Navy Personnel Research and Development Center

San Diego, CA 92152-6800 TN-90-18 April 1990

Real-time Measurement of Mental Workload Using Psychophysiological Measures

Darryl Humphrey
Erik Sirevaag
Arthur F. Kramer
Axel Mecklinger

[redacted]

NPRDC-TN-90-13 April 1990

Real-time Measurement of Mental Workload Using Psychophysiological Measures

Darryl Humphrey
Erik Sirevaag
Arthur F. Kramer
University of Illinois
Champaign, Illinois 61820

Axel Mecklinger
Universitaet Berlin

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

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

OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
April 1990

3. REPORT TYPE AND DATE COVERED
Interim--Jan 89-Dec 89

4. TITLE AND SUBTITLE
Real-Time Measurement of Mental Workload Using Psychophysiological Measures

5. FUNDING NUMBERS
Program Element 0602234N,
Work Unit PR-RS34H21.01

6. AUTHOR(S)
Darryl Humphrey, Erik Sirevaag, Arthur F. Kramer, and Axel Mecklinger

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Psychology, University of Illinois at Urbana-Champaign
603 East Daniel Street,
Champaign, IL 61820

8. PERFORMING ORGANIZATION REPORT NUMBER
NPRDC-TN-90-18

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

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Presented at the Third Annual Workshop on Space Operations, Automation and Robotics; Houston, TX, 1989. Sponsored in cooperation with the Office of Naval Research, Code 1142PS (Dr. Hawkins). Monitored by Navy Personnel Research and Development Center (Dr. L. Trejo).

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

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
The primary goal of the research was to explore the utility of event-related potentials (ERPs) as real-time measures of mei workload. Subjects performed a six-gauge monitoring task and a mental arithmetic task concurrently and at several difficulty lev Difficulty was varied in terms of gauge predictability and by using two or three column arithmetic problems. A bootstrapping appro was adopted in which 1,000 samples of n trials (n = 1, 3, 5, . . . 65 single trials) were classified as to mental workload level using E measures (P300 and slow wave amplitude). Classification accuracies of 85 percent were achieved with 25 trials. Results are discus in terms of potential enhancements for real-time recording of performance monitoring. Keywords: ( )

14. SUBJECT TERMS
Mental workload, event-related potentials, real-time metrics

15. NUMBER OF PAGES
20

16. PRICE CODE

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

FOREWORD

This study explored the utility of event-related potentials (ERPs) as real-time measures of mental workload. Classification accuracies of 86 percent were achieved with 25 trials. Results are discussed in terms of potential enhancements for real-time recording.

This research was sponsored by the Office of Naval Technology (Code 222) under Program Element 0602234N and Work Unit PR-RS34H21.01. Dr. L. J. Trejo served as the contracting officer's representative.

J. C. McLACHLAN
Director, Training Systems Department

SUMMARY

The research presented here derives from an extensive series of investigations that have demonstrated the utility of event-related brain potentials (ERP) in the assessment of residual capacity during the acquisition and performance of a variety of perceptual-motor and cognitive tasks. The primary goal of this study was to explore the utility of ERPs as real-time measures of mental workload. If physiological data, and ERPs in particular, are to serve as real-time measures of operator mental load, the amount of data necessary to reliably discriminate among levels of workload must be determined. To this end, subjects performed two different tasks both separately and together. One task required that subjects monitor a bank of constantly changing gauges and detect critical deviations. The second task was mental arithmetic. Difficulty was varied by requiring subjects to perform operations on two or three columns of numbers. Two conditions that could easily be distinguished on the basis of performance measures were selected for the real-time evaluation of ERPs. A bootstrapping approach was adopted in which 2,000 samples of n trails (n = 1, 3, 5, . . . 65 single trials) were classified using several measures of P300 and slow wave amplitude. Classification accuracies of 85 percent were achieved with 25 trials. The results are discussed in terms of enhancing real-time recording of physiological measures.

CONTENTS

LIST OF FIGURES

INTRODUCTION

The research presented here derives from an extensive series of investigations that have demonstrated the utility of event-related brain potentials (ERPs) in the assessment of residual capacity during the acquisition and performance of a variety of perceptual-motor and cognitive tasks (Donchin, Kramer, & Wickens, 1986; Kramer, 1987). The focus of the present study was to assess the feasibility of employing ERPs as on-line measures of mental workload. If physiological data, and ERPs in particular, are to serve as real-time measures of operator mental load, the amount of data (e.g., seconds, minutes) necessary to reliably discriminate among levels of workload must be determined. This question will be addressed in the present study by adopting a bootstrapping approach to examining the classification accuracy of ERP measures with from 1 to 65 seconds of data. However, before describing the experiment in detail, let us briefly discuss the previous research that suggests that ERPs provide a sensitive and reliable measure of mental workload in an off-line context.

Several recent studies have illustrated the usefulness of the ERP, and more specifically the P300 component, as an index of processing resources (Horst, Munson, Ruchkin, 1984; Israel, Chesney, Wickens, & Donchin, 1980; Kramer, 1987; Kramer & Strayer, 1988; Kramer, Wickens, & Donchin, 1983, 1985; Kramer, Wickens, Vanasse, Heffley, & Donchin, 1981; Natani & Gomer, 1981; Sirevaag, Kramer, Coles, & Donchin, 1989). The general paradigm employed in these studies requires subjects to perform two tasks concurrently. One task is designated as primary and the other task as secondary. Subjects are instructed to maximize their performance on the primary task and devote any additional resources to the performance of the secondary task.

Primary tasks have included system monitoring, decision making, and manual control. Secondary tasks have required subjects to discriminate between tones of different frequencies or lights of different intensities. In general, the response demands of the secondary probe tasks have been minimal, requiring subjects either to covertly count the total number of one type of event or respond to an occasional target probe.

ERPs are elicited by events in either one or both of the tasks. Increases in the perceptual/cognitive difficulty of the primary task result in a decrease in the amplitude of the P300s elicited by the secondary task. Conversely, P300s elicited by discrete events embedded within the primary task increase in amplitude with increases in primary task difficulty. Furthermore, changes in response-related demands of a task have little influence on the P300 (Israel et al., 1980).

The reciprocal relationship between P300s elicited by primary and secondary task stimuli is consistent with the resource trade-offs presumed to underlie dual-task performance decrements (Kahneman, 1973; Navon & Gopher, 1979; Sanders, 1979; Wickens, 1980, 1985). That is, resource models predict that, as the difficulty of one task is increased, additional resources are reallocated to that task in order to maintain performance, thereby depleting the supply of resources that could have been used in the processing of other tasks. Thus, the P300 appears to provide a measure of resource trade-offs that can only be inferred from more traditional performance measures. Furthermore, P300s elicited by secondary task events are selectively sensitive to the perceptual/cognitive demands imposed upon the operator. This selective sensitivity may be especially useful in decomposing the changing processing requirements of complex tasks (Kramer et al., 1983).

One might ask why ERPs should be used to monitor changes in resource demands given that several technically simpler approaches to the assessment of skill acquisition and mental workload have already been implemented. Although numerous performance-based measures of mental workload exist, they suffer from several drawbacks. First, some of the measurement techniques require subjects to perform a secondary task which frequently interferes with the performance of the task of interest (Knowles, 1963; Rolfe, 1971; Wickens, 1979). This is clearly unacceptable in an operational environment in which the safety of the operator must be assured. Even in the laboratory setting, it is difficult to determine which of the two tasks generated an observed performance decrement since the performance on the two tasks is easily confounded. Second, performance-based measures of mental workload provide an output measure of the operator's information processing activities (e.g., reaction time (RT), accuracy). Thus, at best, performance measures provide only an indirect index of cognitive function. Third, performance measures do not always correlate highly with the workload of the tasks (Brown, 1978; Dornic, 1980; Ogden, Levine, & Eisner, 1979). Fourth, although subjective measures are relatively easy to collect and possess high face validity, they do not reflect the moment-to-moment variations in workload that can be indexed by physiological measures.

The present study is part of a continuing effort to explore the utility of psychophysiological measures of mental workload. A primary aim of the project is to determine the feasibility of on-line uses of integrated psychophysiological and performance data. However, given the magnitude of the project, this report will be confined to a description of a preliminary examination of signal/noise ratio parameters of ERPs. More specifically, the functions that relate the amount of ERP data to discrimination accuracy between workload conditions will be derived. In the future, the general analysis approach will be applied to the performance data as well as integrated performance and physiological data.

METHODS

Subjects

Four dextral subjects (2 female) were paid $4.00/hour plus a $1/day bonus for their participation in two, 2-hour sessions and three, 4-hour sessions. All subjects had normal or corrected-to-normal vision.

Tasks

Two different tasks were performed both separately and together. Both tasks will be described in detail.

Monitoring Task

One task consisted of monitoring six gauges. The behavior of a gauge was determined by the interaction of four properties: speed, noise level, noise frequency, and transients. The cursors moved around the gauges at different speeds, a slower gauge taking longer to reach the critical region. Noise level was the amount that the cursor jumped about. The higher the noise level was, the larger were the jumps. These jumps may have been in either the forward or backward direction, randomly determined with the constraint that the overall motion was toward the critical region.

Noise frequency determined how often a gauge jumped. Noise level and noise frequency interacted such that the higher the noise frequency, the more often jumps of a size determined by the noise level occurred. Transients produced infrequent jumps of widely varying magnitude. These jumps were in addition to those produced by the noise level and noise frequency.

The interaction of these properties produced cursor driving functions of varying predictability. Manipulating the driving functions allowed control over gauge monitoring difficulty. The driving functions employed in the high predictability (HP) conditions were such that within a row of three gauges the driving functions were identical in terms of speed, noise level, and noise frequency; no transient occurred for any gauge. The two rows differed in the speed of cursor movement, speed being constant within a row. For the low predictability (LP) conditions, the average value for all properties was equivalent to the HP conditions; however, the individual values were varied with no established correlation between any set of gauges. The LP conditions contained three gauges with a transient. The frequency of the transient was different for each of the three gauges.

The gauges were presented on a CRT in front of the subject. Each gauge was divided into 12 regions (labelled 1 to 12). In addition, each third of the gauge was distinctly colored (green, yellow, or red). The critical level was designated by the position marked by the numeral 9, which was the first region in the red zone.

The purpose of this task was to reset each gauge as quickly as possible once its cursor had entered the critical region. To reset a gauge, the subjects pressed one of six keys after which the cursor returned to the starting position marked by the numeral 1. The cursors were not continuously visible. To sample a given gauge, the subject pressed one of a set of six keys with their left hand. The cursor remained visible for 1,000 milliseconds. Simultaneous sampling was not possible; the cursor for one gauge only was visible at any given moment.

Mental Arithmetic Task

The center of each gauge served as a display area for the operands and operators of the mental arithmetic trials. All of the operands and operators were presented simultaneously and remained in view until an answer was entered or for a maximum of 30 seconds. An answer window appeared to the right of the gauges. Answers were entered via the numeric keypad of the response keyboard and appeared in the window as they were typed. Completion was signaled by pressing the "enter" key of the numeric keypad. The intertrial interval varied from 4 to 15 seconds. Difficulty was manipulated by varying the number of columns on which operations were necessary to complete the problem. The easy version of the task required operations on two columns while the difficult version of the task required operations on three columns of numbers. Henceforth, these versions of the tasks will be referred to as A2 and A3, respectively. Operations included addition and multiplication.

Subjects participated in five sessions. The first two sessions constituted training. The order of the conditions in the training sessions was single task conditions, starting with the easy conditions progressing to the difficult conditions, followed by the dual task conditions. In the final three sessions, the subject performed the eight conditions in a random order determined by a Latin square design. Only the experimental data (i.e., last three sessions) will be presented in this report.

In all sessions, two blocks of each condition were run consecutively, each block taking 5 minutes. A 5- minute break was imposed at the halfway point in addition to any breaks the subject requested.

Performing the gauge monitoring and mental arithmetic tasks in all possible combinations yields eight conditions: 2 task type X 2 levels of difficulty X 2 task pairings (single or dual task condition).

ERP Recording System

Electroencephalographic (EEG) activity was recorded from three midline sites (Fz, Cz, Pz according to the International 10-20 system; Jasper, 1958) referenced to averaged mastoids. All electrodes were Sensormedics[1] Ag/AgCL electrodes. The scalp electrodes were affixed with Grass EC2 electrode cream. The forehead ground, mastoid, and electrooculogram (EOG) electrodes were affixed with the Grass cream and electrode collars. Vertical EOG was recorded from electrodes above and below the right eye. Horizontal EOG was recorded from electrodes lateral to each eye. All electrode impedances were maintained below 10 kohms.

The EEG and EOG were amplified by Grass 12A5 amplifiers with a 10-second time constant and a low-pass filter of 100 Hz, 3 dB/octave roll-off. The recording epoch was 1,300 milliseconds beginning 100 milliseconds prior to an event. The data channels were digitized every 5 milliseconds and were also filtered off-line (-3 dB at 6.89 Hz, 0 dB at 22.22 Hz) prior to further analysis. The psychophysiological data collection was governed by DEC PDP 11/73 computer system (Heffley, Foote, Mui, & Donchin, 1985). Artifact rejection was based upon the vertical-eye-movement absolute deviation and performed off line. ERPs were recorded during the three experimental sessions.

Subjects were seated in a dimly lit, sound attenuated booth. Stimuli were presented on a color monitor located 80 centimeters in front of the subject. Stimulus presentation and behavioral data collection were performed by an IBM AT computer. A GSC Model 901B noise generator coupled with a Realistic SA 150 amplifier presented white noise at 70 dBA over Realistic Minimus-0.3 speakers located within the booth.

Data Analysis Procedures

ERP eliciting events included the sampling of critical and noncritical gauges and presentation of math trials. ERP measurements included P300 latency, P300 base-to-peak amplitude, P300 base-to-peak area, and slow wave area. Behavioral dependent variables included accuracy and response speed in both the monitoring and arithmetic tasks.

In an effort to determine the amount of physiological data needed to discriminate among different experimental conditions, we applied a bootstrapping approach to single subject ERP data. Given the amount of data collected in our study, we decided to begin by examining the physiological differences between two conditions that could be discriminated on the basis of performance measures: the LP single task gauge condition and the gauge samples from the LP/A3 dual task conditions. One thousand samples of size n (n = 1, 3, 5, . . . 65) were randomly selected

[1]Identification of specific equipment is for documentation only and does not imply endorsement.

from single trial data in each of these conditions. By comparing the single trial samples with the grand average waveforms for that condition, the single trial may be classified as a hit (belonging to the criterion condition), a miss (not belonging to the criterion condition), or unclassifiable. Tabulating the classification results in a 2 X 2 contingency table enabled assessment of the efficiency of a number of ERP measures.

RESULTS AND DISCUSSION

The results will be organized in the following manner. First to be described are the effects of single and dual task manipulations on subjects' performance and ERPs. These analyses will enable determination of the relative differences in performance and workload among the single and dual task conditions. Second, we will select two experimental conditions that can be distinguished on the basis of average performance and ERP measures. A bootstrapping approach will then be applied to the single trial ERP data in these conditions. In the bootstrapping approach, 1,000 samples of n trials (n = 1, 3, 5, . . . 65 single trials) will be classified as having come from the two experimental conditions. The classification accuracy value derived from each sample of 1,000 measures will then be plotted as a function of the number of trials in each of the 1,000 samples. Thus, this procedure enables the determination of how changes in the signal/noise ratio of the ERP as a function of averaging (e. g., averaging from 1 to 65 trials for each of the 1,000 samples) translates into gains in the accuracy of discrimination between workload conditions.

The bootstrapping approach will be applied to several different ERP measures including: base-to-peak measures of P300 amplitude (P3bp), measures of P300 area (P3area), cross-correlation measures of P300 amplitude (P3cross), and area measures of a late slow wave component (SWarea). P3bp was defined as the largest positivity in the waveform between 300 and 800 milliseconds post-stimulus relative to a pre-stimulus baseline. The "stimulus" could be either the presentation of the arithmetic task or the presentation of the gauges depending on the condition. P3area was defined as the area from 300 to 800 millisecond-post-stimulus. P3cross measures were calculated by moving a 300-millisecond wide cosine wave across the period from 300 to 800 milliseconds post-stimulus. The slope of the regression function at the point at which the correlation between the cosine "template" and the ERP waveform was maximized was defined as P3cross. SWarea was defined as the area between 750 and 1,100 milliseconds post-stimulus.

Effects of Experimental Manipulation

Figure 1 presents a measure of the accuracy with which subjects reset the gauges in each of the monitoring conditions. A "hit" was scored when subjects reset a gauge within 10 seconds following the point at which it reached a critical value. As can be seen from the figure, accuracy decreased from single to dual task conditions and again with an increase in the difficulty of the dual task. Accuracy also appeared to differ as a function of the predictability of the gauges (HP vs. LP). These differences were confirmed by a repeated measures 2-way ANOVA, with gauge (two gauge conditions, HP and LP) and task (three arithmetic conditions: none, A2, and A3) as factors. Significant main effects were obtained for both the gauge ($F(1, 3) = 13.2, p < .01$) and task ($F(2, 6) = 21.2, p < .01$) factors. A marginally significant interaction between gauge and task factors was also obtained ($F(2, 6) = 2.9, p < .08$) suggesting a decrease in accuracy at the most difficult level of each of the factors.

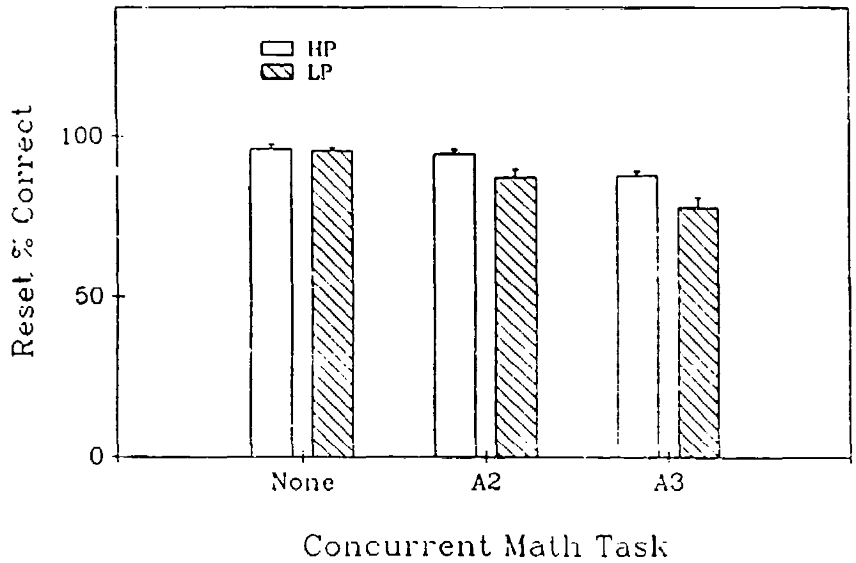

Figure 1. Gauge reset (hit) accuracy in the monitoring task.

Figure 2 presents gauge reset RTs for each of the monitoring conditions. A repeated measures ANOVA performed on this data set revealed a significant main effect for the task factor ($F(2, 6) = 5.4, p < .01$). RT increased from the single to the dual task conditions and again from the A2 to the A3 versions of the arithmetic task. The main effect for the gauge factor did not attain statistical significance.The interaction of the gauge and task factors was not significant.

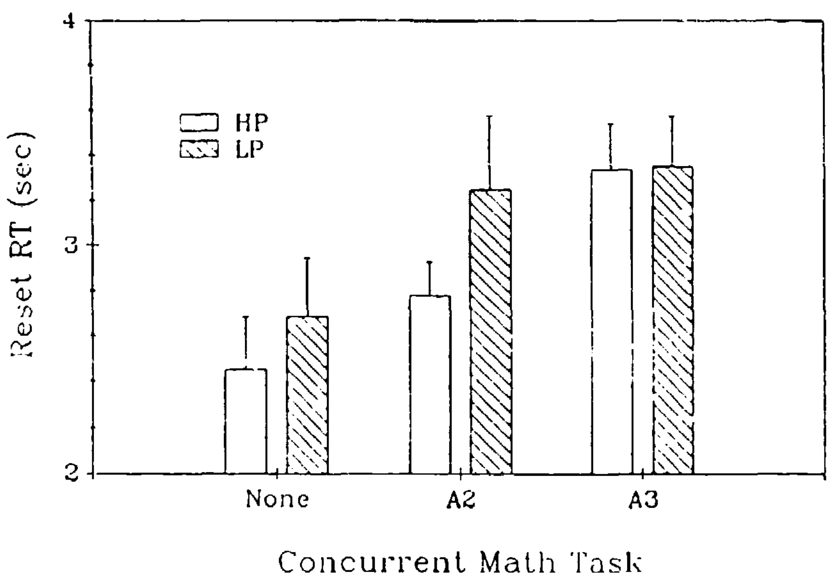

Figure 2. Gauge reset reaction time (RT) in the monitoring task.

Accuracy and RT measure are presented for the arithmetic task in Figures 3 and 4, respectively. Accuracy in the arithmetic task was higher when operations were performed on two columns than when a three column problem was performed ($F(1, 3) = 22.8$, $p < .01$). RT was also faster in the A2 than in the A3 version of the arithmetic task ($F(1, 3) = 26.4$, $p < .01$). Finally, RT in the arithmetic task increased with the transition from the single to dual task conditions and again when the difficulty of the monitoring task was increased.

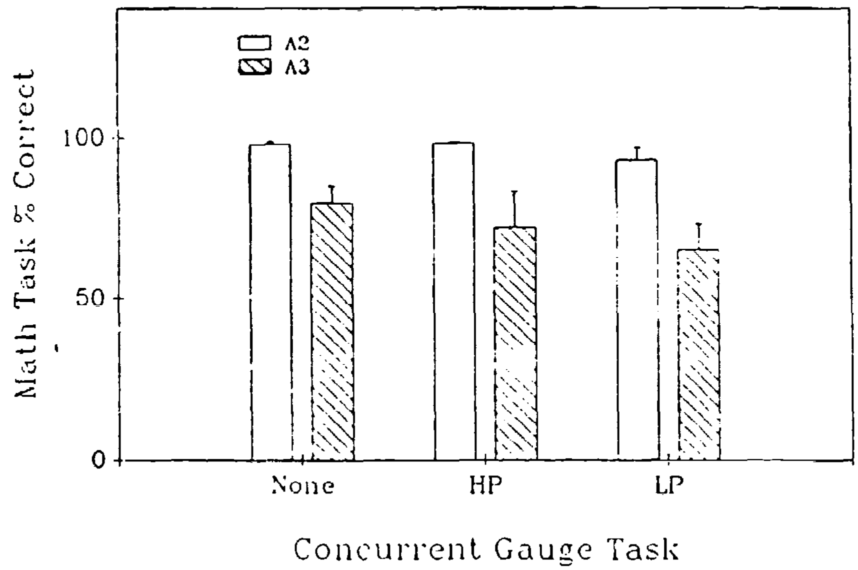

Figure 3. Accuracy in the arithmetic task.

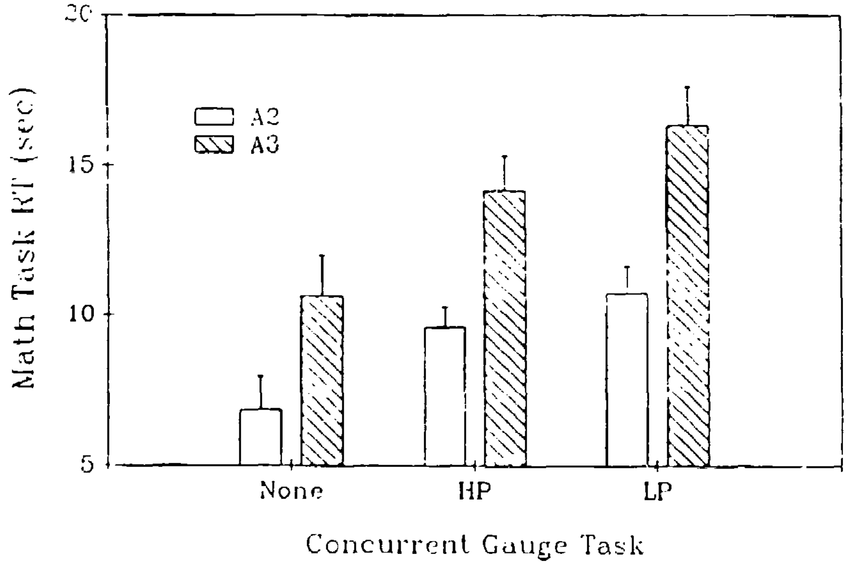

Figure 4. Reaction time (RT) in the arithmetic task.

Of the four ERP measures only two, P3bp and P3area, achieved significance. For the gauge task, a repeated measures 3-way ANOVA was performed with gauge (2 types of gauge, HP and LP), criticality (presentation of a critical cursor or a noncritical cursor), and task (single, dual task with A2 or A3 arithmetic) as factors. Significant main effects were obtained for task (P3bp: $F(2, 3) = 15.7$, $p < .01$; P3area: $F(2, 4) = 14.4$, $p < .01$) and criticality (P3bp: $F(1, 2) = 10.3$, $p < .05$; P3area: $F(1,2) = 18.3$, $p < .05$). There was a marginally significant gauge-by-criticality interaction (P3bp: $F(1, 2) = 9.1$, $p < .09$; P3area: $F(1, 2) = 11.0$ $p < .08$). A 2-way ANOVA for the mental arithmetic data with math type (A2 or A3 arithmetic) and task (single, dual with HP gauges or dual with LP gauges) as factors yielded a significant main effect for the task factor (P3bp: $F(2, 6) = 11.9$, $p < .008$; P3area: $F(2, 6) = 11.13$, $p < .01$).

The analysis of the RT and accuracy data suggests that both the arithmetic and monitoring conditions can be discriminated on the basis of performance measures. Furthermore, since increasing the difficulty of one task influences performance on the other task, one can be confident that both tasks share limited resource(s). Given the demonstrated differences in workload and performance among the experimental conditions, let us now turn to an examination of the feasibility of employing ERPs as real-time measures of mental workload.

Real-time Analysis of Mental Workload

The substantial amount of analysis time required that we select two experimental conditions to analyze further. In order to perform the bootstrapping operation, it was necessary for the experimental conditions to meet two criteria. First, there should be a substantial number of trials available in the selected conditions. This was necessary since repeated samples of 1,000 trials would be selected during the bootstrapping operation. Second, the conditions should be discriminable on the basis of performance measures. Based on these criteria, two easily discriminable conditions were selected from the monitoring task: the single task LP condition and the dual task LP/A3 condition. Later analyses will examine conditions that are less discriminable.

Figure 5 presents the grand average ERPs at Pz across the four subjects for the LP and LP/A3 conditions. It is important to note that the conditions have been further subdivided into waveforms that were elicited during times at which the gauges were in the acceptable range and other times in which the gauges were in the critical region. Since the gauge critical samples were most closely associated with the performance measures, ERPs were employed to discriminate between the LP and LP/A3 conditions during the gauge critical periods. Approximately 70 trials were available in each of these conditions for each of the subjects. The bootstrapping operation was performed separately on the data from two of the original four subjects.

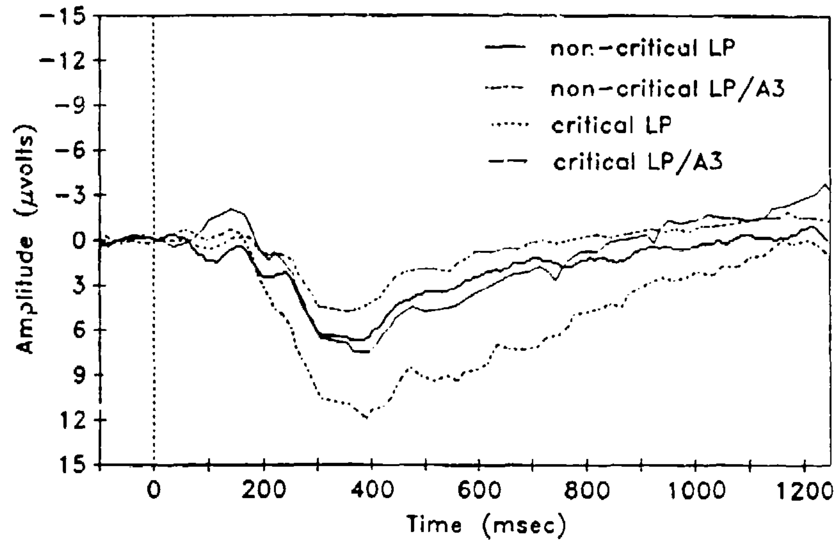

Figure 5. Grand average ERPs recorded at Pz for two gauge events in two conditions.

As described above, the bootstrapping operation involved the repeated selection of single trial ERPs from each of the conditions. Each "sample" comprised 2,000 ERP measures, 1,000 selected from the LP condition and 1,000 selected from the LP/A3 condition. Each of the ERP measures was composed of an average of from 1 to 65 single trial ERP waveforms. Classification accuracy was determined by computing the relative "distance" of each ERP measure from the subject's grand average ERP measures in the LP and LP/A3 conditions. For example, if a subject possessed a grand average P300 amplitude of 50 microvolts in the LP/A3 condition and 10 microvolts in the LP condition, then a single trial measure of 46 microvolts would be classified as LP/A3. This classification procedure was performed for each of the 2,000 ERP samples for each of the different pattern recognition techniques (i.e., P3bp, P3area, P3cross, SWarea).

Figures 6 and 7 present the classification functions for subjects 2 and 3, respectively. The figures depict the accuracy of classification (y-axis) against the number of single trial ERPs that were averaged to produce each of the ERP measures in a sample (each sample included 1,000 ERP measures). Several aspects of the figures are noteworthy. First, for each of the pattern recognition techniques plotted, classification accuracy increased with increases in the number of trials per measure. This continued improvement in classification accuracy represents the increasing signal/noise ratio as additional single trials are averaged to produce each measure. Second, it is clear from the figures that the pattern recognition techniques improved at different rates and achieved different asymptotic levels of accuracy. For both of the subjects, P3bp and P3area improved more quickly and achieved higher levels of performance than SWarea and P3cross. In fact, P3cross is not plotted for subject 2 because it never exceeded 50 percent classification accuracy. Third, for both P3bp and P3area, there was a dramatic improvement in classification accuracy with the addition of the first five single trials followed by a more gradual improvement as additional trials were averaged. Finally, it is interesting to note that classification accuracy improved and reached different asymptotic levels for the two subjects.

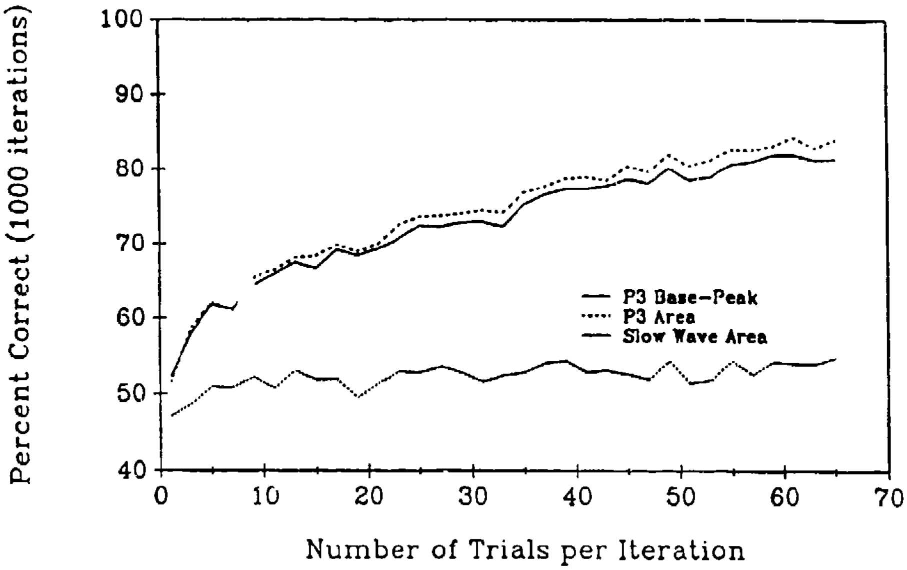

Figure 6. Classification accuracy as a function of the number of trials per measure for subject 2.

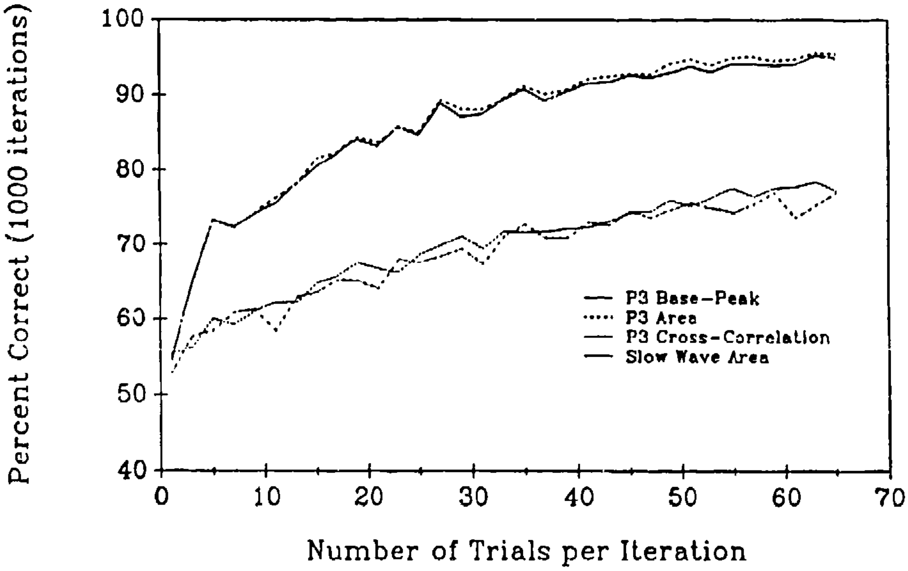

Figure 7. Classification accuracy as a function of the number of trials per measure for subject 3.

CONCLUSIONS AND FUTURE DIRECTIONS

The results of this investigation provide support for the utility of ERPs as real-time measures of mental workload. However, it is important to note that this support is both preliminary and tentative due to the small number of subjects, conditions, and pattern recognition techniques utilized in this study. The result are encouraging, however, and suggest a number of avenues for further exploration.

First, the differential efficiency of the pattern recognition techniques suggests that other techniques may offer improvements over the four thus far examined. The present work used techniques that capitalized on the differences between only one component of the ERP (i.e., either P300 or SWarea). However, a number of other ERP components also appear to be sensitive to variations in mental workload (Horst et al., 1984; Kramer, 1987). Given that these components reflect changes in workload not indexed by P300 and SWarea, the use of multivariate techniques such as discriminant functions should improve the ability to discriminate among different levels of workload. It might also be possible to enhance discriminability by examining changes in the frequency spectra of EEG.

Second, previous examinations of the accuracy of single trial classifications of ERPs have suggested that the efficiency of different pattern recognition techniques is dependent on the characteristics of subject's waveforms (Farwell & Donchin, 1988). For example, base-to-peak measures tend to be most successful when the component of interest is sharply defined while area measures are superior for wider components. Differences in the efficiency of P3cross and SWarea measures for the subjects 2 and 3, also appear to be due to differences in their waveforms. Thus, these analyses suggest that it might be useful to compile a set of heuristics that map waveform characteristics to pattern recognition techniques.

Third, it seems reasonable to suppose that the ability to discriminate among workload levels depends on the homogeneity within workload levels. In the present study, gauge samples in the LP/A3 condition were selected irrespective of whether subjects were performing the arithmetic task (arithmetic tasks were presented with inter-stimulus intervals of from 5 to 15 seconds). Thus, the LP/A3 condition was actually a mixture of single and dual task trials. A comparison of the "dual task" trials in the LP/A3 condition with the LP condition should increase classification accuracy.

Fourth, while it is important to determine classification accuracy in the "best-case" situation, it is also imperative that classification functions are derived for smaller differences in workload. Ongoing efforts are aimed at examining the range of sensitivity of ERP measures to graded differences in workload. Finally, it is clear that classification accuracy can be improved by integrating psychophysiological and performance measures into predicative and descriptive equations. Therefore, it is necessary to determine how the relative sensitivity of different physiological and performance measures vary with changes in task structure and subject state.

REFERENCES

Brown, I. (1978). Dual-task methods of assessing workload. *Ergonomics, 21*, 22-224.

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

Dornic, S. (1980). Language dominance, spare capacity and perceived effort in bilinguals. *Ergonomics, 23*, 369-377.

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

Heffley, E., Foote, B., Mui, T., & Donchin, E. (1985). PEARL II: Portable laboratory computer system for psychophysiological assessment using event related brain potentials. *Neurobehavioral Toxicology and Teratology, 7*, 399-407.

Horst, R., Munson, R., & Ruchkin, D. (1984). Event related brain potential indices of workload in a single task paradigm. *Proceedings of the Human Factors Society, 28th Annual Meeting*. San Antonio, TX.

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

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

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

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

Kramer, A. (1978). Event-related brain potentials. In A. Gale & B. Christie (Eds.), *Psychophysiology and the electronic workplace*. Sussex, England: Wiley.

Kramer, A., & Strayer, D. (1988). Assessing the development of automatic processing: An application of dual-task and event-related brain potential methodologies. *Biological Psychology, 26*, 231-267.

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

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

Kramer, A., Wickens, C., Vanasse, L., Heffley, E., & Donchin, E. (1981). Primary and secondary task analysis of step tracking: An event related brain potentials approach. *Proceedings of the Human Factors Society, 25th Annual Meeting*. Rochester, NY.

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

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

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

Rolfe, J. (1971). The secondary task as a measure of mental load. In W. Singleton, J. Fox, & D. Witfield (Eds.), *Measurements of man at work*. London: Taylor and Francis.

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

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

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

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

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

DISTRIBUTION LIST

Distribution:
Office of Naval Technology (Code 20)
Office of Naval Technology (Code 222) (3)
Director, Office of Naval Research (OCNR-10)
Commanding Officer, Naval Aerospace Medical Research Laboratory, Pensacola, FL
Naval Health Research Center (Code 60)
Commander, Air Force Human Resources Laboratory, Brooks Air Force Base, TX
Program Manager, Manpower Research and Advisory Service, Smithsonian Institution
Defense Technical Information Center (DTIC) (2)