increased manual reallocations and not to the adaptive method of task allocation, increasing the automation cycle schedule should result in improved performance and lowered workload for participants in the yoked condition under the negative feedback condition than under the positive feedback condition. Additionally, there should be no differences in performance or subjective workload scores between the three group conditions.

5. For participants in the adaptive automation group, the derived EEG engagement index is hypothesized to vary as a function of which feedback condition and which task mode the system was operating under. Under positive feedback, when the EEG patterns reflect a low task engagement state, the system automates the tracking task that theoretically further lowers engagement levels. However, if the EEG patterns reflected increasing engagement, the system allocates the tracking task to the manual task mode. Therefore, for positive feedback, it is hypothesized that the EEG engagement index would be highest during manual task mode and lowest during the automatic task mode.

The opposite pattern is expected for negative feedback. Negative feedback is designed to induce optimal states of task engagement. The system does this by allocating tasks to the operator when the EEG shows that the engagement state

is below baseline levels and automates the task when the engagement state is above baseline levels. Therefore, for negative feedback, the EEG engagement index is expected to be significantly lower during the manual task mode than during the automatic task mode (Prinzel, Freeman, Scerbo, & Mikulka, 1997; Prinzel, Hitt, Scerbo, Freeman, & Mikulka, 1995; Prinzel, Scerbo, Freeman, & Mikulka, 1995; Prinzel, Scerbo, Freeman, & Mikulka, 1997).

6. Numerous studies have demonstrated that the P300 amplitude and latency reflects workload levels wherein the amplitude decreases and latency increases with increases in workload demands. Therefore, it is predicted that the amplitude of the P300 component to infrequent, high tones in a secondary, auditory oddball task is predicted to be significantly smaller and P300 latency longer under the higher workload, positive feedback condition than under the lower workload, negative feedback condition.

7. There should be a significant feedback condition X group condition interaction with the P300 discriminating between the two feedback conditions only for those participants in the adaptive automation group. No differences are expected to be evident between the yoked condition and control conditions regardless of which feedback condition the biocybernetic system is operating under.

However, as the results of Scallen, Hancock, and Duley (1995) suggest, if increasing the number of task allocations between manual and automated operating modes results in lowered performance errors and workload scores, then differences in the P300 components may be seen between the two feedback conditions for participants in the yoked condition. If performance and workload differences are seen, it is predicted that the P300 amplitude should be smaller and latency longer under the positive feedback condition than under the negative feedback condition. Also, no differences would be expected between the three group conditions for P300 amplitude and latency.

8. Efficient task performance requires selective attention to task-relevant events. Attention to these events amplifies a range of ERP components, including P100, N100, P200, and N200 as well as slower, broad negativities. Therefore, in addition to the P300 component, the study will also examine the relationship of other ERP components. Although most research has focused on the P300 component of the ERP, a number of researchers have suggested that these other components can also be used in the assessment of workload (e.g., Lindholm, Cheatham, & Koriath, 1984). The consensus is that attention in high workload situations requires allocation of both common, nonspecific resources (e.g., N100 component) and task-specific resources (i.e.,

P300 component). Generally, the amplitude decreases and latency increases as workload demands are increased (Parasuraman, 1990). Therefore, there should be a significant difference in amplitude and latency of these ERP components between negative and positive feedback conditions.

9. Furthermore, it is expected that there will be a significant feedback X experimental group interaction for these different waveforms. The amplitudes are predicted to be greater and the latencies shorter for infrequent, high tones under the negative feedback condition only for those participants in the adaptive automation group. No differences are expected between any other group, feedback condition combination.

However, if increased manual reallocations are responsible for the lowered task load under negative feedback, no differences in these ERP components would be expected between the three group conditions. Rather, only a main effect should be found between negative and positive feedback.

CHAPTER II

METHOD

Participants

Thirty-six undergraduate and graduate students served as participants for this experiment. The ages of the participants ranged from 18 to 40. Participants were given monetary compensation or extra course credit for their voluntary participation. All participants were right-handed as measured by the Edinburgh handedness survey (Oldfield, 1971) and had normal or corrected-to-normal vision.

Apparatus

Electrical cortical activity was recorded with an Electro-Cap International sensor cap. The lycra sensor cap consists of 22 recessed tin electrodes arranged according to the International 10-20 system (Jasper, 1958). One mastoid electrode was used for a reference. Conductive gel was placed into each of the four electrode sites, the reference, and the ground using a dispenser tube and a blunt-tipped hypodermic needle.

The NeuroScan SynAmps is a AC/DC amplifier that provides both a broadband amplifier and a high speed digital acquisition system. The system has four high speed digital signal processors (DSPs) with 1 MByte of RAM per DSPs for data acquisition. The SynAmps has a 33 MHz 486 DX processor with 4 MBytes of RAM and an electronic flash disk dedicated

to management of DSPs. It provides for real-time digital filtering by the DSPs allowing filter settings from DC to 10kHz. Sampling rates can be set between 100 Hz to 20 kHz from 1 to 32 channels. Also, the system has 28 monopolar and 4 bipolar channels provided through a NeuroScan SynAmps headbox connector. The SynAmps amplifier has tracking anti-aliasing filters, first stage amplification to reduce S/N ratio, and an on-line DC offset correction. All impedance calibration is built-in and the input signal is managed through SCAN software. The system was used for ERP acquisition and analyses.

The SynAmps amplifier was connected via an analog output board to a Biopac EEG100A A/D converter through a four-line buffered cable. The analog output board takes the output signal from the SynAmps prior to the sample and hold (S/H) circuits. The analog output board filters the signal and then routes the output to a D-37 connector on the SynAmps back panel. Band-limiting is gathered from single-pole high-pass (1 Hz) and low-pass (70 Hz) filters. The anti-aliasing filters are set for 0.2 times the sample frequency.

The system was also connected to a PC computer through the parallel port on the back panel of the SynAmps amplifier. The Biopac system consists of a four channel, high gain, differential input, bio-potential amplifier. The

frequency response is 1 to 100 Hz. The gain setting is x5000 that allows an input signal range of 4000uV (peak-to-peak). However, for the present study, only the Biopac A/D converter was used.

The Biopac A/D converter was connected to the Macintosh Virtual Instrument (VI). The software designed to run the VI is the Real Time Cognitive Load Evaluation System (RCLES v 3.3.1). It calculates the total EEG power in four bands: theta (4-8 Hz), alpha (8-13 Hz), beta (13-22 Hz), and high beta (38-42 Hz). The VI also performs the engagement index calculations and commands the task mode changes through serial port connections to the task computer.

The Macintosh Virtual Instrument was connected to a PC WIN 486 DX computer that was used to run the MAT (see below). Data was binned according to assigned bit numbers placed in the data record from the PC computer. Auditory oddball tone sequencing and gating was controlled by the VI software and these event signals were also placed in the data record as ERP synchronization triggers.

The monitor was a NEC MultiSync 2A color monitor. A joystick was used for the compensatory tracking task. The gain on the joystick was set to 60% of its maximum and had a bandwidth of 0.8 Hz. A graphical depiction of the experimental set-up is shown in Figure 1.

Experimental Design

A 2 feedback condition (positive or negative feedback) X 2 task mode (automatic or manual mode) X 3 experimental group condition (yoked, control, or adaptive automation) mixed-subjects design was employed. The experimental group condition represented the only nested variable. All other conditions were counterbalanced.

Automation Cycle Sequencing. Each of the thirty-six participants was randomly assigned to either the adaptive automation group (n = 12), the yoked (n = 12), or the control (n = 12) group. The adaptive automation condition required the participants to perform the compensatory tracking task and auditory oddball task under the closed-loop configuration. The data records of switches between task modes were then used to determine the pattern of task allocations to be made between automatic and manual task modes for participants in the yoked condition. Therefore, these participants performed the tracking task under the exact same schedule of manual and automatic task modes as their experimental complement. The control group, on the other hand, consisted of participants who performed a random assignment of task allocations between task modes. The

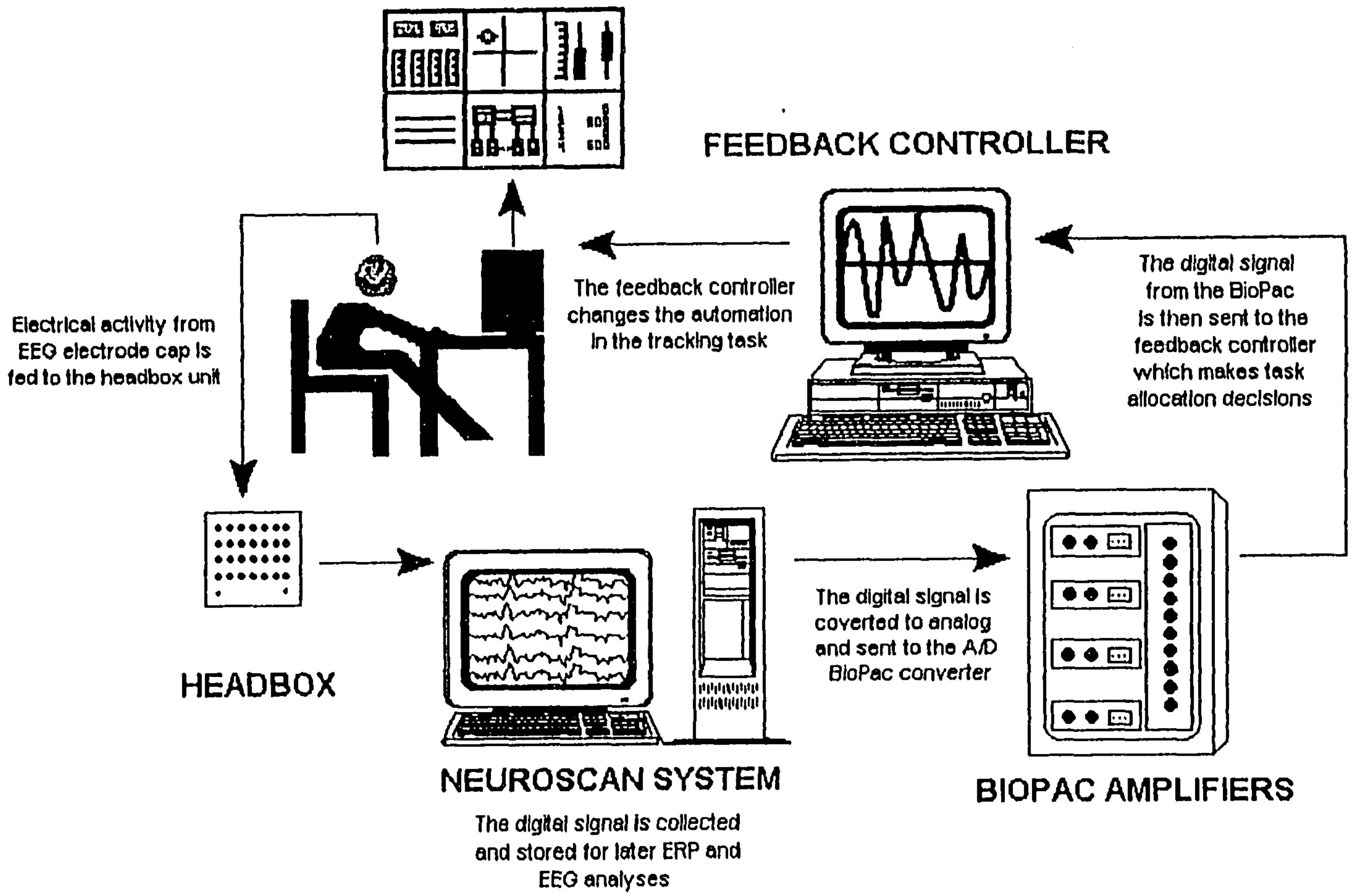

Figure 1. Experimental Hardware Configuration

schedule of task allocations was determined for each control participant based upon the average number of switches in both the positive and negative feedback conditions for the adaptive automation group. For example, control participant number one received a random schedule of task allocations based upon the average number of task allocations that adaptive automation participant number one experienced. All participants, however, had the same sequence of high and low tones in the auditory oddball task.

Dependent Variables. The dependent variables included: (a) the EEG engagement index defined as 20 beta / (alpha+theta); (b) the amplitude and latency of the ERP waveform was analyzed; (c) the number of switches, or task allocations, under each feedback condition; (d) tracking performance as measured by root-mean-squared-error (RMSE); (e) the number of counted high tones in the oddball task; and (7) subjective workload assessed by the NASA-TLX (task load index; Hart & Staveland, 1988; Byers, Bittner, & Hill, 1989).

Statistical Tests and Criterion. All ANOVAs using a repeated measures variable were corrected with the Greenhouse-Geisser procedure (Greenhouse & Geisser, 1959). Alpha level was set at .05. All post hoc comparisons used simple effects analyses and the Tukey post hoc procedure.

Experimental Tasks

Tracking Task. Participants were run using a modified version of the NASA Multi-Attribute Task (MAT) battery (Comstock & Arnegard, 1992). The MAT battery is composed of four separate task areas, or windows, constituting the monitoring, compensatory tracking, communication, and resource management tasks. These different tasks were designed to simulate the tasks that airplane crew members often perform during flight. Only the compensatory tracking task was used in the present study. The task requires participants to use a joystick to maintain a moving circle, approximately 1 cm in diameter, centered on a .5 cm by .5 cm cross located in the center of the screen. Failure to control the circle results in its drifting away from the center cross. The MAT battery is shown in Figure 2.

Auditory Oddball Task. The auditory oddball secondary task consisted of high and low tones at 1100 Hz and 900 Hz, respectively. The frequency of the tone presentation was once per second, and the probability of a high tone was .10 msecs which was randomly assigned for presentation. The inter-stimulus interval was kept uniform across the experimental conditions. Therefore, over a 16-minute trial there were 96 high tone signals and 864 low tone signals.

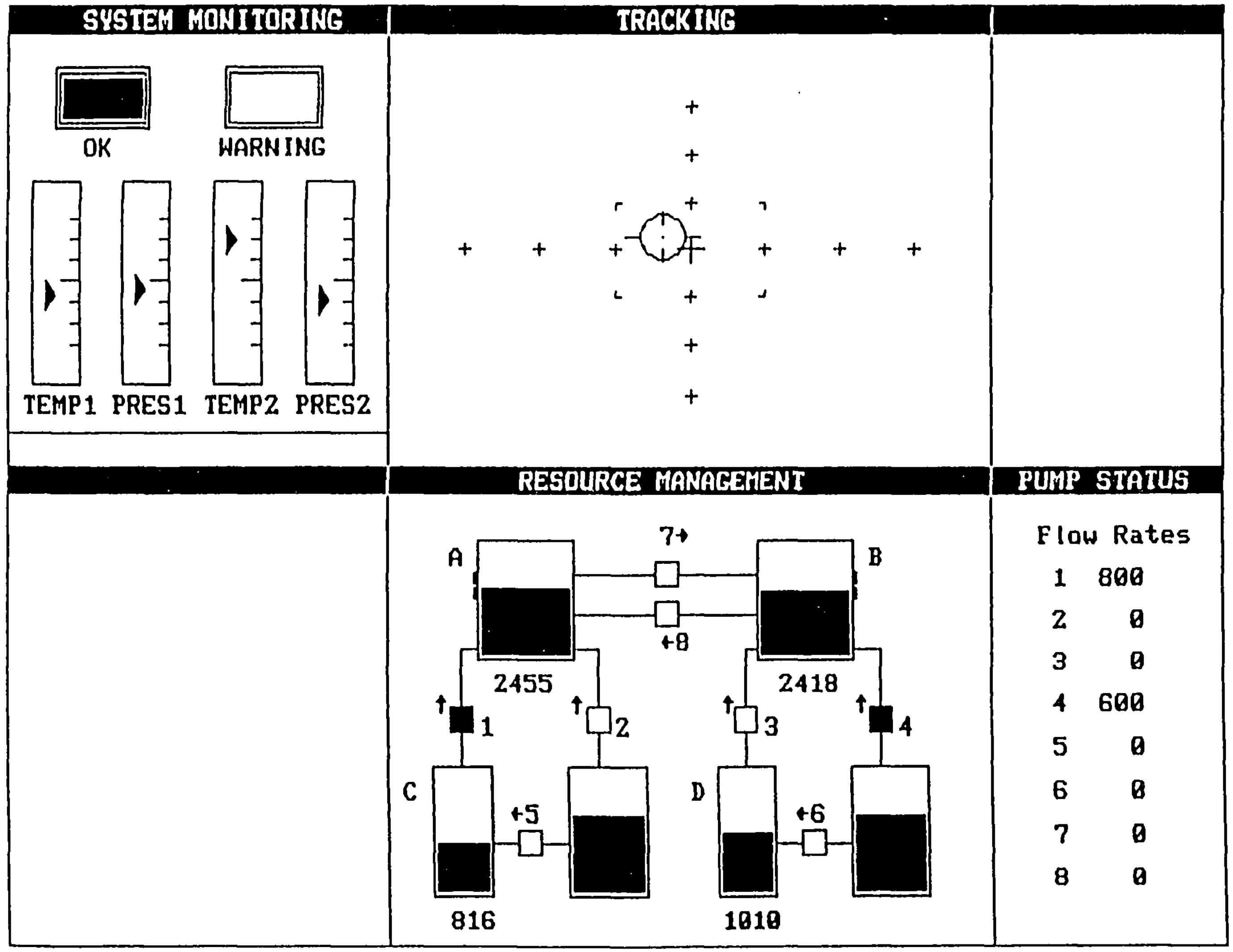

Figure 2. The Multi-Attribute Task Battery

The ordering of the onset of tones was held consistent across participants. The tones were gated to provide a rise and fall time of .10 shaping a square wave signal. The tones were presented to both of the participant's ears through stereo KOSS head phones at 60 dB SPL.

EEG Recording and Analysis

The EEG was recorded from sites Pz, Cz, P3, and P4. A ground site was used located midway between Fpz and Fz. Each site was referenced to the left mastoid. The 10-20 system and the location of electrode sites used in the present experiment is shown in Figure 3.

The EEG was routed through a SynAmps amplifier from an analog output board to the Biopac A/D converter. The outputed analog signal was converted by the BioPac A/D converter to digital, and the digital signals were arranged into epochs of 1024 data points (roughly two and one half seconds). Digitized input channels were then converted back to analog and then routed to an EEG interface with a LabVIEW Virtual Instrument (VI). The VI calculated total EEG power from the bands of theta, alpha, and beta for each of the four sites and converted the signal into a spectral power form using a Fast Fourier Transform (FFT). The EEG frequency bands were set as follows: alpha (8-13 Hz), beta (13-22 Hz), theta (4-8 Hz), and high beta (38-42 Hz). The

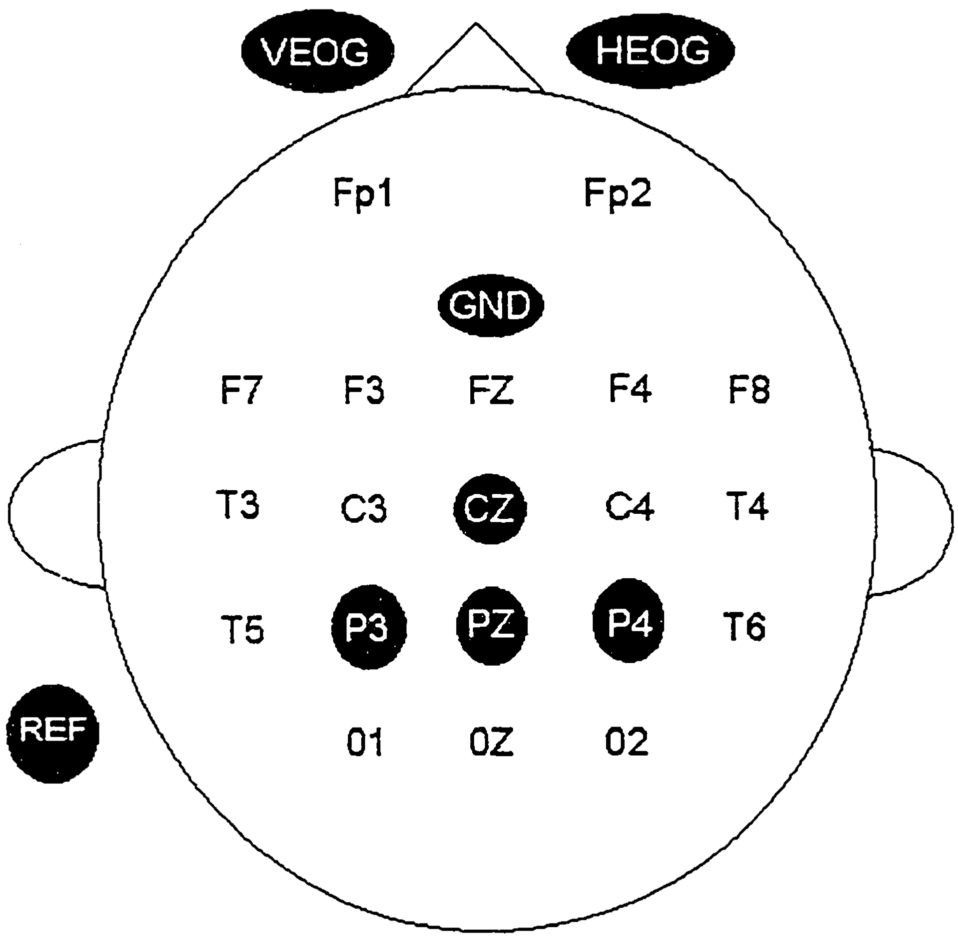

Figure 3. International 10-20 Electrode System

VI also calculated the EEG engagement index that determines the MAT Battery task mode changes. Automation task mode was switched between manual and automatic depending upon the feedback condition. The EEG index was calculated every 2 sec with a moving 20-sec window. The window was then advanced two seconds and a new average was derived. This moving window process continued for the duration of the trial. At each epoch, the index was compared to the mean value determined during a five-minute baseline period for each participant. An EEG index above baseline (see below) indicated that the participant's engagement level was high while an EEG index below baseline indicated that engagement level was low. An artifact rejection subroutine examined the amplitudes of each epoch from the four channels of digitized EEG and compared them with a preset threshold. If the voltage in any channel exceeded the threshold for more than 25% of the epoch (about two-thirds of a second) the epoch was marked as artifact and the calculated index was replaced with a value of zero. These epochs were then ignored when computing the value of the index. The data record resulting from an epoch containing an artifact was marked when it was written to the data file so that it could be ignored during later data analyses.

ERP Recording and Analyses

The NeuroScan SynAmps amplifier system was used for ERP acquisition and analyses. The software package for gathering ERPs was the Acquire386 SCAN software version 3.00. Data was acquired based upon assigned bit numbers placed in the data record from the MAT computer. The signal was gathered with 500 sweeps and points in the time domain providing an A/D rate of 500. All corrections and artifactual rejection were done off-line. The amplifier had a gain setting of 500 with a range of 11 mV and an accuracy rate of 0.168 uV/bit. The low pass filter was 30 Hz and the high pass filter was set at 1.0 Hz. EEG electrodes had an impedance of below 5 KOhms.

The continuous EEG data file was analyzed to reduced ocular artifact through VEOG and HEOG electrodes. These channels were assigned weights according to a sweep duration of 40 ms and minimum sweep criteria of 20. The continuous EEG data file then transformed into an EEG epoch file based on a setting of 500 points per data file. The epoch file was then baseline corrected in the range of -100 to 0 msec from the onset of the signal. ERPs were acquired through a sorting procedure based upon the assigned bit numbers in the data file. The signal was then further filtered with a low pass frequency of 62.5 and a low pass slope of 24 db/oct. The high pass frequency was 5.00 Hz with a high pass slope

of 24 db/oct. All filtering was performed in the time domain. All EEG was referenced to a common average and was smoothed by the SCAN software.

The criteria for ERP component classification was determined by the largest base-peak amplitude and latency within a pre-set window (Kramer, Trejo, & Humphrey, 1996): N100 (0-150 msec), N200 (150-250 msec), P100 (0-150 msec), P200 (150-250 msec), and P300 (275-750 msec).

Experimental Procedure

The participant's scalp was prepared with rubbing alcohol and electrolyte gel. A reference electrode was then affixed to the participant's left mastoid by means of electrode tape and an adhesive pad. ECI Electro-Gel conductive gel was then placed in the reference electrode with a blunt-tip hypodermic needle. Electrode gel was also placed in each of the four electrode sites (Pz, Cz, P3, P4), the ground site, and VEOG and HEOG electrodes. Using the blunt-tip hypodermic needle, the scalp was lightly abraded to reduce the impedance level at each site, relative to the ground, to less than five KOhms.

Participants were then instructed on how to perform the auditory oddball task and the compensatory tracking task. Once the participant had an understanding of these tasks, the EEG electrode cap was connected to the SynAmps headbox connector. Participants were then asked to sit quietly with

their eyes open and then with their eyes closed for five minutes each. EEG was gathered during this time to establish baseline parameters. The mean EEG value during this time represented the baseline criteria for determining task allocations during the experimental session.

After gathering baseline data, participants were given a five-minute break and, thereafter, the experimental session began. For participants in the adaptive automation group, there were two experimental trials consisting of 16 minutes of either positive or negative feedback. Participants in the yoked and control conditions also had two 16-minute trials. However, the yoked participants performed the tasks based upon the schedule of task allocations of their yoked counterparts. For the control group, the two 16-minute trials consisted of a random assignment of the same number of task allocations between manual and automatic task modes for both positive and negative feedback that participants in the adaptive automation group experienced (see above).

After each experimental trial, all participant were asked to fill out the NASA-TLX (see Appendix A). After the experimental session is completed, all participants were debriefed.

CHAPTER III

RESULTS

The data from the study were analyzed using a series of MANOVAs (multivariate analysis of variance) and ANOVAs (analysis of variance) statistical procedures. In all cases, alpha level was set at .05 and was used to determine statistical significance. The Greenhouse-Geisser procedure was used to correct psychophysiological data (Greenhouse & Geisser, 1971). Analyses of simple effects and Student Newman-Keuls (SNK) post-hoc tests were used to examine significant interaction effects.

Task Allocations

A simple ANOVA procedure was performed on the task allocation data for feedback condition for the adaptive group only. The negative feedback condition ($\underline{M}$ = 68.92) produced more task allocations than the positive feedback condition ($\underline{M}$ = 50.83), $F(1, 11) = 6.50$ (see Table 1). An ANOVA also revealed that the amounts of time participants performed the tracking task in the automatic and manual task modes was not significantly different regardless of feedback condition, $F(1, 11) = 0.97$.

Table 1

Analysis of Variance for Task Allocations

Source	*df*	*SS*	*MS*	*F*
Feedback Condition	1	1962.0416	1962.0416	6.50*

Note. *$p < .05$

Tracking Performance

A 3 (group) X 2 (feedback) ANOVA revealed significant main effects for feedback condition, F (1, 33) = 9.01; and group condition, F (2, 33) = 3.31 (see Table 2). Participants performed significantly better under the negative feedback condition (*M* = 8.91) than under the positive feedback condition (*M* = 11.14). Additionally, participants in the adaptive automation group did significantly better on the tracking task (*M* = 8.55) than those participants in the yoked condition (*M* = 11.06) or in the control condition (*M* = 10.45).

There was also a group X feedback condition interaction for tracking performance, F (2, 33) = 4.84 (see Table 2). Participants in the adaptive automation group had significantly lower tracking error when performing the task under the negative feedback condition than under any of the other group, feedback condition combinations (see Figure 4).

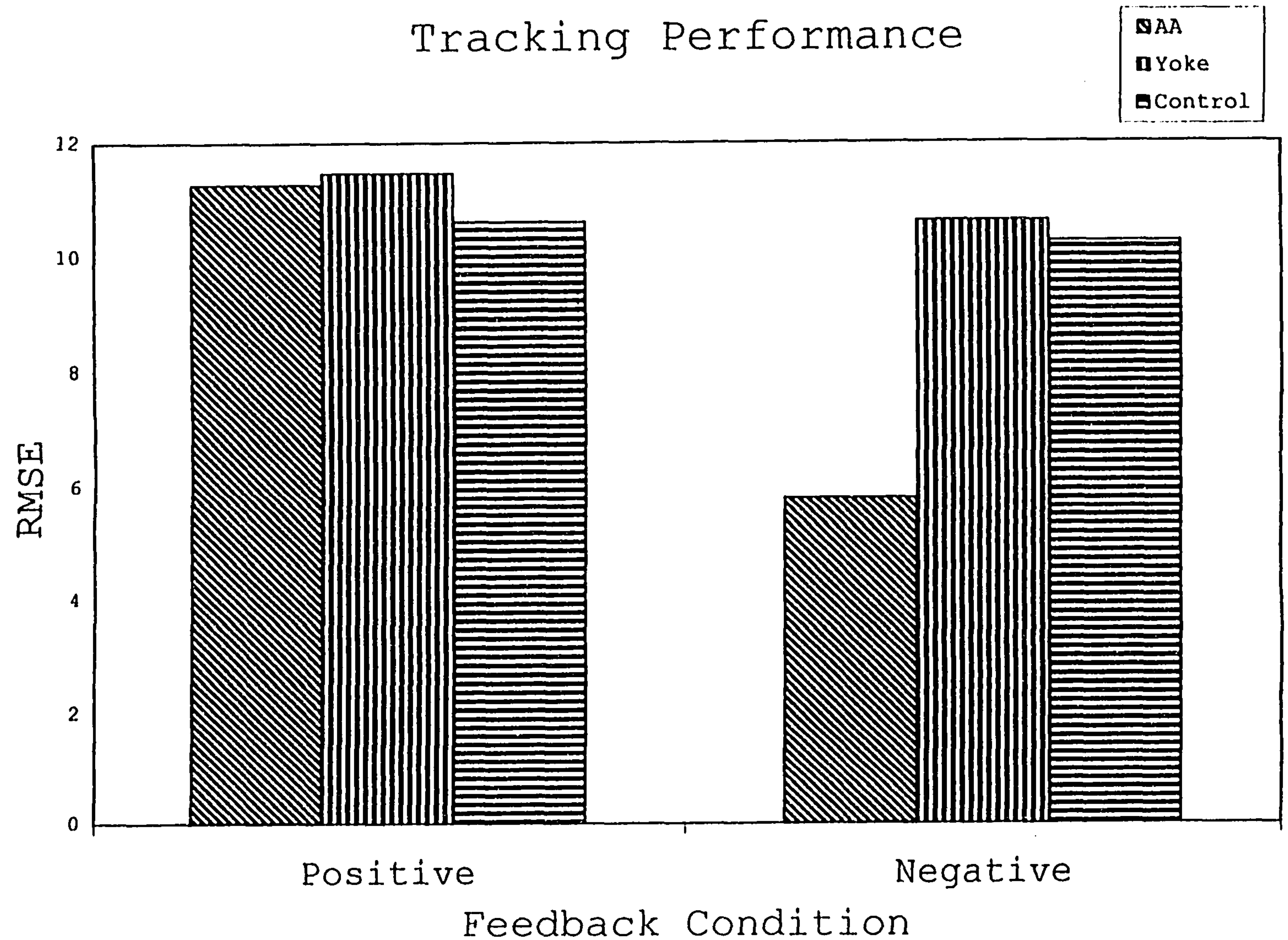

Figure 4. Mean RMSE Scores for Tracking Performance

Table 2

Analysis of Variance for Tracking Performance

Source	*df*	*SS*	*MS*	*F*
Feedback Condition	1	357.1981	357.1981	9.01*
Group Condition	2	327.9033	163.9561	3.31*
Group X Feedback	2	383.4233	191.7116	4.84*

Note. *$p < .05$

Subjective Workload

A significant main effect was found for feedback condition, $F\ (1, 11) = 39.83$ (see Table 3). Participants in the adaptive automation group rated the negative feedback condition to be lower in workload ($M = 72.50$) than the positive feedback condition ($M = 87.66$). There was also a main effect for group condition, $F\ (2, 33) = 13.76$. Those participants in the adaptive automation group reported overall workload ($M = 63.70$) to be much lower than those participants in the yoked condition ($M = 88.04$) or in the control condition ($M = 88.50$).

A group X feedback condition interaction was also found, $F\ (2, 33) = 27.67$. A simple effects analysis showed that participants in the adaptive automation group rated the negative feedback to be much lower in workload than under any of the other group, feedback condition combinations. No

other differences were found to be significant (see Figure 5).

Table 3

Analysis of Variance for Subjective Workload

Source	*df*	*SS*	*MS*	*F*
Feedback Condition	1	4140.500	4140.500	39.83*
Group Condition	2	9655.583	4827.791	13.76*
Group X Feedback	2	5752.583	2876.291	27.67*

Note. *p < .05

Auditory Oddball Task Performance

There was a significant group X feedback condition interaction for secondary task performance, F (2,33) = 4.12 (see Table 5). Participants, in the adaptive automation group, were more accurate in counting the number of high tones presented when they performed the task under the negative feedback condition (*M* = 94.32) than under the positive feedback condition (*M* = 83.29). Also, performance under the adaptive automation, negative feedback condition was significantly better than performance under the yoked group condition for positive feedback (*M* = 85.32) or

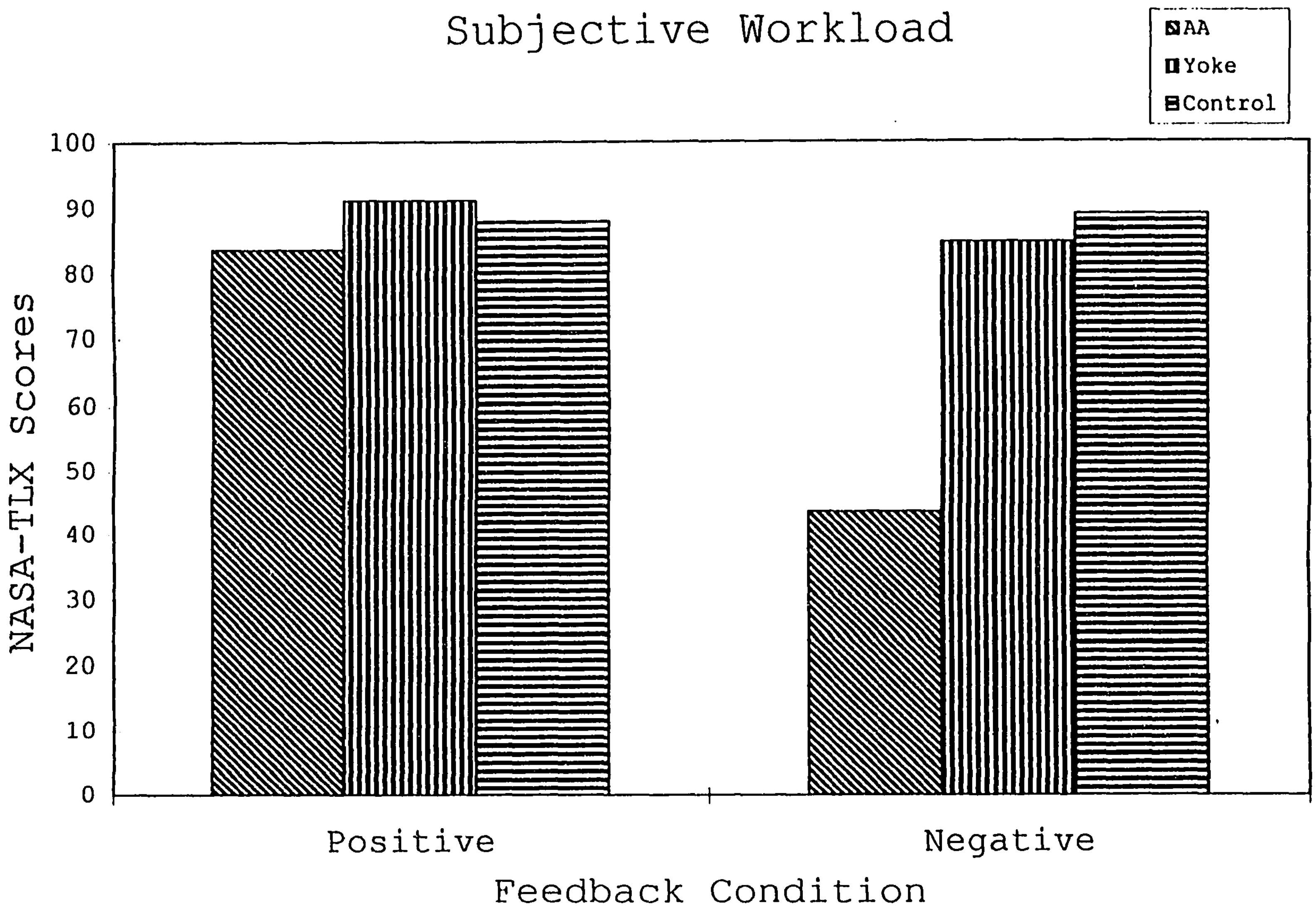

Figure 5. Mean NASA-TLX Scores

negative feedback (*M* = 87.32). Additionally, performance for participants in the control condition for positive feedback (*M* = 84.32) or negative feedback (*M* = 84.98) was significantly poorer than when performing the task under the adaptive automation, negative feedback condition. Simple effects analyses found no differences between the yoked group or control group conditions. Furthermore, performance was not significantly different between these two group conditions and the adaptive automation, positive feedback condition.

Table 4

Analysis of Variance for Secondary Task Performance

Source	*df*	*SS*	*MS*	*F*
Feedback Condition	1	309.3410	309.3410	3.94
Group Condition	2	198.3400	99.1700	2.97
Group X Feedback	2	420.3420	210.1710	5.25*

Note. $^*p < .05$

Electroencephalogram

An ANOVA on the EEG engagement index for the adaptive automation condition revealed no main effects for feedback condition, $F\ (1,11) = 0.89$; or task mode, $F\ (1,11) = 0.34$. There was, however, a significant feedback condition X task

mode interaction for the EEG engagement index, F (1, 11) = 201.32 (see Table 7). A simple effects analysis found that the EEG engagement was higher during positive feedback, manual task mode (M = 11.91) and lower during negative feedback, manual task mode (M = 8.23). Also, the EEG engagement index was larger under the negative feedback, automatic task mode (M = 11.45) than under the positive feedback, automatic task mode (M = 8.10). No differences were found between the negative feedback, automatic task mode and the positive feedback, manual task mode. Additionally, there were no differences found between the negative feedback, manual task mode and the positive feedback, automatic task mode (see Table 6).

Table 5

Analysis of Variance for EEG Engagement Index

Source	*df*	*SS*	*MS*	*F*
Feedback	1	75.3421	75.3421	0.89
Task	1	18.2532	18.2532	0.34
Feedback X Task	1	976.5401	976.5401	201.31*

Note. $*p < .05$

Table 6

Means for EEG Engagement Index

	Task Mode	
	Manual	Automatic
Negative Feedback	8.12	11.83
Positive Feedback	11.98	8.05

Event-Related Potentials

Wilk's Lambda MANOVAs were performed on the base-peak amplitude and latency data for N100, P200, and P300 ERP components for electrodes Cz, Pz, P3, and P4. There were no significant effects found across the four electrodes, F (3, 33) = 1.12. Therefore, subsequent analyses were on collapsed data across electrode sites.

Significant effects were found for feedback condition, F (6, 28) = 13.64; group condition, F (12, 56) = 6.29; and group X feedback condition, F (12, 56) = 8.31. Therefore, subsequent ANOVAs were performed on these main effects and interaction for both ERP amplitude and latency.

N100 Amplitude

There was a significant main effect found for feedback condition, F (1, 11) = 4.93. The N100 amplitude tended to be larger under the negative feedback condition (M = -4.97)

than under the positive feedback condition (*M* = -4.01). There was also a main effect found for group condition, $F(2, 33) = 17.58$. A Tukey post hoc test revealed that the amplitude was larger for those participants in the adaptive automation group (*M* = -4.49) and yoked group (*M* = -4.15) than in the control group (*M* = -3.15).

In addition to main effects, there was a group X feedback condition interaction, $F(2, 33) = 13.00$. N100 amplitude was significantly larger under the adaptive automation, negative feedback condition than under any other group X feedback conditions. Simple effects analyses revealed no other significant effects for this interaction. The group X feedback condition interaction is presented in Table 7. Figures 6-10 presents the ERP across all four electrode sites. Please note that no effects were found to be significant between negative and positive feedback for the control condition for any of the ERP components. Therefore, to facilitate graphical presentation of the ERP results, Figure 10 represents the averaged ERP waveform of these two feedback contingencies for the control conditions.

Table 7

Means for ERP Components

		N1 Amplitude	*N1 Latency*
Group	Feedback		
a	p	-5.39	136.33
a	n	-3.60	140.16
y	p	-4.94	147.66
y	n	-3.35	142.00
c	p	-3.08	139.33
c	n	-3.21	141.91
		P2 Amplitude	*P2 Latency*
Group	Feedback		
a	p	3.38	239.91
a	n	3.55	210.00
y	p	3.90	212.00
y	n	3.80	213.91
c	p	3.22	210.83
c	n	3.19	215.66
		P3 Amplitude	*P3 Latency*
Group	Feedback		
a	p	1.75	350.41
a	n	4.40	306.91
y	p	1.99	348.75
y	n	2.20	331.00
c	p	2.10	338.00
c	n	2.18	329.66

Note. a = adaptive; y = yoke; c = control; n = negative; p = positive

Table 8

Analysis of Variance for N100 Amplitude

Source	*df*	*SS*	*MS*	*F*
Feedback Condition	1	3.2200	3.2200	4.93*
Group Condition	2	23.6000	11.8000	17.58*
Group X Feedback	2	34.1733	17.0866	13.00*

Note. $^{*}p < .05$

N100 Latency

No main effects or interactions were found for feedback condition, $F\ (1,\ 11) = 0.67$; group condition, $F\ (2,\ 33) = 0.94$; or the group X feedback condition interaction, $F\ (2,\ 33) = 0.79$.

Table 9

Analysis of Variance for N100 Latency

Source	*df*	*SS*	*MS*	*F*
Feedback Condition	1	95.6805	95.6805	0.67
Group Condition	2	533.5277	266.7638	0.94
Group X Feedback	2	225.1944	112.5972	0.79

Note. $^{*}p < .05$

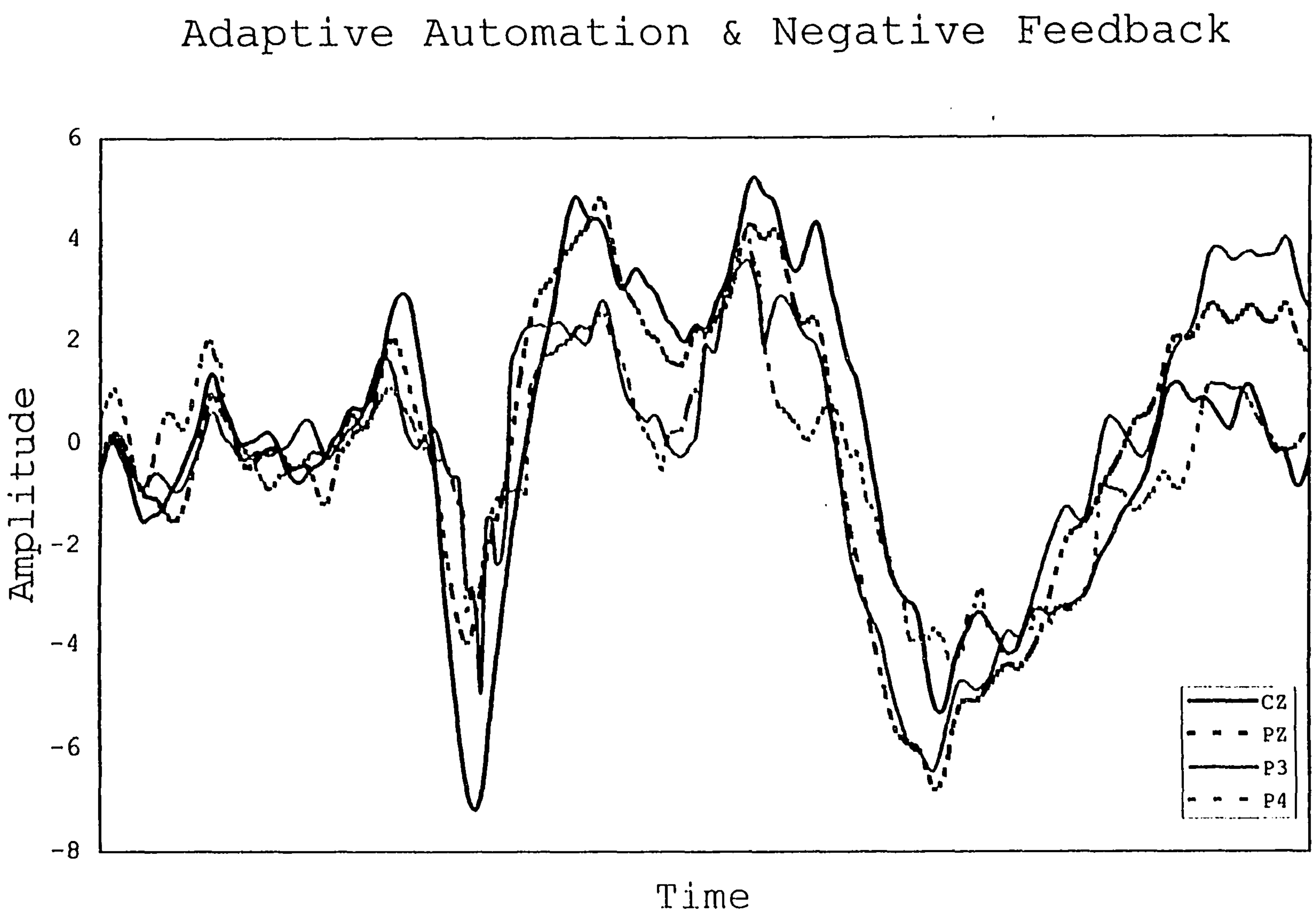

Figure 6. ERP for Adaptive Automation, Negative Feedback Group

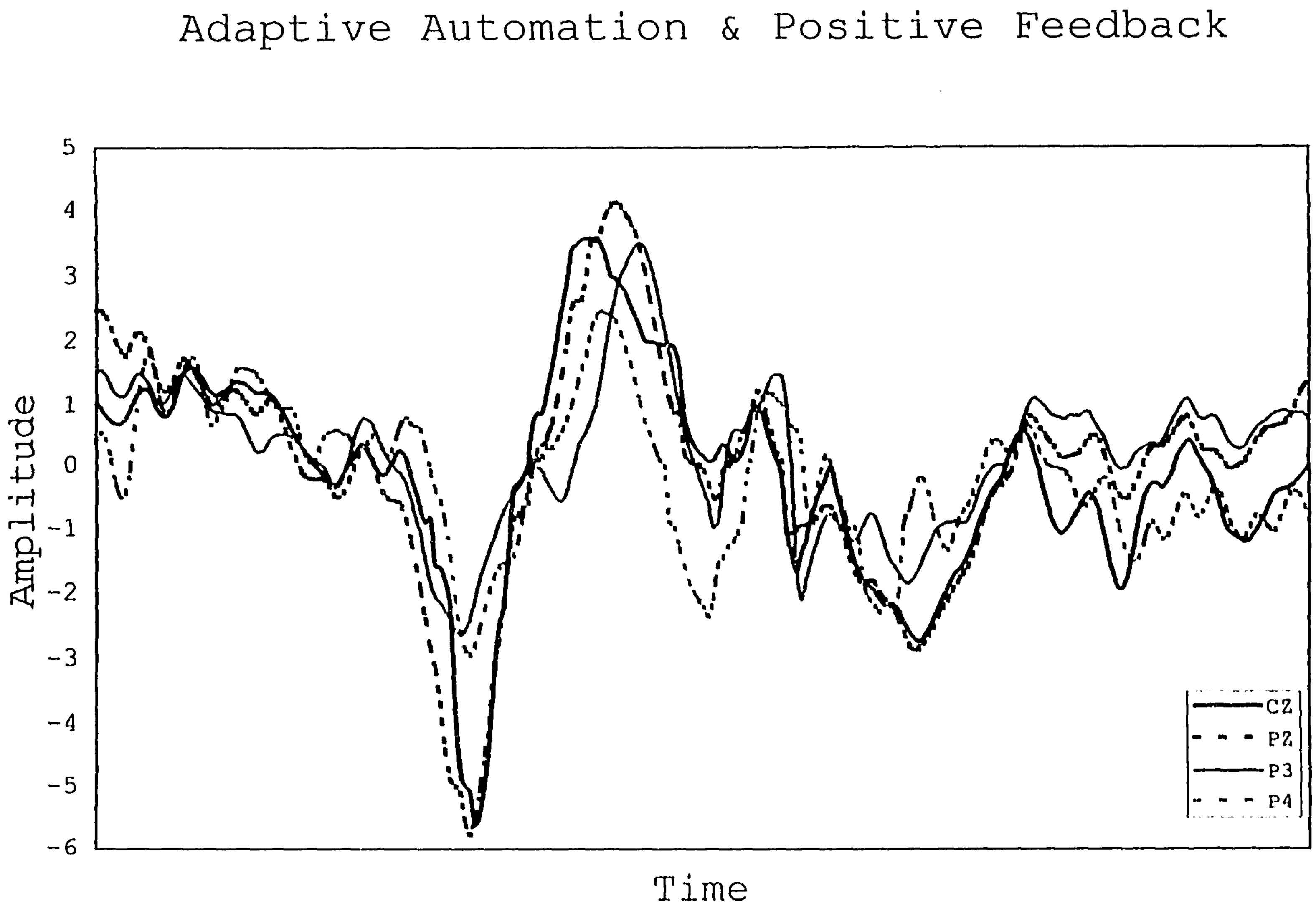

Figure 7. ERP for Adaptive Automation, Positive Feedback Group

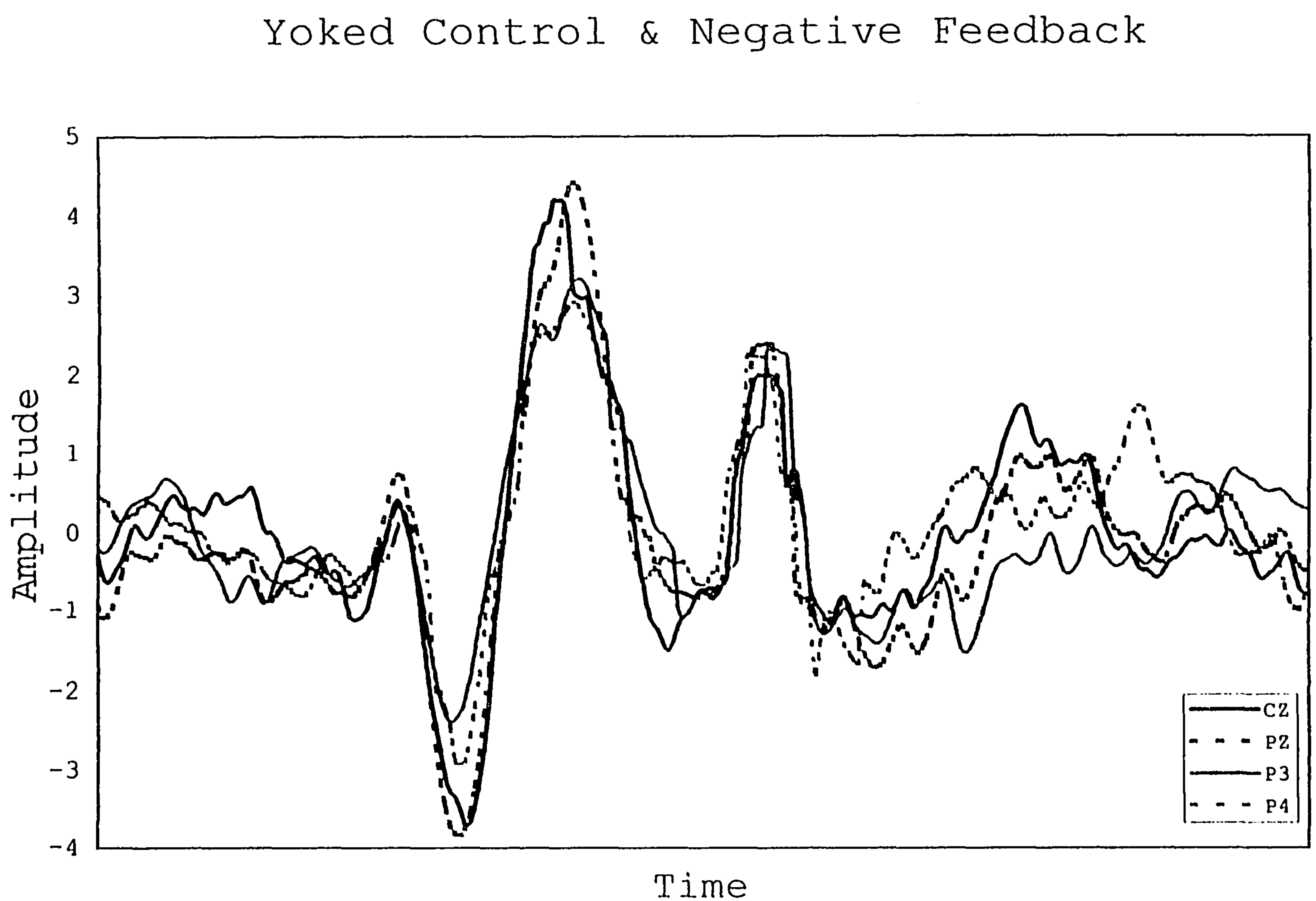

Figure 8. ERP for Yoked Control, Negative Feedback Group

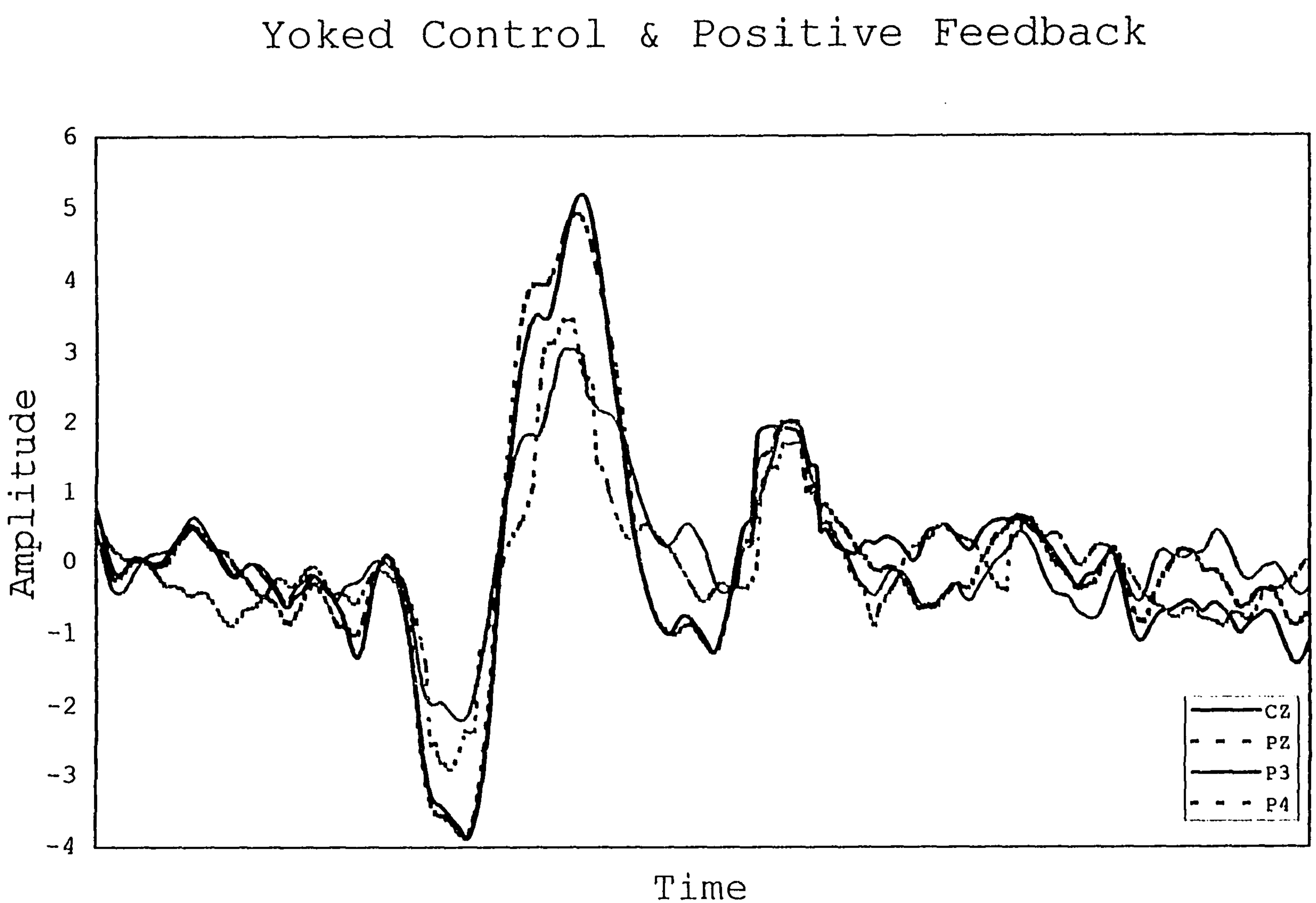

Figure 9. ERP for Yoked Control, Positive Feedback Group

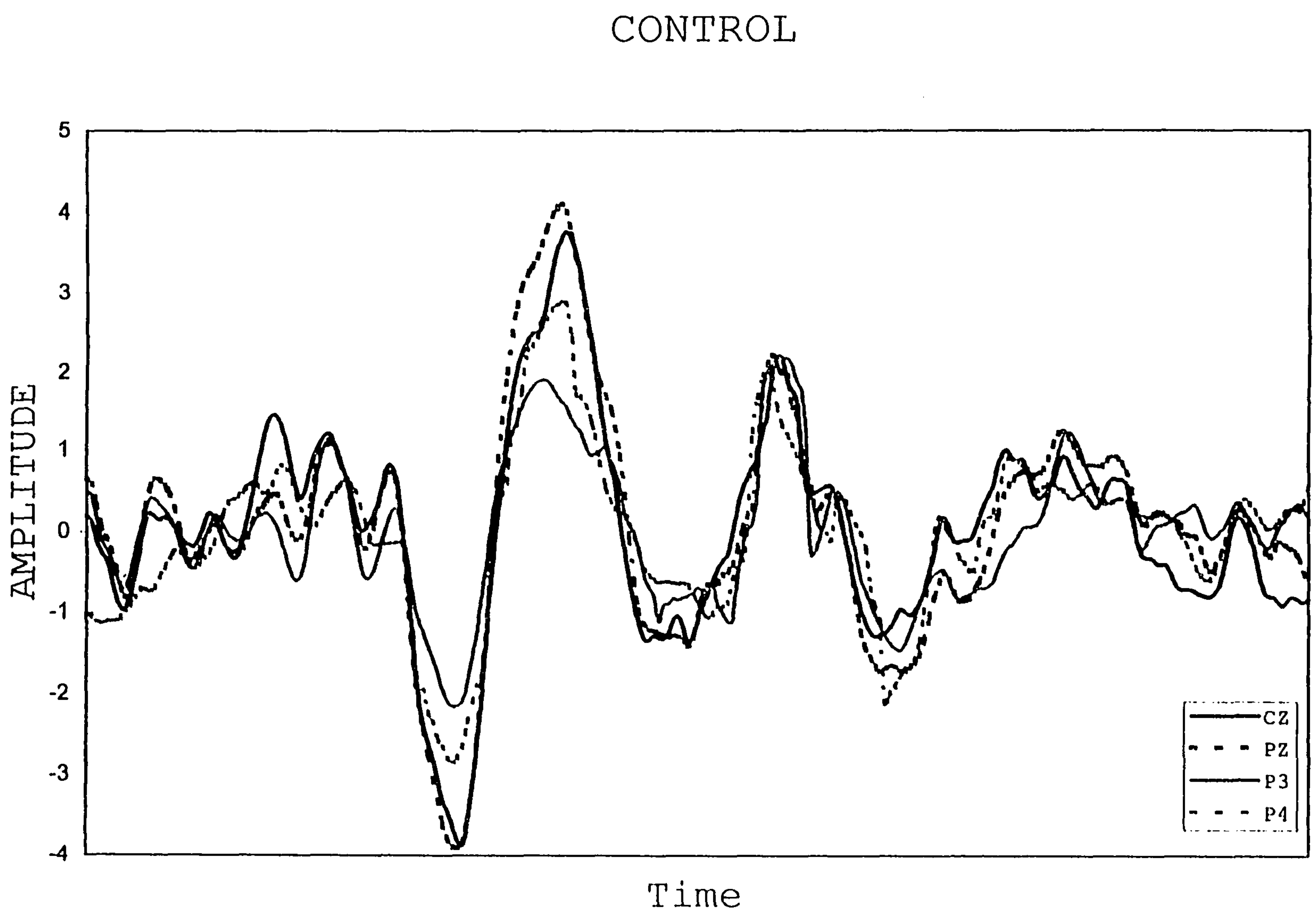

Figure 10. ERP for Combined Control Groups

P200 Amplitude

No effects were found for feedback condition, F (1, 11) = 0.01; group condition, F (2, 33) = 2.87; or the group X feedback condition interaction, F (2, 33) = 0.19.

Table 10

Analysis of Variance for P200 Amplitude

Source	*df*	*SS*	*MS*	*F*
Feedback Condition	1	0.0037	0.0037	0.01
Group Condition	2	5.0512	2.5256	2.87
Group X Feedback	2	0.2391	0.1195	0.19

Note. $*p < .05$

P200 Latency

Significant main effects were found for feedback condition, F (1, 11) = 7.40; and for group condition, F (2, 33) = 4.18. P200 latency to attended tones were longer when participants performed the auditory oddball task under the positive feedback condition (M = 220.91) than under the negative feedback condition (M = 213.19). Also, P200 latency was longer for participants in the adaptive automation group (M = 224.95) than for participants in the

yoked condition (*M* = 212.95) or in the control condition (*M* = 213.25).

The results found for P200 latency for group condition must be viewed in consideration of the group X feedback interaction, $F(2, 33) = 15.37$. A simple effects analysis shows that only the adaptive automation, positive feedback combination (*M* = 239.19) was significantly different from the other group, feedback conditions. The other group, feedback condition combinations averaged approximately 212 msec in latency. Therefore, the differences found for the main effect of group condition are due to the increased P200 latency in the positive feedback condition for participants in the adaptive automation group.

Table 11

Analysis of Variance for P200 Latency

Source	*df*	*SS*	*MS*	*F*
Feedback Condition	1	1073.3888	1073.3888	7.40*
Group Condition	2	2249.3611	1124.6805	4.18*
Group X Feedback	2	4458.8611	2229.4305	15.37*

Note. *$p < .05$

P300 Amplitude

An ANOVA yielded significant main effects for feedback condition, F (1, 11) = 78.72; and for group condition, F (2, 33) = 20.40. P300 amplitude was significantly larger when participants performed the task under the negative feedback condition (*M* = 2.93) than under the positive feedback condition (*M* = 1.94). Also, P300 amplitude was higher for those participants in the adaptive automation group (*M* = 3.08) than for those participants in the yoked condition (*M* = 2.09) or the control condition (*M* = 2.14).

There was also a feedback condition X group interaction, F (2, 33) = 57.21. P300 amplitude was significantly higher under the negative feedback condition for participants in the adaptive automation group than under any other group, feedback combination.

Table 12

Analysis of Variance for P300 Amplitude

Source	*df*	*SS*	*MS*	*F*
Feedback Condition	1	17.2872	17.2872	78.72*
Group Condition	2	14.8793	7.4396	20.42*
Group X Feedback	2	25.1251	12.5625	57.21*

Note. *$p < .05$

P300 Latency

P300 latency was found to be significant only for feedback condition, $F\ (1, 33) = 13.91$. P300 latency was significantly longer under the positive feedback condition (M = 345.72) than under the negative feedback condition (M = 322.52). Neither group condition, $F\ (2, 33) = 0.99$; or group X feedback condition interaction, $F\ (2, 33) = 2.86$ were significant.

Table 13

Analysis of Variance for P300 Latency

Source	*df*	*SS*	*MS*	*F*
Feedback Condition	1	9683.6805	9683.6805	13.91*
Group Condition	2	1510.5833	755.2916	0.99
Group X Feedback	2	3976.8611	3976.4305	2.86

Note. *$p < .05$

CHAPTER IV

FINDINGS AND INTERPRETATION

The present study was conducted to examine the efficacy of using event-related potentials and electroencephalogram for use in adaptive automation technology. Because psychophysiology is likely to be an essential aspect in the development of adaptive automation systems, it is necessary to research the issues that surround the use of these metrics. Furthermore, the present study sought to remedy a short-coming in the literature concerning the impact that adaptive automation has on behavioral, subjective, and psychophysiological measures of workload and task engagement.

To accomplish these research goals, a multi-group design was used composed of adaptive automation, yoked, and control group conditions. Participants in the adaptive automation group were asked to perform a compensatory tracking task and an auditory oddball task while their EEG was continuously monitored. The tracking task was switched between manual and automatic task modes based upon whether their EEG was above or below baseline levels of task engagement and which feedback condition the system operated under. The automation schedule for each participant in the adaptive automation group was presented to a participant in the yoked condition. Therefore, each participant performed

the tasks in the exact cycle sequence as their yoked counterpart. Additionally, a control group was employed that received a random assignment of task mode allocations.

The design was intended to enable the assessment of whether the adaptive automation method of task mode allocation represents a significantly better way of keeping operators "in-the-loop." If so, performance, subjective workload estimates, and psychophysiological correlates of workload would be better moderated for participants in the adaptive automation group, and no differences witnessed between the yoked or control group conditions. However, if adaptive automation does not significantly enhance the human-automation interaction, then no differences would be expected between the three experimental groups. Additionally, the design allowed for a determination to be made as to the utility of using EEG and ERPs in adaptive task allocation.

Task Allocations

If there was a functional relationship between the EEG engagement index and task mode, the index should demonstrate stable short-cycle oscillation under negative feedback and longer and more variable periods of oscillation under positive feedback. The strength of the relationship would be reflected in the degree of contrast between the behavior of the index under the two feedback contingencies. This

should be reflected in significantly more task allocations under the negative feedback condition than under the positive feedback condition. The results showed that indeed the system made more switches between manual and automatic task modes in the negative feedback condition than in the positive feedback condition. Therefore, the system demonstrated expected feedback control behavior under these two feedback contingencies and supports Pope, Bogart, and Bartolome's (1995) finding that the 20 beta/(alpha+theta) EEG engagement index possesses utility as part of an adaptive algorithm for controlling automation task allocation.

Performance and Subjective Workload

A number of researchers have found that manual reallocation can serve as a countermeasure to performance decrements that often accompany the use of automation. For example, researchers have found that short periods of return-to-manual control reduced ommissions and lowered workload ratings (Hadley et al., 1998; Parasuraman, Molloy, & Singh, 1994; Scallen, Hancock, & Duley, 1995). Additionally, increasing the number of manual reallocations resulted in even better performance and lower subjective workload. Therefore, because the negative feedback condition produces more task allocations, the increase in manual reallocations was predicted to result in

significantly better performance and modulated workload than in the positive feedback condition.

It was found that participants performed the tracking task and auditory oddball task significantly better under the negative feedback condition than under the positive feedback condition. Also, subjective workload ratings were found to be significantly lower under the negative feedback condition. Furthermore, an analysis revealed that, although there were more task allocations under the negative feedback condition, participants spent the same amount of time in the automatic (M = 7.45 min) and manual task (M = 8.15 min) modes. Therefore, these results can not be attributed to an inequality in task mode duration between the two feedback contingencies.

The present study was also designed to determine how behavioral and subjective measures are moderated by the use of adaptive automation methods. The rationale behind adaptive automation is that a balance is made between task load and levels of automation. That is, an assessment is made of operator state and changes in task mode are made in response to high or low workload levels. The changes are made in real-time and should produce better performance and lowered workload ratings because of the regulation of workload and maintenance of operator engagement (Hancock & Chignell, 1989; Scerbo, 1996). Therefore, participants in

the adaptive automation group should do significantly better and rate subjective workload lower under the negative feedback condition than under any other group and feedback condition combination. However, if the benefits found with adaptive automation are due solely to an increase in manual reallocations, there should be no differences between the three group conditions in terms of performance or subjective workload ratings because all groups experienced the same number of manual reallocations.

The group X feedback condition interaction for performance and workload ratings support the contention that the benefits found with adaptive automation are not due solely to increased manual reallocations. Participants in the adaptive automation group did significantly better and had lower subjective workload ratings while performing the tasks under the negative feedback condition compared to any other group, feedback condition. Although across all three group conditions participants had lower performance errors and workload ratings in the negative feedback condition, the finding is tempered by the overwhelming results for the adaptive automation group, negative feedback condition for both performance and subjective workload. Therefore, these results support the logic of adaptive algorithms for dynamic task allocation based upon psychophysiological indices with demonstrated behavioral and subjective workload outcomes.

Implications for Adaptive Automation

Perhaps, the most fundamental reason for introducing automation is to lessen the workload demands placed on human operators who must interact with often complex systems. Although the evidence to support this assertion has not always been found (e.g., Riley, 1994), those who use such systems often cite excessive workload as a factor in their choice of automation. For example, Riley, Lyall, and Wiener (1993) reported that urgency of the situation and workload were the two most important factors in pilots' choice to use automated functions, such as autopilot, Flight Management System (FMS), and flight director. Furthermore, Wiener (1988) noted that automated systems tend be "clumsy" in that the automation requires interaction at times when workload is already high; the effect of which is to further increase workload demands. Therefore, it is of high importance to assess how any form of automation task allocation, including adaptive automation, impacts task load and subjective impressions of workload.

Adaptive automation has been suggested as a remedy to the "out-of-the-loop" problems that often are associated with human-automation interaction. Some of these problems include increased performance errors and cognitive workload (Parasuraman, & Riley, 1997). However, few empirical studies are available that have demonstrated that adaptive

task allocation does indeed improve performance and lower workload.

Adaptive aiding has been found to improve performance and workload in studies of aerial search (Morris & Rouse, 1986), flight management (Hilburn, Parasuraman, & Mouloua, 1995; Parasuraman, 1993), monitoring (Parasuraman, Mouloua, & Molloy, 1996), and air traffic control (Hilburn, Jorna, & Parasuraman, 1995). However, none of these studies changed levels of automation based upon real-time measures of workload. The research here used psychophysiological indices and made task allocations in real time based upon whether the EEG characterized low or high task engagement and workload. Therefore, the present study provides support for our previous studies in demonstrating that adaptive task allocation using a real-time approach improves performance and lowers workload demands. Future research, however, is needed to determine whether these effects are transferable to other areas of human and system performance (e.g., monitoring performance).

Electroencephalogram

Byrne and Parasuraman (1996) stated that the use of any candidate psychophysiological metric must be predicated on how well it aids the development of adaptive automation. Although numerous psychophysiological measures are available for use in adaptive automation, only the EEG has been found

to be useful as a measure of operator state under both low task engagement and high task engagement (Kramer, 1991). Therefore, the present study sought to examine the use of the EEG (i.e., EEG engagement index) as an adaptive mechanism for task allocation.

Generally, research has shown that with increases in task engagement, theta is suppressed and alpha is blocked while beta increases in relative power. As task engagement decreases, the EEG decreases in beta and shows concomitant increases in both theta and alpha (Kramer, 1991). Therefore, such EEG characteristics allowed for predictions to be made based upon whether the EEG engagement index operated under positive or negative feedback control.

Positive feedback mechanisms react to "disturbances" in a system, in this case high or low engagement states, by amplifying the magnitude of the effect (Smith & Smith, 1987). When EEG patterns were below baseline levels of engagement, the system was designed to automate the tracking task which should further lower the engagement state. However, when the EEG patterns were above baseline levels of engagement characterized by high beta, alpha blocking, and theta suppression, the system allocated the tracking task to the manual task mode. Therefore, for positive feedback, the EEG engagement index should be lower under the automatic task mode and higher under the manual task mode.

Negative feedback should contrast that of positive feedback control behavior. The reason is that this feedback contingency takes corrective action to keep system behavior within operational limits (Smith & Smith, 1987). To accomplish this, the biocybernetic system, under negative feedback control, automated the tracking task when the EEG engagement index was above baseline levels of engagement and allocated manual control when the EEG engagement index was below baseline levels of engagement. The EEG engagement index should, therefore, be higher under the automatic task mode and lower under the manual task mode.

The feedback condition X task mode interaction confirmed that the EEG demonstrated these characteristics. The value of the EEG engagement index was contingent upon which feedback condition and task mode the system was operating under. Under the manual task mode, the EEG engagement index was higher for positive feedback and lower for negative feedback. Conversely, the index was higher under the automatic task mode for negative feedback, and it was significantly lower under the automatic task mode for positive feedback.

Implications for Adaptive Automation

These results support other studies that have demonstrated the efficacy of EEG for the modulation of mental state in a closed-loop environment. For example,

Schwilden, Stoeckel, and Schuttler (1989) have developed medical models for the closed-loop regulation of anesthetic state using EEG metrics. Such findings are important as Byrne and Parasuraman (1996) noted that assessment of candidate psychophysiological measures for adaptive automation requires iterative closed-loop testing.

Another implication of these results concerns the emphasis in adaptive automation that has been placed on the prevention of task overload. However, a more beneficial application of adaptive automation may be the prevention of task underload in which psychophysiological measures will play a key role (Byrne & Parasuraman, 1996). The present study demonstrated that the EEG was capable of discriminating between different levels of task load and, therefore, suggests its efficacy as an adaptive mechanism for adaptive automation. Although the negative consequences of task underload have not always been appreciated (e.g., Redondo & Del Valle-Inclan, 1992), because of the uniqueness of the EEG as a measure of task underload, the use of this psychophysiological metric should continue to find application in the development and use of adaptive task allocation (Byrne & Parasuraman, 1996; Kramer, 1991).

Event-Related Potentials

A number of researchers (Billings, 1997; Sheridan, 1997; Wickens, 1992; Wiener & Nagel, 1988) have noted that

automation has changed the nature rather than reduced the workload demands placed on human operators. For example, pilots now focus on monitoring system controls and intervene only to detect, assess, and correct system failures. An important by-product of this role shift is the decreased ability to infer operator state because of limited interaction with the automated system. The use of advanced automation concepts, such as adaptive automation, would only increase such role transfer prompting the need for more diagnostic measures for the regulation of mental workload and other psychological constructs.

Byrne and Parasuraman (1996) discussed the role that various psychophysiological measures can play in the development of adaptive automation technology. They stated that ERPs possess a number of characteristics that make them ideal as candidate indices for adaptive task allocation. These include diagnostic specificity, sensitivity, and reliability (see Eggemeier, 1988). However, Parasuraman (Byrne & Parasuraman, 1996; Parasuraman, 1990) concluded that, although many proposals have been made concerning the use of ERPs in adaptive automation, little empirical evidence has been collected to support its efficacy.

The present study sought to address this limitation and assess whether ERPs can be used to make task allocations in an adaptive fashion. Specifically, it was designed to

examine whether the ERP can discriminate between positive and negative feedback conditions. Furthermore, the study sought to determine whether differences were evident between the adaptive automation, yoked, and control group conditions in terms of ERP component waveforms. Finally, because any approach to adaptive automation requires multiple measures of operator state, another goal was to measure the degree of congruence that ERPs have with other workload metrics.

The ERP waveform components to the infrequent, high tones demonstrated significant differences in amplitude and latency between positive and negative feedback conditions. N100 and the P300 ERP components were significantly higher in amplitude under the negative feedback condition than under the positive feedback condition. Additionally, the P300 component was significantly shorter in latency under the negative feedback condition. These results support the findings for performance and subjective workload and demonstrate that the ERP was capable of discriminating between levels of task load in an adaptive environment. Therefore, they support other studies that have found that ERPs can be useful in the development and application of adaptive automation technology (Kramer, 1991; Humphrey & Kramer, 1994; Trejo, Humphrey, & Kramer, 1996).

There was also an experimental group X feedback condition interaction for N100 and P300 amplitude. The

adaptive automation, negative feedback condition produced P3s that were significantly larger in amplitude than any other group, feedback condition. The N100 was also found to be significantly higher in amplitude under the adaptive automation, negative feedback condition. There were no differences found between the yoked and control group conditions. Additionally, positive feedback for the adaptive automation group did not produce ERP waveforms that were significantly different from the yoked or control group conditions in either amplitude or latency measures.

Implications for Adaptive Automation

Mental Models. These findings for the ERP are important for two reasons. First, the P300 is thought to index a context updating of our mental model of the environment (Donchin, Ritter, & McCallum, 1978). Donchin, McCarthy, Kutas, and Ritter (1983) stated that the P300 is a representation of neural action for updating the user's "mental model" that seems to underlie the ability of the nervous system to control behavior. The mental model then is an assessment of deviations from expected inputs and is, therefore, revised whenever discrepancies are found. The frequency of such revisions is dependent upon the "surprise value" and task relevance of the attended stimuli (e.g., high tones). Donchin (1981) noted that ERP components are associated with specific information processing functions,

and the P300 "subroutine" is activated whenever there exists a need to evaluate unusual, task-relevant events (Gopher & Donchin, 1986; Kramer, 1991). Therefore, the group X feedback condition interaction for P300 amplitude suggests that participants in the adaptive automation group may have been better able to predict the "state" of system operation, develop control strategies, select appropriate actions, and interpret the effects of selected actions (Gentner & Stevens, 1983; Johnson-Laird, 1983; Wickens, 1992; Wilson & Rutherford, 1989). The outcomes of such an improved mental model were improved performance and lowered workload and evidenced by larger amplitudes for the P300 ERP component.

The recent interest in mental models is due to changing technology and there is a growing need for metaphors to describe the increasingly "black box" nature of systems (Howell, 1990; Wickens, 1992; Wilson & Rutherford, 1989). It is commonly accepted that people form mental models of tasks and systems, and that these models are used to guide behavior at the interface. Norman (1983) explains that people form internal, mental models of themselves and of the things with which they are interacting with. These extent to which the mental models provide a good fit determines whether users can understand the nature of this interaction. Therefore, automated processes must be made compatible with the users' internal representation of the system (Kantowitz

& Campbell, 1996; Norman, 1983; Parasuraman & Riley, 1997; Scerbo, 1996).

The National Research Council (1982) further noted that the effectiveness of automation depends on matching the designs of automated systems to user's representations of the tasks they perform. The lack of a "match" between the operating characteristics of a system, the user's mental model of the system, and designer's conceptual model of the system can lead to increased errors, workload, response times, and so forth. As Kantowitz and Campbell (1996) suggest, automated design should provide timely, consistent, and accurate feedback, match task demands to environmental demands, design high stimulus-response compatibility, and develop appropriate operator training that facilitates the development of an accurate mental model.

The use of the mental model metaphor then is likely to be of continued service in the design of automated systems. Moreover, the development of advanced automation concepts should only increase the need for accessing the "black box" of the human operator. The need arises, therefore, for ways of measuring the degree of disparity between a user's mental model and the designer's conceptual model. The present results suggest that such can be supplied by the use of ERP measures although additional research would be needed to specify the nature of the ERP, its relation to user mental

models, and how it could be used in adaptive automation design.

Resource Allocation. Another implication of these results concerns how the ERP relates to cognitive workload. As stated previously, the P300 is thought to represent the context updating of our mental model whenever a novel event occurs. Such an updating only occurs if the stimuli associated with a task requires that it be processed; that is, task-irrelevant stimuli that are ignored do not elicit a P300. However, consider the situation in which a participant is instructed to only partially ignore a stimulus, or a participant is asked to perform an oddball task while concurrently performing a tracking task as in the present study. Will the P300 measures reflect these graded changes in task difficulty? If so, then the P300 may serve as an index of the resource demands and, therefore, the cognitive workload imposed on the human operator (Gopher & Donchin, 1986; Kramer, 1987).

Research has consistently demonstrated that the P300 amplitude reflects the amount of expenditure of perceptual/central processing resources associated with performing a task(s) (Gopher & Donchin, 1986; Kramer, 1991; Parasuraman, 1990). The characteristics of the P300 exhibit a decrease in amplitude and an increase in latency to secondary task performance as the difficulty of the primary

task is increased ("amplitude reciprocity hypothesis"; Isreal et al., 1977). The results of this study revealed that the P300 did indeed decrease in amplitude and increase in latency as the workload demands in the task increased. Furthermore, the group X feedback condition interaction for P300 supports the findings for performance and subjective workload and demonstrated that the use of adaptive task allocation reduced the workload for those participants performing the tasks in the negative feedback condition. In addition, the N100 and P200 waveforms further support these results because they are thought to represent the early processes of selective attention and resource allocation (Hackley, Woldoroff, & Hillyard, 1990; Hillyard, Hink, Schwent, & Picton, 1973).

Parasuraman, Bahri, Deaton, Morrison, and Barnes (1992) argued that adaptive automation represents the coupling of levels of automation to levels of operator workload. Therefore, candidate indices which serve as adaptive mechanisms must be capable of discriminating between various levels of task load. Although a number of measures have been proposed, Morrison and Gluckman (1994) suggested the use of psychophysiological metrics because of their potential to yield real-time estimates of mental state with little or no impact on operator performance.

There are many psychophysiological measures available to system designers seeking to use them in adaptive automation design. Such measures include heart rate, heart-rate variability, EEG, EDA, pupillometry, ERP, and others. However, because of the multidimensional nature of mental workload and other psychological constructs (e.g., memory, attention, language processes) that require attention in the design of automated systems, only the ERP has been found to be sensitive to these different information processing activities (Kramer, 1991; Kramer, Trejo, & Humphrey, 1996).

Although the biocybernetic system did not predicate task allocation on the basis of ERP data, the results showed that the ERP was capable of discriminating between levels of taskload in an adaptive environment. Therefore, a next step would require the development of an adaptive algorithm that uses the components of the ERP waveform as an adaptive mechanism for allocating tasks between the operator and automated system. The research by Humphrey and Kramer (1994) as well as the present results demonstrates that such a biopsychometric system is capable of development. Despite the fact that such a system may be years from fruition, at the very least these results demonstrate that the ERP can serve in the developmental role (see Byrne & Parasuraman, 1996) of adaptive automation design. Taken together, then, the results of the ERP data support the conclusion of many

human factors professionals that ERPs possess the adaptive capabilities for determining optimal human-automation interaction (Byrne & Parasuraman, 1996; Defayolle et al., 1971; Donchin, 1980; Farwell & Donchin, 1988; Gomer, 1981; Kramer & Humphrey, 1994; Kramer, Humphrey, Sirevaag, & Mecklinger, 1989; Kramer, Trejo, & Humphrey, 1996; Sem-Jacobsen, 1981; Scerbo, 1996).

Conclusions

The field of human factors has been traditionally defined as the design and evaluation of systems and tools for human use. The goal of human factors is directed at how people, machines, and the environment interact, and what can be done to make certain that productivity, efficiency, and safety are ensured. The idea that one should account for the human during the design process often seems too obvious to deserve much attention. Recently, however, several known disasters, such as Three Mile Island, Challenger space shuttle, and Ralph Nader's consumer product crusades, have challenged such prevailing attitudes towards human factors research. The idea has certainly relevant for the use of automation especially in light of several disastrous accidents that have happened in the past few years in aviation transportation (e.g., Bangalore, India, 2/14/1990; Charlotte, North Carolina, 1994; Nagoya, Japan, 4/26/1994; Roselawn, Indiana, 10/31/1994). The concern is very

relevant for adaptive automation when one considers that aid-initiated adaptation was a factor in the Charlotte wind shear accident (1994).

Scerbo (1996) noted that automation is neither inherently good nor bad. He stated that automation does, however, change the nature of work; it solves some problems while it creates others. Adaptive automation represents the next phase in the development of automated systems. To date, it is not known how this type of technology will impact work performance (Billings, 1997; Scerbo, 1996; Woods, 1996). However, it is clear that automation will continue to impact our lives requiring humans to co-evolve with the technology; this is what Hancock (1996) calls "techneology." Therefore, professionals involved with adaptive automation are incumbent to investigate the issues surrounding the use of adaptive automation technology. As Weiner and Curry (1980) conclude:

> The rapid pace of automation is outstripping one's ability to comprehend all the implications for crew performance. It is unrealistic to call for a halt to cockpit automation until the manifestations are completely understood. We do, however, call for those designing, analyzing, and installing automatic systems in the cockpit to do so carefully; to recognize the behavioral effects

> of automation; to avail themselves of present and future guidelines; and to be watchful for symptoms that might appear in training and operational settings (p.7)

The concerns they raised are as valid today as they were 18 years ago. Fortunately, at present, adaptive automation represents only a conceptual view of how automation can be advanced to improve the human-automation interaction. We now have an opportunity to research the technology before large-scale implementation of adaptive automation becomes available (Scerbo, 1996).

There are a number of issues that must be addressed before adaptive automation can move forward in the design of automated systems. To do otherwise, would be to risk repeating the fatal lessons of the past. As Billings and Woods (1994) noted,

> In high-risk, dynamic environments... technology-centered automation has tended to decrease human involvement in system tasks, and has thus impaired human situational awareness; both are unwanted consequences of today's system designs, but both are dangerous in high-risk systems. [At it's present state of development,] adaptive ("self-adapting") automation represents a potentially serious threat... to the authority that the human

> pilot must have to fulfill his or her responsibility for flight safety (p. 265).

Such a strong cautionary voice points to the need for more research in this area. The present study examined but a small share of these issues. These issues included the use of psychophysiological measures in adaptive automation design as well as a comparison of adaptive task allocation to static task allocation.

Byrne and Parasuraman (1996) stated that psychophysiology is an integral component of adaptive automation as a non-invasive method used to assess operator state. They suggested that such measures could be used not only as an input signal for the regulation of automation, but also to assess underlying changes accompanying performance changes during development of adaptive automation systems. The results support such a conclusion. The ERP and EEG were found to discriminate between positive and negative feedback controls and these were associated with other workload measures. Byrne and Parasuraman noted that any psychophysiological measure must be used in conjunction with other metrics of operator state and any candidate indices must be capable of such an association. Indeed, the EEG and ERP measures accorded well with the performance and subjective workload measures and, therefore, support Byrne and Parasuraman's assessment that

biopsychometrics will play an important role in advanced automation.

Furthermore, this study represents on of the first experiments to demonstrate conclusively the advantages of the adaptive automation paradigm using a real-time approach. Parasuraman, Mouloua, & Molloy (1996) also examined the effects of adaptive task allocation, but they used model-based and performance-based approaches. These adaptive methods do not represent an adaptive aiding mechanism based on real-time measurements of operator workload. Furthermore, these researchers used only performance measures (i.e., reaction time, false alarms, hit rate, omissions). Kramer, Trejo, and Humphrey (1996) also examined the use of adaptive automation and provided both performance and psychophysiological measures. However, their study was a de facto assessment of how much ERP data is needed to discriminate different levels of mental workload and, therefore, was not adaptive automation in the truest sense. Therefore, the present study provides one of the first controlled, empirical studies to evaluate the conjunctive effects of adaptive task allocation on behavioral, subjective, and psychophysiological correlates of workload.

Future Directions. Although the findings presented here give strong support for the benefits of adaptive automation and the use of psychophysiology in the design of

this technology, the study only examined some of the many issues that need consideration. Parasuraman and his colleagues (Byrne & Parasuraman, 1996; Parasuraman, 1993; Parasuraman, Bahri, & Molloy, 1991; Parasuraman et al., 1992; Parasuraman, Mustapha, & Molloy, 1996) have noted a number of variables and factors that should be researched in adaptive automation design. These include the frequency of adaptive changes, adaptive algorithms, automation reliability and consistency, the type of interface, and contextual factors that are unique to specific systems. Scerbo (1996) also added system responsiveness, timing, and authority and invocation to this list. He further stated that research should branch out to other areas that are likely to be of concern for adaptive automation technology, such as mental models, teams, training, and communication. Moreover, if one considers the concerns of Woods (1996) that automation represents what he calls, "apparent simplicity, real complexity," one cannot leave without an impression that there is a considerable amount of work that is needed. However, research must begin somewhere and our work here and the works of others in the field are hoped to stimulate additional research in this new but exciting area of automation technology.

REFERENCES

Alluisi, E.A., Coates, G.D., & Morgan, B.B., Jr. (1977). Effects of temporal stressors on vigilance and information processing. In R.R. Mackie (Ed.), *Vigilance: Theory, operational performance, and physiological correlates* (pp. 361-421). NY: Plenum.

Andreassi, J.L. (1995). *Psychophysiology: Human behavior & physiological response* (3rd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Backs, R.W., Ryan, A.M., & Wilson, G.F. (1994). Psychophysiological measures of workload during continuous manual performance. *Human Factors, 36,* 514-531.

Barrett, G., Neshige, R., & Shibasaki, H. (1987). Human auditory and somatosensory event-related potentials. *Electroencephalography and Clinical Neurophysiology, 66,* 409-419.

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

Beatty, J. (1986). The pupillary system. In M.Coles, E. Donchin, & S. Porges (Eds.), *Psychophysiology: Systems, processes, and applications* (pp. 43-50). New York: Guilford.

Beatty, J. (1988). Pupillometric signs of selective attention. In G. Galbraith, M. Kietzman, & E. Donchin (Eds.), *Neurophysiology and Psychophysiology* (pp. 85-110). Hillsdale, NJ: Lawrence Erlbaum, Associates.

Beatty, J., Greenberg, A., Diebler, W.P., & O'Hanlon, J.F. (1974). Operant control of occipital theta rhythm affects performance in a radar monitoring task. *Science, 183,* 871-873.

Beatty, J., & O'Hanlon, J.F. (1979). Operant control of posterior theta rhythms and vigilance performance: Repeated treatments and transfer of training. In N. Birbaumer and H.D. Kimmel (Eds.), *Progress in operant conditioning of physiological events* (pp. 235-246). Hillsdale, NJ: Lawrence Erlbaum Associates.

Beck, E.C. (1975). Electrophysiology and behavior. *Annual Review of Psychology, 26,* 233-262.

Biferno, M.A. (1985). *Mental workload measurement: Event-related potentials and ratings of workload and fatigue.* (NASA-11860). Ames, CA: NASA.

Billings, C.E. (1991). *Human-centered aircraft automation philosophy: A concept and guidelines.* (NASA-103885). Moffett Field, CA: NASA.

Billings, C.E. (1997). *Automation aviation.* Mahwah, NJ: Lawrence Erlbaum Associates.

Billings, C.E., & Woods, D.D. (1994). Concerns about adaptive automation in aviation systems. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 264-269). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

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

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

Comstock, J.R., & Arnegard, R.J. (1991). *The multi-attribute task battery for human operator workload and strategic behavior research* (NASA Technical Memorandum No. 104174). Hampton, VA: NASA.

Davidson, R.J. (1988). EEG measures of cerebral asymmetry: Conceptual and methodological issues. *International Journal of Neuroscience, 39,* 71-89.

Davidson, F.J., Chapman, J.P., Chapman, L.J., & Henriques, J.B. (1990). Asymmetrical brain electrical activity discriminates between psychometrically matched verbal and spatial cognitive tasks. *Psychophysiology, 27,* 528-543.

Deecke, L. (1976). Potential changes associated with motor action, reflex responses, and readiness. In W.C. McCallum & J.R. Knott (Eds.), *The responsive brain* (pp. 91-93). Bristol, England: Wright.

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* (pp. 141-152). London: Taylor & Francis.

De Weerd, J.P.C., & Kap, J.I. (1981). A posteriori time-varying filtering of averaged evoked potentials II: Mathematical and computational aspects. *Biological Cybernetics, 41,* 223-234.

Donchin, E. (1979). Event-related brain potentials: A tool in the study of human information processing. In H. Begleiter (Ed.), *Evoked potentials and behavior* (pp. 13-75). New York: Plenum.

Donchin, E. (1980). Event-related potentials -- Inferring cognitive activity in operational settings. In F.E. Gomer (Ed.), *Biocybernetic applications for military systems.* (McDonnell Douglas Technical Report MDC EB1911, pp. 35-42). Long Beach, CA: McDonnell Douglas.

Donchin, E. (1981). Suprise!...suprise? *Psychophysiology, 18,* 493-513.

Donchin, E., & Heffley, E. (1978). Multivariate analysis of event-related potential data: A tuturorial review. In D. Otto (Ed.), *Multidisciplinary perspectives in event-related potentials research* (pp. 555-572). Washington, D.C.: US Government Printing Office.

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

Donchin, E., McCarthy, G., Kutas, M., & Ritter, W. (1983). Event-related brain potentials in the study of consciousness. In R.J. Davidson, G.E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Vol. 3* (pp. 81-121). New York: Plenum.

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* (pp.247-265). New York: Academic Press.

Eggemeier, F.T. (1988). Properties of workload assessment techniques. In P.A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 41-62). North-Holland: Elsevier Science Publishers.

Elbert, T., Rockstroh, B., Canavan, A., Birbaumer, N., Lutzenberger, W., von Bulow, I., & Linden, A. (1991). Self-regulation of slow cortical potentials an dits role in epileptogenesis. In J.G. Carlson & A.R. Seifert (Eds.), *International perspectives on self-regulation and health* (pp. 65-94). New York: Plenum.

Endsley, M.R. (1996). Automation and situational awareness. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 163-181). Mahwah, NJ: Lawrence Erlbaum, Assoc.

Endsley, M.R., & Kiris, E.O. (1994). The out-of-the-loop performance problem: Impact of level of automation and situation awareness. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 50-56). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Ford, J.M., Roth, W.T., & Koppell, B.S. (1976). Attention effects on auditory evoked potentials to infrequent events. *Biological Psychology, 4*, 65-77.

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

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

Gale, A. (1987). The electroencephalogram. In A. Gale and B. Christie (Eds.), *Psychophysiology and the electronic workplace* (pp.223-238). London: John Wiley and Sons.

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

Gale, A., & Edwards, J. (1983). The EEG and human behavior. In A. Gale & B. Christie (Eds.). *Physiological correlates of human behavior: Attention and performance* (pp. 99-127). New York: Academic Press.

Gentner, D., & Stevens, A.L. (Eds.), *Mental models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Glenn, F., Barba, C., Wherry, R.J., Morrison, J., Hitchcock, E., & Gluckman, J.P. (1994). Adaptive automation effects on flight management task performance. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 33-39). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

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

Gopher, D., & Donchin, E. (1986). Workload - an examination of the concept. In K.Boff & L.Kaufman (Eds.), *Handbook of perception and human performance: Sensory processes and perception* (pp. 41.1-41.49). New York: John Wiley.

Greenhouse, S.W., & Geisser, S.(1959). On methods in the analysis of profile data. *Psychometrika, 24*, 95-112.

Haagh, S.A.V.M., & Brunia, C.H.M. (1985). Anticipatory response-relevant muscle activity: CNV amplitude and simple reaction time. *Electroencephalography and Clinical Neurophysiology, 61*, 30-39.

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

Hadley, J., Mikulka, P.J., Freeman, F.G., Scerbo, M.W., & Prinzel III, L.J. (1997). *Psychophysiological evaluation of a biocybernetic system for adaptive automation*. Paper presented at the 9th Symposium on Aviation Psychology, Columbus, OH.

Hadley, J., Prinzel III, L.J., Freeman, F.G., & Mikulka, P.J. (in press). Behavioral, subjective, and psychophysiological correlates of various schedules of short-cycle automation. In M.Scerbo & K.Krahl (Eds.), *Automation technology and human performance: Current research and trends*. Mahwah, NJ: Lawrence Erlbaum Associates.

Haider, M., Sprong, P., & Lindsley, D.B. (1964). Attention, vigilance, and cortical evoked potentials. *Science, 145*, 180-182.

Hancock, P.A. (1996). Teleology and technology. In R.Parasuraman & M.Mouloua (Eds.), *Automation and human performance* (pp. 461-498). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Hancock, P.A., Chignell, M.H., & Lowenthal, A. (1985). An adaptive human-machine system. *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics, 15*, 627-629.

Hancock, P.A., & Chignell, M.H. (1988). Mental workload dynamics in adaptive interface design. *IEEE Transactions on Systems, Man, and Cybernetics, 18*, 647-658.

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

Hilburn, B., Jorna, P.G.A.M., & Parasuraman, R. (1995). The effects of advanced ATC automation on mental workload and monitoring performance: An empirical investigation in Dutch airspace. *Proceedings of the 8th international symposium on aviation psychology* (pp. 387-391). Columbus OH, Ohio State University.

Hilburn, B., Molloy, R., Wong, D., & Parasuraman, R. (1993). *Operator versus computer control of adaptive automation* (Technical Report NAWACADWAR-93031-60). Naval Air Warfare Center, Warminster, PA.

Hilburn, B., Parasuraman, R., & Mouloua, M. (1995). Effects of short- and long-cycle adaptive function allocation on performance of flight-related tasks. In N.Johnston, R.Fuller, & N.McDonald (Eds.), *Aviation psychology: Training and selection* (pp. 347-353). Hampshire, England: Ashgate.

Hillyard, S.A., & Picton, T.W. (1979). Event-related brain potentials and selective information in man. In J.E. Desmedt (Ed.), *Cognitive components in cerebral event-related potentials and selective attention - Progress in clinical neurophysiology* (pp. 110-187). Basel: Karger.

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

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

Hoffman, J.E. (1990). Event-related potentials and automatic and controlled processes. In J.W.Rohrbaugh, R.Parasuraman, & R.Johnson, Jr. (Eds.), *Event-related brain potentials: Basic issues and applications* (pp. 145-157). New York: Oxford University Press.

Hoffman, J.E., 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.

Hoffman, J.E., Simons, R.F., & Houck, M.R. (1983). The role of attentional resources in automatic detection. *Cognitive Psychology, 15*, 625-631.

Horst, R.L. (1987). An overview of current approaches and future challenges in physiological monitoring. *Mental State Examination* (NASA CP-2504, pp.25-42).

Horst, R.L., Johnson, R., & Donchin, E. (1980). Event-related brain potentials and subjective probability in a learning task. *Memory & Cognition, 8*, 476-488.

Howell, W.C. (1990). Human factors in the workplace. In M.Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 209-269). Palo Alto, CA: Consulting Psychologist Press, Inc.

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

Isreal, J.B., Chesney, G.L., Wickens, C.D., & Donchin, E. (1980a). P300 and tracking difficulty: Evidence for multiple resources in dual task performance. *Psychophysiology, 17*, 57-70.

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

Itoh, Y., Hayashi, Y., Tsukui, I., & Saito, S., (1989). Heart rate variability and subjective mental workload in flight task validity of mental workload measurement using HRV method. In M.J. Smith, & G. Salvendy (Eds.), *Work with computers: Organizational, management, stress, and health aspects* (pp. 209-216). Amsterdam: Elsevier Science Publishers.

Janssen, W.H., & Gaillard, A.W.K. (1985). EEG and heart rate correlates of task load in car driving. In A. Gundel (Ed.), *Proceedings of the workshop: Electroencephalography in transport operations* (pp. 39-53). Cologne, Germany: Elsevier.

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

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

Johnson, R., Miltner, W., & Braune, C. (1991). Auditory and somatosensory event-related potentials: Effects of attention. *Journal of Psychophysiology, 5*, 11-25.

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

Kahneman, D., & Peavler, W.S. (1969). Insensitive effects and pupillarity changes in association learning. *Journal of Experimental Psychology, 79*, 312-318.

Kahneman, D., Peavler, W.S., & Onuska, L. (1968). Effects of verbalization on incentive on the pupil response to mental activity. *Canadian Journal of Psychology, 22*, 186-196.

Kahneman, D., Tversky, B., Shapiro, D., & Crider, A. (1967). Pupillary, heart rate, and skin resistance changes during a mental task. *Journal of Experimental Psychology, 27*, 187-196.

Kantowitz, B.H., & Campbell, J.L. (1996). Pilot workload and flightdeck automation. In R.Parasuraman & M.Mouloua (Eds.), *Automation and human performance* (pp. 117-136). Mahwah, NJ: Lawrence Erlbaum Associates.

Kramer, A.F. (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* (pp. 197-222). New York: John Wiley & Sons.

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

Kramer, A.F., 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* (pp. 83-102). Houston, TX: NASA Johnson Space Center.

Kramer, A.F., Schneider, W., Fisk, A., & Donchin, E. (1986). The effects of practice and task structure on components of the event-related brain potential. *Psychophysiology, 23*, 33-47.

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

Kramer, A.F., Wickens, C.D., & Donchin, E. (1983). An analysis of the processing requirements 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., & Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203-205.

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

Lindholm, E., & Cheatham, C.M. (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.M., Koriath. (1984). *Physiological assessment of aircraft pilot workload in simulated landing and simulated hostile threat environments.* (AFHRL-TR-83-49). Williams Air Force Base, Arizona: US Air Force.

Lindqvist, A., Keskinen, E., Antila, K., Halkola, L., Peltonen, T., & Valimaki, I. (1983). Heart rate variability, cardiac mechanics, and subjectively evaluated stress during simulator flight. *Aviation, Space, and Environmental Medicine, 54*, 685-690.

Lubar, J.F. (1991). Discourse on the development of EEG diagnostics and biofeedback for Attention-Deficit/Hyperactivity Disorders. *Biofeedback and Self-Regulation, 16*, 201-225.

McCallum, C., Farmer, S.F., & Pocock, P.V. (1984). The effects of physical and semantic incongruities on auditory event-related potentials. *Electroencephalography and Clinical Neurophysiology, 59*, 477-488.

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science, 211*, 77-79.

McGillem, C.D., & Aunon, J.I. (1977). Measurements of signal components in single visually evoked brain potentials. *IEEE Transactions in Biomedical Engineering, BME-24*, 232-241.

Morris, N.M., & Rouse, W.B. (1986). *Adaptive aiding for human-computer control: Experimental studies of dynamic task allocation* (Report AAMRL-TR-86-005). Dayton, OH: Wright-Patterson Air Force Base.

Morrison, J.G., & Gluckman, J.P. (1994). Definitions and prospective guidelines for the application of adaptive automation. In M.Mouloua & R.Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 256-263). Hillsdale, NJ: Lawrence Erlbaum, Assoc.

Morrison, J.G., Gluckman, J.P., & Deaton, J.E. (1991). *Program plan for the adaptive function allocation for intelligent cockpits (AFAIC) program.* (Final Report No. NADC-91028-60). Warminster, PA: Naval Air Development Center.

Mouloua, M., & Parasuraman, R. (Eds.) (1995). *Human performance in automated systems: Current research and trends*. New Jersey: Lawrence Erlbaum, Assoc.

Naatanen, R., & Picton, T.W. (1987). The N1 wave fo the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology, 24,* 375-425.

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

National Research Council (1982). *Automation in combat aircraft*. Washington, D.C.: National Academy Press.

Neville, H.J. (1985). Biological constraints on semantic processing: A comparison of spoken and signed languages. *Psychophysiology, 22,* 576.

Norman, D.A. (1983). Some observations on mental models. In D.Gentner & A.L. Stevens (Eds.), *Mental models.* Hillsdale, NJ: Lawrence Erlbaum Associates.

O'Donnell, R., & Eggemeier, F.T. (1986). Workload assessment methodology. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of perception and human performance, Vol. II: Cognitive processes and performance* (pp. 42-1 -- 42-47). New York: John Wiley.

Offenloch, K., & Zahner, G. (1990). Computer aided physiological assessment of the functional state of pilots during simulated flight. *NATO Advisory Group for Aerospace Research and Development Conference Proceedings, 490,* 9-1 to 9-9.

O'Hanlon, J.F., & Beatty, J. (1979). Concurrence of electroencephalographic and performance changes during a simulated radar watch and some implications for the arousal theory of vigilance. In R.R. Mackie (Ed.), *Vigilance: Theory, operational performance, and physiological correlates* (pp. 189-201). NY: Plenum.

O'Hanlon, J.F., & Beatty, J. (1977). Concurrence of electroencephalographic and performance changes during a simulated radar watch and some implications for the arousal theory of vigilance. In R.R. Mackie (Ed.), *Vigilance* (pp. 189-201). New York: Plenum Press.

Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia, 9*, 97-113.

Opmeer, C.H.J.M., & Krol, J.P. (1973). Towards an objective assessment of cockpit workload: Physiological variables during different flight phases. *Aerospace Medicine, 44*, 527-532.

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

Parasurman, R. (1993). Effects of adaptive function allocation on human performance. In D.J.Garland & J.A.Wise (Eds.), *Human factors and advanced aviation technologies* (pp. 147-157). Dayton Beach, FL: Embry-Riddle Aeronautical University Press.

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

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

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors, 38*, 665-679.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*, 230-253.

Parsons, H.M. (1985). Automation and the individual: Comprehensive and comparative views. *Human Factors, 27*, 99-112.

Picton, T.W. (1988). Introduction. In T.W. Picton (Ed.), *Handbook of encephalography and neurophysiology: Human event-related potentials* (pp. 1-5). Amsterdam: Elsevier.

Pope, A.T., Bogart, E.H., & Bartolome, D.S (1995). Biocybernetic system validates index of operator engagement in automated task. *Biological Psychology, 40*, 187-195.

Prinzel III, L.J., Hitt, J., Scerbo, M.W., & Freeman, F.G. (abstract, 1995). Feedback contingencies and biocybernetic regulation of operator workload. *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society, 39*, 947.

Prinzel III, L.J., Scerbo, M.W., Freeman, F.G, & Mikulka, P.J. (1995). A biocybernetic system for adaptive automation. *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society, 39*, 1365-1369.

Prinzel III, L.J., Freeman, F.G., Scerbo, M.W., & Mikulka, P.J. (1997). A biocybernetic system for examining psychophysiological correlates for adaptive automation. *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society, 41*, 1378.

Prinzel III, L.J., Scerbo, M.W., Freeman, F.G., & Mikulka, P.J. (1997). Behavioral and physiological correlates of a bio-cybernetic, closed-loop system for adaptive automation. In M. Mouloua & J.M. Koonce (Eds.), *Human-Automation interaction: Research and practice* (pp. 66-75). Mahwah, NJ: Lawrence Erlbaum Associates.

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.

Ray, W., & Cole, H. (1985). EEG alpha activity during cognitive processing: Influence of attentional factors. *International Journal of Psychophysiology, 3*, 43-48.

Rebert, C.S., & Tecce, J.J. (1973). A summary of CNV and reactio time. In W.C. McCallum & J.R. Knott (Eds.), *Event-related slow potentials of the brain : Their relations to behavior* (pp. 173-178). Amsterdam: Elsevier.

Redondo, M., & Del Valle-Inclan, F. (1992). Decrements in heart rate variability during memory search. *International Journal of Psychophysiology, 18*, 477-482.

Reising, J.M. (1985). 2010: The symbionic cockpit. *Proceedings of the IEEE National Aerospace and Electronics Conference, 3*, 1050-1054.

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

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., 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.

Riley, V. (1985). Monitoring the monitor: Some possible effects of embedding human models in highly automated manned systems. *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics, 15*, 6-9.

Riley, V. (1994). A theory of operator reliance of automation. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 8-14). Hillsdale, NJ: Lawrence Erlbaum Associates.

Riley, V., Lyall, B., & Wiener, E. (1993). *Analytic methods for flight-deck automation design and evaluation. Phase two report: Pilot use of automation.* Minneapolis, MN: Honeywell Technology Center.

Rouse, W.B. (1977). Human-computer interaction in multitask situations. *IEEE Transactions Systems, Man, and Cybernetics, SMC-7*, 293-300.

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

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

Rouse, W.B. (1994). Twenty years of adaptive aiding: Origin of the concept and lessons learned. In M.Mouloua & R.Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 28-32). Hillsdale, NJ: Lawrence Erlbaum, Assoc.

Ruchkin, D.S. (1988). Measurement of ERPs. In T.W. Picton (Ed.), *Handbook of encephalography and neurophysiology: Human event-related potentials* (pp. 7-43). Amsterdam: Elsevier.

Ruchkin, D.S., Sutton, S., & Teuting, P. (1975). Emitted and evoked P300 potentials and variation in stimulus probability. *Psychophysiology, 12,* 591-595.

Rugg, M.D. (1985). The effects of semantic priming and word repetition on event-related potentials. *Psychophysiology, 22,* 642-647.

Sanders, M.S. & McCormick, E.J. (1993). *Human Factors in engineering and design* (7th Edition). NY: McGraw-Hill.

Sayers, B.M.C.A. (1973). Analysis of heart rate variability. *Ergonomics, 16,* 17-32.

Scallen, S.F., Hancock, P.A., & Duley, J.A. (1995). Pilot performance and preference for short cycles of automation in adaptive function allocation. *Applied Ergonomics, 26,* 397-403.

Scallen, S.F., Duley, J.A., & Hancock, P.A. (1994). Pilot performance and preference for cycles of automation in adaptive function allocation. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 154-160). Hillsdale, NJ: Lawrence Erlbaum Associates.

Scerbo, M.W. (1994). Implementing adaptive automation in aviation: The pilot-cockpit team. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 249-255). Hillsdale, NJ: Lawrence Erlbaum Associates.

Scerbo, M.W. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M.Mouloua (Eds.), Automation and human performance: Theory and applications (pp. 37-64). Mahwah, New Jersey: Lawrence Erlbaum Assoc.

Scherg, M., Vajsar, J., Picton, T.W. (1989). A source analysis of the late human auditory evoked potential. *Journal of Cognitive Neuroscience, 1*, 336-355.

Schwilden, H., Stoeckel, H., & Schuttler, J. (1989). Closed-loop feedback control of propofol anaesthesia by quantitative EEG analysis in humans. *British Journal of Anesthesiology, 62*, 290-296.

Sekiguchi, C., Handa, Y., Gotoh, M., Kurihara, Y., Nagasawa, Y., & Kuroda, I. (1978). Evaluation method of mental workload under flight conditions. *Aviation, Space, and Environmental Medicine, 49*, 920-925.

Sem-Jacobsen, C.W. (1981). Brain/computer communication to reduce human error: A perspective. *Aviation, Space, and Environmental Medicine, 52*, 31-38.

Sheridan, T.B. (1997). Supervisory control. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 1295-1327). New York: John Wiley & Sons.

Sherg, M. (1990). Fundamentals of dipole soure potential analysis. In F. Grandori, M. Hoke, & G.L. Romani (Eds.), *Auditory evoked magnetic fields and electric potentials. Advances in audiology* (pp. 40-69). Karger, Basel.

Singh, I.L., Molloy, R., & Parasuraman, R. (1993). Automation-induced "complacency": Development of the complacency-potential rating scale. *International Journal of Aviation Psychology, 3*, 111-121.

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

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

Sirevaag, E., Kramer, A., Wickens, C.D., Reisweber, M., Strayer, D., & Grenell, J. (1993). Assessment of pilot performance and mental workload in rotary wing aircraft. *Ergonomics, 36*, 1121-1140.

Smith, T.J., & Smith, K.U. (1987). Feedback control mechanisms of human behavior. In G. Salvendy(Ed.), *Handbook of human factors* (pp. 251-293). New York: John Wiley & Sons.

Spyker, D., Stackhouse, A., Khalafall, A., & McLane, R. (1971). *Development of techniques for measuring pilot workload.* (NASA report 1886). Minneapolis, MN: Honeywell Technology Center.

Squires, N.K., Squires, K.C., & Hillyard, S.A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology, 38*, 387-401.

Sterman, M.B., Kaiser, D.A., Mann, C.A., Suyenobu, B.Y., Beyma, D.C., & Francis, J.R. (1993). Application of quantitative EEG analysis to workload assessment in an advanced aircraft simulator. *Proceedings of Annual Meeting of the Human Factors and Ergonomics Society,* 37,118-121.

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

Sterman, M.B., Schummer, G.J., Dushenko, T.W., & Smith, J.C. (1987). *Electroencephalographic correlates of pilot performance: Simulation and in-flight studies* (pp. 1-16). Neuilly sur Seine, France: AGARD.

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

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

Streitberg, B., Röhmel, J., Herrmann, W.M., & Kubicki, S. (1987). COMSTAT rule for vigilance classification based on spontaneous EEG activity. *Neuropsychobiology, 17*, 105-117.

Surwillo, W.W. (1990). *Psychophysiology for clinical psychologists.* Norwood, NJ: Ablex publishing company.

Sutton, S., Braren, M., Zubin, J., John, E.R. (1965). Evoked potential correlates of stimulus uncertainty. *Science, 150*, 1187-1188.

Sutton, S., Teuting, P., Zubin, J., & John, E.R. (1967). Information delivery and the sensory evoked potential. *Science, 155*, 1436-1439.

Teece, J.J. (1972). Contingent negative variation. In E. Niedermeyer & F. Lopes da Silva (Eds.), *Electroencephalography: Basic principles, clinical applications, and related fields* (pp. 657-679). Baltimore: Urban & Schwarzenberg.

Teece, J.J., & Cattanach, L. (1993). Contingent negative variation. In E. Niedermeyer & F. Lopes da Silva (Eds.), *Electroencephalography: Basic principles, clinical applications, and related fields* (pp. 887-910). Baltimore: Urban & Schwarzenberg.

Teuting, P., Sutton, S., & Zubin, J. (1970). Quantitative evoked potential correlates of the probability of events. *Psychophysiology, 7*, 385-394.

Theissen, M.S., Lay, J.E., & Stern, J.A. (1986). *Neuropsychological workload test battery validation study*. (FZM 7446). Fort Worth, TX: General Dynamics.

Torsvall, L., & Akerstedt, T. (1987). Sleepiness on the job: Continuously measured EEG changes in train drivers. *Electroencephalography and Clinical Neurophysiology, 66*, 502-511.

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

Vaughan, H.G., & Arezzo, J.C. (1988). The neural basis of event-related potentials. In T.W. Picton (Ed.), *Handbook of encephalography and neurophysiology: Human event-related potentials* (pp. 45-96). Amsterdam: Elsevier.

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

Wesensten, N.J., Crowley, J., Balkin, T., & Kamimori, G. (1993). Effects of simulated high altitude exposure on long-latency event-related brain potentials and performance. *Aviation, Space, and Environmental Medicine, 64*, 30-36.

Wickens, C.D. (1990). Applications of event-related potential research to problems in human factors. In J.W.Rohrbaugh, R.Parasuraman, & R.Johnson, Jr. (Eds.), *Event-related brain potentials: Basic issues and applications* (pp. 301-309). New York: Oxford University Press.

Wickens, C.D. (1992). *Engineering psychology and human performance, 2nd Ed.* New York: Harper Collins.

Wickens, C.D., Isreal, J.B., & Donchin, E. (1977). The event related cortical potential as an index of workload. *Proceedings of the Human Factors and Ergonomics Society, 21*, 282-287.

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

Wiener, E.L. (1988). Cockpit automation. In E.L. Wiener & D.C. Nagel (Eds.), *Human factors in aviation* (pp. 433-461). San Diego: Academic Press.

Wiener, E.L. (1989). *Human factors of advanced technology ("glass cockpit") transport aircraft* (Technical Report 117528). Moffett Field, CA: NASA-Ames Research Center.

Wiener, E.L., & Curry, R.E. (1980). Flight deck automation: Promises and problems. *Ergonomics, 23*, 995-1011.

Wiener, N. (1949). *Extrapolation, interpolation, and smoothing of stationary time series.* New York: J. Wiley & Sons

Wiener, E.L. (1989). *Human factors of advanced technology ("glass cockpit") transport aircraft* (Technical Report 117528). Moffett Field, CA: NASA-Ames Research Center.

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

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

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

Wilson, J.R., & Rutherford, A. (1989). Mental models: Theory and applications in human factors. *Human Factors, 31*, 617-633.

Woods, D.D. (1996). Decomposing automation: Apparent simplicity, real complexity. In R.Parasuraman & M.Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 3-18). Mahwah, NJ: Lawrence Erlbaum Assoc.

Woody, C.D. (1967). Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. *Medical Biological Engineering, 5*, 529-553.

APPENDIX A
NASA-TASK LOAD INDEX

MENTAL DEMAND

Low High

PHYSICAL DEMAND

Low High

TEMPORAL DEMAND

Low High

PERFORMANCE

Good Poor

EFFORT

Low High

FRUSTRATION

Low High

VITA

Lawrence (Lance) Jacob Prinzel III was born on June 25, 1970. He graduated summa cum laude from Old Dominion University in May 1993 with a Bachelor of Science degree in psychology. Lance later earned his Master of Science in psychology in August 1995. In August 1998, he completed the requirements for the Doctor of Philosophy in industrial/organizational psychology with a major in human factors and a minor in physiological psychology.

Lance was a recipient of NASA Graduate Student Research Fellowships and Space Consortium Research Scholarships. During his graduate studies at Old Dominion University, he had presented over 25 papers and published 15+ research articles in some of the leading periodicals in the field. Lance has accepted a position with the NASA Langley Research Center as a Cognitive Neuropsychologist.

Correspondence may be sent to:

Old Dominion University - Psychology Department
Mills Godwin Building, Room 250
Norfolk, VA 23529

IMAGE EVALUATION
TEST TARGET (QA-3)

150mm

6"

APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989