NASA/TP-2001-211018

The Efficacy of Psychophysiological Measures for Implementing Adaptive Technology

Mark W. Scerbo, Frederick G. Freeman, Peter J. Mikulka
Old Dominion University, Norfolk, Virginia

Raja Parasuraman and Francesco Di Nocero
The Catholic University of America, Washington, DC

Lawrence J. Prinzel III
Langley Research Center, Hampton, Virginia

June 2001

The NASA STI Program Office ... in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA Scientific and Technical Information (STI) Program Office plays a key part in helping NASA maintain this important role.

The NASA STI Program Office is operated by Langley Research Center, the lead center for NASA's scientific and technical information. The NASA STI Program Office provides access to the NASA STI Database, the largest collection of aeronautical and space science STI in the world. The Program Office is also NASA's institutional mechanism for disseminating the results of its research and development activities. These results are published by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers, but having less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

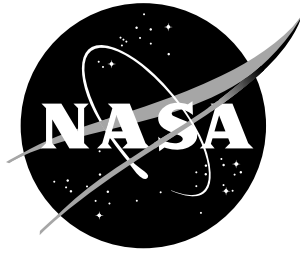- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services that complement the STI Program Office's diverse offerings include creating custom thesauri, building customized databases, organizing and publishing research results ... even providing videos.

For more information about the NASA STI Program Office, see the following:

- Access the NASA STI Program Home Page at *http://www.sti.nasa.gov*

- E-mail your question via the Internet to help@sti.nasa.gov

- Fax your question to the NASA STI Help Desk at (301) 621-0134

- Phone the NASA STI Help Desk at (301) 621-0390

- Write to:
  NASA STI Help Desk
  NASA Center for AeroSpace Information
  7121 Standard Drive
  Hanover, MD 21076-1320

NASA/TP-2001-211018

The Efficacy of Psychophysiological Measures for Implementing Adaptive Technology

Mark W. Scerbo, Frederick G. Freeman, Peter J. Mikulka
Old Dominion University, Norfolk, Virginia

Raja Parasuraman and Francesco Di Nocero
The Catholic University of America, Washington, DC

Lawrence J. Prinzel III
Langley Research Center, Hampton, Virginia

June 2001

ABSTRACT

Adaptive automation refers to technology that can change its mode of operation dynamically. Further, both the technology and the operator can initiate changes in the level or mode of automation. The present paper reviews research on adaptive technology. The paper is intended as a guide and review for those seeking to use psychophysiological measures in design and assessing adaptively automated systems. It is divided into four primary sections. In the first section, issues surrounding the development and implementation of adaptive automation are presented. Because physiological-based measures show much promise for implementing adaptive automation, the second section is devoted to examining candidate indices and reviews some of the current research on these measures as they relate to workload. In the third section, detailed discussion is devoted to electroencephalogram (EEG) and event-related potentials (ERPs) measures of workload. The final section provides an example of how psychophysiological measures can be used in adaptive automation design.

TABLE OF CONTENTS

4

Symbols and Abbreviations

ADHD	Attention-Deficit / Hyperactivity Disorder
BESA	Brain Electric Source Analysis
CFIT	Controlled Flight Into Terrain
CMT	Cognitive Motor Test
CNV	Contingent Negative Variation
DARPA	Defense Advanced Research Projects Agency
EEG	Electroencephalogram
EMG	Electromyogram
ERD	Event-Related Desynchronization
ERN	Event-Related Negativity
ERP	Event-Related Potential
fMRI	Functional Magnetic Resonance Imagery
GCAS	Ground Collision-Avoidance System
GPS	Global Positioning System
HP	Heart Period
HR	Heart Rate
HRV	Heart-Rate Variability
IBI	Inter-Beat Interval
IFR	Instrument Flight Rules
LED	Light Emitting Diodes
MATB	Multi-Attribute Task Battery
NTSB	National Transportation Safety Board
PCA	Principal Components Analysis
PET	Positron Emission Tomography
RHP	Residual Heart Period
RSA	Respiratory Sinus Arrythmia
RMSE	Root-Mean-Squared-Error
SCP	Slow Cortical Potentials
SMR	Sensory-Motor Response
TLX	Task-Load-Index (NASA-TLX)
VFR	Visual Flight Rules
VI	Virtual Instrument

SECTION I

Summary

The evolution of technology has generated modern machines and systems that expand the range of human capabilities enormously. The benefits afforded by such technological progress necessitate machines and systems of greater sophistication and complexity. Consequently, the demands placed upon the operators of these systems increase along with the growth in complexity. The introduction of automation into systems has helped operators manage the complexity, but has not necessarily relieved the burden of interacting with such systems. Thus, one of the major challenges facing designers today concerns the best way to utilize technology to serve the needs of society without exceeding the limits of those individuals who must operate the technology.

The purpose of the present paper is to examine some of the issues surrounding the use of automation in complex systems and its effect on the human operator. More specifically, this paper focuses on *adaptive* automation, a form of automation that is dynamic and can adjust to the needs of the operator in real time. One of the critical issues for any adaptive system concerns how changes among the modes of operation will be accomplished. There are a variety of ways to trigger changes among modes including critical events, operator models, and real time measures of performance. One of the more promising methods, however, may be the use of physiological measures that reflect changes in operator workload. These measures can be obtained continuously and with little or no interference in the operator's task. The merits of many of these measures can be found in several reviews (see Byrne & Parasuraman, 1996; Kahneman, 1973; Kramer, Trejo, & Humphrey, 1996; Parasuraman, 1990). The primary purpose of the present paper, however, is to review the most recent research on these measures and evaluate their potential for adaptive automation.

Automation

Automation has been described as a machine agent that can execute functions normally carried out by humans (Parasuraman & Riley, 1997). These can be entire functions, activities, or subsets thereof. Automation serves several purposes (Wickens, 1992). It can perform functions that are beyond the ability of humans, it can perform functions for which humans are ill-suited, and it can perform those functions that humans find bothersome or a nuisance.

The level of automation in a system can vary. Sheridan and Verplank (1978) proposed a model where differences range from completely manual to fully automatic (see Table 1). Several examples of degrees of automation can be found in a typical automobile. At the lowest level, virtually all automobiles require the driver to put the car into gear. At the other extreme, the antilock braking system calculates how much pressure to apply to each wheel to bring the car to a halt without locking up any wheels. It does so without communicating any of its calculations or actions. All the driver has to do is apply the brakes. The presets on the car's audio system allow individuals to automatically tune to their favorite stations. The system limits the range of available frequencies to a select few and presents these choices to the user on separate buttons.

Table 1. 10 Levels of Human-Automation Interaction (Sheridan & Verplank, 1978)

1) Whole task done by human except for actual machine operation	Manual
2) . . .	Semiautomatic
3) . . .	Semiautomatic
4) Computer suggests options and proposes one of them	Semiautomatic
5) Computer chooses an action and performs it if human approves	Semiautomatic
6) Computer chooses an action and performs unless human disapproves	Semiautomatic
7) . . .	Semiautomatic
8) . . .	Semiautomatic
9) . . .	Semiautomatic
10) Computer does everything autonomously	Automatic

Recently, Parasuraman, Sheridan, and Wickens (2000) expanded upon this model to provide designers with a framework for considering what types and levels of automation ought to be implemented in a given system. This expanded model allows for various levels of automation within different functions. The four functions they describe are system analogs of different stages of human information processing: information acquisition, information analysis, decision selection, and action implementation (see Table 2).

Table 2. Information processing functions (based on Parasuraman, Sheridan, & Wickens, 2000)

Stage of Processing	Functions
Information Acquisition	Detecting and registering input data
Information Analysis	Applying cognitive functions to the information (e.g., analyzing and summarizing, making predictions, inferences, modifying and augmenting information displays, etc.)
Decision Selection	Augmenting or replacing human selection of decision options
Action Implementation	Executing functions or choices of actions

Advantages and Disadvantages of Automation

Automation is touted as having numerous benefits. As machines assume greater responsibilities there are fewer activities for humans to do. Thus, automation can reduce workload. Automation can also

afford operators greater control over more complex systems (Woods, 1994) or reduce the variability of human performance and thereby reduce errors. As Wiener (1988) indicates, the use of automation in aviation has helped to increase fuel efficiency, reduce flight times, and navigate more effectively.

It should be noted, however, that the advantages of automation come at a price. Automation *changes* the way activities are carried out and therefore creates a different set of problems (Billings, 1991; Wiener & Curry, 1980; Woods, 1996). For instance, higher degrees of automation leave fewer activities for the operator to perform thereby changing his or her role from active participant to passive observer. Parasuraman, Mouloua, Molloy, and Hilburn (1996) describe a program of research demonstrating that automation can inhibit one's ability to detect critical signals or warning conditions. This change in operator roles can also result in the deterioration of manual skills in the presence of long periods of automation (Wickens, 1992). In addition, several investigators have commented that automation does not necessarily reduce workload. In some instances it may even increase workload and generate new types of errors (Kurlik, 1993; Sarter & Woods, 1995; Wiener, 1989). Woods (1996) has also suggested that automation can lead to incongruent goals between operators and system components. Further, he argues that in systems where subcomponents are tightly coupled, problems may propagate more quickly and be more difficult to isolate. Thus, it should not be surprising that the introduction of automation generates a good deal of skepticism among its users. Several researchers have shown that confidence in automation and in oneself affect how and when it is used (Lee & Moray, 1992; Muir, 1987; Riley, 1996).

Clearly, there are both advantages and disadvantages associated with automation. Woods (1996) suggests that the costs and benefits of automation are the result of changes in the nature of work. The introduction of automation does not necessarily eliminate work, it redistributes it and this has important consequences for how humans interact with this type of technology. Parasuraman and Riley (1997) have argued that successful applications of automation technology require an understanding of operator decisions about when to use the automation as well as the conditions under which operators will come to rely too heavily upon automation or neglect it altogether. Further, they also argue that operators cannot be held solely responsible for their interactions with automation. The designers of automated systems and the organizational climate under which the technology is created and used also contribute to its effectiveness.

Adaptable and Adaptive Technology

The issues surrounding automation discussed above take on greater importance when we turn our attention to automation that is adaptive. In this type of automation, the level of automation or the number of systems operating under automation can be modified or the format of the interface can be modified in real time (Hammer & Small, 1995; Scerbo, 1996). More important, changes in the state of automation can be initiated by either the human or the system (Hancock & Chignell, 1987; Morrison, Gluckman, & Deaton, 1991; Rouse, 1976). Parasuraman, Bahri, Deaton, Morrison, and Barnes (1992) have argued that adaptive automation allows for a tighter coupling between the level of automation and the level of operator workload.

Research on adaptive technology has resulted in some confusion in the literature between systems that are *adaptable* and those that are *adaptive*. Adaptive technology can be discussed within a two-fold taxonomy of adaptive technology. The first dimension addresses the source of flexibility in the system. In systems that have dynamic displays, it is primarily the presentation format that changes. In other systems,

it is the functionality that is modifiable.

The second dimension concerns how changes among states or modes of automation are invoked. In *adaptable* systems, the user initiates changes among presentation modes or functionality. In truly *adaptive* systems, as noted previously, *both* the user and the system can initiate changes among system states or modes. The distinction between adaptable and adaptive technology is one of authority. Adaptable systems reflect a superordinate-subordinate relationship between the operator and the system. In this arrangement, the human always maintains the authority to invoke or change the automation. In adaptive systems, on the other hand, the authority over invocation is shared. Each member of the "human-system team" can initiate changes in state of modes of operation.

At this point it might prove useful to compare the two-fold taxonomy description to the 4-stage processing model proposed by Parasuraman et al. (2000). The taxonomy description of systems with dynamic displays represents automation in the information analysis stage described by Parasuraman and his colleagues. Those systems in which functionality is flexible are analogous to the action implementation systems described by Parasuraman et al. The taxonomy does not include systems in which information acquisition is flexible; however, work on the Rotorcraft Pilot's Associate (see below) does include a system that can modify the kind of information it seeks based upon specific criteria. The taxonomy also does not include a category for flexible decision selection. However, a closer look at these functions in the Parasuraman et al. model show that they really reflect differences in authority and communication (i.e., whether the system has the authority to changes states of operation and how that is communicated to the operator).

Research on Adaptive Technology

Research on adaptive automation grew out of work in artificial intelligence during the 1970's. Much of this effort was directed toward developing adaptive aids to help allocate tasks between humans and computers (Rouse, 1976; 1977). A significant step forward came with an attempt to use state of the art intelligent systems to assist pilots of advanced fighter aircraft. This program, called the Pilot's Associate, was a joint effort among the Defense Advanced Research Projects Agency (DARPA), Lockheed Aeronautical Systems Company, McDonnell Aircraft Company, and the Wright Research and Development Center. The objective behind the Pilot's Associate was to provide pilots with an "assistant" that would supply them with information in the appropriate format when they needed it. The system was a network of cooperative knowledge-based subsystems that could monitor and assess events and then formulate plans to respond to problems (Hammer & Small, 1995).

The U.S. Army has continued this development effort with their Rotorcraft Pilot's Associate (RPA) program (Colucci, 1995). The goal of this program is to develop an intelligent "crew member" for the next generation of attack helicopters. Miller, Guerlain, and Hannen (1999) argue that because helicopter missions are less sequential in nature than those carried out by fixed wing aircraft they pose a greater challenge for designers and developers.

A Case for Adaptive Automation

As Scerbo (1996) noted, adaptive technology represents the next step in the evolution of automation. Users of this technology will be faced with systems that are qualitatively different from

those available today. Thus, it is not difficult to find arguments against the development of this technology. Often, these arguments concern authority and responsibility. The availability of adaptive automation usurps some of the control over operating a system from the user. Many operators are reluctant to give up their control. There may be two reasons for this. First, many people believe that they possess better skills and more expertise than the systems they operate. Second, they believe that they are responsible for the safety of the system they operate and the lives of others affected by the system (Billings & Woods, 1994; Malin & Schrenckenghost, 1992). Pilots too, have argued that because they have responsibility for the aircraft, themselves, and any passengers, they should have the authority to initiate changes in automation (Billings, 1991).

Although it would be unwise to condone the development of any technology unchecked, there are several good arguments for pursuing adaptive automation. As noted above, one of the advantages of automation is that it can perform activities that are beyond human capabilities. Thus one potential benefit of adaptive automation is that it could automate functions at the precisely the instant they are needed most.

Let's consider commercial aviation. Inagaki and his colleagues (1999, 2000) have been investigating the application of adaptive technology to decisions surrounding aborted take-offs. Should an engine fail during take-off, the pilot has but seconds to decide whether to continue climbing or abort the take-off. Indeed, the NTSB (1990) has reported that pilots do not always make the correct decision under these circumstances. Inagaki, Takae and Moray (1999) have shown mathematically that the optimal approach to this problem is not one where the human pilot maintains full control over this decision. Nor is it one where full control is delegated to the avioncs. In fact, the best decisions are made when the pilot and automation share control depending upon critical factors such as actual airspeed, desired airspeed, the reliability of warnings, pilot response time, etc. In a study designed to examine decision making under these conditions, Inagaki et al. (1999) found that fewer errors were made when control over the decisions was traded between humans and the automation. Moreover, these investigators found that improvements in interface design alone were insufficient to bolster decision-making accuracy to levels that could be obtained with adaptive technology.

Another serious issue affecting both commercial and military aviation is the problem of Controlled Flight Into the Terrain (CFIT). It has been reported that CFIT is one of the leading categories of accidents in commercial aviation (Khatwa & Roelen, 1996) and Shappell and Wiegmann (1997) found that within the U.S. Navy and Marines Corps an average of 10 aircraft per year were lost to CFIT accidents.

Scott (1999) describes an adaptive system being developed by the USAF, Lockheed Martin, NASA, and the Swedish Air Force to combat this problem. The automatic Ground Collision-Avoidance System (GCAS) is being tested on the F-16D. The system assesses both internal and external sources of information and calculates the time it will take until the aircraft breaks through a pilot determined minimum altitude. Approximately 5 sec beforehand, the pilot is warned that the GCAS is about to take over. If no action is taken, a break-X warning is presented when the aircraft descends to the critical altitude, an audio "fly up" warning is presented, and the GCAS usurps control of the aircraft. When the system has maneuvered the aircraft in a heading out of the way of the terrain, it returns control of the aircraft to the pilot with the message, "You got it". The intervention is designed to right the aircraft quicker than any human pilot can respond. Indeed, test pilots acknowledged the rapid intervention. Moreover, test pilots who were given the authority to override GCAS eventually conceded control to the adaptive system. Scott believes that GCAS may soon find its way onto the Swedish JAS 39 Gripens and

F-16 and the F-22 Joint Strike Fighter in the U.S.

Scerbo (1996) also noted that there are situations where it might be critical for the system to have authority over automation invocation. For example, it is not uncommon for many of today's fighter aircraft to sustain G force levels that can exceed the physiological tolerances of the pilot (Buick, 1989). Conditions such as these can render the pilot of an armed and fast moving aircraft unconscious for periods of up to 12 seconds (Whinnery, 1989). Obviously, in this context an adaptive system could not only save lives, but protect the aircraft as well.

Outside of aviation, research efforts have been aimed at testing and evaluating adaptive cruise control for automobiles (Stanton & Young, 2000; Young & Stanton, 2000). In traditional cruise control systems, the speed of the vehicle is maintained by automatic control of the accelerator. In adaptive cruise control systems, the speed of the vehicle can be adjusted if an obstacle is detected in the road ahead. Future systems will address lateral deviations as well. Because millions more people travel by automobiles than by air and because the fatality rates on U.S. highways are 25 times higher than in the air, the necessity to explore alternative technologies to increase highway safety is undeniable.

Adaptive Strategies

Although technical demonstrations of adaptive automation exist, they were not necessarily efforts guided by how the technology *ought* to be implemented. Morrison and Gluckman (1994) described a program of research aimed at understanding how adaptive automation might be implemented. Strategies for invoking automation were based upon two primary factors. The first of these concerns how functions might be changed. For example, Rouse and Rouse (1983) described three different ways in which automation could assist the operator. First, whole tasks could be allocated to either the system or the operator to perform. Second, a specific task could be partitioned or divided so that the system and operator each share responsibility for unique portions of the task. Third, a task could be transformed or represented in an alternative format to make it easier for the operator to perform.

The second factor described by Morrison and Gluckman (1994) concerns the triggering mechanism for shifting among modes or levels of automation. In other words, to what properties (of the human operator, the task environment, or both) should the system adapt? A number of methods for adaptive automation have been proposed. Parasuraman et al. (1992) reviewed the major techniques and found that they fell into five main categories: critical environmental events, operator performance measurement, operator modeling, physiological assessment, and hybrid methods.

Critical Events. In the critical environment events method, the implementation of automation is tied to the occurrence of specific tactical events that occur in the task environment. For example, in aviation, the take-off and landing are considered the most demanding phases of flight. A goal-based adaptive system might change its mode of operation to address the additional demands during these specific operations (Barnes & Grossman, 1985). Alternatively, a system could monitor ongoing activities within a mission for the occurrence of critical events. Automation would be invoked when these events were detected, such as in an air traffic control system; a rapid rise in traffic density or complexity could lead to the presentation to the controller of automated decision aids for conflict detection and resolution (Hilburn, Jorna, Byrne, & Parasuraman, 1997). This method of automation is adaptive because if the critical events do not occur, the automation is not invoked. Such an adaptive automation method is inherently flexible because it can be tied to system operational procedures. Although this strategy might

be the most straight forward to implement, Parasuraman et al. (1992) have argued that systems of this nature would not be very sensitive to operator workload or performance. Consequently, this method will invoke automation irrespective of whether or not the operator needs assistance at the time.

Operator Modeling. The operator modeling and performance measurement techniques attempt to overcome the loose coupling problems associated with the critical events method. In this technique, human operator states or performance may be modeled theoretically, with the adaptive logic being driven by the model parameters.

Regarding operator modeling, an operator's current level of performance is compared against models of operator performance with the system under various levels of workload. The ability to predict future demands allows the system to be proactive in invoking automation changes to meet current needs in a dynamic environment (see Parasuraman et al., 1992 for a review). For example, in the system described by Rouse, Geddes, and Curry (1987-1988), the operator model is designed to estimate current and future states of an operator's activities, intentions, resources, and performance. Inputs include information about the operator, the system, and the outside world. An intent module interprets the operator's actions within the context of the inputs and the operator's goals and plans. A resource module estimates current and future demands based upon the operator's activities and the outputs of the intent module. The performance module uses this information to predict current and future levels of performance and to determine the need and format for adaptive aiding. This operator model fits within an architecture that includes an error monitor, adaptive aiding module, and interface manager to not only help operators overcome their limitations, but enhance their abilities as well. Other intelligent systems that incorporate human intent inferencing models have been proposed (Geddes, 1985; Hancock & Chignell 1987).

Operator performance measures can also be used to invoke automation based upon real-time measures of the operator's performance. For example, performance could be measured continuously and deviations from some specified criteria could trigger the automation.

Performance Measurement. Recently, several investigators have approached adaptive automation from this perspective by studying study motor skill performance in teams using a simple tracking task. For instance, Scerbo, Ceplenski, Krahl, and Eischeid (1996) had participants perform a pursuit tracking task in which a target traced a figure 8 on a computer screen. The task was partitioned, however, such that one individual controlled the vertical movement of the cursor and another controlled the horizontal movement. The participants were assigned to three different teams. One of these was a human-human team in which two participants worked together to perform the task. In the other two teams, participants shared control with the computer. Half of the participants worked with a computer that exhibited expert-level skills and the remaining individuals worked with a computer exhibiting novice-level skill. The skill level of the "computer teammate" was generated from another set of human performance data.

The results of that study are shown in Figure 1. The RMSE scores for the X axis are plotted over blocks of trials. The extreme levels of performance are exhibited by the computer teammate, i.e., the worst performance is that of the computer novice and the best performance is that of the computer expert. The performance of the human teammates is determined largely by group assignment. Those paired with the novice computer performed more poorly than the others. By contrast, those paired with the expert computer performed quite well initially, and eventually reached the level of the computer expert.

The results of this study are important for two reasons. First, the skill level of one's partner clearly

affected overall levels of performance. Those humans paired with a computer of expert-level skill quickly reached the level of their partner. Moreover, two humans working together outperformed the participants who were paired with the novice computer. These findings are particularly noteworthy because control over either axis was *independent*. Thus, the skill level of one's partner exerted considerable influence over one's performance even though their partner had no effect on the ability to minimize one's own RMSE scores.

The second important result from this study suggests that task partitioning may be a viable strategy in adaptive technology even where motor skills are involved. All participants showed improvement over the session regardless of whether their partner was human or computer. Further, these results suggest that automation modeled after expert-level performance may optimize the human partner's performance.
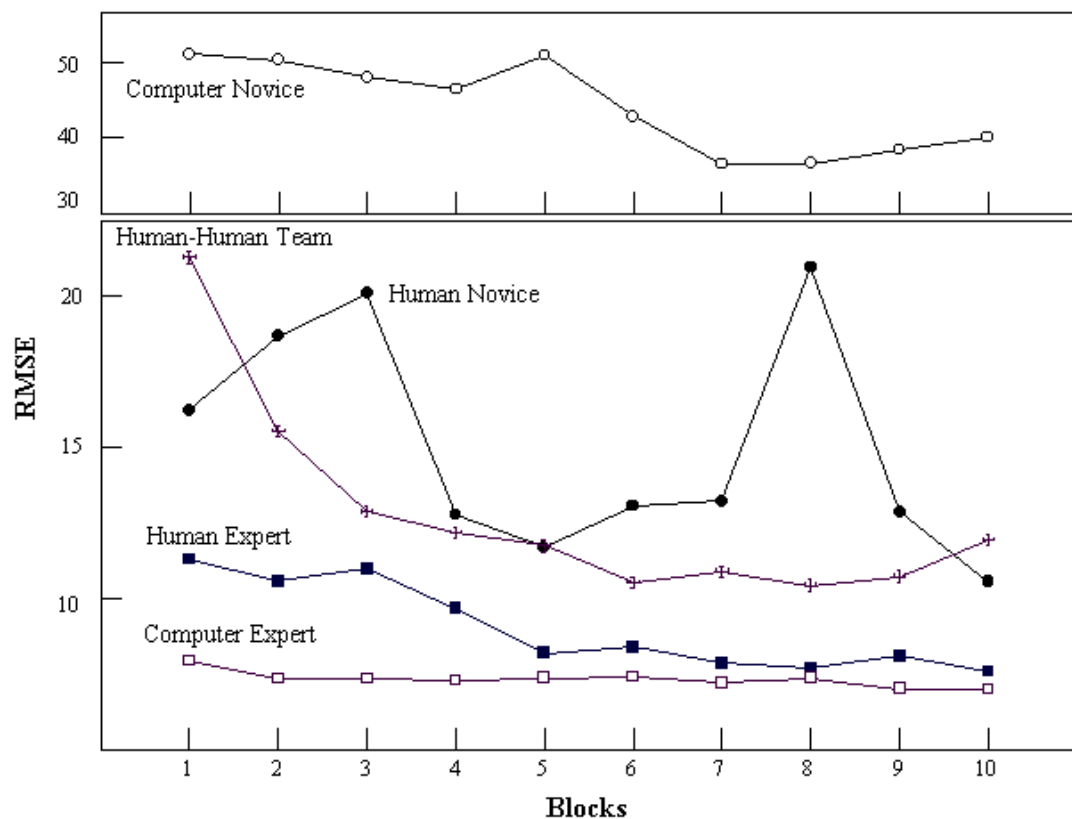
Figure 1. RMSE scores for human and computer teammates over blocks of trials.
(Note: Compressed ordinate in top section of chart)

The results of Scerbo et al. (1996) showed promise for task partitioning with a simple tracking task. That task, however, was not adaptive. In a subsequent study, Krahl and Scerbo (1997) revisited this issue with a truly adaptive task. In this experiment, participants were again assigned to work with either a human or computer teammate of different skill level and were asked to perform a pursuit tracking task separated into horizontal and vertical axes. The participants were instructed that their goal was to achieve

their lowest overall team score (based on a combination of RMSE scores from both teammates). On a given trial, if the participants thought that they could do better than their partner, they could press a button on the top of their joystick and attempt to take control of both axes. Participants would gain control over both axes *only* if they had demonstrated superior performance on the previous trial. Otherwise, each partner retained control of his or her axis. If control had changed hands on a given trial, it would revert back to both partners on the subsequent trial. Thus, the task can be considered adaptive because the level of automation on a given trial was determined by the overall level of team performance.

As in the earlier study, Krahl and Scerbo (1997) found that, in general, performance improved over blocks of trials. Again, performance was moderated by group assignment. Those assigned to the expert computer outperformed those assigned to the novice computer. In this study, however, the differences between the expert and novice computer conditions are particularly noteworthy. Because of the adaptive nature of the task, if one's partner took control of both axes on some trials, that participant would not necessarily get the chance to work at the task on every trial. The results showed that in the novice condition, the human and computer teammates took control of both axes equally often. By contrast, in the expert condition, humans managed to usurp control from their computer teammate on only 3% of the trials. The computer expert, however, took control from the human teammate on 34% of the trials. Thus, the human teammates in this condition attained their superior level of performance with 1/3 less opportunity to practice the task.

Similar benefits of task partitioning have also been reported by Scallen and Hancock (1997). These investigators had their participants perform a tracking task in addition to a monitoring and targeting task. Participants performed under automatic, partitioned (horizontal and vertical), and manual modes. Under the appropriate conditions, the tracking task was automated during peak workload periods of the targeting task. Scallen and Hancock found that the availability of automation improved tracking performance during the nonautomated periods of the task and that the level of improvement in performance was comparable in the fully automatic and task partitioning conditions. Moreover, performance on the targeting task also benefited from both the automatic and task partitioning conditions.

Taken together, the findings of Scerbo et al. (1996), Krahl and Scerbo (1997), and Scallen and Hancock (1997) demonstrate positive effects for task partitioning of motor skills. Further, these results indicate that task partitioning can be an effective strategy in a truly adaptive environment. In addition, the results of Krahl and Scerbo (1997) show that optimal performance may be obtained with less practice from operators in an adaptive environment when paired with a more skilled partner.

Collectively, both the operator measurement and modeling methodologies each have merits and disadvantages. Measurement has the advantage of being an "on-line" technique that can potentially respond to unpredictable changes in the operator's cognitive states. However, this method is only as good as the sensitivity and diagnosticity of the measurement technology. Performance measurement also occurs "after the fact", i.e. after a point in time when adaptation may be needed to compensate for substandard performance. Modeling techniques have the advantage that they can be implemented off-line and easily incorporated into rule-based expert systems. However, this method requires a valid model, and many models may be required to deal with all aspects of human operator performance in complex task environments. Physiological methods are considered below. Finally, because each of these methods have advantages and limitations, hybrid methods that combine aspects of each have been proposed (Parasuraman et al., 1992).

Psychophysiological Assessment. The last method that has been proposed for implementing adaptive automation involves the use of psychophysiological measures and represents the primary focus

of the present paper. Although the use of other methods such as operator modeling, performance measurement, etc. have merits, there are several advantages to such a system (Byrne, & Parasuraman, 1996; Gomer, 1981; Parasuraman et al., 1992). In certain applications, these advantages may be sufficient to overcome the disadvantages of cost, user acceptance, etc. associated with the use of these measures. First, psychophysiological measures, unlike most behavioral measures (with the exception of continuous motor tasks) can be obtained continuously. In many systems where the operator is placed in a supervisory role, very few overt responses (e.g. button presses) may be made even though the operator is engaged in considerable cognitive activity. In such a situation the behavioral measure provides an impoverished sample of the mental activity of the operator. Psychophysiological measures, on the other hand, may be recorded continuously without respect to overt responses and may provide a measure of the covert activities of the human operator. In other words, psychophysiological measures have higher *bandwidth* than behavioral or performance measures. Second, in some instances, psychophysiological measures may provide more information when coupled with behavioral measures than behavioral measures alone.   For example, changes in reaction time may reflect contributions of both central processing (working memory) and response-related processing to workload. However, when coupled with P300 amplitude and latency changes of the ERP, (discussed in later section) such changes may be more precisely localized to central processing stages than to response-related processing (Donchin et al., 1986). Furthermore, measures of brain function can indicate not only *when* an operator is overloaded, drowsy, or fatigued, but also *which* brain networks and circuits may be affected. This could potentially offer new avenues for adaptive "intervention" to optimize performance.

Despite these advantages, it must be recognized that several critical conceptual and technical issues must be tackled before psychophysiological adaptive systems could be fielded. The criteria of sensitivity and diagnosticity that apply to behavioral measures apply as forcefully to psychophysiological measures.   In fact one could make the argument that the sensitivity issue applies more stringently. This is because it is generally possible to attach some meaning to absolute values of behavioral measures, even with only limited knowledge of the stimulus context. For example, a reaction time of 200 msec or an accuracy score of 95% can be taken to represent highly efficient performance, without having to know the task context. The meaning of a P300 amplitude of 15 µV, on the other hand, cannot be determined without details of the experimental and recording conditions. Also, as Kramer (1991) indicates, they are often confounded with other sources of noise. Major technical problems (e.g. artifact-free recording in noisy cockpit environments; reliable single-trial recordings, etc.) also have to be solved before psychophysiological measures could be used routinely in real working environments. In addition, factors such as reliability, cost, and user inconvenience and mistrust, must be dealt with.

In the following sections, current mental workload research with psychophysiological measures is reviewed. A survey of general physiological indices is presented first. Next, a review of recent research on EEG is presented. This is followed by a review of more advanced cortical measures.

SECTION II

Physiological Measures

The operator's workload and specifically, mental workload is a critical factor in the allotment of system control. Mental workload is generally defined as the difference between task workload demands and the capacity of the operator (Kantowitz, 1988; O'Donnell & Eggemeier, 1986). When mental workload is high there is little remaining operator capacity to perform other tasks. A critical aspect in the design of an adaptive automation system is the "decision criterion" for the shifting control of the system between the human operator and the system. Three potential sources for such a decision criterion are: 1) subjective assessments of the operator's workload (e.g., NASA-TLX), 2) performance of the operator on the primary task, and sometimes a secondary tasks, and 3) the physiological state of the operator. The majority of the studies reviewed here used all three of these sources for the assessment of mental workload, but this review focused on the effectiveness of current physiological indexes of workload.

There are several important considerations in the choice of an appropriate measure or a composite of measures to serve as the decision criterion. A primary consideration is the validity of any measure to be used as an index of workload (this presumes that the measure in question is reliable). Does it provide an accurate reflection of workload and is it able to effectively discriminate the levels of workload and the types of workload demands made on the operator? This reflects the notion of measurement sensitivity and diagnosticity (O'Donnell & Eggemeier, 1986). Sensitivity refers to the ability of a potential measure to discriminate between levels of workload. This is a rough assessment and is often done by comparing baseline-resting states with the work state. Diagnosticity, refers to a finer grain assessment of whether the measure can differentiate the levels and types of workload demands (e.g., physical, automatic tasks, cognitive tasks). Also, the ease of measurement is important, especially when recording in an operational environment, and the operator must be willing to tolerate the measurement procedure. Related to this is the operator's acceptance of the measure as a criterion for system control. Additionally, the assessment should allow a timely assessment of the operator's state, so that system decisions are not delayed. A general advantage of many physiological measures is the ability to have continuous, on-line recordings. Lastly, the measurement procedure should not interfere with the operator's performance.

Candidate Physiological Measures

The present review is based on recent evidence for the use of physiological measures to assess workload. Specifically, it addresses the empirical literature published since 1995. This approach was adopted to complement earlier reviews (see for example, Byrne & Parasurman, 1996; Kahneman, 1973; Kramer, Trejo, & Humphrey, 1996; Parasuraman, 1990) and to focus on those measures that appear promising for adaptive automation.

Eye blink. A number of studies have examined the usefulness of the eye blink as an index of mental workload. This work is concerned with the reflexive eye blink as opposed to the voluntary blink. The reflexive nature of the eye blink is thought to reflect general arousal due to 1) the proximity of the facial nerves responsible for the eye blink and the medullary structures responsible for arousal, 2) the suggestion that these midbrain reticular formation structures have some role in the integration of ocular activities, and 3) the lack of identifiable triggers for reflexive blinks (Morris & Miller, 1996).

A range of measures have been derived from the eye blink recordings. Eye blink rate, variability of blink rate, blink amplitude, and closure duration are the most commonly used. Seven studies examined the value of the eye blink in the assessment of operator workload. Three studies used operational environments (flying, driving) and four simulated tasks. They all used blink rate as a measure of workload. Prior research has suggested that eye blink rate decreases as visual workload increases (Fogarty & Stern, 1989; Stern, Walrath, & Goldstein, 1984).

Eye blinks during laboratory tasks. Veltman and Gaillard (1996) used a pursuit task in a flight simulator with a secondary continuous auditory memory task (CMT). Each flight scenario was broken into discrete segments for analysis: rest, flight, flight with CMT, landing, and after landing.   A comparison of these segments found that blink interval for the landing segment, the most demanding segment, was longer than for all other flight and rest segments, which were not different. Also, the duration of blinks was longer during rest than during the flight segments and blink duration was shortest during the landing segment.   In a second study they manipulated the difficulty of a pursuit/tunnel task by varying the angle of the horizontal and vertical turn requirements which produced four workload conditions (Veltman & Gaillard,1998). Four measures of eyeblink were analyzed: interval, duration, time to close, and amplitude. A comparison between a resting baseline and the tracking task found that the blink interval was shorter (high blink rate) during rest and the time to close the eye was the longest during rest. A comparison within levels of task difficulty showed that blink interval and heart period were responsive and both measures decreased (higher blink rate and faster HR) as task difficulty increased.

Fournier, Wilson, and Swain (1999) used the Multiple Attribute Task Battery (MATB; Comstock & Arnegard, 1992). They had a single task (communication task) condition and three multiple task conditions that differed in workload. There was no resting baseline comparison, but the comparison of the single task to the multi-task conditions indicated that blink duration, rate and amplitude differentiated between these two conditions. Essentially, with multi-task workload the blink duration was shorter, the blink amplitude was greater and the blink rate was slower. However, a comparison among the three levels of multiple task workload found no differences. Performance measures discriminated among the multi-task workload level (as did cardiovascular indexes), so the lack of eye blink measures indicated poor diagnosticity and not a design failure to create distinct workloads.

 Backs, Ryan, and Wilson (1994) used a tracking task that factorially combined two levels of physical workload with three levels of perceptual/cognitive workload. Blink rate decreased from baseline levels during task performance. However, blink rate did not differ among the six tracking workload conditions. The authors also recorded respiration and heart activity measures which did differentiate among the workload conditions. This suggests that blink rate may be most useful as an index of the presence of workload, but not a good diagnostic choice for discriminating among levels of demand.

Eye blinks during operational tasks. Verwey and Veltman (1996) had subjects drive an automobile over a 40km route. During this task, they introduced a continuous auditory memory task (CMT), which required the subject to keep a running tally of the number of targets detected. The duration of the secondary task was varied. A comparison of blink intervals during the auditory CMT with a control condition, found that blink interval increased (lower blink rate) with task duration (10s, 30s and 60s). This finding is partly due to the opposite trend for the control condition, where eye blink interval decreased with CMT duration. The authors suggest that eyeblink had limited sensitivity to the CMT.

Hankins and Wilson (1998) recorded eye blinks during a flight in a single engine aircraft. The flight was divided into 19 phases comprised of four basic categories: ground based preflight, VFR, IFR, and IFR at high speeds. Blink rate varied over the phases of the flight, with blink rate lower during all of

the IFR segments when the pilot wore goggles to block out visual input from outside of the cockpit. There was approximately a 50 percent decrease in blink rate under these conditions (10 blink/min difference). In comparison, the range of variation in the Verwey and Veltman (1996) study was two blinks/min. Also, the one segment requiring the pilot to perform a touch-and-go landing showed a lower blink rate, which was comparable to the IFR phases. This shows that eye blink is a moderately sensitive measure of workload. By comparison, a similar analysis of pilot heart rate more clearly differentiated among the flight segments and presumably different workload demands.

Wilson, Fullenkamp and Davis (1994) recorded eye blinks in a laboratory setting and during flight. In flight, during one phase the pilot flew the plane and during another phase only observed. In the laboratory, the pilots performed a tracking task with two levels of difficulty. Eye blink rate and duration of eye closure were measured. For the flight, eye blink rates were higher during flight than during ground baseline testing. This finding is inconsistent with above results (Verwey & Weltman, 1996), in that blink rate decreases with increases in visual workload. However, pilot blink rate while flying versus observing showed a trend for lower blink rates when the pilot was flying. In the laboratory phase, blink rates decreased during the tracking task compared to a baseline condition. There was no difference between the two levels of tracking difficulty. Eye blink closure duration was shorter during the tracking task than baseline. There were no other task differences in closure duration.

Yamada (1998) recorded eye blink rates while subjects searched a visual array of 400 stimuli for target stimuli. Workload was manipulated by varying memory set size for the search task (1, 2, or 4). There was a major reduction (approximately 15 blinks/min) in eye blink rate from a rest baseline to the search task. There was a small linear trend for an increase in eye blink rates with memory set size (5 blinks/min). In a second study, school children were shown a boring animated video, performed a color Stroop task, and played a video game. Eye blink rate decreased as the visual and task demands increased, from the passive viewing of the video (15 blinks/min) to active performance during the video game (4 blinks/min).

Lastly, in the most comprehensive examination to date of eye blink measures, Morris and Miller (1996) studied the effect of fatigue on pilot performance in a flight simulator. Each pilot flew a 4.5h flight scenario which manipulated levels of workload. Over the course of the scenario there was an increase in performance error score. The authors used seven ocular measures: blink amplitude, blink duration, blink rate, long closure rate (eye closures of more than 500ms), peak saccade velocity, saccade rate, and saccade velocity. They used a stepwise multiple regression analysis to find which of these variables significantly predicted performance error. In two separate analyses, blink amplitude and long closure rate were the best predictors and accounted for over 50 percent of the variance. Blink amplitude decreased as the error score increased, since the eyelid droops with fatigue and there is a shorter distance to travel for closure. Long closure rates increased with the error score. This finding is important because it suggests a potentially valuable use of the eye blink amplitude and long closures in the evaluation of operator fatigue. This may indicate an extremely useful method for the assessment of driver fatigue.

Generally, the above results suggest that there is a slowing of blink rate with increased visual demand. Although this is a simple measure, sensitivity was consistently seen between a baseline and workload conditions. However, comparisons between levels of types of workload demands were not well differentiated. There is little evidence of measure diagnosticity.

Respiration. Respiration as an index of workload/performance has often been a secondary measure associated with the spectral analysis of cardiovascular function and usually as a control condition. In this section, the focus will be on the diagnosticity and sensitivity of respiration. The prior

literature has generally indicated that respiratory rate increases and depth decreases with increasing cognitive demands (Wientjes, 1992; Wilson & Eggemeier, 1991).

Respiration during laboratory tasks. A recent example of the use of respiration as a cardiovascular control variable is the work of Backs and his colleagues (Backs, 1997; Backs, Ryan, & Wilson, 1994). These two studies employed a compensatory tracking task and used a spectral analysis of heart rate variability. A portion of the heart rate variability spectrum (0.14-0.40Hz) is termed the Respiratory Sinus Arrythmia (RSA), and thought to reflect the changes in heart rate associated with respiration. Inspiration causes an acceleration and exhalation causes a slowing of heart rate (Porges & Byrne, 1992). Backs, et al.(1994) factorially combined the degree of tracking disturbance (low and high) and tracking order of control (velocity, mixed and acceleration) resulting in 6 workload conditions. Respiration was measured by thoracic and abdominal strain gauges. Two days of training on the tracking task preceded testing. Initial analysis indicated that respiration rate increased and depth decreased with the introduction of the task (an index of sensitivity) compared to a resting baseline. More important, is the ability of respiration to differentiate among the levels of tracking difficulty (diagnosticity). Respiration rate increased with increasing order of control difficulty, but depth was unchanged, while an increase in the tracking disturbance produced an *increase* in respiration depth, from the already shallow level, but there was no comparable change in rate. It is important to stress that analysis of the tracking task yielded significant performance effects for both factors and their interaction. If this were not the case, any discussion of diagnosticity would be moot. A principal components analysis (PCA) of the eleven dependent variables used in this study yielded five factors with both respiration measures loading on the primary factor with eye blink rate and accounting for approximately 26% of the variance.

In a follow-up study (Backs, 1997), the intermediate level of the order-of-control dimension of the tracking task was dropped and an auditory oddball task was added. The latter change was to add cognitive workload. Again, a comparison with a baseline-resting condition found an increase in respiratory rate and a decrease in depth. A comparison of the task manipulations found no effect of the order-of-control manipulation, but increasing the task disturbance did result in an increase in respiratory depth. When the requirement to attend to the secondary task (oddball) was added to the current workload, the respiration rate decreased with the higher level of tracking task disturbance. Analysis of tracking performance found greater error with increasing levels of disturbance and order of control and their interaction.

These two studies with comparable methodologies do not offer strong evidence of diagnostic reliability. In the first study, the order-of-control manipulation (perceptual/central processing workload demand) of tracking difficulty produced increases in respiration rates, but neither rate nor depth were affected in the second study. The studies found a common effect of tracking disturbance (physical workload demand) with an increased depth of respiration under increased difficulty. Also, both studies found that respiration rate was unaffected by increased tracking disturbance. However, the addition of cognitive load in the second study resulted in a decrease in respiration rate with increasing task disturbance.

Fournier, et al.(1999) used the Multiple Attribute Task Battery and varied the workload using a single communication task to which additional tasks were added to produce three multi-task conditions of low, medium and high workload. Performance on the multi-tasks indicated that they were significantly different in difficulty. Respiration rate and amplitude were measured. There were no effects for respiration amplitude. Respiration rate was higher for the multi-task conditions than the single task condition, but the three multi-task conditions were not different. Respiration rate was sensitive to the

major shift in workload from the single communication task to the multi-tasks, but was not able to differentiate among the three multi-tasks.

Sammer (1998) examined respiratory changes in response to a physical lever moving task, a mental arithmetic task, and the combination of both tasks. Respiration was recorded with a thoracic strain gauge. The analyses for this study focused on the three task conditions without a baseline condition. The analysis of respiratory power (a linearly detrended measure of respiration, 0.2-0.5Hz, which removed changes due to time) found no differences among the three conditions. An analysis of non-linear effects suggested that respiratory activity was affected by the physical demands of the task; the physical and combined conditions were equivalent with greater respiratory activity than the mental workload condition.

Respiration during operational tasks. Wilson, et al. (1994) recorded respiration rates from F-4 pilots during a laboratory tracking task with two levels of difficulty and an actual flight. For the flight component, the major comparison was between a baseline ground segment, a demanding low-level flight segment flown by the pilot and a less demanding cruising flight segment with the pilot as an observer. There were no differences in respiration rates for the three flight phases. A caveat for the lack of results for the flight data was that the aircraft's breathing system may have impeded respiration activity when compared to the preflight ground baseline. For the laboratory part of the study, respiration rates were higher than those seen during the flight phase. Also, respiration rates were not different between the two levels of tracking difficulty, although there was a trend for higher respiration rates in the more difficult condition. Unfortunately, the study reports no performance data for the two levels of difficulty.

A study by Veltman and Gaillard (1996) is included under this heading, because their flight simulator and flight scenarios approximated an operational environment more so than the much simpler laboratory tasks included in the above section. They used different flight scenarios combined with a brief (4 min) secondary auditory continuous memory task (CMT). This CMT required the subject to recognize a target in a series of non-targets and to keep a running tally of the number of targets detected. Performance in the simulator was a derived error score. Flight performance was poorer during the last minute of the CMT compared to the flight task alone. The flight scenarios were broken down into four components. Respiration was decomposed into two spectral bandwidths, but their results were the same. Respiration was deeper and slower after landing, but there were no differences from the other segments: rest, flight, flight plus the CMT, and landing. This study provides little evidence of sensitivity or diagnosticity for respiration.

Overall, these studies provide support for the notion that respiration (rate, depth) is sensitive to workload demands when compared to baseline-resting conditions. There is an increase in respiratory rate and a decrease in depth when workloads are compared to rest. Also, there is some evidence that respiration may be diagnostic for levels of workload. However, it is not clear that it can differentiate among the types of workload (e.g., cognitive, physical).

Cardiovascular Activity. Cardiovascular activity is the most commonly used index of cognitive workload. It is a relatively unobtrusive physiological measure and it appears to be readily accepted by subjects in an operational environment. In a recent review of applied physiological measurement techniques, Fahrenberg and Wientjes (2000) ranked cardiovascular measurement as the most suitable for field studies due to its reliability, unobtrusiveness and ease of recording. Of the studies in this review, 21 used one or more indexes derived from heart activity, and many studies combined this with other physiological indexes.

The earlier literature reports a consistent pattern of cardiovascular activity from laboratory and field studies; heart rate increases and heart rate variability (HRV) decreases as a function of increases in cognitive workload (Wilson, 1992).

One trend in the use of cardiovascular function as a measure of workload, specifically mental workload, is the assertion that heart rate is not a sensitive or an especially diagnostic measure. There are two reasons for this. First, it is affected by physical exertion and second, it does not provide information about the underlying functioning of the sympathetic and parasympathetic nervous systems. Several authors feel that it is only through an understanding of the relative contributions of the autonomic nervous system on cardiovascular functioning that good diagnosticity of mental workload can be achieved (Backs, 1995; Berntson, Cacioppo, & Quigley, 1993; Jorna, 1992; Mulder, Mulder, Meijman, Veldman, & van Roon, 2000).

Spectral analysis of variations in heart rhythm is proposed to provide an index of the relative contributions of the underlying components: parasympathetic inhibition and sympathetic activation. Spectral analysis of heart rhythm is typically segmented into three distinct bandwidths: 1) low frequency (0.02-0.06Hz), which is associated with temperature regulation; 2) mid-frequency (0.07-0.14Hz), which is affected by blood pressure regulation and cognitive effort; 3) hi-frequency (0.15-0.50Hz) which is associated with the effects of respiration on heart rate, the respiratory sinus arrhythmia (RSA). The mid-frequency bandwidth, is associated with the combined activity of the parasympathetic and sympathetic systems, while the RSA is influenced by parasympathetic activity. Mulder, et al. (2000) suggest that suppression of the mid-frequency bandwidth is "very diagnostic" of the operation of attention-demanding cognitive control mechanisms (i.e., mental workload). Another measure, residual heart rate (RHR), has been developed to reflect the impact of sympathetic activation on heart rhythm. Residual heart rate is the heart rate that remains after removing the part linearly related to respiratory activity, RSA.

The ultimate value of these complex measures will be resolved empirically. Does the present research or will the future research indicate that the more complex component analyses are better predictors of cognitive workload? Are they better *diagnostic* indexes of mental workload?

Cardiovascular activity in laboratory tasks. Boutcher, Nugent, McClaren and Weltman (1998) challenged aerobically fit men and two control groups with the Stroop task and an arithmetic task (subtraction of a series of spoken numbers). The premise for this study was that fit males have a greater vagal tone, increased parasympathetic activity, which may affect reactivity to mental challenge. Of relevance to the present review was the effect of the two cognitive tasks on cardiovascular function as measured by HRV in mid- and high-bands. The relevant comparison was between baseline and the given task. For the arithmetic task there were no significant changes for either HRV band, although there was a trend for a reduction in variability during the task. However, the same comparison of the Stroop task revealed a significant reduction of HRV in both bands. Sammer (1998) compared a physical task (moving a lever when a cue appears), a cognitive task (counting target letters appearing in a serial array) and a combination of both task (dual task). Heart period (IBI), and HRV in the low (0.01-0.05Hz), mid- (0.06-0.16Hz), and high (0.2-0.4Hz) bands were computed. A comparison among the three tasks (no baseline comparison was included) found significant effects for all four measures. Heart period was largest (slowest HR) for the cognitive task, intermediate for the physical task, and smallest for the dual task (faster HR). Over the spectral bands, HRV was less for the dual task and greater for the physical and cognitive tasks, which were not different. Simply, heart period differentiated among the tasks better than the HRV measures. Fournier, et al., (1999) used the Multiple Attribute Task Battery and created four discrete tasks: a single task and three multiple tasks of increasing difficulty. HR and HRV, in the mid-

and high- bands, were the dependent variables. In an initial comparison of the single task condition to the multiple tasks, all three measures were different: HR was higher and HRV in both bands was reduced in the multi-task conditions. A subsequent comparison among the three multiple tasks found that HR differentiated between the highest difficulty task (higher HR) and the other two multiple tasks, whereas only the mid-band HRV was different between the high and low difficulty multiple tasks.

The above studies suggest that the simple measure of HR was more sensitive and diagnostic that the HRV measure. Also, there was little evidence that the HRV mid- band was more sensitive to mental challenges than the other spectral bands.

Backs and his colleagues (Backs, 1995; 1997; Backs, Lenneman, & Sicard, 1999; Backs, Ryan, & Wilson, 1994) have proposed a complex decomposition of cardiovascular activity into autonomic dimensions (parasympathetic and sympathetic activity) in order to generate a more sensitive and diagnostic measure of workload. They conducted a series of studies using a single-axis, compensatory tracking task that varied physical demand by either 1) requiring different amounts of force to move the joystick, or 2) varying the disturbance value of the cursor movement, and varied cognitive/perceptual load by manipulating order-of-control (velocity, acceleration, mixed). Also, secondary tasks were added to increase discrete workloads (eg., target recognition varying set size, mathematical tasks, oddball counting tasks).

Backs claims that HR does not fare well as a diagnostic indicator of workload. By employing a principal components analysis, it is possible to use the more or less standard measures of cardiovascular activity: heart rate, or inversely, heart period, the heart rate variability spectrum broken down into three frequency bandwidths thought to correspond to sources of autonomic activation, and residual heart period. The latter, RHP is usually a poor index of workload. The other measures have been shown to have reasonable value in detecting extremes in workload (eg., resting vs. work), as there is some evidence for diagnosticity, especially for HP and HR and occasionally, mid-band HRV.   The PCA generally produces one factor associated with parasympathetic activity. The most consistent findings indicate that the four variables load on two factors, typically accounting for approximately 50% and 30% of the variance. The first factor is associated with parasympathetic activity and loads mid-band HRV and RSA, while the second factor is associated with sympathetic activity and loads HP and Residual HP.   The factor loadings of these four variables are used to produce parasympathetic and sympathetic component scores, which are then subjected to the same analyses used for the original variables. To the extent that these composite scores produce more consistent outcomes, they will be valuable as workload diagnostic tools.

Cardiovascular activity in quasi-operational tasks. Rau (1996) used simulations of an electrical distribution system (electroenergy network) with trained operators. Two operators worked during each scenario, one as the shift leader and the other as a co-operator. Three types of tasks performed during system operation were chosen to reflect different levels of cognitive workload. Comparisons were made among these three workload conditions using HR. Heart rate was lower for the least demanding condition and increased during the more demanding conditions, which were not different. Also, the shift leader showed higher HR during the most demanding task than the co-operator.

Veltman and Gaillard (1996) analyzed IBI and mid- and high- band HRV from subjects working in a flight simulator. A secondary CMT was included to increase cognitive workload. For analysis, the flight scenario was divided into five segments: rest periods, flight, flight with CMT, landing, post landing. IBI was longer (slower HR) during the rest periods than all flight segments, but no effect was seen for HRV bands. A comparison among the four remaining "flight" segments found that IBI was

shorter (faster HR) for the flight with CMT and landing segments than for flight alone (diagnostic), while HRV in both bands was lower and equal for the three flight segments, than during the post-landing segment, which showed greater variability. Veltman and Gaillard (1998) used pilots in a flight simulator with a flight scenario with 4 levels of maneuvering/pursuit difficulty. They measured heart period IBI and mid- and high-band HRV. The IBI was longer and HRVs were greater during a resting baseline than all flight segments. Comparisons among the levels of task difficulty found that IBI was diagnostic, with IBI decreasing (faster HR) as the task difficulty increased. HRV was not sensitive to task differences.

Tattersall and Hockey (1995) examined flight engineers in a flight simulator using HR and the mid- and high-bands of the HRV spectrum. The flight phase was divided into the takeoff/landing segment, and three levels of cognitive task demands during the cruising segment: system monitoring, routine fault correction, and problem solving. Compared to a baseline condition, HR increased and HRVs decreased during flight segments. During the flight segments, HR was higher during takeoff/landing than the in-flight cognitive tasks, which were not different. For HRV, only the mid-band was significant with more suppression of variability for the demanding problem solving tasks than for the other two task types.

Backs, et al.(1999) used pilots in a Boeing 747 simulator with low and high workload scenarios. Five segments of the two flight scenarios (takeoff, top of climb, cruise, approach, and landing) were analyzed.   Four cardiovascular measures were derived: Heart Period (interbeat interval), mid-band HRV, high- band HRV or Respiratory Sinus Arrhythmia (RSA), and Residual Heart Period. RHP is the heart period that remains after removing RSA, resulting in an index of sympathetic input to the heart. This measure is related to Residual Heart Rate, which removes the linearly related effect of respiratory activity on heart rate (Mulder, et al., 2000). A principal components analysis of these four variables estimated the relative contribution of the parasympathetic and sympathetic nervous systems and produced a score for each component. Importantly, the authors present reliabilities for each of the six measures in this design and HP was clearly the only statistically and clinically reliable measure. HP was shorter (faster HR) for the high workload scenario. Additionally, HP increased (slower HR) from takeoff to the cruise segment. HRV changes across flight segments are consistent with HP with suppression of HRV with higher workloads.

In summary, the work in simulators indicates that heart rate increases, interbeat interval decreases and heart rate variability decreases with increased workload demands. This is clear when a resting baseline is contrasted with workloads. So, these measures are sensitive to major differences in workload. Less clear is the pattern with levels of mental task demands. HR and IBI seem to show evidence of differentiating among task demands, while the mid-band HRV shows less compelling evidence for differentiating among task workloads.

Cardiovascular activity in operational tasks. Seven studies were reviewed that used operational environments.   All used either heart rate, or interbeat interval, and heart rate variability as the cardiovascular indexes of workload. The more complex analyses using principal components were not used in these studies. In general, the results of these studies produced small effects and showed little evidence of diagnosticity. However, overall they suggest a pattern of decreasing HRV and increasing HR (shorter IBI) with increasing workloads. Some of the inherent difficulties of real work environments is that many of the tasks are very brief, the work environment is complex making it hard to isolate task components (mental, physical), and the need to control movement artifacts.

Wilson, et al. (1994) analyzed interbeat interval (IBI), and HRV in the mid- and high- frequency ranges (0.06-0.12Hz, 0.12-0.40Hz). They found that IBI was shorter (faster HR) in flight than on the

ground. During the segment flown by the pilot, high-band HRV was decreased compared to a segment flown by the weapons officer. There was no effect in the HRV-mid band, which is thought to reflect differences in workload. Hankins and Wilson (1998) compared 19 discrete flight segments flown in a single engine plane. They used HR, IBI, HRV in mid- and high- bands (0.06-0.14Hz, 0.15-0.40Hz). HR was the highest during takeoff and landing, lower for the instrument flight segments, and lowest when climbing and on the ground. HRV for both bands were highly correlated (0.91) and showed a mirror image pattern to that of HR, though with fewer statistical differences among flight segments. They conclude that HRV was not sensitive to variations of in-flight mental workload demands and that greater differences in task load may be necessary to see effects.

Egelund (1982) compared standard deviation of HR (total HRV), mid-band HRV, and HR for automobile drivers over a 340km highway circuit. Mid-band HRV showed a significant linear increase as a function of distance, although HRV decreased in the last two segments. There was no effect for the total HRV which suggest the sensitivity of the mid-band measure of HRV.   HR showed a significant quadratic effect as a function of distance, although HR was essentially unchanged until a slight 1-2 bpm change for the last two segments of a 12 segment route. Gobel, et al. (1998) analyzed HR and HRV of bus drivers on an in-city route, with a range of different activities (e.g., turning, opening doors). Most of these events were very brief, less than 10 seconds in duration, which is less than the 10-20 seconds used to process the HR and HRV measures. HR was lowest for the rest period and greatest for the ticket invalidating, closing doors, and making notes. Conversely, HRV was greatest for rest and least for the ticket invalidation and activating the windshield wipers. Myrtek, Deutschmann-Janicke, Strohmaier, Zimmerman, Lawrenz, Brugner, and Muller (1994) did a similar study to Gobel, et al. (1998) using train drivers and analyzed HRV and HR. They compared different ride segments (e.g., fast and slow speed segments, braking, standstill). HRV while moving was lower than at "standstill" except for the highest speeds (100-200kmh) which was not different. Also, HR was the lowest at the highest speeds. The inverse pattern suggests that the high speed condition is a low workload (monotonous) condition. Finally, Verwey and Veltman (1996) used automobile driving as a platform for the well controlled introduction of two demanding secondary tasks (visual counting task, continuous auditory memory task) for 10, 30, or 60 second intervals. They analyzed IBI and HRV. Driving performance was degraded by the introduction of these tasks. Generally, there were no main effects for either cadiovascular measure. However, during the longest presentation of the auditory CMT, IBI decreased (HR increased) and HRV decreased. Also, comparisons between standing still and driving revealed less HRV while driving. There was no comparable effect for IBI. For this design, IBI and HRV were only sensitive to the longer workloads; shorter workloads produced no apparent effects.

Lastly, in an interesting and very different study, Myrtek, Weber, Brugner, and Muller (1996) had university students wear a cardiovascular recorder for a day and indicate specific activities on an event recorded. They analyzed HR and HRV. A comparison between global academic activities and leisure activities indicated that HRV was greater for leisure activities, but there was no comparable effect for HR. Also, students classified as chronically stressed, evidenced higher HR and lower HRV at the university compared to being home.

Speech Measures. Within the parameters of this section one article argued for the use of speech as an index of workload. Brenner, Doherty, and Shipp (1994) suggested that speech would be a valuable assessment index of workload because of its unobtrusive nature. A discrete trial tracking task was employed with two levels of tracking difficulty produced by the instability of the target's movements. On half of the trials a tone signaled the subjects to count out loud from "90 to 100" as quickly as possible.

This tone occurred at 10-second intervals during the trial and resulted in nine repetitions of the spoken number sequence. These repeated sequences served as the basis of the speech-workload measure. The authors used several speech measures (e.g., rate, loudness, fundamental frequency, changes in fundamental). A direct comparison of speech under the two levels of task difficulty found that speech rate, fundamental speech frequency and speech loudness discriminated between the task levels. The speech rate was faster, the fundamental speech frequency was higher and speech was louder for the difficult tracking condition.   Although statistically significant, these effects were small; loudness increased by 1 decibel, fundamental frequency by 2 Hz and the speech rate by 4%.

The speech measures may serve as useful indexes of workload. The present work was based on a contrived counting task and only two discrete trials served as the basis for this analysis. Also, there was a baseline condition where the subjects repeated the number sequence without the tracking task, but due to missing data it was not included in the analysis. Lastly, the value of these measures rests on the natural occurrence of speaking in the target operational environment. The environment would need to generate a sufficient amount of speech to serve as a continuous assessment of workload.

Multiple Measures

Several of the studies in this review directly compared the value of two or more of the above measures. It is instructive to use their findings as a simple scoring system of the sensitivity and diagnosticity of the physiological measures. For instance, Verwey and Veltman (1996) evaluated the relative sensitivity of heart rate (IBI), HRV, and eyeblink for the two secondary loading (auditory CMT and visual CMT) tasks. Of the three measures, HRV appears to discriminate between the presence and absence of both secondary tasks and between driving and standing still. Backs, et al. (1994) compared blink rate, RSA, mid-band HRV, HP, respiration depth and rate, and electromyogram using tracking performance. A factor analysis of all of the experimental physiological, subjective, and performance dependent variables yielded five factors. The first factor loaded blink rate (.70), respiration depth (-.94) and respiration rate (.93). The second factor loaded RSA (.89), HRV (.63), and HP (.79). The first two factors account for 46.3% of the variance. EMG (.63) loaded on the fourth factor. All of the physiological measures were sensitive to the imposition of workload. Respiration rate and depth differentiated among the two tracking workload factors (physical and perceptual/central processing), and HRV discriminated the physical dimension of the tracking task, but not the perceptual/central processing dimension. Hankins and Wilson (1998) compared eyeblink and cardiac activity during flight. Although, not a well controlled design (flight segments overlap in task requirements and vary in length), the results suggest that HR and HRV in mid- and high- bands are more responsive to variation in flight segments than eye blink. Veltman and Gaillard (1998) used HP, HRV, respiration, eye blink, and blood pressure to examine the effect of flight simulator scenarios of four levels of difficulty and the impact of an auditory CMT. In a comparison between rest conditions and workloads, all of the variables showed significant changes. In a comparison among the task difficulty levels, HP and blink interval proved to be diagnostic, both decreasing with additional tracking/visual workload. Lastly, Fournier, et al. (1999) compared HR, HRV mid- and high band, respiration amplitude and rate, and eye blink rate, duration and amplitude. They examined single and multiple tasks on the MATB. All measures with the exception of respiration rate differentiated between a single task workload and a multi-task workload. In diagnostic comparisons

among the three multi-task conditions, only the cardiovascular measures were helpful. HR and mid-band HRV were effective at differentiating among the tasks.

This summary indicates that cardiovascular measures, typically HR, HP and HRV provide the clearest evidence of diagnostic value. Other measures have shown less reliable indications of the same. Although, this summary is far from exhaustive, it provides a review of the major research techniques and analyses used in the area. These studies employed a range of tasks, with manipulations of task loads and the types of task demand. Some of the failures to find diagnostic effects of the physiological measures may be due to the adequacy of these manipulations of task load. Despite this, the present evidence supports the use of cardiovascular activity. It is a relatively simple measure to record and is minimally intrusive. Further, the attempts to better define the dimensions of cardiovascular activity (Backs, 1995; Backs, et al., 1999; Mulder, et al., 2000) may provide the most fruitful area for the development of adaptive automation systems.

SECTION III

The Efficacy of Cortical Measures for Adaptive Automation

A basic assumption in the use of EEG for adaptive automation is that some aspect(s) of the EEG may be used as an index of mental workload which in turn may be employed as a modulator of task parameters. In order for this to occur there must be a demonstrable correlation between various EEG parameters and such psychological variables as arousal, attention, and workload. Since these variables are multidimensional and typically not easy to define (cf. Scerbo, et al., 1998), and since there are a wide variety of EEG parameters that might be employed, the task of correlating the two is somewhat daunting. For example, Vidulich, Stratton, Crabtree, and Wilson (1994) attempted to correlate different physiological measures with changes in workload and *situational awareness* in a simulated air-to-ground flight mission. Theta power increased and alpha power decreased in a GPS Night Display condition. The authors noted that the changes in EEG suggested greater cognitive demand for maintaining situational awareness. However, they raised the question: would the effects be better conceptualized as workload measures or as measures of situational awareness? Thus, unambiguous definitions of constructs is a continuing problem.

Another potential problem in looking for relationships between EEG and behavior involves how well performance reflects mental effort. As task demands increase, maintaining task performance may require higher levels of physiological activation, "subjective strain", and enhanced attention (Hockey, 1997). Task performance, however, often remains stable. As a consequence, changes in physiological measures would inappropriately be assumed not to correlate with behavioral measures of workload. Thus, while the EEG may provide a unique measure of mental workload, one must be careful when attempting to validate it against behavior.

The following is a selective review of the literature in which different approaches to this question were addressed. First, the function of different EEG bandwidths is reviewed. Next, studies which attempt to control these bandwidths through neurofeedback (see Evans & Abarbanel, 1999 for more detailed discussion) in order to alter behavior is evaluated since the ability to alter these bandwidths has direct application to their incorporation into an adaptive automation system. Finally, the potential application of these techniques to an adaptive automation environment is examined.

EEG Bandwidths: Arousal, Attention, and Workload

Variables that may affect the recorded signal. There are a number of variables to consider when measuring EEG that may affect the nature of the recording and subsequent interpretation of results. First, the number and location of the recording sites typically varies from study to study. While just about everyone uses the international 10-20 system for electrode placement, the number of recording sites may vary from two to over 30. Some experiments only record activity over a specific area such as the occipital lobe (e.g. sites O1 and O2) while others record activity over each lobe in each hemisphere. Second, signal processing may vary across studies. Typically, experiments will examine EEG bandwidths (i.e. number of waves per second or Hz) with traditional divisions being approximately as follows: delta – 0.5-3 Hz, theta – 4-7 Hz, alpha – 8-12 Hz, beta – 13-30 Hz. However, there are a number of variations of these categories. For example, the alpha bandwidth may be divided into two or even three subcategories.

Beta waves may also be divided into a number of categories. Some investigators examine a special type of beta found over the sensorimotor cortex at 12-15 Hz. Sometimes high beta or beta over 30 Hz is referred to as gamma waves. Finally, a number of investigators argue for the analysis of very narrow bandwidths (e.g. 1 Hz; see below).

The techniques used for deriving these different bandwidths may also vary across studies. Typically, a Fast Fourier transform is performed on the digitized (and windowed) EEG signal. This is a time series technique designed to decompose a signal into its frequency components (Ray, 1990). This procedure is also referred to as a power spectral analysis. Most studies report the amount of power within a specific bandwidth. Power is defined as μV^2/cycle/sec. Absolute power reflects not only the amplitude of the brain generated signal but also nonbrain factors such as scalp resistance, skull thickness, and different conductance properties of the skull, dura, and scalp. To control for these variables, relative power, defined as the power in a frequency band divided by total power, is calculated (Abarbanel, 1999). Some studies report absolute power and some report relative power.

Another important consideration when examining the results of different studies is whether bipolar or monopolar recordings were used. With bipolar recordings the EEG signal is recorded by comparing the signal between two active sites, the resulting signal being that which is not common to both sites. With monopolar recordings one electrode serves as a reference for all other electrodes. The reference electrode is suppose to be over an inactive or neutral site such as the ear lobes (sometimes linked ears are used), the nasion, mastoid, or vertex. However, none of these sites are completely neutral and the actual reference used varies across studies.

Basic assumptions. When studying the relationship between EEG and mental workload several basic assumptions are made: 1) EEG measures can be used as an index of arousal and attention, 2) variations in arousal and attention reflect variations in mental workload and 3) variations in task parameters which affect mental workload can be related to variations in EEG. Research conducted in the 1960's and 1970's attempted to demonstrate such a relationship. Unfortunately, this relatively straightforward assumption proved difficult to confirm. For example, while differences in arousal level can account for differences in overall vigilance performance, reductions in arousal may occur without corresponding reductions in vigilance. Conversely, reductions in vigilance may occur even when arousal states are maintained (Gale, 1977; Parasuraman, 1983).

Numerous problems occur when trying to interpret much of the early work in this area. Often researchers recorded from only a few electrode sites (e.g., Davies & Krkovic, 1965; Gale, Davies, & Smallbone, 1977). Also, recording methodology varied from study to study (e.g. sites used for reference) as did the nature of the task (e.g. auditory vs. visual signals). Varying EEG definitions of what constitutes arousal and attention (cf. Scerbo, et al., 1998) have added to the confusion. Typically, changes in EEG from frequencies in the beta (13-30 Hz.) range to frequencies in alpha (8-12 Hz) or theta (4-7 Hz) range were assumed to reflect decreases in arousal. Unfortunately, there does not seem to be a universal agreement on the dividing line between different bandwidths or on what aspects of the EEG across these frequency ranges reflect different levels of alertness. This confusion may be due, in part, to the assumption that arousal is a unidimensional construct that varies from sleep to high states of alertness. If there are in fact qualitatively different states of arousal reflected by different patterns of EEG measured over different cortical sites, then trying to demonstrate a simple relationship between arousal, attention, and/or workload (as measured by EEG) and variations in performance would prove to be difficult (Streitberg, Röhmel, Herrmann, & Kubick, 1987). Further, some investigators have defined the bandwidths based on a Principal Components Analysis (PCA) while others have argued that the

bandwidths relevant for specific behavioral functions should be defined on an individual basis (cf. Klimesch, 1999). Even for those studies using a PCA, there is a lack of agreement as to the bandwidth cutoffs. Klimesch points out that divergent results regarding PCA defined bandwidth limits could be due to 1) some studies using absolute power while others use relative power, 2) electrode placement, 3) mono- vs bipolar recording, and 4) task type.

      Theta and alpha rhythms and workload. Theta rhythms have typically been defined as the EEG bandwidth ranging between 4-7 Hz., though some definitions may vary slightly (e.g. 5-7 or 4-6 Hz.). In 1977, Schacter reviewed the relationship between theta and psychological phenomena. One major category of behavior which he related to theta is a hypnagogic state in which individuals are drowsy and have a marked decrease in awareness of their environment. Theta during this state was characterized as low voltage, irregular activity "…carrying superimposed faster components in the beta range…" and spread diffusely over the cortex. Schacter stated "…that the theta activity observed during the hypnagogic period is indicative of a lowered pre-stimulus level of alertness which is accompanied by impaired ability to process and respond to environmental information." Sleep deprivation, for example, is associated with extreme drowsiness during which theta activity is dominant. Errors on signal detection tasks by sleep deprived subjects have been found to be associated with the occurrence of theta in a variety of studies.

      In marked contrast to the relationship between theta and a hypnagogic state, Schacter reported on numerous studies demonstrating a positive relationship between the occurrence of theta and problem solving, perceptual processing, and, in some cases, learning and memory. *Increases* in theta were typically associated with increases in mental workload as defined by task difficulty and stimulus complexity. Schacter felt that the enhanced theta seen with increased task difficulty was not related to non-specific increments in alertness. Interestingly, *decreased* theta was found to be associated with incorrect responses on a signal detection task (Daniel, 1967). Many of the studies reviewed by Schacter found increased theta over frontal sites. This contrasts with the association of theta with hypnagogic and sleep-deprived states, in which theta is diffusely spread over the scalp. Schacter reported that little information was available at the time on the amplitude and regularity of theta during problem solving tasks. Also, few studies had been done relating theta to learning and memory. This is no longer the case and will be discussed in a later section.

      Recent reports in which theta and alpha are related to performance seem to have some of the same problems seen in earlier studies, though in many cases the results do appear to be more consistent. Different studies record from different sites; some use monopolar and some bipolar recording; some studies report absolute power and some relative power; and different cutoffs/definitions of what constitutes a specific bandwidth are employed. However, as reported by Schacter over 20 years ago, recent studies which have examined the relationship between EEG, sleep deprivation, workload, and vigilance commonly report that long periods without sleep result in a vigilance decrement associated with an increase in theta and a decrease in alpha activity. Such results have been reported for sleep deprivation studies in the laboratory (e.g. Hasan, Kirvonen, Varri, Hakkiner, & Loula, 1993; Lorenzo, Ramos, Guevara, & Corsi-Cabrera, 1995) and in real world occupational settings such as truck drivers, train drivers, and airline pilots (Cabon, Coblentz, Mollard, & Fouillot, 1993; Gundel, Drescher, Maas, Samel, & Vejvoda, 1995; Miller, 1995). Nevertheless, a variety of studies also have reported that theta, especially when recorded from frontal sites, is related to increases in either attention, workload, or memory load.

      Problems of interpretation with regard to the relationship between physiological and psychological

variables continue to exist. Paus, Zatorre, Hofle, Caramanos, Gotman, Petrides, and Evans (1997) for example, reported a progressive increase in theta over a 60-minute auditory vigilance task. They argued, based on the EEG and cerebral blood flow, that the right hemisphere, especially the frontal and parietal cortex, is important in attentional processes. In this instance, increases in theta could be interpreted as being due to increased drowsiness. However, it is plausible that, as the task progressed, the workload for the subjects, in terms of maintaining attention, increased and this was what the increase in theta reflected.

Pennekamp, Bosel, Mecklinger and Ott (1994) had subjects perform a Mackworth clock test for 60 minutes and examined the EEG using a bipolar recording from P4 referenced to Cz and a Principal Components Analysis (PCA) to define the bandwidths. Theta power (5.5-7.0 Hz) was higher in the interval prior to a detected as compared to a non-detected target. No relationship between alpha and performance was reported. Unfortunately, the study examined EEG from only one site, which may account for the absence of a significant effect for alpha. Finally, although the authors examined the EEG relative to the time of stimulus presentation, no information was presented regarding changes in bandwidth power over the 60 minutes of the task. Comparisons of recordings from multiple sites, especially frontal and occipital sites, would have been of interest as would the use of a much lower event rate.

Valentino, Arruda, and Gold (1993) compared the EEG of good and poor performers on an auditory continuous performance task that lasted only 10 minutes (though they described it as a vigilance task; typical vigilance tasks last 30-60 minutes or longer). Bipolar recordings were taken from eight sites and absolute power in traditional bandwidths was analyzed. Good performers had higher levels of beta2 (17.5-25.0 Hz), especially in fronto-temporal and temporal left-hemisphere. A decrease in beta2 from the first to the second 5-minute block was the major EEG change that was consistent with changes in performance. Frontal theta increased from the resting condition to the first five minutes of the task, though the authors attributed this change to eye movement artifact. Given that the task only lasted for 10 minutes and that both event rate and target rate were relatively high (i.e. task difficulty was not manipulated), this study is difficult to evaluate in terms of vigilance performance.

Makeig and Inlow (1993) examined the correlation between EEG bandwidths and "local error rate" throughout a 60-minute auditory vigilance task that was performed with eyes closed. Local error rate was a technique devised to assess momentary lapses in alertness. EEG was recorded from 13 sites and referenced to the right mastoid. In contrast to Pennekamp, et al. (1994) EEG absolute power below 6-7 Hz was *positively* correlated with local error rate while power near 10 Hz was negatively correlated. Makeig and Jung (1995) replicated these results and argued that all performance-related changes in the EEG spectrum are confined to one principal component of spectral variance. The discrepancy with Pennekamp, et al.'s results could be due to task variables, method for assessing errors, or recording methodology, including number and location of electrode sites. For example, while Pennekamp, et al. used traditional vigilance task parameters in which targets were infrequent, Makeig and Inlow, in order to track local error rates presented 10 targets per minute. While they still observed vigilance decrements, it seems reasonable to assume that such high rates would produce differences in attention in comparison to one target per minute. Also, the local error rate was calculated over a time window of 32.8 seconds that advanced in steps of 1.64 seconds. Pennekamp, et al. examined the EEG related to each specific correct and incorrect response but did not look at EEG changes as a function of time on task.

Barcelo, Gale, and Hall (1995) examined EEG absolute power during visual orienting to stimuli varying in complexity and number. EEG was recorded from eight sites using linked mastoids as a reference. Recording began three seconds prior to stimulus onset and continued for 24 seconds. Although

changes were seen in all bands across time, the authors described a sharp increase in occipital theta during the first three seconds of stimulus presentation as an "outstanding feature." Number of stimuli caused a power reduction in alpha and beta but not theta. Complexity did not produce any significant effects. The authors argued that occipital but not frontal theta was related to attention as measured by the orienting response. Since this was a visual task and did not involve high workload cognitive processing (i.e. just visual orientation), it is not surprising that occipital and not frontal theta was affected.

The studies just described were concerned with sustained (i.e., vigilance) and short-term attention. Based on the Makeig studies *and* earlier research it seems reasonable to conclude that long periods of vigilance will induce a hypnagogic state characterized by drowsiness, decreased attention, and a tonic (Kramer, 1991) increase in theta activity. The other studies either did not measure vigilance or did not manipulate workload. However, it is not clear, based on the vigilance studies just described, what the relationship is between theta and correct vs incorrect responses. A large number of studies have examined the relationship between *phasic* increases in memory/workload and alpha and theta. It is possible that conflicts in reported results such as those between Makeig and associates and those of Pennekamp, et al. (1994) may be due to this distinction.

In a recent review, Klimesch (1999) examined the relationship between oscillations in the alpha and theta bandwidths and cognitive performance. In contrast to commonly used definitions of different bandwidths, Klimesch argues that there are three subdivisions of the alpha bandwidth which should be defined on an individual basis as a function of a person's alpha "peak". Each subdivision consists of a 2-Hz window, two windows below and one above the peak alpha level. Theta is defined as a 2-Hz window below the lowest alpha level. He argues that the upper alpha band responds selectively to semantic long-term memory demands while the lower two alpha bands reflect different types of attentional demands. Traditional, fixed band analyses, he argues, should be "abandoned." While factor analytic studies of EEG power often result in three bandwidths in the theta/alpha range, he points out that their exact ranges vary from study to study, due in part to differences in recording techniques. Klimesch further argues that bandwidths should be adjusted for each recording site for each individual.

With regard to hypnagogic states, sleep, and sleep deprivation, Klimesch (1999), in agreement with previous data, notes that drowsiness, sleep onset, and sleep deprivation are associated with increased theta and lower alpha power. He states that increased efforts to maintain a state of alertness are related to an **increase** in *tonic* lower alpha. Klimesch also makes a distinction between tonic and phasic power in the alpha and theta bands. With regard to **tonic** EEG, individuals with greater absolute power in the upper alpha band and less absolute power in the theta frequency evince better cognitive performance. He bases the conclusion on tonic EEG differences related to age, intelligence, hypnagogic states, and neurological damage.

In contrast to tonic levels of alpha and theta power, phasic changes in response to task demands are characterized by a decrease in (desynchronization of) high alpha. Some authors refer to this as *Event Related Desynchronization* (ERD), though not necessarily just for high alpha, and argue for its use as an index of workload over such measures as event related potentials because it can be employed in real time (e.g. Dujardin, Derambure, Defebvre, Bourriez, Jacquesson, & Guieu, 1993; Pfurtscheller, 1992; Pfurtscheller & Klimesch, 1991). According to Klimesch (1999), "lower alpha desynchronization (in the range of about 6-10 Hz) is obtained in response to a variety of non-task and non-stimulus specific factors which may be subsumed under the term 'attention'. It is topographically widespread over the entire scalp…". In contrast upper alpha (10-12 Hz) desynchronization is topographically restricted and occurs in response to task specific (semantic/memory) demands. Although he is not very specific as to where these

changes occur, there is a strong suggestion that they are most commonly seen over the left hemisphere. However, this may be a function of the nature of the task (verbal) used. Studies which did not support this relationship were described by Klimesch as employing broad band definitions of alpha and/or use of stimuli (e.g. simple tones) that affect other rhythms (e.g. Krause, Lang, Laine, Kuusisto, & Porn, 1996). In agreement with previously described studies, he suggests that there is greater high alpha desynchronization in good compared to poor performers.

Theta *sychronization* (i.e. an increase in theta power) according to Klimesch (1999), reflects episodic memory and encoding of new information. The importance of adjusting the bandwidths for individuals was further stressed with regard to a study on episodic memory in which absolute theta power reached significance, "…but only if frequency bands were adjusted individually." In an effort to distinguish between theta seen during hypnagogic states and that seen in response to task demands, Klimesch suggests that a broad band of large irregular slow activity is characteristic of drowsiness and inattention. A narrow band (e.g. 2 Hz) of regular, rhythmic theta activity in the range of the peak theta frequency reflects encoding of new information. As with high alpha, it is not clear exactly where this theta is topographically located, though there is a suggestion by Klimesch, based on the work of Gevins (e.g., 1995), that increases in frontal midline theta are associated with increased memory load. Interestingly, Gevins, Leong, Du, Smith, Le, DuRousseau, Zhang, and Libove (1995) have reported a change in topography in evoked potentials after 7-8 hours of performance from one focused on midline central and precentral sites to one focused primarily on right hemisphere precentral and parietal sites.

Frontal theta, often recorded along the midline, has been reported to increase with increased memory load and work load in general (e.g., Gundel & Wilson, 1992; Lang, Lang, Kornhuber, Diekmann, & Kornhuber, 1988; Mecklinger, Dramer, & Strayer, 1992). Wilson, Swain, and Ullsperger (1999) examined the relationship between EEG and different levels of memory load. EEG was recorded from 19 sites using monopolar recordings with a linked ear reference. Traditional bandwidths were used; beta was divided into beta 1 (12.3-15.8 Hz) and beta 2 (16.2 –24.9 Hz). Only one bandwidth was used for alpha (8.3-11.9 Hz). EEG was recorded for four 1-second intervals of the retention interval immediately after presentation of the memory set. Memory load was varied in three different experiments: 1) a weighted condition in which 60% of the trials contained a memory set of only one item and the remaining 40% of the trials which were evenly distributed between 3, 5, 7 or 8 items; 2) a random condition in which the number of items were equally represented and randomly presented; and 3) a blocked condition in which the number of items was constant for a block of trials, but changed for each block. Although theta increased with increased memory load it did so for only the weighted condition. Wilson, et al. point out that, since variations in set size across conditions remained the same (i.e. number of items to be kept in memory), increases in theta cannot be a simple function of memory load as argued by Klimesch (1999). In contrast to theta, alpha decreased with increases in memory load regardless of task condition. Also, in all three conditions, significant reductions in alpha were found only in the left hemisphere at P3, Pz, T3, T5, and C3. Power for beta2 increased from stimulus onset to the end of the interval for the smaller memory sets; however, for the larger memory sets beta2 decreased over the interval. Interestingly, in Wilson, et al.'s (1999) discussion of theta they suggest that the paradox of two separate classes of theta activity (i.e. reflecting a hypnagogic state and also increased workload) has yet to be resolved.

Gevins, et al. (1998) also evaluated the effects of memory load on EEG. They recorded from 27 sites using an electronically linked mastoid reference. Two processes were used to analyze the EEG. First, measurements were obtained for each subject using traditional bandwidths and taking the average amplitude in a 1-Hz window around the peak of the power spectra within each band. Second, a neural

network-based pattern recognition was applied to the EEG data. Subjects were tested on verbal and spatial versions of a working memory task at low-moderate-, and high-loads. Theta activity was largest at frontal midline and increased with increases in memory load while alpha decreased with increases in memory load. There was a small asymmetry for alpha as a function of task type: alpha was lower at P8 for the spatial compared to the verbal; there were no differences at P7 as a function of task type. Beta was largest at temporal sites but only decreased as a function of load at Cz. The neural network weighted alpha features at occipital and parietal sites most heavily followed by frontal theta. Gevins suggests that neural network pattern recognition techniques hold great promise in evaluating the effects of workload on EEG (Gevins & Cutillo, 1993; Gevins, Leong, Du, Smith, Le, DuRousseau, Zhang, & Libove, 1995).

In contrast to Klimesch (1999) who argued that decreases in high alpha and increases in theta are related to increases in memory load, Gevins, et al. argued that the EEG is sensitive to variations in difficulty in a wide variety of tasks. Despite Klimesch's argument regarding the necessity of using an individual's peak alpha level to establish bandwidths, numerous studies using more traditional determinations of bandwidths, including high and low alpha, have supported the view of Gevins, et al. regarding both theta (Gevins, Smith, McEvoy, & Yu, 1997; Gundel & Wilson, 1992; Hankins & Wilson, 1998; Laukka, Jarvilehto, Alexandrov, & Lindqvist, 1995; Nakashima & Sato, 1992; Yamada, 1998) and alpha (Gevins, et al., 1997; Gevins, Zeitlin, Doyle, Schaffer, & Callaway, 1979; Gevins, Zeitlin, Doyle, et al., 1979; Gundel & Wilson, 1992; Hankins & Wilson, 1998; John & Easton, 1995; Sterman, Mann, Kaiser, & Suyenobu, 1994). Furthermore, as a general measure of workload, decreases in alpha and/or increases in theta have been reported in ATC simulators, flight simulators, and in pilots during actual flight (Brookings, Wilson, & Swain, 1996; Hankins & Wilson, 1998; Sterman, Mann, Kaiser, & Suyenobu, 1994; Sterman & Mann, 1995). It seems appropriate to reiterate that the type of task and stimulus modality may have a significant effect on power in the alpha and theta bandwidths (cf. Krause, et al., 1991; Pfurtscheller & Aranibar, 1977; Ray & Cole, 1985; Wilson, et al., 1999). Also, there may be topographical differences in alpha as a function of task type and stimulus modality. Theta, conversely, seems to be less dependent, though not completely, on these variables. Finally, a number of investigators (cf. John & Easton, 1995) have pointed out that the use of very narrow bandwidths (i.e. less than 1 Hz) would increase the ability to detect differences in workload that may be obscured by the more traditional bandwidths (e.g. theta: 4-7 Hz)

Little reference has been made here to delta, beta and gamma (approximately 35-50 Hz) activity. While some studies have found beta to be related to task load, the vast majority of studies have found that alpha and theta are much more responsive. Beta activity appears to be related more to different aspects of cognition. Ray and Cole (1985) found that beta and alpha are differentially responsive to task type as well as to what specifically is required of the subject (i.e. tasks which require intake vs rejection; spatial vs verbal tasks). Fernandez, Harmony, Rodriguez, Bernal, Silva, Reyes, and Marosi (1995) also reported changes in beta as a function of task, as did Brookings, et al. (1996). Differences in beta attributable to task type do not appear to specifically reflect variations in workload.

Slow waves, primarily in the delta bandwidth, have been related to increased inhibition. Harmony, Fernandez, Silva, Bernal, Diaz-Comas, Reyes, Marosi, Rodriguez, and Rodriguez (1996), citing Vogel, Broverman, and Klaiber (1968), suggest that there are two types of inhibition represented by delta waves: 1) "…a gross inactivation of an entire excitatory process, resulting in a relaxed, less active state, as in sleep" and 2) a selective suppression of "…inappropriate or non-relevant neural activity during the performance of a mental task". Such a distinction is similar to that made for theta, though Harmony, et al. include delta and low theta in this category. Harmony, et al., recording from 20 sites with reference to

linked ears and controlling for eye movement artifact, found an increase in power from 1.56 to 5.46 Hz during the performance of a mental arithmetic task and from 1.56 to 3.90 during the performance of a Sternberg memory task. Slow wave power increased with increases in task difficulty. Some changes, both increases and decreases, in the beta band were observed, depending on the site and specific frequency. They argued that increases in delta activity occur only in those tasks requiring attention to internal processing. Attention to external stimuli decreases delta activity.

Investigations of gamma activity appear to suggest a relationship with sensory and cognitive functions (Basar, Basar-Eroglu, Karakas, & Schuermann, 2000). Increases in gamma, especially over the parietal cortex, have been found to reflect changes in attention (Gruber, Mueller, Keil, & Elbert, 1999; Shibata, Shimoyama, et al., 1999). However, unlike the literature on alpha and theta bandwidths, there does not appear to be much data on the relationship between gamma and workload.

In summary, it would appear that the theta and alpha bandwidths may be used to index mental workload, attention, and perhaps performance levels. Topographically, increases in theta recorded from midline frontal sites would appear to be a relatively consistent indicator of workload and attentional effort. Conversely, decreases in alpha are related to increases in mental workload and attention. In contrast to theta, topographical changes in alpha appear to be somewhat more related to the type of information the individual is required to process. Also, high alpha (e.g. 10-12 Hz) appears to be more consistently related to variations in workload than lower alpha. While changes in beta have been related to cognitive processing and to overall arousal, reported research is not as strong as far as its relationship to variations in workload. Delta activity, like theta, also appears to be inversely related to workload. However, recording delta waves is especially susceptible to eye movement artifacts. Thus, attempts to either index or to manipulate experienced workload would seem to have the greatest potential for success by addressing changes in theta and alpha activity. One such attempt involves the use of bio/neurofeedback.

EEG and Biofeedback

A number of researchers and clinical practitioners have attempted to apply information concerning the relationship between EEG and attention/arousal to performance enhancement through the use of biofeedback/neurofeedback techniques. In such applications, individuals have been trained to produce those EEG patterns that were assumed to reflect greater or lesser degrees of arousal. The effect of such training on performance was then evaluated. Early studies in this area involved reinforcing increases and decreases in occipital theta activity (Beatty, Greenberg, Diebler, & O'Hanlon, 1974; Beatty & O'Hanlon, 1979; O'Hanlon & Beatty, 1977; O'Hanlon, Royal, & Beatty, 1979). While occipital theta regulation has been shown to affect vigilance performance, the effects are not particularly strong (Alluisi, Coates, & Morgan, 1977), nor have they been shown to transfer readily to other situations (Beatty & O'Hanlon, 1979).

Investigations of the ability to control EEG rhythms through neurofeedback have been applied to a variety of other behaviors. Much of the impetus for investigating neurofeedback for specific EEG rhythms came from the research of Sterman and his associates. Sterman found that training epileptic subjects to produce a "sensorimotor rhythm" or SMR of 12-15 Hz recorded from the scalp over the sensorimotor cortex had the effect of elevating seizure thresholds (Sterman, 1982, 1986). Kuhlman (1978) argued that the human analog of the SMR, originally observed by Sterman in cats, was really an

EEG of 8-13 Hz, that he referred to as "Mu", localized over the sensorimotor area. Kuhlman considered this rhythm to differ from alpha, although the frequency bands are the same.

Control of seizures in epileptics has also been attempted by providing neurofeedback for slow cortical potentials (SCPs) such as contingent negative variation (CNV). Elbert and his associates have demonstrated that normal subjects also can learn, through neurofeedback, to control their SCPs (Elbert, Rockstroh, Lutzenberger, & Birbaumer, 1980; Lutzenberger, Elbert, Rockstroh, & Birbaumer, 1983; Rockstroh, Elbert, Birbaumer, Lutzenberger, 1982). Roberts, Rockstroh, Lutzenberger, Elbert, and Birbaumer (1989) found that normals could achieve about 10-20 uV control of SCPs in as few as two sessions. Attempts to demonstrate similar control in epileptics has been shown to take considerably longer, but has been successful (Kotchoubey, Schleichert, Lutzenberger, & Birbaumer, 1997). However, the ability to control SCPs through neurofeedback is not related to specific changes in EEG power spectra, at least in epileptics (Kotchoubey, Busch, Strehl, & Biebaumer, 1999).

Numerous studies have found that the amplitude and topography of SCPs (including the preparatory aspect of the Bereitschaftspotential (BP), a slow negative potential seen prior to making a response) reflect both task type and task demands (cf. Freude & Ullsperger, 1999), with greater negativity reflecting better performance. SCPs have been shown to increase with increases in task difficulty, in dual task paradigms, and with increases in time pressure. Higher SCP amplitudes were also found prior to correct compared to incorrect responses. Brody, Rau, Kohler, Schupp, and Lutzenberger (1994) found that teaching subjects to increase the magnitude of the negativity of the SCP resulted in a larger EMG startle response, which the authors related to enhanced cortical arousal. Finally, practice, which may be argued ultimately to result in less effort, is related to a decrease in SCPs (cf. Freude & Ullsperger, 1999).

EEG Biofeedback and Task Performance. A number of attempts have been made to determine whether biofeedback for specific characteristics of EEG might affect performance on various tasks. Early research in this area involved training subjects to increase or decrease alpha waves. Alpha feedback training was claimed to be able to decrease needed sleep time, facilitate task performance, increase pain thresholds, and improve memory. Subsequent evaluation of these early studies found them to be fraught with methodological problems (see Petruzzello, Landers, & Salazar, 1991 for a review). Biofeedback for SCPs, on the other hand, has been found to produce faster reaction times, to improve performance on mental arithmetic, and to show less performance decrements on a vigilance task (Bauer, 1984; Birbaumer, Elbert, Canavan, & Rockstroh, 1990; Lutzenberger, Elbert, Rockstroh, Birbaumer, 1979). With regard to the vigilance study, there appears to be an inverted U relationship between amount of SCP produced and performance efficiency, as high SCP amplitudes resulted in an increase in error rates (Lutzenberger, et al., 1979).

A number of attempts have been made to enhance athletic performance using neurofeedback. Much of this research has been conducted by Landers and his associates. Crews and Landers (1993) analyzed slow potentials, traditional bandwidth activity, and 40 Hz activity as measures of attentional patterns prior to golf putts. They found a progressive increase in alpha power in the left hemisphere prior to the actual putt. Alpha power in the right hemisphere remained relatively stable. This effect has been found during the preparatory period prior to the execution of a response in golf, archery, and riflery (Hatfield, Landers, & Ray, 1984; Lawton, Hung, Saarela, & Hatfield, 1998; Salazar, Landers, Petruzzello, Han Crews, & Kubitz, 1990). Hillman, Apparies, Janelle, and Hatfield (2000) examined both alpha and beta power in skilled marksmen four seconds prior to either the execution or rejection of shots. Rejected shots produced a progressive increase in both alpha and beta power compared to executed shots, though

35

there was greater power in the left hemisphere for both responses.The authors felt the results reflected appropriate allocation of resources underlying the achievement of a focused state. In skilled marksmen this constituted a decrease in verbal processing (left hemisphere) and an increase in visuo-spatial processing (right hemisphere). Several studies have demonstrated that biofeedback training for increasing the negativity of SCPs in the left hemisphere increased the amount of beta in the left hemisphere and resulted in better performance scores (e.g. archery scores) compared to control subjects and subjects trained to increase SCP negativity in the right hemisphere (Landers, Petruzzello, Salazar, Crews, Kubitz, Gannon, & Han, 1991; Petruzzello, et al., 1991).

Neurofeedback training to control other EEG rhythms has been used in a variety of other paradigms. Wolpaw and his associates have trained individuals to control the mu rhythm, an 8-12 Hz rhythm focused over the sensorimotor cortex (McFarland, Neat, Read, & Wolpaw, 1993; Wolpaw & McFarland, 1994; Wolpaw, McFarland, Neat, & Forneris, 1991). Subjects were trained to move a cursor around a computer screen by altering the mu rhythm amplitude. The authors stated that the control demonstrated by the subjects was not due to covert changes in motor behavior. Soroko and Musuraliev (1995) also claimed to be able to bring different EEG bandwidths under voluntary control within three to four training sessions. Research by Sheer and his associates has involved studying 40 Hz EEG (with a frequency window between 36 and 44 Hz) and attention (Loring & Sheer, 1984; Spydell & Sheer, 1982). Feedback training for 40 Hz EEG was associated with subjective states of attention, concentration, vigilance, and effortfulness (Ford, Bird, Newton, & Sheer, 1980) and has been described as producing a focused arousal.

One area of continuing research where EEG biofeedback has been applied (apparently successfully) involves the treatment of Attention Deficit/Hyperactivity Disorder, although the technique has also been applied to numerous psychiatric disorders (cf. Abarbanel, 1999; Thompson & Thompson, 1998). Research by Lubar and his associates has involved training subjects to increase certain EEG rhythms (e.g. sensorimotor response (SMR) or beta) and to decrease others (e.g. theta). He has reported dramatic changes in the EEG as well as in the behaviors of children with ADHD following biofeedback training, including improvements on psychometric tests and in school performance (Lubar, 1991; 1997; Lubar, Swartwood, Swartwood, & O'Donnell,1995). At an empirical level, practitioners of neurofeedback training argue that an elevated theta/beta ratio correlates with the presence of ADHD symptoms while a reduced theta/beta ratio correlates with the resolution of these symptoms (Abarbanel, 1999).

The experimental basis for the use of neurofeedback training to treat ADHD relies mainly on the view that theta activity is related to increased drowsiness and lowered ability to attend to the environment while beta is related to increased arousal and enhanced ability to attend to environmental stimuli. As has been pointed out, such a characterization is only partially valid and relates to Klimesch's (1999) characterization of *tonic* theta. While theorizing about the neurophysiological basis of neurofeedback training has involved a discussion of cortical EEG oscillations being driven by thalamo-cortical and hippocampal-cortical loops (Abarbanel, 1999; Sterman, 1996), a discussion of the relationship between increased theta and increases in attention have often been ignored. In a study by Benham, Rasey, Lubar, Frederick and Zoffuto (1997), subjects were asked to listen to a story involving vivid action scenes about dinosaurs attacking people and to press a switch whenever they became engaged in the story. Many of the subjects produced increased 4-8 Hz activity when they were engaged in the story. The authors attributed this increase to visualization processes related to the story. As the data on workload has shown, increased workload/attention is associated with increased theta activity, especially in the frontal lobes. Thus, their results could be attributed to an increase in attention.

Some other relevant and/or interesting aspects of Lubar's research concern the placement of electrodes and the potential use of external stimuli to drive the EEG within certain bandwidths. According to Lubar, one does not have to place electrodes over the entire scalp. He places electrodes halfway between Cz and Fz or Cz and Pz for feedback training because those areas represent the highest ratios of theta to beta activity in ADHD patients. Also, training just using these sites still produces changes across the cortex (Lubar & Lubar, 1999).

In recent years, "neurotherapists" have claimed that audio and visual stimulation may be employed in conjunction with neurofeedback training to enhance the effect of the feedback (e.g. Lubar, 1997; Frederick, Lubar, Rasey, Brim & Blackburn, 1999; Patrick, 1997). Lubar (1997) claimed that stimulation at an individual's "dominant EEG frequency, e.g. 10 Hz" or at twice the dominant frequency increased the spectral power in the beta range (13-21 Hz) up to 18%. Lubar did not report whether such stimulation had actually been employed as a supplement to neurofeedback training for ADHD. Rosenfeld, Reinhart and Srivastava (1997) summarized the rationale for using *entrainment* of EEG as: 1) it may lead to faster effects than neurofeedback alone and 2) it may drive the EEG away from an individual's dominant frequency that represents a pathological state. Both assume "…that evoked or driven rhythms involve the same pathways, mechanisms, and overall physiology as true spontaneous EEG rhythms." (Rosenfeld, et al., 1997; p. 4). They state that clinical reports of improvement are not accompanied by any concomitant EEG monitoring. Further, clinical studies typically employ low-intensity LED-based stimulator goggles to entrain EEG. Rosenfeld, et al. questioned whether such stimulation was capable of driving specific EEG bandwidths. In their study they employed such stimulation generated by a commercially available audio-visual stimulation unit. EEG was recorded from Cz and Pz referenced to linked mastoids. One group of subjects received alpha stimulation while a second received beta stimulation. Subjects kept their eyes closed throughout training. Alpha stimulation produced either no entrainment or prolonged entrainment in subjects with high alpha baseline. Low alpha baseline subjects produced only transient entrainment. Some beta stimulation subjects showed prolonged beta enhancement some transient enhancement, and some beta inhibition, which could be predicted from baseline beta levels. No attempt was made to relate any of these EEG changes to behavior. It would also be of interest to record from more than just two sites and to look at theta enhancement.

Swingle (1996), using normal college students, reported being able to decrease theta power, recorded from Cz referenced to the ear lobes, by presenting the subjects two equal amplitude sinusoidal tones with a frequency difference of 10 Hz embedded in pink noise. Theta decreased significantly by an average of 13% relative to baseline for the experimental subjects while that for a control group increased slightly relative to the baseline. In succeeding experiments Swingle was also able to decrease theta in child and adult ADHD patients. However, no control groups were used in the latter experiments. Also, Swingle only evaluated theta suppression for the five minutes of tone presentation and did not report any behavioral evaluations (Swingle, 1998). Lane, Kasian, Owens, and Marsh (1998) examined the effect of binaural auditory beats on vigilance performance and moods. Binaural auditory beats in the delta range, in unpublished reports by the Monroe Institute, are claimed to be associated with enhanced creativity and improved sleep. In the beta range they are claimed to be able to enhance attention and performance on memory tasks (cf. Lane, et al., 1998). Lane, et al. found that beta-frequency binaural auditory beats produced a marginally significant (i.e. using a one-tailed test) enhancement of vigilance performance relative to delta/theta binaural beats. Although the study was interesting, its effects were weak and it did not record any EEG.

Although the research on theta/beta neuorfeedback training by Lubar and others would seem to be

applicable to workload, vigilance performance and adaptive automation, it is important to note that 1., he was studying children and adults with specific attentional disorders 2., biofeedback training involved extensive time expenditures, with subjects being given anywhere from 40 to 80 sessions of training, and 3., it is quite possible that the need for extended training was a function of the subject population. Similar extended training was reported necessary by Elbert, et al. (1991) when trying to get epileptics to control SCPs. Normal subjects learned to control such activity in only two sessions. Data concerning theta/beta neurofeedback training in normals is lacking. Lubar and his associates have reported attempts to do so, but the results were mixed at best and open to alternative interpretations (Rasey, Lubar, McIntyre, Zoffuto, & Abbott, 1995). Use of auditory and visual stimuli to drive certain EEG bandwidths is intriguing, but again there is little research demonstrating this effect and little or no behavioral evaluations have been reported. If such a technique were to be proven valid, it conceivably could be employed as a noninvasive countermeasure to various forms of "hazardous states of awareness" (Pope & Bogart, 1992).

      Application of neurofeedback to adaptive automation. Attempts to investigate the potential for applying EEG biofeedback techniques to attention and workload in an adaptive automation setting should consider the following:

1. Petruzzello, et al. (1991) have pointed out that in research investigating biofeedback training in EEG, often there is no prior knowledge of what are optimal levels of activity with regard to performance.

2. A great deal of recent research has demonstrated that hemispheric specialization may result in different patterns of EEG activation depending on the nature of the stimuli used in a task. Since early studies on the relationship between EEG and vigilance employed a variety of stimuli, it is reasonable to assume that some of the variability across studies might be attributable to hemispheric differences in processing. The study by Landers, et al. (1990) suggested that feedback for left hemisphere SCP changes enhance performance relative to right hemisphere and control groups.

3. Differences in the nature of the dependent variables across studies may have added to the variability of results. Some investigators measure absolute power while other measure relative power. Furthermore, differences in reference sites might also account for between study differences in results. While many early studies used mastoid references, others used the linked ear technique or bipolar recordings.

4. Biofeedback for a variety of different EEG measures has been employed. While Lubar has provided the most recent evidence that such techniques might improve attentional skills in individuals with ADHD, it is not clear how such training might affect normal adults. Sheer has argued for 40 Hz EEG biofeedback while Elbert and his associates have argued for SCP biofeedback. Lubar has argued for increasing beta while decreasing theta. Each of these groups of researchers not only examines different aspects of the EEG, they record the EEG from different electrode sites.

5. The use of auditory and visual stimulation to drive specific EEG bandwidths, though popular

in clinical neurofeedback, appears to be a relatively unexplored field from a controlled, experimental point of view. Positive results could provide a cheap, efficient, and nonintrusive mechanism for affecting arousal, attentional states, and, perhaps, even performance.

Event-Related Potentials

The EEG thus provides a sensitive index of workload and can be used in adaptive automation systems, as reviewed previously. However, the processing characteristics of specific components underlying a task can be assessed more directly by recording task-evoked ERPs, which reflect the neural activity in response to a stimulus with millisecond precision.

Among ERPs, probably the most widely supported measures are those based on the P300 and N100 potentials of the brain (Donchin et al., 1986; Parasuraman, 1990). We describe these measures, as well as more recent work on the error-related negativity (ERN; Gehring et al., 1993), which can potentially provide a sensitive index for use in adaptive systems.

N100 and P300. Both the N100 and P300 ERP measures have been shown to provide fairly sensitive measures of mental workload in multi-task situations. In addition, these brain potentials have a measure of diagnosticity as well. The P300 has been shown to reflect primarily the allocation of perceptual-cognitive resources and not response-related processes (Donchin et al., 1986). The N100 brain potential, on the other hand, has been found to reflect attentional resources associated with early information-processing stages (Mangun, 1995; Rugg & Coles, 1995).

Most of the N100 and P300 studies of workload have examined dual-task performance, although working memory studies have also been carried out (Kramer, 1991; Rugg & Coles, 1995). The amplitude of the P300 component to a secondary task of counting infrequent tones among more frequent tones (an "oddball" task) decreases when combined with a primary task such as visual discrimination or psychomotor tracking. Thus, P300 amplitude shows a dual-task decrement, i.e. it is reduced in amplitude when the eliciting task is combined with another task. Importantly, only changes in the difficulty of visual discrimination and not motor tracking affect P300 amplitude on the secondary task (Israel et al., 1980). Because the P300 component has been shown to be more sensitive to central stages of information processing than to the response-selection stage, this finding provides strong supporting evidence for a modular theory of resources based on stages of processing (Wickens, 1984).

A strong prediction of theories that postulate sharing of scarce resources is one of *resource reciprocity*, or an inverse relationship between primary and secondary task resource allocation and performance. As resources are withdrawn from the primary task, they are simultaneously allocated to the secondary task; and vice versa. If P300 amplitude reflects allocation of resources to a central processing stage, its amplitude should vary accordingly. Consistent with this prediction, P300 amplitude to the primary or secondary task has been found to increase or decrease appropriately as resources are applied or withdrawn (Kramer, 1991). Resource allocation between two tasks can also be manipulated endogenously (through instructions and a payoff scheme) such that resources are allocated in differing proportions between two tasks (e.g., 25-75%, 50-50%, or 75-25%). Such studies have the advantage that the stimuli and responses remain constant across conditions, as opposed to single-dual task comparisons, which confound resource allocation with stimulus and response variation. In the resource reciprocity studies, the amplitudes of the early-latency N100 components has been found to vary in a graded manner with resource allocation (Parasuraman, 1985, 1990).

The significance of these results should be emphasized. The finding of a dual-task decrement in P300 or N100 amplitude is insufficient by itself to support a resource model, because factors other than resource scarcity may contribute to dual-task interference. Consequently, the demonstration that ERP components show *graded* changes between tasks as resources are dynamically traded off between each other, with invariant stimuli and response requirements, provides strong support for resource theories and for the utility of ERPs as a dynamic measure of mental workload. Thus, for an adaptive system that is needed to regulate operator workload around some optimal value could use on-line measures of N100 and P300 to a secondary task probe as an adaptive trigger. Kaber and Riley (1999) have used behavioral measures of secondary-task performance in adaptive control, with generally positive results. However, to date no study has used ERP measures in a similar way. The evidence strongly supports their use in this way, especially since they may be more sensitive than behavioral measures to central (as opposed to response-related) sources of mental workload.

Error-Related Negativity. Recently a new ERP component has been identified, the error-related negativity or ERN. As the name suggests, this component is elicited when subjects make an error in a task. As such, measures based on this ERP component could potentially be used to control user performance in a complex automated environment. Because this is a new area of research, we review some of the basic findings on the ERN.

In most ERP research, ERPs are averaged across trials in which the subject responds correctly. Until recently, therefore, ERPs associated with errors were not widely studied. One reason for this neglect may simply have been that error rates are often low, so that researchers may have lacked sufficient numbers of trials to generate a robust ERP. However, as for any other aspect of human performance, it is reasonable to hypothesize that whenever an individual commits an error, a specific neural mechanism is activated. Studies suggest that this mechanism can be identified as a negative ERP component that has a frontocentral distribution over the scalp (Gehring et al., 1993). The ERN has an amplitude of about 10 μV, reaches a peak about 100-150 ms after the onset of the erroneous response (as revealed by measures of electromyographic activity), and is smaller or absent during trials in which the subject makes a correct response.

In an early study by Falkenstein et al. (1990), ERPs were recorded for trials in which subjects made errors in a reaction time task under time pressure conditions. This allowed the researchers to generate sufficient numbers of trials to compare correct and error trials. High time pressure, however, has been demonstrated to reduce the amplitude of the ERN component (Falkenstein et al., 2000). The ERN component can also be obtained from trials in which subjects provided incorrect responses in a number of other experimental paradigms, such as go/no-go tasks, the Eriksen letter flanker task, and time-estimation tasks.

Bernstein et al. (1995) showed that a larger ERN is observed as a function of the difference between the error response and the correct response (number of movements parameters not shared by the two responses). The ERN also occurs not only when subjects respond using the incorrect hand in a go/no-go paradigm, but also when they respond to no-go stimuli. Furthermore, ERN amplitude is larger when task instructions emphasize response accuracy over speed (Gehring et al., 1993)

ERN amplitude is also related to perceived accuracy, that is the extent to which subjects are aware of their errors. Scheffers and colleagues (Scheffers et al., 1996; Scheffers & Coles, 2000) showed that error awareness and ERN amplitude covary directly in an experiment where subjects were required to judge their responses along different levels of correctness (from "sure correct" to "sure incorrect"). ERN is also associated with remedial actions performed as a consequence of an error. In fact, as reported by

Gehring et al. (1993), ERN amplitudes are larger when errors are made with less force, and when they are more likely followed by correct responses.

Several authors (Dehaene et al., 1994; Falkenstein et al., 1995; Gehring et al., 1993) have suggested that the ERN component appears selectively on error trials. For this reason, the ERN has been considered to be a sensitive index of an error detection and/or compensation mechanism, supposedly occurring when an error is detected in time to attempt a correction of the response. However, the interpretation of the ERN component is still hotly debated. Early interpretations of the psychological meaning of this component suggested that the component represents the outcome of a comparison process between intended and actual response. The ERN would be elicited when the neural representation of the actual (erroneous) response is compared with the representation of the required (correct) response, and a discrepancy is found. In this sense the ERN would belong to the same family of ERP components such as the Mismatch Negativity and the N400 (Näätänen, 1992).

Other interpretations have been proposed. For example Kopp et al. (1996), suggested that the ERN is an index of an error tendency inhibition, while Carter et al. (1998) proposed that it is the outcome of a conflict detection process. Using fMRI, Carter et al. (1998) found activity in the anterior cingulate cortex, not only related to errors, but also to correct trials, under conditions of increased response competition. However, they did not record electrical brain activity, and therefore a direct comparison with electrophysiological studies is not possible. Falkenstein et al. (2000) compared the ERN recorded from tasks with a strong conflict and without conflict, and found that the averaged components had the same amplitude. Furthermore, Scheffers and Coles (2000) showed how their own data did not support a conflict hypothesis since incompatible stimuli in their experiment did not lead to a larger ERN on correct trials.

Other authors have also concluded that the ERN is the manifestation of the activity of a "generic" neural system involved in error-detection. For example, Miltner et al. (1997) investigated the error detection process in a situation where detection was not performed on the basis of a representation of the correct response (there was no processing of the task stimulus). They used a time estimation task, and the occurrence of the component was observed, even in this case, after incorrect estimation.

There are, however, other interpretations. In a recent study using EMG, Vidal et al. (2000) proposed that the ERN could be interpreted in terms of emotions related to the emission of the response. Their results are intriguing, since they would suggest that error detection occurs well before the response is emitted. Furthermore, a link between emotions and ERN would, once again, support the idea that this component is originated by the activity of the anterior cingulate cortex.

Finally, whatever the neural mechanism involved in the generation of the ERN component, it seems, that the process is output-independent. Holroyd et al. (1998) tested subjects performing a choice reaction time task using either their hands or feet. Using brain electric source analysis (BESA) to compare the ERNs elicited by hand and foot errors they found identical scalp distributions of these error potentials.

The relevance of this measure to human factors applications, including adaptive automation, is straightforward. Most physiological measures (including ERPs) that have been considered as candidate triggers for adaptive automation have focused on adaptation based on mental workload. While workload is an important aspect of human interaction with automation, the ERN allows identification (and perhaps prevention) of operator errors in real time. Implementing such a measure provides another approach to adaptive automation. For example, the ERN could be used to identify the human operator tendency to either commit, recognize, or correct an error. This could potentially be detected covertly by on-line measurement of ERN, prior to the actual occurrence of the error. Theoretically a system could be

activated by an ERN detector in order to either take control of the situation (for example in those cases where time to act is an issue), or notifying the operator about the error he/she committed, even providing an adaptive interface which selectively presents the critical sub-systems or function. As noted above, Inagaki (1999) has proposed an adaptive automation system in which the system takes over control from the operator when there is insufficient time for the human to react in a time-critical situation, e.g., engine malfunction near the critical V1 speed during takeoff. The criterion for adaptive control was simply the minimum amount of time left for human perception and decision making. He proposed that when the time available exceeds this minimum, human control is possible. But of course human decision making (e.g., to continue the takeoff or to abort) could be erroneous. In principle, ERN detection during this period could be used to trigger machine control instead of an arbitrary time criterion.

A system such as the one above outlined provides at least two levels of automation (control/suggest) and it would have the advantage of keeping the operator still in control of the entire system, while providing however an anchor for troubleshooting, when the error actually occurs (and having the possibility, for the system, to correct it by itself if needed). In this way the control by the system would be limited to the extreme conditions (temporal window too short for the human decision-making process, for example). Other measures could be based on the ERN complex in order to provide to the system as more information as possible regarding the state of the operator. For example, the Pe component, that occurs right after the ERN, has been associated with a "subjective/emotional error assessment process modulated by the individual significance of an error" (Falkenstein et al., 2000). Subjects who commit errors often show a Pe of bigger amplitude compared to those who commit fewer errors. Even if the authors admit that this interpretation is not entirely satisfactory, this component could turn out to be very useful to monitor when the operator develops a sort of habituation to his/her own errors.

Finally, an important issue has to be resolved prior to use of the ERN in adaptive automation. In order to trigger a system in real time, such a component should be viewable not only in the averaged waveform, but also in single trial recording. Up to now, no studies have provided information about single trial ERN. N100 and P300 are detectable in single trials (Kramer, 1991), and so in principle should the ERN. However one of the main obstacles could be the variability in the latency of this component, and that could affect the stability of the measure. Additional research on single trial identification of the ERN is needed.

Cerebral Metabolism and Blood Flow

PET. In addition to EEG and ERPs, measures of brain function based on newer imaging technologies have become prominent in recent years. Foremost among these are PET, fMRI, and optical imaging. Each of these techniques is designed to assess different aspects of cerebral metabolism and blood flow. It is therefore instructive to examine the characteristics of these aspects of brain physiology in relation to mental states such as attention, workload and alertness (Parasuraman, 1998), which may be assessed and adapted to adaptive systems.

Mental workload can be intuitively characterized as reflecting how hard one's mind is working at any given moment. Given that the mind is a function of the brain, it follows that mental workload should be associated with brain work. How can brain work be assessed? Over a century ago Charles Sherrington suggested that brain work was related to the regulation of the blood supply of the brain (Roy &

Sherrington, 1890).   Sherrington demonstrated that there is a close coupling between the electrical activity of neuronal cells, the energy demands of the associated cellular processes, and regional blood flow in the brain. His pioneering work suggested that if mental activity results in increased neuronal response in localized regions of the brain, then in principle it should be possible to measure mental workload by assessing regional cerebral metabolism and blood flow.

The development of PET paved the way for measurement of regional cerebral metabolism and blood flow in humans. PET is an adaptation of autoradiographic techniques originally developed for measuring blood flow in animals. Regional cerebral glucose metabolism can be non-invasively determined using PET and radioactively labeled glucose (18-fluoro-deoxyglucose), while regional cerebral blood flow may be assessed with PET and radioactively-labeled oxygen (O-15) in water. PET is also more accurate than the older methods in localizing the specific cortical regions activated by cognitive task demands.

Several studies have shown that PET can be used to index the attentional demands of both single (Corbetta, 1998) and multiple-task performance (Nestor et al., 1991). In particular, PET studies of divided attention consistently point to right frontal lobe activation (Parasuraman & Caggiano, in press). This suggests that the volume and extent of activation in this region could potentially be used as an index of mental workload.

Despite its sensitivity, PET has a number of disadvantages. First, the spatial resolution of PET, particularly in individual subjects, leaves much to be desired. Second, the need for ionizing radiation, although safe when used within exposure limits, is an impediment against frequent use in studies with normal human subjects.   Third, the technique is expensive and requires the use of a scanner and a high-energy physics facility. Fourth, PET imposes a degree of immobilization on the subject that severely limits its use in complex task environments. For all these reasons, PET is likely to be of limited utility as a physiological measure for adaptive automation research.

fMRI. The recent development of fMRI has overcome some of the limitations of PET. fMRI provides noninvasive, high-resolution assessment of regional cerebral blood flow. No ionizing radiation is used, thus permitting extensive and repeat testing of subjects.   Subjects must remain relatively immobilized in a scanner, so that this disadvantage that is shared with PET remains. However, portable fMRI systems that use lower strength magnetic fields are being researched and may become available in a few years. Furthermore techniques for detecting and correcting for artifacts from movement are being improved. It is possible, therefore, that "ambulatory" fMRI systems that permit subject movement will become available in the near future.

Much of the fMRI work on workload stems from studies of the neural substrates of working memory. This is a type of memory involved in keeping and maintaining information "on line" so that it can be used in the service of other processing activities—in language, decision making, and problem solving (Baddeley, 1992). A general finding is that active maintenance of information in working memory is associated with activation of both frontal and posterior (parietal) cortical regions, depending on the type of material encoded and the specific operation in working memory probed. For example, it is well known that perceptual operations can be divided into object and spatial components and that these operations are mediated by cortical processing streams that activate regions in the inferior temporal (ventral stream) and parietal cortices (dorsal stream), respectively (Ungerleider & Mishkin, 1982). This cortical subdivision of labor during perception has been postulated to result in similar division "upstream" in frontal cortex. In support of this prediction, working memory for objects such as faces have been found to activate lateral prefrontal cortex whereas spatial working memory has been shown to recruit more dorsal regions of the

frontal cortex in the premotor region (Smith & Jonides, 1997). Jiang et al. (2000) also showed that sustained use of working memory in a multiple-target discrimination task activated both frontal and posterior regions, but that only the frontal activation was maintained across time as targets were repeatedly encountered, whereas the posterior activation declined. This suggests that the frontal activation serves as the mechanism for the maintenance in working memory of the target representation.

Recent divided attention (dual-task) studies using fMRI have obtained results relevant to the distinction between unitary and modular theories of workload. In these studies, brain regions activated by concurrent execution of two tasks are compared to those during the execution of either task in isolation. Dual-task performance ostensibly requires a "central executive" because of the need for coordination (although this must be empirically demonstrated for any given pair of tasks, and not all studies have done this). Therefore, any brain region activated by dual-task but *not* by single-task performance would potentially provide evidence for a specialized central executive region, such as the prefrontal cortex or the anterior cingulate cortex (Posner & DiGirilamo, 1998). Two recent fMRI studies failed to provide support for this view. In tasks involving verbal and face working memory, no new brain area was activated for dual-task performance. Instead, activation increased with dual-task performance but in the same regions active during performance of each task individually. Although these findings do not rule out the possibility that specific executive processes are mediated anatomically by specialized modules, they do concur with other findings and are consistent with a modular view of workload with content-specific slave buffers but in which there is no separate central executive control center.

fMRI is a promising new brain imaging technique. Although there are only a few fMRI studies of workload to date, more studies are likely to emerge soon. Furthermore, as discussed previously, portable fMRI systems might become a reality in the future. Thus, although the evidence does not yet support the use of this method for adaptive automation research, it may do so in the near future.

Optical Imaging. Finally, optical imaging of cerebral oximetry is a relatively cheap technique that can be used for user state assessment. Cerebral oximetry refers to the regional measurement of oxygen saturation of hemoglobin in human brain. In this respect the technique follows the Sherrington procedure outlined earlier in the discussion of PET, but the technique is considerably less sensitive than PET or fMRI.

Several optical imaging systems have been developed and are now marketed. In the typical system, an infrared sensor is attached to the head and the absorption coefficient is determined (Klose et al., 1992). The sensor is attached to either the right or left forehead of the subject, thereby providing saturation values for either hemisphere. A new bilateral sensor system has recently been developed which allows for simultaneous measurement of left and right hemisphere values. However, values are obtained for the entire hemisphere, and the technique cannot therefore discriminate between saturation values *within* a hemisphere, e.g., between frontal and posterior regions. In essence, this procedure produces 1 "voxel" of activation for each hemisphere, as opposed to the many thousands that PET and fMRI can provide.

The optical imaging technique is relatively new and few studies of workload have been conducted. However, a promising series of studies by Warm and colleagues using vigilance tasks have been conducted (Hitchcock et al., 2000; Mayleben, 1998). In one study, reduced right hemispheric activation over time was found for a long-duration vigilance task. This effect interacted with memory load (Mayleben, 1998). In another vigilance study in which targets were precued with cues of varying reliability, the right hemisphere activation again declined and interacted with cue reliability (Hitchcock et al., 2000). These results, while not providing for localization of function in the brain, do indicate that

optical imaging measures can be sensitive to workload. Because the procedure is relatively cheap, completely noninvasive, and permits subject movement, it may be particularly useful in adaptive automation research. However, additional basic studies examining its sensitivity to variations in workload are needed.

Summary

In the previous three sections of this paper, a variety of physiological measures were reviewed. A summary of this portion of the paper is presented in Table 3. The table lists each of the measures discussed. For each measure, information regarding its sensitivity, diagnosticity, ease of use, current real world/real time feasibility, intrusiveness, and expense associated with obtaining the measure is presented. The question marks indicate that there is insufficient data for evaluating some aspects of that measure. Given the wide range of issues associated with each measure discussed above, the table should be viewed as a guide in considering candidate measures for adaptive automation. In the final section, a program of research using EEG power band ratios to trigger changes in adaptive automation is presented as an example of how psychophysiological measures can be employed in the development of adaptive automation systems.

Table 3. Summary of psychophysiological measures and their current applicability to adaptive automation.

	Sensitivity[1]	Diagnosticity[2]	Ease of use	Current real world/ real time feasibility	Cost	Intrusiveness
EYE BLINKS						
Eye blink rate	high	moderate	high	good	moderate	low
Eye blink ampl.	high	low	high	good	moderate	low
RESPIRATION						
Resp. rate	high	moderate	high	good	moderate	low/moderate
Resp. depth	high	low	high	good	moderate	low/moderate
CARDIOVASCULAR ACTIVITIY						
Heart rate/ period	high	high	high	very good	low	low/moderate
Heart rate Variability	high	moderate	fair	fair	moderate	low/moderate
EEG			moderate	very good	moderate	moderate
Delta	moderate	moderate				
Theta	high	high				
Alpha	high	high				
Beta	moderate	moderate				
EEG Power Band Ratios						
Beta/(Alpha +Theta)	high	high	moderate	very good	moderate	moderate
ERPs			fair	fair	moderate	moderate
N100	high	high				
P300	high	high				
ERN	moderate	?	low	low	moderate	moderate
PET	high	?	low	low	high	high
fMRI	high	?	low	low	high	high
OPTICAL IMAGERY	?	?	moderate	fair	moderate	low

1. Sensitivity refers to whether the measure differentiates baseline from workload.
2. Diagnosticity refers to whether the measure differentiates different levels of workload.

SECTION IV

NASA-Developed Biocybernetic System for Adaptive Task Allocation

To date, the most promising work using psychophysiological measures for adaptive automation centers around a biocybernetic system developed at NASA (Pope, Bogart, & Bartolome; 1995). In their system, EEG signals are recorded and sent to a LabView Virtual Instrument (VI) that determines the power in the alpha, beta, and theta bands for all sites. The VI also calculates an engagement index (see below) and according to the value of that index triggers changes between automatic and manual modes of the computerized task being performed by the operator.

The engagement index adopted by Pope and his colleagues (1995) is based upon the idea that various ratios of EEG power bands (alpha, beta, theta, etc.) can be particularly sensitive to differences in attention and arousal. As noted above, Streitberg, Röhmel, Herrmann, and Kubicki (1987) showed that the collective activity among multiple power bands was useful in distinguishing among stages of vigilance and wakefulness. Also, as noted above Lubar and his associates (Lubar, 1991; Lubar, Swartwood, Swartwood, & O'Donnell, 1995) observed higher theta to beta ratios, particularly over the frontal cortex, for individuals with Attention Deficit/Hyperactivity Disorder compared to controls. More recently, Cunningham, Scerbo, and Freeman (2000) found a relationship between various EEG ratio band indices, daydreaming, and sustained attention. They examined two power band ratios, beta/(alpha+theta) and beta/alpha, in a group of individuals performing a target detection task who were asked to press a button whenever they realized they had been daydreaming. Cunningham et al. observed significant differences in the value of both power band ratios in the intervals before and after the reported daydreams.

Pope et al. (1995) initially studied four different engagement indices: beta/alpha and beta/(alpha+theta) combined from sites, Cz, Pz, P3, and P4, alpha at T5 & P3/alpha at Cz & Pz, and alpha at O1/alpha at O2. Although different indices were being studied, the system always operated the same way. The EEG signals were recorded for 40 sec to obtain a value of the engagement index. Once the session began, the 40-sec window was advanced 2 sec and a new value of the index was calculated.

The task performed by the participants in the Pope et al. (1995) study was the compensatory tracking task from the Multiple-Attribute Task (MAT) Battery (Comstock & Arnegard, 1992). The MAT is a PC-based group of tasks that represent the kind of activities typically performed by pilots. Each task, monitoring, resource management, and compensatory tracking, is displayed in a separate window on the screen. In their study, all of the tasks remained in automatic mode except the tracking task that shifted between automatic and manual modes.

The participants in the study performed under both positive and negative feedback conditions. Under negative feedback, the tracking task was switched to or maintained in automatic mode when the slope of the index derived from two successive 40-sec windows increased. Pope et al. (1995) argued that an increase in the slope of two successive values of the index would reflect an increase in engagement. Likewise, a decrease in slope would reflect a decrease in engagement causing the tracking task to be switched to or maintained in manual mode. The system operated in the opposite manner under positive feedback. Pope and his colleagues (1995) reasoned that under negative feedback, the system would oscillate back and forth between automatic and manual modes more frequently in order to maintain a stable level of engagement. By contrast, under positive feedback the system should migrate toward

47

extreme values of the index and remain in each task mode for longer periods of time resulting in fewer switches between modes.

Pope et al. (1995) had participants perform under both positive and negative feedback conditions and found that more switches occurred between automatic and manual modes under negative as opposed to positive feedback conditions. Further, the system operated best under the index, beta/(alpha+theta).

In a subsequent set of studies, Freeman, Milkula, Prinzel and Scerbo (1999) repeated the experiment by Pope et al. (1995) but with a different set of goals in mind. First, they were interested in validating system operation by examining the values of three candidate engagement indices: 1/alpha, beta/alpha, and beta/(alpha+theta). Second, they were interested in whether the negative feedback condition which was designed to stabilize engagement would result in superior tracking performance. Third, they reasoned that using the slope of the index might not be the best representation of engagement. Specifically, they argued that any change in the polarity of the slope irrespective of magnitude could produce a switch in task modes. Thus, if an operator's level of engagement was substantially below his or her mean for the session, the system could change task modes after only a slight increase in the value of the index even though the overall level of engagement was still quite low. A better approach would be to obtain a stable baseline of the index and use any deviation from the absolute value of the baseline index to trigger the switches between task modes. Freeman et al. modified the system to switch modes accordingly and had their participants perform under both positive and negative feedback conditions.

The results of this experiment showed that the system performed as expected. Data for the index, beta/(alpha+theta), are shown in Table 4. Under negative feedback, when the value of the index was high (reflecting higher engagement) the task was switched to automatic mode and when the value was low (i.e., lower engagement) the task was switched to manual mode. The opposite pattern occurred under positive feedback. A similar pattern was observed for the other two indices although the differences between automatic and manual modes within each feedback condition were more pronounced.

Freeman et al. (1999) also examined performance on those periods where the subject manually operated the tracking task. They found that tracking performance improved under negative as compared to positive feedback. Moreover, this improvement was greater when the absolute value of the index was used to trigger changes between task modes. A subsequent study showed that this advantage for better tracking performance under negative feedback appears to be quite stable. Freeman, Mikulka, Scerbo, Prinzel, and Clouatre (2000) observed similar results with individuals who performed the task over much longer intervals.

Table 4. Engagement Index (Beta/(Alpha+Theta))Values for Automatic and Manual Modes under Negative and Positive Feedback.

Task Mode	Automatic	Manual
Negative Feedback	15.50	12.20
Positive Feedback	10.01	15.32

One important parameter that might affect the sensitivity of the biocybernetic, adaptive system is the size of the measurement interval or window used to compute the engagement index. It is conceivable that better sensitivity might be achieved with a smaller window. Hadley, Mikulka, Freeman, Scerbo, and Prinzel (1997) examined this possibility by comparing window sizes of 40 and 4 seconds. These investigators, once again, observed an interaction between feedback condition and task mode similar to the one shown in Table 2; however, the differences between task mode and feedback condition were more pronounced under the 4-sec window. In addition, not only did the smaller window generate more switches between task modes it also resulted in better tracking performance. Hadley et al. concluded that a narrower window may not only improve the sensitivity of the system to changes in engagement, but may facilitate performance as well.

Workload. In another study, Prinzel, Freeman, Scerbo, Mikulka, and Pope (2000) examined the effects of task load on performance. The low workload condition replicated the procedure of Pope et al. (1995) in that participants performed only the tracking portion of the MAT task, i.e., the monitoring and resource management tasks remained in automatic mode. Under the high workload condition, however, the monitoring and resource management tasks remained in manual mode and the participants had to perform all three tasks simultaneously. In both workload conditions, only the tracking task switched between automatic and manual modes. These investigators expected that the system would make more task allocations under the high workload condition because of the operator's need to address the unpredictable demands of three different tasks. In addition, the investigators assessed subjective estimates of workload with the NASA-Task Load Index (TLX; Hart & Staveland, 1988). In addition, the data from participants operating within the closed-loop system were compared to those of a control group who performed the same tasks without the closed-loop system.

Prinzel et al. (2000) found that once again, more switches were made between automatic and manual modes for the tracking task under negative as compared to positive feedback. Further, more switches between task modes were also observed in the high workload condition. The data for the tracking task revealed that performance was better under negative as opposed to positive feedback and that further, performance was better in the low workload condition. An analysis of the TLX scores confirmed that the high workload condition was indeed rated higher in subjective workload than the low workload condition. Perhaps the most important finding was that performance was better and workload was rated lower for those participants operating within the closed loop system than those in the control group. Collectively, these results suggest that under negative feedback conditions, the biocybernetic adaptive automation system does indeed moderate workload and bolster performance.

Vigilance. One of the limitations to this line of research was that all of the experimentation had

been done on a continuous psychomotor tracking task. Obviously, the idea of an adaptive automation system driven by physiological measures has little merit if it only applies to one type of task. Mikulka, Hadley, Freeman, and Scerbo (1999) investigated how the biocybernetic, closed loop system functioned using a task that required sustained attention. Several researchers have noted that when individuals are required to monitor events for critical signals over extended periods of time, the presentation rate of events has a significant effect on performance. Specifically, as the event rate increases the ability to detect critical signals declines (Dember & Warm, 1979; See, Howe, Warm, & Dember, 1995).

In their study, Mikulka et al. (1999) asked observers to monitor the repetitive presentation of a pair of lines for an occasional increase in line length over a 40-min vigil. The presentation rate varied among 6, 20, or 60 events per minute. Using the negative feedback contingency with the biocybernetic system, when the EEG index reflected higher levels of engagement the event rate was lowered. By contrast, under lower levels of engagement event rates were increased. The investigators compared the performance of participants working within the closed loop system to that of a yoked control group who received the same pattern of increases and decreased in event rate, but whose EEG was not used to drive those changes. The results showed that the ability to detect critical signals declined over the session for both groups, but less so for the individuals using the biocybernetic system. These results are important because they suggest that the biocybernetic, closed loop system may help to bolster performance on activities outside of those for which it was originally designed.

Task partitioning with a physiologically-based system. Eischeid, Scerbo, and Freeman (1998) examined the effects of task partitioning and computer skill using the biocybernetic, closed-loop system. In this experiment, a compensatory tracking task was used and partitioned into horizontal and vertical axes. This permitted three modes of operation. In the manual mode, the participant controlled both axes while in the automatic mode the computer controlled both axes. In the third or partitioned mode, the participant and computer each controlled one axis. The index, beta/(alpha+theta), was used to invoke changes in automation mode and participants were assigned to teams with a computer partner that performed at either an expert or novice level.

The results showed that the skill level of the computer teammate interacted with the automation mode. Specifically, those assigned to work with the expert computer had similar tracking scores in the manual and partitioned modes. On the other hand, the performance of those who worked with the novice-level teammate was worse in the partitioned mode than in the manual mode. Once again, these findings support the idea that there is a disadvantage to working with a teammate of lesser skill. In this instance, the skill level of the novice computer was so poor that the participants' tracking performance along their own *independent* axis in the partitioned mode was lower than what they could achieve if they were required to track both axes in the manual mode. Eischeid, Scerbo, and Freeman (1998) argued that for task partitioning to be beneficial to operator performance in an adaptive environment, the skill level of the computer teammate would have to be equal to or greater than that of the operator.

Other Psychophysiological Measures. Prinzel et al. (1998) further explored the "developmental" (Byrne & Parasuraman, 1996) capabilities of the system for adaptive automation design. Such a system would not have much utility if the potential of the system was limited to the use of EEG measures. Therefore, considerable research effort has been directed towards examining the use of the system with other psychophysiological measures. Prinzel et al. demonstrated that the system could also make task allocation decision on the basis of the P300 and N100 components of the ERP. The results support other research (e.g., Humphrey & Kramer, 1994) that also reported on the possibility for ERP-based measures of workload as a "trigger" for implementing adaptive technologies. Currently, research at

the NASA Langley Research Center (Physiological / Psychological Stressors & Factors project, Dr. Lawrence J. Prinzel, manager) has been directed towards the development of neural network algorithms that will utilize performance and EEG and HRV measures in determining when a pilot may be in a "hazardous states of awareness" (Pope & Bogart, 1992). On the basis of normative research and subject-matter expert assessments, pilot state awareness profiles are developed that consider mission (e.g., phase of flight), environmental (e.g., turbulence), and aircraft states (e.g., operational configuration). Pilot state is then modulated through adaptive task allocation and adaptive interface methods to help bring the pilot back "in-the-loop". NASA Langley Research Center is also supporting similar research at other universities that will examine the efficacy of combining these measures to make task allocation decisions.

References

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

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). New York: Plenum.

Backs, R. W. (1995). Going beyond heart rate: Modes of autonomic control in the cardiovascular assessment of mental workload. *The International Journal of Aviation Psychology, 5*, 25-48.

Backs, R. W. (1997). Psychophysiological aspects of selective and divided attention during continuous manual tracking. *Acta Psychologia, 96*, 167-191.

Backs, R. W., Lenneman, J. K., & Sicard, J. L. (1999). The use of autonomic components to improve cardiovascular assessment of mental workload in flight simulation. *The International Journal of Aviation Psychology, 9*, 33-47.

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

Baddeley, A. D. (1992). Working memory. *Science, 255*, 556-559.

Barcelo, F., Gale, A., & Hall, M. (1995). Multichannel EEG power reflects information processing and attentional demands during visual orienting. *Journal of Psychophysiology, 9*, 32-44.

Barnes, M., & Grossman, J. (1985). *The intelligent assistant concept for electronic warfare systems.* China Lake, CA. NAWC TP 5585.

Basar, E., Basar-Eroglu, C., Karakas, S., & Schuermann, M. (2000). Brain oscillations in perception and memory. *International Journal of Psychophysiology, 35*, 95-124.

Bauer, H. (1984). Regulation of slow brain potentials affects task performance. In T.Elbert, B. Rockstroh, W. Lutzenberger, & N. Birbaumer (Eds.), *Self-regulation of the Brain and Behavior*. Berlin: Springer.

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 & H.D. Kimmel (Eds.), *Biofeedback and self-regulation* (pp. 235-246). Hillsdale, NJ: Erlbaum.

Benham, G., Rasey, H.W., Lubar, J. F., Frederick, J. A., & Zoffuto, A. C. (1997). EEG power-spectral and coherence differences between attentional states during a complex auditory task. *Journal of Neurotherapy, 2,* 1-9.

Bernstein, P. S., Scheffers, M. K., & Coles, M. G. H. (1995). "Where did I go wrong?" A psychophysiological analysis of error detection. *Journal of Experimental Psychology: Human Perception and Performance*, *21,* 1312-1322.

Berntson, G. G., Cacioppo, J. T., & Quigley, K. S. (1993). Cardiac psychophysiology and autonomic space in humans: empirical perspectives and conceptual implications. *Psychological Bulletin, 114*, 296-322.

Billings, C.E. (1991). *Human-centered aircraft automation philosophy: A concept and guidelines*. Moffett Field, CA: NASA Technical Memorandum No. 103885.

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.

Birbaumer, N., Elbert, T., Canavan, A.G.M., & Rockstroh, B. (1990). Slow potentials of the cerebral cortex and behavior. *Physiological Reviews, 70,* 1-41.

Boutcher, S. H., Nugent, F. W., McClaren, P. F., & Weltman, A. L. (1998). Heart period variability of trained and untrained men at rest and during mental challenge. *Psychophysiology, 35*, 16-22.

Brenner, M., Doherty, E. T., & Shipp, T. (1994). Speech measures indicating workload demand. *Aviation, Space, and Environmental Medicine, January,* 21-26.

Brody, S., Rau, H., Kohler, F., Schupp, H., & Lutzenberger, W. (1994). Slow cortical potential biofeedback and the startle reflex. *Biofeedback and Self-Regulation, 19*, 1-11

Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology, 42,* 361-377.

Buick, F. (1989). *+Gz protection in the future - review of scientific literature*. Technical report DCIEM 89-RR-47. Defense and Civil Institute of Environmental Medicine, Downsview, Ontario, Canada.

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

Cabon, P., Coblentz, A., Mollard, R., & Fouillot, J. P. (1993). Human vigilance in railway and long-haul flight operation. *Ergonomics, 36*, 1019-1033.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, *280*, 747-749.

Colucci, F. (1995). Rotorcraft pilot's associate update: The army's largest science and technology program, *Vertiflite*, March/April.

Comstock, J.R., Jr., & Arnegard, R.J. (1992). *The multi-attribute task battery for human operator workload and strategic behavior research*. National Aeronautics and Space Administration Technical Memorandum No. 104174.

Corbetta, M. (1998). Functional anatomy of visual attention in the human brain: Studies with positron emission tomography. In R. Parasuraman (Ed.), *The attentive brain* (pp. 95-122). Cambridge, MA: MIT Press.

Crews, D. J., Landers, D. M. (1993). Electroencephalographic measures of attentional patterns prior to the golf putt. *Medicine & Science in Sports & Exercise, 25*, 116-126.

Cunningham, S., Scerbo, M.W., & Freeman, F.G. (2000). The electrocortical correlates of daydreaming during vigilance tasks. *Journal of Mental Imagery, 24*, 61-72.

Daniel, R.S. (1967). Alpha and theta EEG in vigilance. *Perceptual and Motor Skills, 25*, 697-703.

Davies, D.R. & Krkovic, A. (1965). Skin conductance, alpha-activity, and vigilance. *American Journal of Psychology, 78*, 304-306.

Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a neural system for error detection and compensation. *Psychological Science*, *5*, 303-305.

Dember, W.N., & Warm, J.S. (1979). *Psychology of perception, Second Edition*. New York: Holt, Rinehart, and Winston.

Donchin, E., Kramer, A. F., & Wickens, C. (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, N.J: Till & Till.

Dujardin, K., Derambure, P., Defebvre, L; Bourriez, J. L, et al. (1993). Evaluation of event-related desynchronization (ERD) during a recognition task: Effect of attention. *Electroencephalography & Clinical Neurophysiology, 86*, 353-356.

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

Eischeid, T.M., Scerbo, M.W., & Freeman, F.G. (1998). The effects of task partitioning and computer skill on engagement and performance with an adaptive, biocybernetic system. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, 133-137.

Elbert, T., Rockstroh, B., Lutzenberger, W., & Birbaumer, N. (1980). Biofeedback of slow cortical potentials. I. *Electroencephalography & Clinical Neurophysiology, 48*, 293-301.

Endsley, M. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32-64.

Evans, J.R., & Abarbanel, A. (1999). Quantitative EEG and Neurofeedback. New York: Academic Press.

Fahrenberg, J., & Wientjes, C. J. E. (2000). Recording methods in applied environments. In R. W. Backs & W. Boucsein (Eds.), *Engineering Psychology: Issues and Applications* (pp. 111-136). London: Lawrence Erlbaum Associates.

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1990). Effects of errors in choice reaction tasks on the ERP under focused and divided attention. In C.H.M. Brunia, A.W.K. Gaillard, & A. Kok (Eds.), *Psychophysiological brain research*, (pp. 192-195). Tilburg: Tilburg University Press.

Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: a tutorial. *Biological Psychology*, *51*, 87-107.

Falkenstein, M., Koshlykova, N.A., Kiroj, V.N., Hoormann, J., & Hohnsbein, J. (1995). Late ERP components in visual and auditory Go/Nogo tasks. *Electroencephalography and Clinical Neurophysiology*, *96*, 36-43.

Fernandez, T., Harmony, T., Rodriguez, M., Bernal, J., Silva, J., Reyes, A., & Marosi, E. (1995). EEG activation patterns during the performance of tasks involving different components of mental calculation. *Electroencephalography & Clinical Neurophysiology, 94,* 175-182.

Fogarty, C., & Stern, J. A. (1989). Eye movements and blinks: their relationship to higher cognitive processes. *International Journal of Psychophysiology, 8*, 35-42.

Ford, M., Bird, B.L., Newton, F.A., & Sheer, D. (1980). Maintenance and regulation of 40-Hz EEG: Biofeedback effects. *Biofeedback & Self-Regulation, 5*, 193-203.

Fournier, L. R., Wilson, G. F., & Swain, C. R. (1999). Electrophysiological, behavioral, and subjective indexes of workload when performing multiple tasks: manipulations of task difficulty and training. *International Journal of Psychophysiology, 31,* 129-145.

Frederick, J. A., Lubar, J.F., Rasey, H.W., Brim, S.A., & Blackburn, J. (1999). Effects of 18.5 Hz auditory and visual stimulation on EEG amplitude at the vertex. *Journal of Neurotherapy, 3*, 23-28.

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

Freeman, F. G., Mikulka, P. J., Scerbo, M. W., Prinzel, L. J., & Clouatre, K. (2000). Evaluation of a psychophysiologically controlled adaptive automation system, using performance on a tracking task. *Applied Psychophysiology and Biofeedback, 25*, 103-115.

Freude, G. & Ullsperger, P. (2000). Slow brain potentials as a measure of effort? Applications in mental workload studies in laboratory settings. In R. W. Backs, & W. Boucsein, (Eds.), *Engineering psychophysiology: Issues and applications* (pp. 255-267). Mahwah, NJ: Lawrence Erlbaum Associates.

Gale, A. (1977). Some EEG correlates of sustained attention. In R. R. Mackie (Ed.), *Vigilance: Theory, Operational Performance, and Physiological Correlates* (pp. 263-283). New York: Plenum.

Gale, A., Davies, R., & Smallbone, A. (1977). EEG correlates of signal rate, time in task, and individual differences in reaction time during a five-stage sustained attention task. *Ergonomics, 20,* 363-376.

Gale, A., & Edwards, J. (1983). The EEG and human behavior. In A. Gale & J. Edwards (Eds.), *Physiological correlates of human behavior* (pp. 156-185). New York: Academic Press.

Geddes, N. (1985). Intent inferencing using scripts and plans. In *Proceedings of the First Annual Aerospace Applications of Artificial Intelligence Conference* (pp. 123-127). Washington DC: AAAI.

Gehring, W.J., Goss, B., Coles, M.G.H., Meyer, D.E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science*, *4*, 385-390.

Gehring, W. J., & Knight, R. T. (2000). Prefrontal-cingulate interactions in action monitoring. *Nature Neuroscience*, *3*, 516-520.

Gevins, A.S. & Cutillo, B.A. (1993) Spatiotemporal dynamics of component processes in human working memory. *Electroencephalography & Clinical Neurophysiology, 87*, 128-143.

Gevins, A., Leong, H., Du, R., Smith, M. E., et al. (1995). Towards measurement of brain function in operational environments. *Biological Psychology, 40*, 169-186.

Gevins, A. S., Smith, M.E., Leong, H., McEvoy, L., Whitfield, S., Du, R., & Rush, G. (1998). Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human Factors, 40*, 79-91.

Gevins, A. S., Smith, McEvoy, L., & Yu, D. (1997). High resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice. *Cerebral Cortex, 7*, 374-385.

Gevins, A.S., Zeitlin, G.M., Doyle, J.C., Schaffer, R.E., & Callaway, E. (1979). EEG patterns during "cognitive" tasks: II. Analysis of controlled tasks. *Electroencephalography & Clinical Neurophysiology, 47*, 704-710.

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

Gobel, M., Springer, J., & Schreff, J. (1998). Stress and strain of short haul bus drivers: psychophysiology as a design oriented method for analysis. *Ergonomics, 41,* 563-580.

Gomer, F.E. (1981). Physiological monitoring and the concept of adaptive systems. In J. Moraal & K.-F. Kraiss (Eds.), Manned systems design: Methods, equipment, and applications (pp. 271-287). New York: Plenum Press.

Gruber, T., Mueller, M.M., Keil, A., & Elbert, T. (1999). Selective visual-spatial attention alters induced gamma band responses in the human. *Electroencephalography & Clinical Neurophysiology, 110*, 2074-2085.

Gundel, A., Drescher, J., Maas, H., Samel, A., & Vejvoda, M. (1995). Sleepiness of civil airline pilots during two consecutive night flights of extended duration. *Biological Psychology, 40,* 131-141.

Gundel, A., & Wilson, G.F. (1992). Topographical changes in the ongoing EEG related to the difficulty of mental tasks. *Brain Topography, 5*, 17-25.

Hadley, G., Mikulka, P.J., Freeman, F.G., & Scerbo, M.W. (1997). An examination of system sensitivity in a biocybernetic adaptive system. *Proceedings of the Ninth symposium on Aviation Psychology*, (pp. 305-309).

Hammer, J.M., & Small, R.L. (1995). An intelligent interface in an associate system. In W.B. Rouse (Ed.) *Human/technology interaction in complex systems, Vol. 7*, (pp. 1-44). Greenwich, CT: JAI Press.

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.

Hankins, T.C. & Wilson, G. F. (1998). A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. *Aviation, Space & Environmental Medicine, 69*, 360-367.

Harmony, T., Fernandez, T., Silva, J., Bernal, J., Diaz-Comas, L., Reyes, A., Marosi, E., Rodriguez, M., & Rodriguez, M. (1996). EEG delta activity: An indicator of attention to internal processing during performance of mental tasks. *International Journal of Psychophysiology, 24*, 161-171.

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.

Hasan, J., Hirvonen, K., Varri, A., Hakkinen, V., & Loula, P. (1993). Validation of computer analysed polygraphic patterns during drowsiness and sleep onset. *Electroencephalography & Clinical Neurophysiology, 87*, 117-127.

Hatfield, B.D., Landers, D.M., & Ray, W.J. (1984). Cognitive processes during self-paced motor performance: an electroencephalographic profile of skilled marksmen. *Journal of Sports Exercise Psychology, 6*, 42-59.

Hilburn, B., Jorna, P. G., Byrne, E. A., & Parasuraman, R. (1997). The effect of adaptive air traffic control (ATC) decision aiding on controller mental workload. In M. Mouloua & J. Koonce (Eds.), *Human-automation interaction: Research and practice* (pp. 84-91). Mahwah, NJ: Erlbaum Associates.

Hillman, C.H., Apparies, R.J., Janelle, C.M., & Hatfield, B.D. (2000). An electrocortical comparison of executed and rejected shots in skilled marksmen. *Biological Psychology, 52,* 71-83.

Hitchcock, E. M., Warm, J. S., Dember, W. N., Matthews, G. M., Shear, P. K., Rosa, R. R., Tripp, L., Mayleben, D. W., & Parasuraman, R. (2000). Effects of signal salience and cueing on cerebral blood flow velocity during sustained attention. *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting*, 3.382-3.385.

Hockey, G.R.J. (1997). Compensatory control in the regulation of human performance under stress and high workload: a cognitive-energetical framework. *Biological Psychology, 45,* 73-93.

Holroyd, C. B., Dien, J., & Coles, M. G. H. (1998). Error-related scalp potentials elicited by hand and foot movements: Evidence for an output-independent error-processing system in humans. *Neuroscience Letters*, *242*, 65-68.

Inagaki, T. (2000). Situation-adaptive autonomy: Dynamic trading of authority between human and automation. *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting*, 3.13-3.16.

Inagaki, T., Takae, Y., & Moray, N. (1999). Automation and human-interface for takeoff safety. *Proceedings of the 10th International Symposium on Aviation Psychology*, 402-407.

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

Jiang, Y., Haxby, J. V., Martin, A., Ungerleider, L. G., & Parasuraman, R. (2000). Complementary neural mechanisms for tracking items in human working memory. *Science, 287,* 643-646.

John, E.R. & Easton, P. (1995). Quantitative electrophysiological studies of mental tasks. *Biological Psychology, 40*, 101-113.

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

Kaber, D. B., & Riley, J. M. (1999). Adaptive automation of a dynamic control task based on secondary task workload measurement. *International Journal of Cognitive Ergonomics, 3*, 169-187.

Kantowitz, B. H. (1988). Mental workload. In P. A. Hancock (Ed.), *Human Factors Psychology*. Amsterdam: Elsevier.

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

Khatwa, R., & Roelen, A. L. C. (1996). An analysis of controlled-flight-into-terrain (CFIT) accidents of commercial operators, 1988 through 1994. *Flight Safety Digest, 15* (4/5), 1-45.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews, 29*, 169-195.

Klose, P. H., Lewis, G. D., Messing, W. P., Kasperski, R. R., & Flemming, J. M. (1992). Noninvasive infrared cerebral oximetry. In *SPIE Proceedings (Physiological monitoring and early detection method, 164,* 202-207.

Kopp, B., Rist, F., & Mattler, U. (1996). N200 in the flanker task as a neurobehavioral tool for investigating executive control. *Psychophysiology*, *33*, 282-294.

Kotchoubey, B., Busch, S., Strehl, U., & Birbaumer, N. (1999). Changes in EEG power spectra during biofeedback of slow cortical potentials in epilepsy. *Applied Psychophysiology and Biofeedback, 24*, 213-233.

Kotchoubey, B., Schleichert, H., Lutzenberger, W., & Birbaumer, N. (1997). A new method for self-regulation of slow cortical potentials in a timed paradigm. *Applied Psychophysiology and Biofeedback, 22*, 77-93.

Krahl, K., & Scerbo, M.W. (1997). Performance on an Adaptive Automated Team Tracking Task with Human and Computer Teammates. *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting*, 551-555.

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

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

Krause, C.M., Lang, H.A., Laine, M., Kuusisto, M.J., & Porn, B. (1996). Event-related EEG desynchronization and synchronization during an auditory task. *Electroencephalography & Clinical Neurophysiology, 98*, 319-326.

Kuhlman, W.N. (1978). Functional topography of the human mu rhythm. *Electroencephalography & Clinical Neurophysiology, 43*, 83-93.

Kurlik, A. (1993). Modeling strategic behavior in human-automation interaction: Why and "aid" can (and should) go unused. *Human Factors, 35,* 221-242.

Landers, D.M., Petruzzello, S.J., Salazar, W. Crews, D.J., et al. (1991). The influence of electrocortical biofeedback on performance in pre-elite archers. *Medicine & Science in Sports & Exercise, 23*, 123-129.

Lane, J.D., Kasian, S.J., Owens, J.E., & Marsh, G.R. (1998). Binaural auditory beats affect vigilance performance and mood. *Physiology and Behavior, 63*, 249-252.

Lang, W., Lang, M., Kornhuber, V., Diekmann, V., & Kornhuber, H.H. (1988). Even-related EEG-spectra in a concept formation task. *Human Neurobiology, 6*, 295-301.

Laukka, S.J., Jarvilehto, T., Alexandrov, Y.I., & Lindqvist, J. (1995). Frontal midline theta related to learning in a simulated driving task. *Biological Psychology, 40*, 313-320.

Lawton, G.W., Hung, T.M. Saarela, P., & Hatfield, B.D. (1998). Electroencephalography and mental states associated with elite performance. *Journal of Sport and Exercise Psychology, 20*, 35-53.

Lee, J.D., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics, 35*, 1243-1270.

Lorenzo, I., Ramos, C.A., Guevara, M.A., & Corsi-Cabrrera, M. (1995). Effect of total sleep deprivation on reaction time and waking EEG activity in man. *Sleep, 18*, 346-354.

Loring, D.W. & Sheer, D.E. (1984). Laterality of 40 Hz EEG and EMG during cognitive performance. *Psychophysiology, 21*, 34-38.

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

Lubar, J.F. (1997). Neocortical dynamics: Implications for understanding the role of neurofeedback and related techniques for the enhancement of attention. *Applied Psychophysiology and Biofeedback, 22*, 111-126.

Lubar, J.F. & Lubar, J.O. (1999). Neurofeedback assessment and treatment for Attention Deficit/Hyperactivity Disorders. In J. R. Evans, & A. Abarbanel (Eds.), *Introduction to quantitative EEG and neurofeedback* (pp. 103-143). San Diego, CA: Academic Press.

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

Lutzenberger, W., Elbert, T., Rockstroh, B, & Birbaumer, N. (1979). Effects of slow cortical potentials on performance in a signal detection task. *International Journal of Neuroscience, 9*, 175-183.

Lutzenberger, W., Elbert, T., Rockstroh, B, & Birbaumer, N. (1983). Biofeedback of slow cortical potentials and its effects on performance in mental arithmetic task. *Biological Psychology, 14*, 99-111.

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

Makeig, S., & Jung, T. (1995). Changes in alertness are a principal component of variance in the EEG spectrum. *Neuroreport: an International Journal for the Rapid Communication of Research in Neuroscience, 7*, 213-216.

Malin, J. T., & Schreckenghost, D.L. (1992). *Making intelligent systems team players: Overview for designers.* NASA Technical Memorandum 104751, Johnson Space Center, Houston, TX.

Mangun G R. (1995). Neural mechanism of visual selective attention. *Psychophysiology*, *32*, 4-18.

Mayleben, D. (1998). *Cerebral blood flow during sustained attention*. Unpublished doctoral dissertation. University of Cincinnati, Cincinnati, OH.

McFarland, D.J., Neat, G.W. Read, F.F., & Wolpaw, J.R. (1993). An EEG-based method for graded cursor control. *Psychobiology, 21*, 77-81.

Mecklinger, A., Dramer, A.F., & Strayer, D.L. (1992). Event related potentials and EEG components in a semantic memory search task. *Psychophysiology, 29*, 104-119.

Mikulka, P., Hadley, G., Freeman, F., & Scerbo, M. (1999). The effects of a biocybernetic system on vigilance decrement. *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting*, 1410.

Miller, C. A., Guerlain, S., & Hannen, M. D. (1999). The Rotorcraft Pilot's Associate cockpit information manager: Acceptable behavior from a new crew member. Paper presented at the American Helicopter Society 55th Annual Forum, Montreal, Quebec. May 25-27.

Miller, J.C. (1995). Batch processing of 10000 h of truck driver EEG data. *Biological Psychology, 40*, 209-222.

Miltner, W. H. R., Braun C., & Coles, M. G. H. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a "generic" neural system for error-detection. *Brain Topography*, *10*, 165-166.

Morris, T. L., & Miller, J. C. (1996). Electrooculographic and performance indices of fatigue during simulated flight. *Biological Psychology, 42*, 343-360.

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 Associates.

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.

Muir, B.M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies, 27*, 527-539.

Mulder, G., Mulder, L. J. M., Meijman, T. F., Veldman, J. B. P., & van Roon, A. M. (2000). A psychophysiological approach to working conditions. In R. W. Backs & W. Boucsein (Eds.), *Engineering psychology: Issues and applications* (pp. 139-159). Mahwah, NJ: Lawrence Erlbaum Associates.

Myrtek, M., Weber, D., Brugner, G., & Muller, W. (1996). Occupational stress and strain for female students: results of physiological, behavioral, and psychological monitoring. *Biological Psychology, 42*, 379-391.

Näätänen, R. (1992). *Attention and brain function.* Mahwah,NJ: Erlbaum.

Nakashima, K., & Sato, H. (1992). The effects of various mental tasks on appearance of frontal midline theta activity in EEG. *Journal of Human Ergology, 21*, 201-206.

National Transportation Safety Board (1990). Special Investigation Report - Runway overruns following high speed rejected takeoffs. SIR-90/02.

Nestor, P., Parasuraman, R., Haxby, J. V. , & Grady, C. L. (1991). Divided attention and brain metabolic dysfunction in mild dementia of the Alzheimer type. *Neuropsychologia, 29*, 379-387.

O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Vol. II. Cognitive processes and performance* (pp. 42-1 – 42-49). New York: Wiley.

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: Theory, operational performance and physiological correlates* (pp. 189-201). New York: Plenum.

O'Hanlon, J.F., Royal, J.W., & Beatty, J. (1977). EEG theta regulation and radar vigilance performance of professional radar operators and air traffic controllers under laboratory and controlled field conditions. In J. Beatty, & H. Legewie (Eds.), *Biofeedback and behavior* (pp. 47-165). New York: Plenum.

Parasuraman, R. (1983). Vigilance, arousal, and the brain. In A. Gale, & J. Edwards (Eds.), *Physiological correlates of human behavior* (pp. 139-143). New York: Academic Press.

Parasuraman, R. (1985). Event-related brain potentials and intermodal divided attention. In *Proceedings of the Human Factors Society, 29th Annual Meeting*, 971-975.

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

Parasuraman, R. (1998). *The attentive brain.* Cambridge, MA: MIT Press.

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

Parasuraman, R., & Caggiano, D. (in press). Mental workload. In V. S. Ramachandran (Ed.) *Encyclopedia of the human brain.* San Diego: Academic Press.

Parasuraman, R., & Hancock. P. A. (2000). Adaptive control of workload. In P. A. Hancock & P. E. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 305-320). Mahwah, NJ: Erlbaum.

Parasuraman, R., Mouloua, M., & Hilburn, B. (1999). Adaptive aiding and adaptive allocation enhance human-machine interaction. In Scerbo, M.W., & Mouloua, M. (Eds.), *Automation technology and human performance: Current research and trends* (pp. 119-123). Mahwah, NJ: Lawrence Erlbaum.

Parasuraman, R., Mouloua, M., Molloy, R., & Hilburn, B. (1996). Monitoring of automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 91-115). Mahwah, NJ: Lawrence Erlbaum Associates.

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

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

Patrick, G.P. (1996). Improved neuronal regulation in ADHD: an application of fifteen sessions of photic-driven EEG neurotherapy. *Journal of Neurotherapy,* 1.

Paus, T., Zatorre, R.J., Hofle, N., Caramanos, Z., et al. (1997). Time-related changes in neural systems underlying attention and arousal during the performance of an auditory vigilance task. *Journal of Cognitive Neuroscience, 9*, 392-408.

Pennekamp, P., Boesel, R., Mecklinger, A., & Ott, H. (1994). Differences in EEG-theta for responded and omitted targets in a sustained attention task. *Journal of Psychophysiology, 8*, 131-141.

Petruzzello, S.J., Landers, D.M., & Salazar, W. (1991). Biofeedback and sport/exercise performance: Applications and limitations. *Behavior Therapy, 22*, 379-392.

Pfurtscheller, G. (1992). Event-related synchronization (ERS): an electrophysiological correlate of cortical areas at rest. EEG. *Electroencephalography & Clinical Neurophysiology, 83*, 62-69.

Pfurtscheller, G. & Aranibar, A. (1977). Event-related cortical desynchronization detected by power measurement of scalp EEG. *Electroencephalography & Clinical Neurophysiology, 42*, 817-826.

Pfurtscheller, G. & Klimesch, W. (1991). Event related desynchronization during motor behavior and visual information processing. In C.H.M. Brunia, G. Mulder, & M.N. Verbaten (Eds.), *Event-related Brain Research* (EEG Suppl. 42), pp. 58-65.

Pope, A.T., & Bogart, E.H. (1992). Identification of hazardous awareness states in monitoring environments. *SAE 1992 Transactions: Journal of Aerospace, 101*, 449-456.

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

Porges, S. W., & Byrne, E. A. (1992). Research methods for measurement of heart rate and respiration.

Biological Psychology, 34, 93-130.

Posner, M. I., & DiGirolamo, G. J. (1998). Conflict, target detection and cognitive control. In R. Parasuraman (Ed.), *The attentive brain*. Cambridge: MIT Press.

Posner, M. I., & Tudela, P. (1997). Imaging resources. *Biological Psychology, 45,* 95-107.

Prinzel, L.J., Freeman, F.G., Mikulka, P.J., Scerbo, M.W., & Pope, A.T. (1998). An assessment of the efficacy of event-related potentials for adaptive automation. *Proceedings of the Human Factors & Ergonomics Society, 42*, 1602.

Prinzel, L.J., Freeman, F.G., Scerbo, M.W., Mikulka, P.J., & Pope, A.T. (2000). A closed-loop system for examining psychophysiological measures for adaptive automation. *International Journal of Aviation Psychology, 10,* 393-410.

Rasey, H.W., Lubar, J.F., McIntyre, A., Zoffuto, A.C., & Abbott, P.L. (1996). EEG biofeedback for the enhancement of attentional processing in normal college students. *Journal of Neurotherapy*, 1, 15-21.

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

Ray, W.J. (1990). The electrocortical system. In J. T. Cacioppo, & L. G. Tassinary (Eds.), *Principles of psychophysiology: Physical, social, and inferential elements* (pp. 385-412). Cambridge: Cambridge University Press.

Ray, W.J. & Cole, H.W. (1985). EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes. *Science, 228*, 750-752.

Riley, V. (1996). Operator reliance on automation: Theory and data. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 19-35). Mahwah, NJ: Lawrence Erlbaum Associates.

Roberts, L., Rockstroh, B., Lutzenberger, W., Elbert, T., & Birbaumer, N. (1989). Self-reports during feedback regulation of slow cortical potentials. *Psychophysiology, 26*, 392-403.

Rockstroh, B., Elbert, T., Lutzenberger, W., & Birbaumer, N. (1982). The effects of slow cortical potentials on response speed. *Psychophysiology, 19*, 211-217.

Rosenfeld, J.P., Reinhart, A.M., & Srivastava, S. (1997). The effects of alpha (10-Hz) and beta (22-Hz) "entrainment" stimulation on the alpha and beta EEG bands: individual differences are critical to prediction of effects. *Applied Psychophysiology and Biofeedback, 22*, 3-20.

Rouse, W.B. (1976). Adaptive allocation of decision making responsibility between supervisor and computer. In T.B. Sheridan & G. Johannsen (Eds.), *Monitoring behavior and supervisory control* (pp. 295-306). New York: Plenum Press.

Rouse, W.B. (1977). Human-computer interaction in multi-task situations. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-7*, 384-392.

Rouse, W.B., Geddes, N.D., & Curry, R.E. (1987-1988). An architecture for intelligent interfaces: Outline of an approach to supporting operators of complex systems. *Human-Computer Interaction, 3*, 87-122.

Rouse, W.B., & Rouse, S.H. (1983). *A framework for research on adaptive decision aids*. Technical Report AFAMRL-TR-83-082. Wright-Patterson Air Force Base, OH: Air Force Aerospace Medical Research Laboratory.

Roy, C. S., & Sherrington, C. S. (1890). On the regulation of the blood supply of the brain. *Journal of Physiology (London), 11,* 85-108.

Rugg, M. D., & Coles, M. G. H. (1995). *Electrophysiology of mind: Event-related brain potentials and cognition.* Oxford: Oxford University Press.

Salazar, W., Landers, D.M., Petruzzello, S.J., Han, M.W. ,Crews, D.J., & Kubitz, K.A. (1990). Hemispheric asymmetry, cardiac response, and performance in elite archers*. Research Quarterly for Exercise and Sport, 61*, 351-359.

Sammer, G. (1998). Heart period variability and respiratory changes associated with physical and mental load: non-linear analysis*. Ergonomics, 41*, 746-755.

Sarter, N.B., & Woods, D.D. (1995). How in the world did we ever get into that mode? Mode errors and awareness in supervisory control. *Human Factors, 37*, 5-19.

Scallen, S.F., & Hancock, P.A. (1997). Performance effects of adaptive allocation. In R.S. Jensen & L. Rakovan (Eds.), *Proceedings of the Ninth International Symposium on Aviation Psychology* (pp. 316-321). The Department of Aviation, The Aviation Psychology Laboratory, The Ohio State University, Columbus, OH .

Scerbo, M.W. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 37-63). Mahwah, NJ: Lawrence Erlbaum Associates.

Scerbo, M.W., Ceplenski, P.J., Krahl, K., & Eischeid, T. (1996, Sept.). *Performance on a Team Tracking Task with Human and Computer Teammates*. Poster presented at the 40th Annual Meeting of the Human Factors and Ergonomics Society, Philadelphia, PA.

Scerbo, M.W., Freeman, F.G., Mikulka, P.J., Parasuraman, R., Schoenfeld, V. Eischeid, T., Krahl, K., & Stark, J. (1998). Hazardous states of awareness: functional characteristics and measurement. Final Report, NASA Master Contract NAS1-19858, Task Assignment No. 113.

Schacter, D.L. (1977). EEG theta and psychological phenomena: a review and analysis. *Biological Psychology, 5,* 47-82.

Scheffers, M.K., & Coles, M.G.H. (2000). Performance monitoring in a confusing world: error-related brain activity, judgements of response accuracy, and types of errors. *Journal of Experimental Psychology: Human Perception and Performance*, *26,* 141-151.

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

Scott, W. B. (1999). Automatic GCAS: 'You can't fly any lower'. *Aviation Week & Space Technology, February*, 76-79.

See, J. E., Howe, S. R., Warm, J. S., and Dember, W. N. (1995). Meta-analysis of the sensitivity decrement in vigilance. *Psychological Bulletin, 117*, 230-249.

Shappell, S. A., & Wiegmann, D. (1997). Why would an experienced aviator fly a perfectly good aircraft into the ground? In R.S. Jensen & L. Rakovan (Eds.), *Proceedings of the Ninth International Symposium on Aviation Psychology* (pp. 26-32). The Department of Aviation, The Aviation Psychology Laboratory, The Ohio State University, Columbus, OH.

Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators*. MIT Man-Machine Systems laboratory, Cambridge, MA.

Shibata, T., Shimoyama, I., Ito, T, Abla, D. Iwasa, H., Koseki, K., Yamanouchi, N., Sato, T., & Nakajima, Y. (1999). Attention changes the peak latency of the visual gamma-band oscillation of the EEG. *Neuroreport: An International Journal for the Rapid Communication of Research in Neuroscience, 10*, 1167-1170.

Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology 33,* 5-42.

Soroko, S. I, & Musuraliev, T. Zh. (1995). Possibilities of directed reorganization of EEG parameters in man by adaptational biocontrol. *Human Physiology, 21*, 427-434.

Spydell, J.D. & Sheer, D.E. (1982). Effect of problem solving on right and left hemisphere 40 hertz EEG activity. *Psychophysiology, 19*, 420-428.

Stanton, N. A., & Young, M. S. (2000). The role of mental models in using adaptive cruise control. *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting*, 3.298-3.301.

Sterman, M.B. (1982). EEG biofeedback in the treatment of epilepsy: An overview circa 1980. In L. White & B. Tursky (Eds.), *Clinical biofeedback: Efficacy and mechanisms* (pp. 311-330). *New York: Guilford.*

Sterman, M.B. (1986). Epilepsy and its treatment with EEG feedback therapy. *Annals of Behavioral Medicine, 8,* 21-25.

Sterman, M. B. (1996). Physiological origins and functional correlates of EEG rhythmic activities: Implications for self-regulation. *Biofeedback & Self Regulation, 21*, 3-33.

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

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

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

Swingle, P.G. (1996). Subthreshold 10-Hz sound suppresses EEG theta: clinical application for the potentiation of neurotherapeutic treatemnt of ADD/ADHD. *Journal of Neurotherapy, 2*.

Tan, K., Kaber, D. B., Riley, J. M., & Endsley, M. R. (2000). Human factors issues in implementation of AA to complex systems. *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting*, 1-97 – 1-100.

Tattersall, A. J., & Hockey, G. R. J. (1995). Level of operator control and changes in heart rate variability during simulated flight maintenance. *Human Factors, 37,* 682-698.

Thompson, L. & Thompson, M. (1998). Neurofeedback combined with training in metacognitive strategies: effectiveness in students with ADD. *Applied Psychophysiology and Biofeedback, 23*, 243-263.

Ungerleider, L., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. Masfield (Eds.), *Analysis of visual behavior* (pp. 549-586). Cambridge, MA: MIT Press.

Valentino, D.A., Arruda, J.E., & Gold, S. M. (1993). Comparison of QEEG and response accuracy in good vs poorer performers during a vigilance task. *International Journal of Psychophysiology, 15,* 123-133.

Veltman, J. A., & Gaillard, A. W. K. (1996). Physiological indices of workload in a simulated flight task. *Biological Psychology, 42*, 323-342.

Veltman, J. A., & Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics, 41,* 656-669.

Verwey, W. B., & Veltman, H. A. (1996). Detecting short periods of elevated workload: a comparison of nine workload assessment techniques. *Journal of Experimental Psychology: Applied, 2,* 270-285.

Vidal, F., Hasbrouq, T., Grapperon, J., & Bonnet, M. (2000). Is the "error negativity" specific to errors? *Biological Psychology*, 51, 109-128.

Vidulich, M.A., Stratton, M., Crabtree, M., & Wilson, G. (1994). Performance-based and physiological measures of situational awareness. *Aviation, Space & Environmental Medicine, 65*, A7-A12.

Vogel, W., Broverman, D.M., & Klaiber, E.L.(1968). EEG and mental abilities. *Electroencephalography & Clinical Neurophysiology, 24,* 166-175.

Whinnery, J.E. (1989). Observations on the neurophysiologic theory of acceleration (+Gz) induced loss of consciousness. *Aviation, Space, and Environmental Medicine, 6,* 589-593.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman and D. R. Davies (Eds.), *Varieties of attention* (pp. 63-102). San Diego: Academic Press.

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

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.* NASA Contractor 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.

Wientjes, C. J. E. (1992). Respiration in psychophysiology: Methods and applications. *Biological Psychology, 34*, 179-203.

Wilson, G. F. (1992). Applied use of cardiac and respiration measures: Practical considerations and precautions. *Biological Psychology, 34*, 163-178.

Wilson, G. F. & Eggemeier, F. T. (1991). Physiological measures of workload in multi-task environments. In D. Damos (Ed.), *Multiple task performance: Selected topic.* London: Taylor and Francis.

Wilson, G. F., Fullenkamp, P., & Davis, I. (1994). Evoked potential, cardiac, blink, and respiration measures of pilot workload in air-to-ground missions. *Aviation, Space, and Environmental Medicine, 65,* 100-105.

Wilson, G.F., Swain, C.R. & Ullsperger, P. (1999). EEG power changes during a multiple level memory retention task. *International Journal of Psychophysiology, 32*, 107-118.

Wolpaw, J.R. & McFarland, D.J. (1994). Multichannel EEG-based brain-computer communication. *Electroencephalography & Clinical Neurophysiology, 90,* 444-449.

Wolpaw, J. R., McFarland, D. J., Neat, G. W., & Forneris, C. A. (1991). An EEG-based brain-computer interface for cursor control. *Electroencephalography & Clinical Neurophysiology, 78*, 252-259.

Woods, D.D. (1994). Automation: Apparent simplicity, real complexity. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 1-7). Hillsdale, NJ: Lawrence Erlbaum Associates.

Woods, D.D. (1996). Decomposing automation: Apparent simplicity, real complexity. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 3-17). Mahwah, NJ: Lawrence Erlbaum Associates.

Yamada, F. (1998). Frontal midline theta rhythm and eyeblinking activity during a VDT task and a video game: useful tools for psychophysiology. *Ergonomics, 41*, 678-688.

Young, M. S., & Stanton, N. A. (2000). Journey's end: Will vehicle automation make skilled drivers redundant? *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting*, 1.89-1.92.

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE June 2001	3. REPORT TYPE AND DATES COVERED Technical Publication

4. TITLE AND SUBTITLE

The Efficacy of Psychophysiological Measures for Implementing Adaptive Technology

5. FUNDING NUMBERS

WU 711-50-21-01

6. AUTHOR(S)

Mark W. Scerbo, Frederick G. Freeman, Peter J. Mikulka, Raja Parasuraman, Francesco Di Nocero, and Lawrence J. Prinzel III

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

NASA Langley Research Center
Hampton, VA 23681-2199

8. PERFORMING ORGANIZATION REPORT NUMBER

L-18086

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

NASA/TP-2001-211018

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified-Unlimited
Subject Category 54   Distribution: Standard
Availability: NASA CASI (301) 621-0390

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Adaptive automation refers to technology that can change its mode of operation dynamically. Further, both the technology and the operator can initiate changes in the level or mode of automation. The present paper reviews research on adaptive technology. It is divided into three primary sections. In the first section, issues surrounding the development and implementation of adaptive automation are presented. Because physiological-based measures show much promise for implementing adaptive automation, the second section is devoted to examining candidate indices. In the final section, those techniques that show the greatest promise for adaptive automation as well as issues that still need to be resolved are discussed.

14. SUBJECT TERMS

Adaptive Automation; Psychophysiology

15. NUMBER OF PAGES

71

16. PRICE CODE

A04

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL