Navy Personnel Research and Development Center
San Diego, CA 92152-6800 TN-90-9 January 1990

(2)

Brain Activity During Tactical Decision-making: III. Relationships Between Probe-evoked Potentials, Simulation Performance, and On-job Performance

Leonard J. Trejo
Gregory W. Lewis
Mark H. Blankenship

NPRDC-TN-90-9 January 1990

Brain Activity During Tactical Decision-making: III. Relationships Between Probe-evoked Potentials, Simulation Performance, and On-job Performance

Leonard J. Trejo
Gregory W. Lewis
Mark H. Blankenship

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

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

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
January 1990

3. REPORT TYPE AND DATE COVERED
Interim--Sep 1988-Oct 1989

4. TITLE AND SUBTITLE
Brain Activity During Tactical Decision-making: III. Relationships Between Probe-evoked Potentials, Simulation Performance, and On-job Performance

5. FUNDING NUMBERS
PE 0602763N, 521-804-042.03.2
PE 0602131M, 44-521-080-203

6. AUTHOR(S)
Leonard J. Trejo, Gregory W. Lewis, Mark H. Blankenship

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Navy Personnel Research and Development Center
San Diego, California 92152-6800

8. PERFORMING ORGANIZATION REPORT NUMBER
NPRDC-TN-90-9

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Technology (Code 222), Washington, DC 20350

Headquarters, Marine Corps (MA), Quantico, VA

10. SPONSORING/MONITORING

11. SUPPLEMENTARY NOTES

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

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This report, the third in a series, addresses the use of event-related potentials (ERPs) to predict the decision-making performance of combat system operators. We describe the relationships between individual measures of probe-ERP amplitude, and both task and on-job performance in 30 military subjects.

14. SUBJECT TERMS
Brain activity, combat systems, decision making, evoked potentials, performance assessment, workload

15. NUMBER OF PAGES
23

16. PRICE CODE

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

FOREWORD

This report is the third in a series of reports examining the feasibility of using neuroelectric signals to predict decision making of combat system operators under varying workloads. The first report (HFOSL TN 71-86-6) identified assumptions underlying this approach to the study of decision making. The second report (NPRDC TN 88-12) provided detailed analyses of the physiological changes in brain activity that occur in response to an irrelevant visual probe as cognitive workload increased in a combat system simulation.

This report describes relationships between physiological brain activity, and both combat system simulation performance and on-job performance.

Research described in this report was performed under program element 0602763N, work unit 521-804-042.03.2 (Future Technologies-Biopsychometrics), sponsored by the Office of Naval Technology, and program element 0602131M, work unit 44-521-080-203 (Biopsychometric Assessment), sponsored by Headquarters, Marine Corps (MA).

J. C. McLACHLAN
Director, Training Systems Department

SUMMARY

Problem

The demands of modern combat systems have the potential for exceeding the capacity of the human to accurately process information, especially during times of great stress. The capacity of the human to perceive, integrate, remember, and use information may be challenged when the individual is monitoring radar and sonar displays, operating electronic warfare systems, or flying aircraft. Exceeding the capacity of the human operator in such situations may impair decision making and could result in costly tactical errors.

Although much is being done to improve the hardware reliability of combat systems, not enough is being done to improve the performance of system operators. The most unpredictable element in combat systems is often the human operator. Traditional personnel testing and training technologies have not eliminated this unpredictability. In part, this is because traditional methods tend to measure or enhance what a person knows rather than how a person processes information.

The current research is driven by the Navy's and Marine Corps's need for better methods of assessing the performance of combat system operators, particularly for predicting the ability of operators to continue to make accurate decisions under heavy workloads.

Objective

This report, the third in a series, addresses the use of event-related potentials (ERPs) to predict the decision-making performance of combat system operators. We describe the relationships between individual measures of probe-ERP amplitude, and both task and on-job performance in 30 military subjects.

Approach

Our approach was to demonstrate relationships between *first-order* and *second-order* measures of ERPs as correlates of both task- and on-job performance. First-order measures emphasize the central tendency within a group, or within an individual over a period. Such measures include the amplitude and latency of components in the average ERP, average energy, or power within intervals of the average ERP. Second-order measures emphasize changes in first-order measures across time or conditions. They can include either differences between first-order measures obtained under different conditions, or trial-to-trial variability of ERP amplitude within a condition.

We presented irrelevant visual stimuli (also called probes) to 30 male, U. S. Marine Corps volunteers (security guards) during a passive baseline period and during their participation in an air defense radar simulation (AIRDEF). Each subject performed the simulation at two levels of workload, which were defined in terms of the rate at which targets appeared on the radar display. The probe stimuli were diffuse, low-intensity flashes of light with a duration of 16 milliseconds (ms) presented at irregular intervals. These flashes appeared on and filled the same 13-inch color display used by subjects to monitor the simulation, but had a negligible effect on the visibility of the simulation data.

Under each condition, event-related potentials (ERPs) were recorded from eight electrodes covering the left and right frontal, temporal, parietal, and occipital areas of the scalp. One electrode was placed above the eye (FP2) to monitor eye movements. A vertex electrode (placed at Cz) was the reference for all recordings. Each single ERP was first analog-filtered (3 dB bandwidth 0.1-100 Hz), then sampled at 256 Hz, digitized, and stored by a computer. Signal-average waveforms were computed from six artifact-free ERPs for each condition. Each point in the signal-average waveform was the time-indexed average of the six single ERPs. These waveforms were digitally filtered (0.5-25 Hz) and divided into eight adjacent, non-overlapping time windows, approximately 50 ms wide, that spanned the range between 50 and 450 ms after stimulus onset. The root-mean-square (RMS) value of the waveform was computed in each window of the waveform. For brevity, we will refer to this measure as RMS-a. This RMS-a value was used as the dependent variable in a repeated measures analysis of variance. Within-subjects factors were *workload, coronal* electrode position (anterior-posterior), *sagittal* electrode position (left-right), and *time* (window latency).

Results and Conclusions

The results of this study are consistent with an information processing model in which neural responses to irrelevant probe stimuli (probe ERPs) predict and covary with human performance. This model hypothesizes that two major factors contribute to human performance of complex decision-making tasks. First, performance increases in direct relationship to the *total capacity* of the human to process task-relevant information. Second, performance also increases in relationship to the ability to shift processing resources in response to task demands. We call this second factor *allocation range.*

We found that first-order measures of probe ERP amplitude appear to index total capacity by significantly predicting both subsequent AIRDEF task performance and on-job performance of security guards. The pattern of correlations between the workload-sensitive first-order measures we examined and AIRDEF performance measures clearly support a total capacity hypothesis. A direct relationship also held between first-order probe-ERP measures and on-job performance. The high-performance group of subjects exhibited higher mean RMS-a values than the low-performance group in four of the five workload-sensitive windows. Two of these differences, frontal Windows 2 and 6, corresponding to post-stimulus latencies of 227 and 330 ms, were significant.

We also found that changes in first-order measures of probe-ERP amplitude as a function of workload were inversely related to task performance. There was a pattern of significant negative correlations between these changes or second-order measures and AIRDEF task performance, which supported the allocation range hypothesis. For these correlations, the maximal proportion of variance accounted for using second-order measures was higher than that found for first-order measures. Second-order measures also indicated a pattern of differences between high and low on-job performance groups that is consistent with the predictions of the allocation range hypothesis.

The most consistent relationships between performance and RMS-a (both first and second order) were found with frontal Window 6. This window represents average amplitude in the probe-ERP recorded bipolarly at frontal sites F3 and F4 (the reference was at the vertex, Cz). The latency range of this window is 305 to 355 ms. The primary component of the ERP that occupies this latency range is known as the P300. P300 typically exhibits a maximum amplitude on the midline centro-parietal region and may be recorded at sites Cz and Pz. Since a voltage difference between frontal sites and Cz will reflect activity at Cz as well, it is highly probable that our frontal Window 6 represents P300 amplitude.

Future Directions

Our results demonstrate that ERP waveforms recorded for irrelevant probe stimuli, presented either before or during the performance of a complex decision-making task, provide information about the performance of an individual on that task. They also provide some information about the performance level of that individual on the job.

These relationships between individuals' probe-ERP waveforms and their task and on-job performance are predicted by a neural-cognitive model of human information processing. Although the variability in these relationships is high, we expect that refinements in both the predictors (probe-ERP measurement technologies) and criteria (performance variables; e.g., task and simulation performance) will lead to the development of probe-ERP measures that afford the improved prediction and enhancement of human performance in combat systems operations.

CONTENTS

LIST OF TABLES

INTRODUCTION

Problem

The demands of many military occupations have the potential for exceeding the capacity of the human to process information, especially during times of great stress, such as those faced by combat system operators. The capacity of the human to perceive, integrate, remember, and use information may be challenged when the individual is flying aircraft, monitoring radar and sonar displays, or operating electronic warfare systems. Exceeding the capacity of the human operator in such situations may impair decision-making performance and could result in costly tactical errors. Although much is being done to improve the reliability and effectiveness of combat systems, there is an increased need to monitor and improve the performance of system operators. For these reasons, the most unpredictable elements in combat systems are often the operators themselves. Years of personnel selection testing and classification (e.g., Armed Services Vocational Aptitude Battery) have not eliminated this unpredictability. In part, this is because such tests tend to measure what a person *knows* rather than how a person *processes information*.

This research is driven by the Navy's and Marine Corps's need for improved methods of assessing individual combat system operators, particularly for predicting the ability of operators to continue to make good decisions under varying workloads.

Objective

This is the third in a series of reports concerned with use of event-related potentials (ERPs) to predict the performance of combat system operators. The first report (Trejo, 1986) examined hypotheses, assumptions, and experimental design issues. The second report described the effects of workload on group averages of the root-mean-square (RMS) amplitude of probe-ERPs acquired in 30 military subjects during the performance of an air defense radar simulation (Trejo, Lewis, & Blankenship, 1987). In this report, we describe the relationships between individual measures of probe-ERP amplitude and both task and on-job performance in 30 military subjects.

Approach

Our approach was to demonstrate relationships between *first-order* and *second-order* measures of ERPs in a test condition with both task and on-job performance. First-order measures emphasize the central tendency within a group, or within an individual over a period. Such measures include the amplitude and latency of components in the average ERP, average energy, or power within intervals of the average ERP; are relatively stable within individuals (Lewis, 1983); and are usually measured by averaging ERPs over constant test conditions.

Research relating ERPs to task performance has been primarily concerned with using specific ERP components to describe underlying processes in various models of cognitive processing. Direct correlations between individual task performance and ERP measures have not typically been estimated. However, these studies do provide evidence of relationships between ERPs and task performance by inferring lower performance for difficult tasks than for easy tasks. For example, the amplitude of the auditory P300 ERP component has been shown to correlate with the difficulty (and hence performance) of concurrently performed display-monitoring (Israel, Wickens, Chesney, & Donchin, 1980) and visuo-motor tracking tasks (Kramer, Wickens, & Donchin, 1983).

In military personnel research, the emphasis has been on direct measures of individual performance. For example, latency of the P300, which is believed to reflect processes related to stimulus evaluation (Donchin, Ritter, & McCallum, 1978), was positively correlated with reaction time in a Sternberg memory task (Gomer, Spicuzza, & O'Donnell, 1976). Relationships between first-order measures of ERPs and performance of sonar operators (Lewis & Rimland, 1980) and aviators (Lewis & Rimland, 1979; Lindholm, Cheatham, & Koriath, 1984) have also been demonstrated. Relationships between first-order ERP measures and global measures of performance, such as supervisor ratings, have also been observed. For example, the amplitude of the visually evoked magnetic field (VEF), a magnetic analog of the visual ERP, has been found to correlate with supervisor ratings of on-job performance in Marine security guards (Lewis, Trejo, Nunez, Weinberg, & Naitoh, 1988).

Another approach to predicting human performance involves second-order measures of the ERP. By second-order measures, we mean measures that emphasize changes in first-order measures across time or conditions. These measures can include either differences between first-order measures obtained under different conditions or trial-to-trial variability of ERP amplitude within a condition. For example, Lewis et al. (1988) found that, for Marine security guards, the variability in the amplitude of the VEF within a single test session was less for high performers than for low performers. As with first-order measures, second-order measures may be obtained from group or individual data.

We now examine the relationships of first- and second-order ERP measures with performance of an air defense simulation task in which irrelevant visual probes were presented during task performance at varying workload levels. We will assess the construct of an individual's total cognitive resources and resource allocations from measurements of probe-ERP amplitudes.

METHODS

Complete descriptions of the air defense task and our measurement methods appear in Trejo et al. (1987). A brief summary follows.

Procedure

We presented irrelevant visual stimuli (also called probes) to 30 male, U.S. Marine Corps, security-guard volunteers during a passive baseline period and during their participation in an air defense radar simulation (AIRDEF). Each subject performed the simulation at two levels of workload, which were defined in terms of the rate at which targets appeared on the radar display. The probe stimuli were diffuse, low-intensity flashes of light with a duration of 16 milliseconds (ms) presented at irregular intervals. These flashes appeared on and filled the same 13-inch color display used by subjects to monitor the simulation, but had a negligible effect on the visibility of the simulation data.

Under each condition, event-related potentials (ERPs) were recorded from eight electrodes covering the frontal, temporal, parietal, and occipital areas of the scalp. One electrode (FP2) was placed above the eye to monitor eye movements. A vertex electrode (placed at Cz) was the reference for all recordings. Each single ERP was first analog-filtered (3 dB bandwidth 0.1-100 Hz), then sampled at 256 Hz, digitized, and stored by a computer. Signal-average waveforms were computed from six artifact-free ERPs for each condition. Each point in the signal-average waveform was the time-indexed average of the six single ERPs. These waveforms were digitally filtered (0.5-25 Hz) and divided into eight adjacent, non-overlapping time windows, approximately 50 ms wide, that spanned the range between 50 and 450 ms after stimulus onset. The root-mean-square (RMS) value of the waveform was computed in each window of the waveform. For brevity, we will refer to this measure as RMS-a. This RMS-a value was used as the dependent variable in a repeated measures analysis of variance. Within-subjects factors were *workload, coronal* electrode position (anterior-posterior), *sagittal* electrode position (left-right), and *time* (window latency).

The analyses of variance revealed five significant workload-sensitive window-site combinations in which RMS-a was 30 to 40 percent lower during active participation in the AIRDEF task than during a passive baseline period (Trejo et al., 1987). These five measures are listed in Table 1 with the average percentage difference between baseline and active conditions, and the statistics describing the significance of these differences. Trejo et al. (1987) present complete details concerning the derivation and analysis of these measures.

Table 1. Workload Sensitive Probe-ERP Measures

Window number	*Center latency*[1]	*Site*	*Average RMS-a* Baseline	*Average RMS-a* AIRDEF	*Percent change*	*SS*	$F^2_{1,5664}$
4	229	Parietal	4.53	2.46	-45.8	172.49	53.83
4	229	Occipital	5.69	4.02	-29.4	162.42	51.32
2	127	Frontal	4.29	2.58	-39.9	117.11	36.54
5	279	Frontal	3.61	2.41	-33.2	57.19	17.85
6	330	Frontal	3.88	2.32	-40.1	96.70	30.18

[1]Post-stimulus latency of window center in milliseconds.
[2]$p < .001$ for all these effects.

AIRDEF Performance Assessment

Performance on the AIRDEF simulation was assessed using behavioral elements of the task, which included *kills, kill range, hits, splashes,* and *in-flight launches.* Because these performance elements are critical to the analysis of overall AIRDEF performance, we will describe each element in detail. A kill is a successful interception of a hostile target by a weapon that is launched under the direct control of the subject. All kills do not have equal value; instead, the value of a kill increases with the distance from the subject, who is imagined to be on board a ship. The value of a kill is scaled by the *kill range*, which is the distance from the ship to target intercept. Kills at the maximum weapons range (20 miles) have the maximum value. A measure of the average value of a subject's kills is provided by the average kill range, which is the total of the ranges for all targets killed in one AIRDEF engagement divided by the total number of targets killed. It is important to point out that long-range kills imply good decision-making performance by the subject because long-range kills require accurate judgments of relative target and weapon speeds.

An incoming target, after being detected and displayed to the subject on the radar screen, can travel at one of three speeds. Speed is indicated by the spacing between position markers (blips) on the radar screen. Slow targets travel small distances between each radar update, leaving small gaps between each blip. Fast targets travel large distances between updates and leave large gaps between blips. The subject's weapons always travel at the same speed as the fast targets. Medium targets travel at speeds between fast and slow targets.

The subject must judge the speed of an incoming target relative to the speed of his weapon, and, based on that judgment, choose the best time to launch a weapon to kill the target at maximum range. All of this is done in the context of multiple incoming targets.

If a subject fails to fire a weapon at an incoming target in time, the target will hit the ship. The target arrival rates used in this study were chosen to ensure that only the highest performers would complete two engagements without sustaining any hits.

The consequences for firing a weapon too early are less severe than those for firing too late. If a subject fires too early, his weapon will reach the maximum weapons range before the target. Then, his weapon will "splash" ineffectively into the ocean and the target will continue to approach. Although splashes do not have a direct influence on the assessment of a subject's overall AIRDEF performance, they have an indirect negative influence on it. This occurs because the rules of the AIRDEF simulation state that only one weapon at a time can be in-flight at a target. If a weapon is fired early, the subject must wait until that weapon splashes before he can fire another one. When the subject eventually fires a weapon at this target, the target will have traveled closer to the ship and will probably be killed at a low range. If the subject attempts to fire at the target before the in-flight weapon splashes, he will incur an in-flight launch penalty, which has a direct negative influence on his overall performance rating. Thus, splashes reduce the overall performance by indirectly lowering the average kill range and by raising the probability of an in-flight launch penalty.

Finally, in-flight launch penalties may occur not only during flights that turn into splashes but also during flights that result in kills. A description of the basic performance data acquired in AIRDEF is provided by Kelly, Greitzer, and Hershman (1981).

Overall AIRDEF performance is assessed by the *normalized skill rating* (N-skill) (Trejo, 1986), which is a composite measure of task performance that is normalized for task difficulty, as measured by the number of targets (18 or 36) appearing on the screen in one engagement:

$$N\text{-}skill = 5\left(average\ range\right) - 12\left(\frac{hits}{targets}\right) - 2\left(\frac{inflight.\ launch}{targets}\right) \tag{1}$$

This equation provides large rewards for long-range kills, large penalties for hits, and a smaller penalty for in-flight launches. The normalization of hits and in-flight launches for the number of targets produces a measure of skill that enables skill comparisons to be made among different difficulty levels of AIRDEF on a per-target basis. The average range factor in the equation is not normalized because, in forming the average, the number of targets is already accounted for.

RESULTS

First-order ERP Measures and Task Performance

Our first hypothesis states that, other factors being equal, first-order measures of the ERP which directly reflect total resources should be higher for subjects who perform well on the AIRDEF task than for subjects who perform poorly. To test this hypothesis, we examined correlations between task-performance measures and first-order probe-ERP measures.

The first-order measures that we computed in the first study (Trejo et al., 1987) included RMS-a amplitudes for 8 electrode sites and 8 time windows, a total of 64 measures.[1] However, RMS-a changed significantly with workload only at the five site-window combinations listed in Taole 1. Since RMS-a for these five site-window combination measures decreased significantly with workload, they are the most likely to reflect individual cognitive resources. For this reason, we restricted the correlation analysis to these site-window combinations. Furthermore, since we expected that the best estimate of total capacity would be provided under low cognitive load, we used only the values of the RMS-a for the probe-ERPs recorded in the baseline testing condition, instead of those for Levels 1 and 2, where decision making occurred.

Table 2 lists the correlation coefficients between the selected baseline first-order probe-ERP measures and active AIRDEF performance measures (including the N-skill composite score) for RMS-a. Since we expected the sign of the correlation coefficient to be positive for good performance and negative for errors, we used one-tailed t-tests. With 28 degrees of freedom, the critical values of the correlation coefficient for $p < .05$ and $p < .01$ are $\pm$ .306 and $\pm$ .423, respectively (Edwards, 1976).

The pattern of correlations between first-order RMS measures and AIRDEF performance variables confirmed our predictions. In general, baseline RMS-a was positively correlated with good performance and negatively correlated with errors. Of the 50 correlation coefficients computed, 14 were significant and had the predicted sign. Although the magnitudes of the coefficients varied somewhat, the pattern was consistent across AIRDEF performance levels.

Although significant correlations (Table 2) between AIRDEF performance and first-order measures were observed for all five site-window combinations, only those for frontal Window 6 exceeded the $p < .01$ significance criterion.

Although the present data are not precise enough to determine the exact functional relationship between first order measures and performance, we performed a regression analysis to determine whether the relationship was linear or non-linear. We chose the most significant first-order measure, frontal Window 6, as the predictor and used subjects' average N-skill across AIRDEF Levels 1 and 2 as the dependent variable. For the RMS-a measure, shown in Figure 1, significant relationships were found with both linear ($r^2 = .21$, $F_{1,28} = 7.47$, $p < .011$) and logarithmic regressions ($r^2 = .30$, $F_{1,28} = 12.05$, $p < .0017$).

[1] In the first study, a signal-to-noise measure (RMS-s), was also computed. However, conclusions drawn from RMS-a and RMS-s measures in that study and in the present study are identical. Therefore, only RMS-a will be considered here.

Table 2. Correlations of Baseline First-order RMS-a Measures and Active AIRDEF Performance

A. Level 1 AIRDEF Performance

Site-window[1]	*Kills*	*Hits*	*Inflight launches*	*Average range*	*N-skill rating*
Frontal-W2	0.23	-0.16	-0.13	0.24	0.32*
Frontal-W5	0.27	-0.31*	-0.39*	-0.15	0.32*
Frontal-W6	0.30	-0.29	-0.25	0.19	0.45**
Parietal-W4	0.35*	-0.31*	-0.20	0.08	0.40*
Occipital-W4	0.28	-0.23	-0.22	0.10	0.34*

B. Level 2 AIRDEF Performance

Site-window[1]	*Kills*	*Hits*	*Inflight launches*	*Average range*	*N-skill rating*
Frontal-W2	0.25	-0.35*	0.01	0.12	0.31*
Frontal-W5	0.17	-0.22	-0.12	0.03	0.18
Frontal-W6	0.42**	-0.42**	-0.08	0.13	0.38*
Parietal-W4	0.21	-0.24	-0.07	-0.08	0.14
Occipital-W4	0.10	-0.16	-0.07	-0.06	0.09

[1]Average of RMS-a values for homologous sites in both hemispheres.
*One-tailed test, $p < .05$; $^{**}p < .01$.

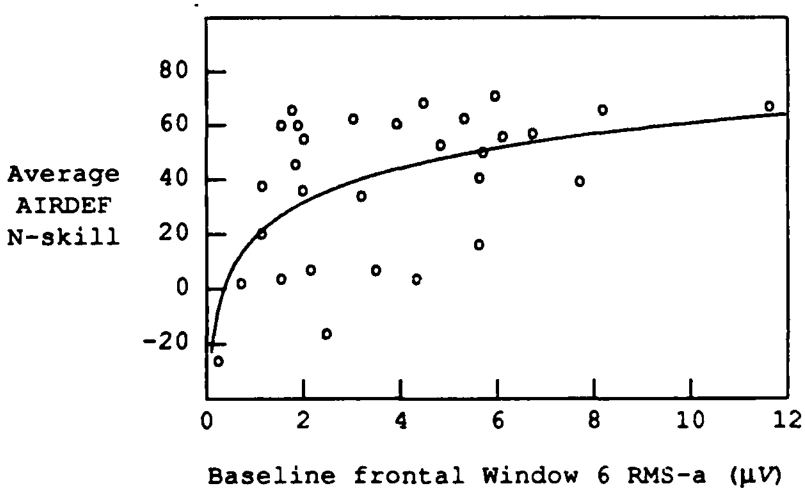

Figure 1. Shown are the 30 subjects' paired values of RMS-a in baseline frontal Window 6 and average N-skill across AIRDEF Levels 1 and 2 (18 and 36 target conditions). A non-linear regression was significant, indicating the presence of a predictive relationship between the first-order measure and average AIRDEF performance ($y = 42.1 \log x + 18.94$, $F_{1,28} = 12.05$, $p < .0017$, $r^2 = .30$).

First-order Measures and On-job Performance

For on-job performance, we chose a supervisor's rating of the subjects (Lewis et al., 1988). Each subject's supervisor was asked to rate the subject on four criteria: job knowledge, job reliability, job performance, and motivation. Each criterion could be rated as superior, highly satisfactory, or satisfactory. Subjects who received a superior rating in each criterion were classified as "high" performers. Subjects who received less than superior for any criterion were classified as "low" performers. Based on this rating scheme, there were 13 high and 11 low performers (6 subjects could not be included because they received no supervisory ratings).

According to the hypothesis that total cognitive capacity is indexed by baseline first-order RMS-a amplitude value, if cognitive capacity contributes greatly to security guard performance, we expected larger mean values of the baseline RMS-a measure for high performers than for low performers. We tested this hypothesis using t-tests of the differences between the means of the baseline RMS-a measures of the high and low on-job performance groups. Prior to computing the t-statistic, the variances of the groups were compared. If the variances differed significantly ($p <$ .05, Folded F-test), the t-statistic and degrees of freedom were evaluated using separate variance estimates for each group (SAS Institute Inc., 1982). The results are shown in Table 3.

Table 3. t-tests of First-order Baseline RMS-a Means for On-job Performance Groups

	Group RMS-a Means			
Site-window	*High*	*Low*	*df*	*t*
Frontal-W2	4.37	2.66	22	2.22*
Frontal-W5	3.58	3.31	15.8[1]	0.33
Frontal-W6	4.26	2.56	22	2.07*
Parietal-W4	4.39	3.18	17.7[1]	1.37
Occipital-W4	5.85	4.06	22	1.27

[1]Adjusted *df* due to unequal variances; folded F test, $p < .05$.
*One sided t-test, $p < .05$.

As predicted by the resource model, the high on-job performance group exhibited larger mean RMS-a values than the low group for the five workload-sensitive site-window combinations examined. For two of these combinations, frontal Windows 2 and 6, the group means were significantly different for the RMS-a measure. These are the same combinations that were most consistently correlated with AIRDEF performance measures (Table 2).

Second-order Measures and Task Performance

Our hypothesis of equating total processing resources with high ERP amplitudes in low-load conditions leads to a prediction about changes in ERP amplitudes between low-load and high-load conditions. Specifically, the range for which ERP amplitude can decrease as a function of workload is limited by the maximum amplitude observed in low-load conditions. Furthermore, our hypothesis of equating good performance in high-load conditions with allocation of cognitive resources to the task predicts that a change in ERP amplitude for the irrelevant probe should accompany an increase in workload in order to maintain good performance. Thus, we predicted that individuals with high total cognitive processing resources whose performance does not change as a function of workload would show large, workload-related decreases in ERP amplitude as compared to individuals whose performance does change. To test this prediction, we examined correlations between the second-order RMS measures (change in workload-sensitive first-order measures) and task performance variables.

We defined normalized difference scores to serve as measures of change in ERP amplitude as a function of workload. To simplify the analyses and reduce the number of correlations to be computed, we first averaged ERP amplitudes for Levels 1 and 2 (18 and 36 target conditions, respectively) to form a single "active" ERP amplitude measure for each site-window computation. We then computed the normalized difference by dividing the difference between active RMS-a and baseline RMS-a (average in Levels 1 and 2) by their sum:

$$\textit{Normalized difference} = \frac{\textit{active ERP RMS - baseline ERP RMS}}{\textit{active ERP RMS + baseline ERP RMS}} \qquad (2)$$

The normalization by the sum in the denominator was chosen to emphasize changes in RMS-a amplitude and de-emphasize the absolute amplitudes exhibited by an individual. This measure is bounded between -1.0 and 1.0, representing 100 percent decrease and 100 percent increase in the RMS-a relative to the sum. The correlations of the normalized difference measures for the five site-window combinations of Table 1 with AIRDEF performance variables for Levels 1 and 2 are shown in Table 4.

As with the first-order measures, the pattern of correlations we obtained is highly consistent with our hypotheses. In this case, the predictions are for negative correlations of the amount of change in the RMS-a measures with good AIRDEF performance and positive correlations with error rates. This arises from the way in which the normalized difference scores are computed: A negative value indicates a *decrease* in RMS-a from baseline to active conditions.

Thirteen correlations were significant and in the predicted direction. All but one of these were correlations between hits, kills, and N-skill (which is heavily weighted for hits and kills) and normalized RMS-a differences at frontal Window 6 (corresponding to 330 ms), parietal Window 4 (corresponding to 229 ms) and occipital Window 4. Again, the most consistently significant site-window combination was frontal Window 6.

As with the first-order measures, we performed a regression analysis to determine the functional relationship between second-order measures and task performance. Again, we chose average N-skill across Levels 1 and 2 as the dependent variable and frontal Window 6 combination as the predictor. The data appear in Figure 2. The linear regression was significant ($r^2 = .27$, $F_{1,28} = 10.47$, $p < .0031$) and is also shown in Figure 2.

Table 4. Correlations of Second-order RMS-a Measures and Active AIRDEF Performance

A. Level 1 (18 target condition) AIRDEF Performance

Site-window[1]	*Kills*	*Hits*	*Inflight launches*	*Average range*	*N-skill rating*
Frontal-W2	-0.12	0.12	-0.09	-0.02	-0.10
Frontal-W5	-0.33	0.38*	0.21	0.24	-0.30
Frontal-W6	-0.47**	0.52**	0.22	0.01	-0.56**
Parietal-W4	-0.43**	0.43**	-0.02	0.06	-0.40*
Occipital-W4	-0.47**	0.47**	0.04	0.23	-0.36*

B. Level 2 (36 target condition) AIRDEF Performance

Site-window[1]	*Kills*	*Hits*	*Inflight launches*	*Average range*	*N-skill rating*
Frontal-W2	-0.16	0.29	-0.02	0.03	-0.20
Frontal-W5	-0.07	0.20	0.08	-0.12	-0.21
Frontal-W6	-0.44**	0.47**	0.02	-0.06	-0.38*
Parietal-W4	-0.27	0.32	-0.07	0.16	-0.15
Occipital-W4	0.01	0.13	-0.03	0.19	0.00

[1]Average of RMS-a values for homologous sites in both hemispheres.
*One-tailed $p < .05$; ** $p < .01$; ***$p < .001$.

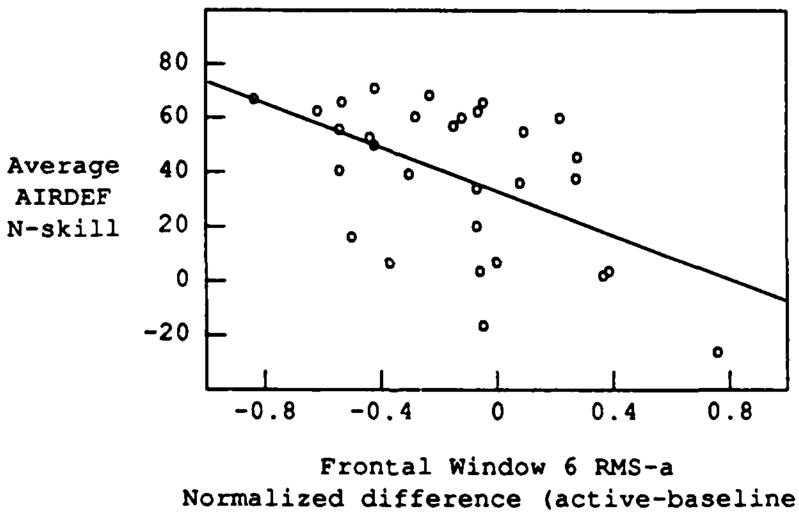

Figure 2. Shown are the 30 subjects' paired values of RMS-a in normalized difference score for frontal Window 6 and average N-skill across AIRDEF Levels 1 and 2 (18 and 36 target conditions). A linear regression was significant, indicating the presence of a direct relationship between the second-order measure and average AIRDEF performance ($y = 40.5$ x+ 33.19, $F_{1,28} = 10.47$, $p < .0031$, $r^2 = .27$).

Second-order Measures and On-job Performance

As for the first-order measures we examined, we compared the high and low performance group means for second order measures. Since high performance is expected to correlate with decreases in workload-sensitive ERP amplitude measures, such as RMS-a, we expected that mean normalized difference scores for the high performance group would be negative with respect to those of the low performance group. Again we examined only the site-window combinations listed in Table 1.

The procedures used for comparing the means (t-tests, variance comparisons) were the same as for the first-order measures as shown Table 3. No significant differences were observed between the means of the high and low job-performance groups on the second-order measures. However, as expected, mean normalized differences were more negative (indicating greater decreases) for the high performance group than for the low performance group for four of the five means compared (frontal-W2, frontal-W5, frontal-W6, and parietal-W4).

DISCUSSION

The results of this study are consistent with an information processing model in which neural responses to irrelevant probe stimuli (probe ERPs) predict and covary with human performance. Specifically, this model postulates that *total capacity* and *allocation range*, will covary with or predict performance in specific directions. Total capacity, measured by probe-ERP amplitudes under low load, is directly related to performance and inversely related to error rates. Allocation range, measured as a decrease in probe-ERP amplitude under load, is inversely related to performance and directly related to error rates. Our analyses were restricted to the root-mean-square value within segments of the average probe-ERP waveform (RMS-a) which had previously been shown to decrease significantly as a function of AIRDEF workload. RMS-a is primarily a measure of signal amplitude. In our studies, the probe stimulus producing this ERP was a visual flash, which during processing shares neural pathways critical for the performance of the highly visual AIRDEF task.

We found that first-order measures of probe ERP amplitude appear to index total capacity by predicting both subsequent AIRDEF task performance and on-job performance of security guards. In a strict sense, performance prediction requires making an inference about future performance based on present data. The correlations and regressions we computed between first-order measures of RMS-a (Tables 2 and 3, Figure 1) satisfied this prediction requirement. The pattern of linear correlations between the workload-sensitive first-order measures we examined and AIRDEF performance measures clearly supported the total capacity hypothesis. Further support for this hypothesis was provided by a non-linear regression model which accounted for 30 percent of future average AIRDEF performance (N-skill, equation (1)) using the logarithm of baseline probe-ERP RMS-a in frontal Window 6.

Although we have no theoretical position concerning the logarithmic model (Figure 1), this model does approximate the asymptotic nature of the N-skill performance measure, which is restricted by a maximum value of 100. Other functions could express this nature more exactly. For example, a saturating exponential function of the form

$$y = P_{max} - be^{\frac{-x}{a}}$$

would never exceed P_{max} (which is 100 for N-skill) and would provide a useful constant, a, which indicates the RMS value at which performance is expected to reach a maximum. We found, however, that the variance in the present data was too high to permit meaningful comparisons between this model and the logrithmic model.

A direct relationship was also found between first-order probe-ERP measures and on-job performance. The high-performing group of security guards exhibited higher mean RMS-a values than the low-performing group in all five of the workload-sensitive windows (Table 3). Two of these differences, frontal Windows 2 and 6, were significant.

Second-order measures, as represented by the normalized difference equation (2), were expected to be negatively correlated with task or on-job performance and positively correlated with error rates. This prediction followed from the *allocation range* hypothesis of our neural information processing model. According to this hypothesis, subjects who exhibit greater allocation range, as shown by greater decreases in first-order measures as a function of load, should exhibit greater selectivity for stimulus processing under high loads, or less distractability, than subjects with less allocation range. In short, less distraction should lead to better performance on a wide variety of tasks.

The linear correlation between second-order measure and task performance (Table 4) support the allocation range hypothesis. For these correlations, the maximal proportion of variance accounted for using second-order measures was higher than that found for first-order measures. For example, the second-order measures for frontal Window 6 accounted for between 14 and 31 percent of the variance in the N-skill performance measure (r_{min} = -.38, r_{max} = -.56), as compared to 14 and 20 percent for the first-order measures (r_{min} = .38, r_{max} = .45). This apparent advantage of second-order measures may partly be due to the inadequacy of linear regression models for the first-order measures. On the other hand, linear models worked satisfactorily for the second-order measures, as shown by the regressions of average AIRDEF N-skill on frontal Window 6 normalized differences in Figure 2. For RMS-a normalized difference scores, a linear model accounted for 27 percent of the variance, comparable to what was accounted for by the non-linear model used for first-order measures.

Second-order measures also indicated a pattern of differences between high and low on-job performance groups that is consistent with the predictions of the allocation range hypothesis. Although no significant differences were found, mean normalized differences were more negative (larger decreases in first-order measures) for high performers than for low performers in four of the five site-window combinations tested. The exception was occipital Window 4, which predicted performance inconsistently in other first- and second-order correlations.

The most consistent relationships between performance and RMS-a (both first and second order) were found with frontal Window 6. This window represents average amplitude in the probe-ERP recorded bipolarly between frontal sites F3 and F4 (10/20 system) and the reference, which was at the vertex Cz. The latency range of the window is 305 to 355 ms. The primary component of the ERP that occupies this latency range is known as the P300 (Donchin et al., 1978). P300 typically exhibits a maximum amplitude on the midline centro-parietal region and may be recorded at sites Cz and Pz. Since a voltage difference between frontal sites and Cz will reflect activity at Cz, it is highly probable that our frontal Window 6 represents P300 amplitude.

Many studies have demonstrated relationships between P300 amplitude or latency measures and various aspects of human performance (reviewed in Gopher & Donchin, 1986). However, the emphasis in the literature has been on correlations between P300 and workload rather than performance itself. In one closely related example (Israel et al., 1980), an unrestricted area measure of auditory probe-ERP P300 amplitude ranging from 300 to 1180 ms decreased between low and high workload levels of a display-monitoring task. In general, workload-related reductions in P300 amplitude occur when the evoking stimulus is not related to the primary (workload-varying) task. For example, in the Israel et al. (1980) study, the stimuli were auditory probes related to a secondary counting task. In our study, the probes were not explicitly related to any task.

Our results show that in addition to sensitivity to workload, P300 may serve as both a predictor and correlate of specific task performance and general (i.e., on-job, performance). To our knowledge, this is the first report in which ERP measures, task performance, and on-job performance were analyzed simultaneously in the same group of subjects.

Our results also indicate that other ERP components may exhibit both workload sensitivity and relationships to human performance. While frontal Window 5 may also reflect P300, clearly frontal Window 2, at a latency of 100 to 148 ms, and parietal and occipital Windows 4, at 203 to 250 ms relate to other ERP components. Frontal Window 2 may reflect ERP components related to selective attention (e.g., the N1 or Nd waves), and the parietal or occipital windows may reflect the N2 or the *mismatch negativity* waves (Ritter, Simpson, Vaughan, & Friedman, 1979; Naatanen, 1985). All of these waves are sensitive to differential processing of environmental stimuli during the performance of a variety of tasks. However, without more attention to the morphology and spatial distribution of the ERP than we employed, it is risky to attempt to match our windows to these components. One point that can be made is that a simple model of neural information processing makes predictions about a variety of ERP measures. The data have shown that some of the model predictions are supported by a measure that almost certainly represents P300, but that the behavior of other measures also supported these predictions.

CONCLUSIONS

1. A neural information processing model which links human performance to factors of total capacity and allocation range was supported by amplitude RMS-a measures of ERPs produced with irrelevant visual probe stimuli.

2. In support of the total capacity factor, baseline RMS-a in frontal Window 6 (the amplitude of the probe-ERP in a 50-ms window recorded about 330 ms after probe onset over frontal brain areas during a baseline condition) was significantly and directly related to subsequent air defense simulation (AIRDEF) performance as well as to on-job performance in a sample of 30 Marine security guards. A non-linear, logarithmic, function provided better performance prediction than did a simple linear model.

3. In support of the allocation range factor, normalized RMS-a difference scores for frontal Window 6 (differences between probe-ERP RMS-a values in active conditions and baseline conditions divided by their sum), were significantly and inversely related to AIRDEF performance. These scores also tended to relate to on-job performance, but were not significantly correlated with it. No advantage was found for a non-linear function as compared to a simple linear model.

FUTURE DIRECTIONS

The present research has demonstrated relationships between simple ERP amplitude measures and global AIRDEF performance criteria. Future experiments should employ greater attention to morphology and spatial distribution of ERP components in order to better identify sources of ERP variance that index human performance. By refining ERP measures to account for individual differences in morphology and spatial distribution, both the reliability and diagnosticity of the measures may be increased. To this end, we plan to investigate baseline-to-peak component measures, single-epoch variability, covariance of components with individual templates, discriminant analysis, and neural network algorithms for ERP classification.

The utility of ERP measures may also be enhanced by refining task performance criteria. Future work will examine different factors contributing to overall AIRDEF performance, including signal detection, short-term visual memory, visual speed and distance estimation, and decision-making strategies. By separately relating diverse performance criteria to a range of ERP measures, in individual subjects, we expect to sharpen the correlations and predictions that may be obtained.

REFERENCES

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

Edwards, A. L. (1976). *An introduction to linear regression and correlation*. San Francisco: W. H. Freeman and Company.

Geisser, S., & Greenhouse, S. W. (1958). An extension of Box's results on the use of the *F* distribution in multivariate analysis. *Annals of Mathematical Statistics, 329*, 885-891.

Gomer, F. E., Spicuzza, R. J., & O'Donnell, R. D. (1976). Evoked potential correlates of visual item recognition during memory-scanning tasks. *Physiological Psychology, 34*, 61-65.

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

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

Kelly, R. T., Greitzer, F. L., & Hershman, R. L. (July 1981). *Air Defense: A computer game for research in human performance* (NPRDC Tech. Rep. 81-15). San Diego: Navy Personnel Research and Development Center.

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

Lewis, G. W. (1983). Event related brain electrical and magnetic activity: Toward predicting on-job performance. *International Journal of Neuroscience, 18*, 159-182.

Lewis, G. W., & Rimland, B. (1979). *Hemispheric asymmetry as related to pilot and radar intercept officer performance* (NPRDC Tech. Rep. 79-13). San Diego: Navy Personnel Research and Development Center.

Lewis, G. W., & Rimland, B. (1980). *Psychobiological measures as predictors of sonar operator performance* (NPRDC Tech. Rep. 80-26). San Diego: Navy Personnel Research and Development Center.

Lewis, G. W., Trejo, L. J., Nunez, P. L., Weinberg, H., & Naitoh, P. (1988). Evoked neuromagnetic fields: Implications for indexing performance. In K. Atsumi, M. Kotani. S Ueno, T. Katila, & S. J. Williamson (Eds.), *Biomagnetism'87* (pp. 266-269). Tokyo, Japan: Tokyo Denki University Press.

Lindholm, E., Cheatham, C., & Koriath, J. (1984). *Physiological assessment of aircraft pilot workload in simulated landing and simulated hostile threat environments* (AFHRL-TR-83-49). Brooks Air Force Base, Texas: Air Force Human Resources Laboratory.

Naatanen, R. (1985). Selective attention and stimulus processing: Reflections in event-related potentials, magnetoencephalogram, and regional cerebral blood flow. In M. I. Posner and O. S. M. Marin (Eds.), *Attention and performance XI*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Ritter, W., Simson, R., Vaughan, Jr., H., & Friedman, D. (1979). A brain event related to the making of a sensory discrimination. *Science, 203*, 1358-1361.

SAS Institute Inc. (1982). The TTEST Procedure. In A. A. Ray (Ed.), *SAS User's Guide: Statistics*, Chapter 13. Cary, NC: SAS Institute Inc.

Trejo, L. J. (1986). *Brain activity during tactical decision-making: I. Hypotheses and experimental design* (HFOSL Tech. Note 71-86-6). San Diego: Navy Personnel Research and Development Center.

Trejo, L. J., Lewis, G. W., & Blankenship, M. H. (1987). B*rain activity during tactical decision-making: II. Probe-evoked potentials and workload* (NPRDC Tech. Note 88-12). San Diego: Navy Personnel Research and Development Center.

DISTRIBUTION LIST

Technology Area Manager, Office of Naval Technology
Director, Office of Naval Research (OCNR-10)
Office of Chief of Naval Operations (OP-933D4)
Naval Medical Command (MEDCOM 02D)
Office of Naval Technology (Code 223)
Naval Medical Research and Development Command (Code 40)
Technical Director, Naval Biodynamics Laboratory
Naval Aerospace Medical Research Laboratory (Code 031)
Naval Health Research Center (Code 60)
Commandant of the Marine Corps, Commanding General Marine Corps Research
 Development and Acquisition Command (MA)
Commanding Officer, Naval Aerospace Medical Research Laboratory, Pensacola, FL
Defense Technical Information Center (DTIC) (2)