DOT/FAA/AM-99/28

Multi-dimensional Characterizations of Operator State: A Validation Of Oculomotor Metrics

Erik J. Sirevaag†
John W. Rohrbaugh†
John A. Stern*
Andrei B. Vedeniapin†
Kevin D. Packingham*
Clara M. LaJonchere*

Washington University

†School of Medicine
Department of Psychiatry
St. Louis, Missouri 63110

*Department of Psychology
St. Louis, Missouri 63130

November 1999

Final Report

U.S. Department of Transportation

Federal Aviation Administration

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents thereof.

1. Report No. DOT/FAA/AM-99/28	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle Multi-dimensional Characterizations of Operator State: A Validation of Oculomotor Metrics	5. Report Date November 1999
	6. Performing Organization Code

7. Author(s) Sirevaag, E.J.†, Rohrbaugh, J.W.†, Stern, J.A.*, Vedeniapin, A.B.†, Packingham, K.D.†, and LaJonchere, C.M.*	8. Performing Organization Report No.

9. Performing Organization Name and Address Washington University †School of Medicine *Department of Psychology Department of Psychiatry St. Louis, MO 63130 St. Louis, MO 63110	10. Work Unit No. (TRAIS)
	11. Contract or Grant No. DTFA Grant 94-6-015

12. Sponsoring Agency name and Address Office of Aviation Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591	13. Type of Report and Period Covered
	14. Sponsoring Agency Code

16. Abstract

Relationships between overt behavioral measures such as Reaction Time (RT) and response accuracy (percent correct, A') and psychophysiological indices of oculomotor, electroencephalographic (EEG), and cardiovascular activity were delineated within the context of a 50 min continuous performance task. Subjects maintained comparable mean performance levels across all task segments. However, variability in response speed and accuracy increased with Time-On-Task. The increased variability was associated with longer blink durations, decreased post-stimulus blink latencies, decreased anticipatory and reactive saccade velocities and amplitudes, and fewer and later reactive saccades. While blinks were inhibited prior to all stimuli, the post-stimulus period of inhibition was longest following imperative stimuli. Target stimuli were also associated with more efficient anticipatory eye-movements. In the absence of a blink, RTs were substantially delayed. When blinks were present, very short latency blinks were associated with more variable RTs and increased errors. If blink latencies were late, RTs were late as well. Trials containing especially long duration blinks were associated with decreases in performance accuracy. Target stimuli followed by reactive saccades were responded to more slowly and with less accuracy than when effective anticipatory eye-movements preceded stimulus onset. Furthermore, the larger the amplitude of the reactive saccade, the greater the increase in RT. Abstracting peripheral information (recheck saccades) also incurred a cost in terms of increased RT -- and the slower the velocity of the recheck saccade, the greater the impact upon RT. These electrooculographic effects were accompanied by systematic changes in EEG and cardiovascular responses and exploratory multi-variate modeling indicated the degree to which both within — as well as between-subject performance variability — could be accounted for by various combinations of the psychophysiological measures.

17. Key Words Blinks, Saccades, Alertness, EOG, EEG, ECG, Modeling	18. Distribution Statement Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 33	22. Price

Form DOT F 1700.7 (8-72)   Reproduction of completed page authorized

i

TABLE OF CONTENTS

Multi-dimensional Characterizations of Operator State: A Validation of Oculomotor Metrics

INTRODUCTION

The present experiment represents one component of a larger project, the ultimate goal of which is to develop an oculomotor-based system capable of real-time alertness/attention monitoring in the field; specifically in aviation operational situations. The general strategy underlying this effort is to elaborate the relationships between the activity of the eyes and performance (as well as the specific signal processing algorithms that best capture this relationship) in laboratory/simulator settings prior to actual field testing.

However, a number of factors, though not insurmountable, complicate the general issue of how to relate oculomotor activity to momentary periods of inattention and/or lapses in vigilance. The foremost problem concerns validation. Ideally, one would like an alertness index, which can be directly related to some aspect of performance. A major difficulty arises from the existence of substantial individual differences in the ability to perform different kinds of tasks at different levels of alertness. Thus, a perfect measure of alertness might predict performance quite well for one individual, while another person may show no overt signs of impairment even though alertness is at unacceptably low levels. Conversely, the same person may be able to perform well on a perceptual motor task at low alertness levels but be incapable of performing a more demanding mental arithmetic task.

One might ask why, in the case of no obvious performance impairment, the level of alertness should be considered unacceptable. The answer to this question lies in the ability to deal with rare but cognitively challenging events. Consider, for example, an airplane pilot during a long-duration flight. Due to the high level of automation in the cockpit and the routine nature of the demands placed upon the pilot, adequate performance, as measured by any parameter of aircraft activity (e.g., altitude/course deviations) can be maintained at even low levels of alertness. The performance consequences of impaired alertness may only be evident when a critical equipment failure or freak atmospheric event increases the demands upon the pilot, for example.

In short, the ultimate goal of an alertness-monitoring system is to detect operator states associated with an increased probability of making a catastrophic error (either in detection or judgment). The validation problem, then, concerns how to prove that a system is capable of reducing the occurrence of catastrophic, yet extremely rare, events. While field studies provide the best information concerning feasibility, reliability and usability, the validity of a device cannot be ascertained in controlled field settings – while equipment can be stressed to the failure point, this strategy cannot be employed when assessing human operators.

Clearly, the best assessment of the efficacy of any device can be obtained after the system has been ubiquitously employed in the field for a substantial period of time. Then five-year accident rates before and after implementation can be evaluated to determine whether a substantial effect upon safety has occurred. In the interim, validity-testing must come from controlled laboratory and simulator environments where the state of the operator, as well as the demands of the task can be varied to include levels that would be prohibitively dangerous in applied settings. While there are some problems of generalization, increased errors and slowed reaction times in the laboratory have been found to predict an increased likelihood of catastrophic performance in the field. The present study represents an attempt to improve the ability to link oculomotor behavior to performance in two ways.

First, traditional measures of "overt" performance have been augmented by the inclusion of "covert," physiological behavioral measures. These measures provide a richer description of operator state against which to check the validity of putative metrics such as measures of oculomotor activity. Thus, because the causes underlying performance lapses are inherently multi-dimensional (encompassing such factors as fatigue and inattention) a multi-dimensional characterization of alertness is desirable.

The second strategy for improving the ability to detect relationships between eye-movements and performance is to recognize that the transition from a state of high alertness to drowsiness does not occur in a smooth (or graded) fashion. Instead, the frequency with which relatively stable periods of performance are punctuated by transient attentional drop-outs increases. Thus, while similar average performance levels may be maintained, performance variability about the mean is likely to increase as a function of both fatigue and lowered levels of attentiveness (Bills, 1934). Attempts to relate alertness measures to behavior should not be restricted to predicting mean performance levels but should also explore whether the measure is premonitory of increased performance variability.

To this end, a modified Bakan Continuous Performance Task (CPT) was employed. Subjects were required to monitor sequentially displayed digits and produce a response when either three odd or three even digits were presented in succession. This task was chosen for several reasons. First, information processing demands varied considerably on a stimulus-by-stimulus basis allowing for momentary shifts in processing efficacy to be evaluated. Second, this task could be presented for long enough periods of time to allow for the manifestation of Time on Task (TOT) effects. Finally, this paradigm produces an environment rich in eye-movements that can be systematically related to factors such as fatigue as well as the expectancies and strategies of individual subjects.

METHODS

Subjects

Twenty individuals ranging in age from 19 to 28 years participated in the study. All subjects were right-handed, in good health, and possessed normal or corrected-to-normal vision and hearing. The subjects were screened for previous history of head trauma or neurological disorders. Monetary remuneration ($30.00) was provided upon completion of the experiment. Data from one of the subjects were discarded due to inadequate levels of responding (error rates in excess of 50%); while the electroencephalographic (EEG) records of three of the subjects contained either clinical EEG abnormalities or unacceptably high levels of movement and/or electrodermal contamination. Thus, the analyses described below were conducted upon a database consisting of 16 subjects (eight male and eight female).

Tasks/Stimuli

A modified Bakan running-memory task was employed. A total of 1,200 stimuli was presented at four predictable locations along a horizontal row separated by 0, 7, 14 and 21 degrees of visual angle from the left-most location. These stimuli, consisting of yellow numerical digits (0.9) flanked by "Xs" (e.g., X6X) presented against a blue background, traversed the screen from left to right and back again with a constant Stimulus Onset Asynchrony (SOA) of 2,511 msec. The central digit was replaced with an "X" (e.g., XXX), after an initial exposure duration of 300 msec. Subjects were instructed to press a mouse button with one hand following the presentation of three consecutive odd digits, while three consecutively displayed even digits required a button press with the opposite hand. Subjects were further instructed that four consecutive even/odd digit sequences would never occur.

Two pseudo-random stimulus lists (A and B) were created, differing only in that the relative positions of even and odd digits were reversed in the two lists. The last 600 stimuli in each list were generated by reversing the order of the first 600 stimuli. Of the 1,200 stimuli, 12.25% required a response (evenly distributed across hands). Target density was controlled so that the number of imperative stimuli was either two or three during each minute of task performance.

Eight groupings created by the various digit (even/odd) by response hand (left/right) by stimulus list (A/B) mappings were equally represented within the male (n=8) and female (n=8) subject groups. The words "EVEN" and "ODD" (oriented vertically) were continuously displayed along the appropriate left/right extremes of the display screen in order to inform subjects which response hand was to be utilized for a given stimulus category.

Apparatus

EEG activity was recorded continuously, from 17 standard sites (Fz, F3, F4, F7, F8, Cz, C3, C4, T3, T4, T5, T6, Pz, P3, P4, O1 and O2) according to the International 10/20 system (see Jasper, 1958), and from two non-standard sites (C3' and C4') located 4 cm to the left and right of Cz, re-

spectively. The Ag/AgCl scalp EEG electrodes were all referred to the two earlobes linked through 5 Kohm resistors. A ground electrode was attached to the subject's forehead.

Horizontal electrooculographic (EOG) activity was recorded from electrodes placed on the outer canthus of each eye. For two of the subjects, vertical EOG measurements were obtained from a bipolar montage located on the outer canthus of the left eye oriented perpendicularly to the plane of horizontal recording. For the remaining subjects, traditional vertical EOG placements above and below the left eye and centered on the pupil were employed.

Electrocardiographic (ECG) activity was recorded from two electrodes — one located immediately below the heart and the other symmetrically placed with respect to midline on the right side of the chest. Thoracic and abdominal respiratory movements were monitored via Grass PT 5A volumetric pressure transducers. All electrode impedances were maintained at or below 5 kohms.

The physiological signals were amplified by Grass Model 8-24E amplifiers with a time constant of 10 sec and a high frequency cutoff of 70 Hz. Continuous paper tracings of the polygraph output were obtained for each session. The data were also continuously sampled at 200 Hz and stored to disk by a PDP11-based data acquisition and recording system. The presentation of visual stimuli and the recording of responses were under the control of a 386-based personal computer. A background of white noise was continuously presented through E.A.R. Tone 3A insert headphones.

Procedure

Each subject participated in a single session that always began at 1:00 pm to capitalize on the early to mid-afternoon dip in alertness (see Broughton, 1994 for a review). Upon arrival at the laboratory, subjects filled out an informed consent form and then endorsed an item on the Stanford Sleepiness Scale (SSS – Hoddes et al., 1973) indicating their current level of drowsiness. A personal history questionnaire, that included items concerning current medical and medication status, recent history of food and alcohol consumption, typical and current sleep patterns, as well as standard requests for information on age, gender and handedness, was then administered.

The next hour consisted of a thorough briefing on the purpose of the experiment; a discussion of the tasks to be performed, as well as various concerns related to the recording of physiological signals; and the actual attachment of the electrodes and respiration transducers. Subjects were then seated in a dimly lit, sound-attenuated and shielded chamber, and the sleepiness scale was administered once again. Earplugs, through which white noise was continuously presented during all recording conditions, were inserted at this time.

Five minutes of data were subsequently recorded under eyes-open resting conditions during which the subjects were asked to minimize eye movements by maintaining fixation on the center of a CRT placed directly in front of them. This condition was followed by an eye-movement calibration procedure that required subjects to visually track the movements of a fixation spot, that horizontally traversed the display screen in discrete steps.

Subjects next received five minutes of practice on the running memory task. Instructions emphasizing both speed and accuracy were delivered (i.e., the subjects were to respond as quickly as possible but not at rates that would generate substantial numbers of errors). After providing an additional self-report of drowsiness, subjects experienced 50 minutes of continuous performance on the running memory task. During this block, experimenter/subject interactions were not permitted. A final eyes-closed resting condition (10-min) then occurred. In addition to the initial SSS administration upon arrival at the laboratory (SS1), additional measures were obtained just prior to the 5-min eyes-open resting condition (SSS2), the 50-min Bakan task (SSS3), and before and after the 10 min eyes-closed resting conditions (SSS4 and SSS5).

Analyses

Questions concerning the consequences of increased Time-On-Task (TOT) were explored by segmenting the 50 min of continuous performance as follows: The first 2 min of the task (considered a "warm-up" period) were not analyzed, the remaining 48 min were divided into three consecutive 16-min blocks, designated "Early," "Middle," and "Late." A total of 1,147 stimuli were presented during these three segments.

Unless otherwise stated, the stimuli were further subdivided into five mutually exclusive categories.

"Imperative" stimuli (i.e., a third consecutively presented odd or even digit) required a button-press response. The stimulus immediately preceding an imperative stimulus (i.e., the second consecutively presented odd or even number) was labelled "2nd-Digit". A stimulus was termed a "disconfirmation" if it was preceded by a 2nd-Digit stimulus (leading to an expectancy to respond) but then alternated (i.e., an even digit preceded by two odd digits or an odd digit preceded by two even digits).

Stimuli immediately following imperative stimuli, labeled "post-imperative," were singled out because subjects could always predict whether they would be even or odd (four consecutive even/odd sequences never occurred). Finally, all digits (except for the special-case disconfirmation stimuli) that represented an alternation (e.g., even preceded by odd and vice versa) were categorized as "1st-Digit" stimuli. Of the 1,147 total stimuli analyzed, 385 were 1st-Digit, 310 were 2nd-Digit, 141 were imperative, 141 were post-imperative, and 170 constituted disconfirmations.

Unless otherwise stated, the stimulus type and TOT effects were statistically evaluated by a repeated measures 2-factor (5 stimulus type x 3 task periods) Analysis of Variance (ANOVA). Planned contrasts of hypotheses concerning specific cell means were conducted using a modified Bonferonni technique (Keppel, 1991) only in the presence of significant main effects and/or interactions.

In general, tests of the following two hypotheses were planned: 1) increased TOT would be associated with overt performance decrements and physiological response patterns indicative of decreased alertness during later task segments; and 2) the distinct information processing demands and anticipation levels imposed by stimuli in the categories detailed above would be associated with differing physiological response patterns.

RESULTS AND DISCUSSION

Subjective Sleepiness Ratings

Data summarizing the SSS estimates of drowsiness at different periods during the experimental session are presented in Figure 1. Higher scores on this scale are indicative of increasing drowsiness. Subjective reports of sleepiness clearly differed during the experimental session [$F(4,60)=26.76; p<0.0001$]. Follow up testing revealed that while SSS ratings remained constant during the three administrations prior to the vigilance block (times 1, 2, and 3), *elevated reports of sleepiness were obtained for both of the tests, that followed 50 min of continuous task performance* (times 4 and 5).
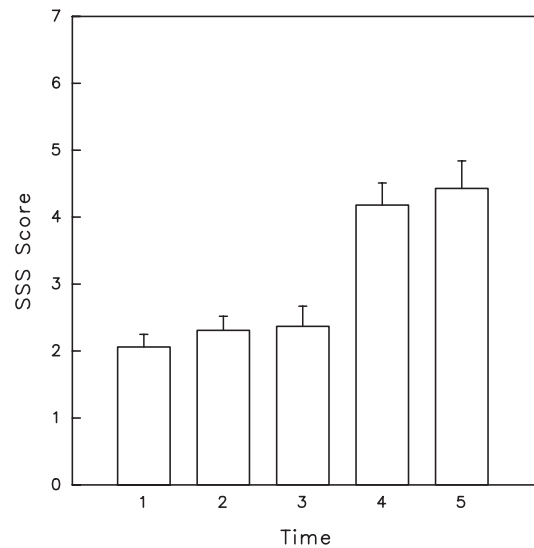

Figure 1. Stanford Sleepiness Scale Ratings.

Overt Performance Data

Measures of response speed, accuracy and operator sensitivity (Calderia, 1980 or Wickens, 1984, for a discussion of using A' as a non-parametric analog of d') are presented in Table 1. While several of these measures showed trends towards poorer performance at later task periods, none of these differences reached statistical significance. Clearly, subjects were able to maintain comparable mean performance levels throughout the 50-min task block. However, inspection of the standard deviation scores associated with these measures indicated, 1) considerable between-subject variability and 2) increased within-subject variability during later task blocks (especially during the middle period).

For these reasons, an analysis strategy was developed that would minimize between-subject differences in baseline response levels and at the same time be sensitive to changes in performance variability as well as central tendency. For each subject, mean RT and percent correct measures were computed for every minute of the task. This distribution was then z-transformed, and local mean and standard deviations of the transformed scores were computed for each minute of the task. These local estimates of central tendency and variability were then averaged across the early, middle, and late task periods.

The z-scores for the RT distributions were multiplied by -1 so that for any given task segment, mean z-scores significantly more negative than the expected value of zero indicate poorer than average

Table I. Mean (and Standard Deviation) of Performance Metrics by Time Segment.

Measures	Task Segment		
	Early	Middle	Late
Reaction Time (msec)	647 (136.8)	667 (147.2)	655 (126.5)
A'	0.978 (0.0215)	0.964 (0.0439)	0.966 (0.0339)
Percent Correct	98.4 (1.61)	97.7 (2.36)	97.9 (1.87)
Target Hit Percentage	91.8 (8.38)	86.3 (16.55)	86.7 (13.08)
False Alarm Frequency	2 (4.0)	3 (3.0)	2 (2.0)

performance with respect to the task as a whole. Similarly, positive segment means indicate better than average performance during the segment. In a similar fashion, segment standard deviations significantly less than or greater than the expected value of 1 indicate reduced and increased performance variability, respectively. One-tailed t-tests were performed to test the hypothesis that mean performance levels would be higher and less variable during the early task segment and would deteriorate and become more variable during the middle and late task segments.

The results of these analyses are displayed in Figure 2. Just as the non-transformed measures of

Figure 2. Mean (Panel A) and Std. Dev. (Panel B) of local performance measures as a function of task segment.

response speed and accuracy did not differ as a function of task segment, the mean values of the z-score distributions (see Panel A) did not differ from the expected value of 0, even though there was a trend towards poorer performance in later task segments. However, Panel B indicates that performance variability (as indexed by average local performance standard deviations) was lowest during the early task segment for both RT ($t=-2.70$; $p<0.02$) and percent correct ($t=-1.82$; $p<0.05$), while RT variability alone increased during the middle segment of the task ($t=2.08$; $p<0.03$).

An additional analysis was performed to ascertain whether the reason that the measures of central tendency failed to discriminate performance during the various task segments was due to differences in the manner in which individual subjects may have opted to trade response speed for accuracy (Wickelgren, 1977). To this end, a composite Performance Index (PI) was created by averaging the means of the z-transformed RT and percent correct measures.

Note that high values of the PI require fast reactions and high accuracy while low PI scores reflect slow responses and low accuracy. Intermediate values can be produced either by speed/accuracy tradeoffs (e.g., fast/inaccurate or slow/accurate) or by intermediate response levels along both dimensions. Mean PI levels for the 3 task segments are presented in Figure 3. PI scores exceeded expected levels for the task as a whole during the early task segment ($t=1.89$; $p<0.04$) – when response speed and accuracy are considered simultaneously, performance did, in fact, degrade with increasing time on task. *Thus, the performance data indicate that the increased subjective reports of drowsiness induced by the vigilance task were associated with both poorer performance levels and increasing performance variability during middle and late task segments.*
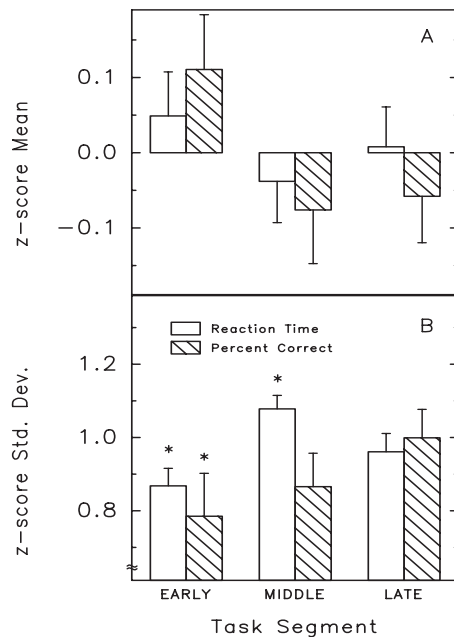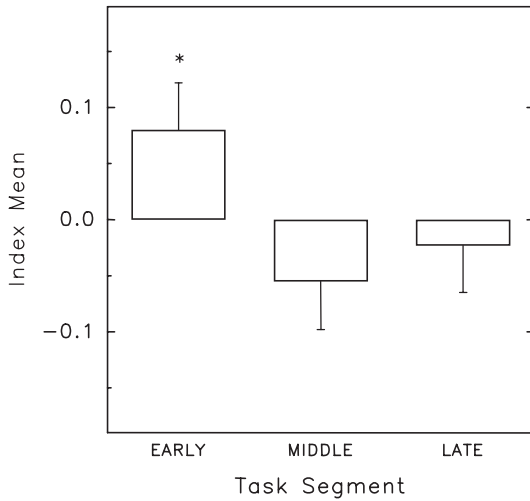
Figure 3. Mean Performance Index Scores as a function of Task Segment (* indicates p<0.05).

Blink Metrics

Blinks in the vertical EOG were automatically scored by a computer algorithm if all of the following criteria were met: a) a slope deviation in excess of 30 µVolts of appropriate polarity in a 30 msec window could be determined (initial point labelled as blink onset); b) an inflection point (termed blink peak) occurred within 200 msec of blink onset; c) the voltage difference between the blink onset and peak points was at least 50 µVolts (termed blink amplitude); and finally, d) the vertical EOG recovered (i.e., returned towards baseline) 30% of the blink amplitude value within 75 msec.
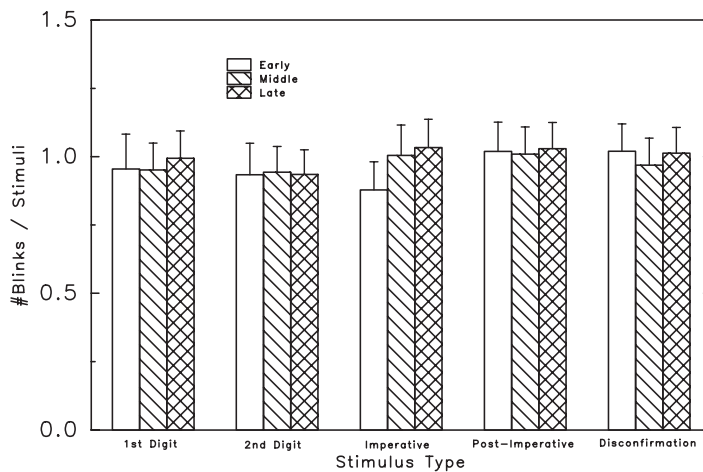

Figure 4. Blink proportions as a function of stimulus type and task segment.

Upon detection of a blink, the 50% closure duration was defined as the time difference between the point on the ascending limb of the blink (when 50% of the blink amplitude was exceeded) and a similar point on the descending limb. The performance of the automated scoring algorithm was monitored by an experimenter who was responsible for rejecting noise artifacts incorrectly identified as blinks, as well as for manually identifying blinks that the computer failed to detect.

On average, subjects produced 23.66 (s.d. = 11.96) blinks/min during the vigilance task, which represents a significant increase over the 15.78 (s.d. = 9.40) blinks/min generated during the 5 min eyes-open no task condition ($t=4.1593$; $p=0.0004$). Figure 4 displays the blink proportions (#blinks/#stimuli) associated with the various stimuli presented during the three task segments. Blink proportions (rather than frequencies) are presented to correct for the unequal trial sizes in the different stimulus categories. Blink production was not sensitive to either the TOT [$F(2,30)=0.51$; $p=0.5330$] or stimulus category [$F(4,60)=0.88$; $p=0.1015$] manipulations. Despite an apparent trend towards decreased blinking following imperative stimuli during the early task segment, the TOT x stimulus type interaction also did not reach statistical significance [$F(8,120)=2.09$; $p=0.1015$]. *Thus, while the vigilance task increased the pressure to produce blinks with respect to the non-task condition; equivalent blink rates were associated with all task periods and stimulus types.*

In contrast, blink durations (see Figure 5) did vary as a function of both TOT [$F(2,30)=9.53$; $p=0.0035$] and stimulus type [$F(4,60)=12.67$; $p=0.0001$]. The lack of a significant interaction [$F(8,120)=0.30$; $p=0.737$] allowed follow-up tests for each of these independent variables to be collapsed across levels of the other factor. These tests confirmed that a) blink durations increased during the middle and late task segments; and b) blink duration was also sensitive to the differing information processing demands associated with the various stimulus categories. *Several stimulus category effects emerged, all of which are consistent with the hypothesis that blink durations would decrease with decreasing temporal proximity to periods where an imperative stimulus was anticipated.*
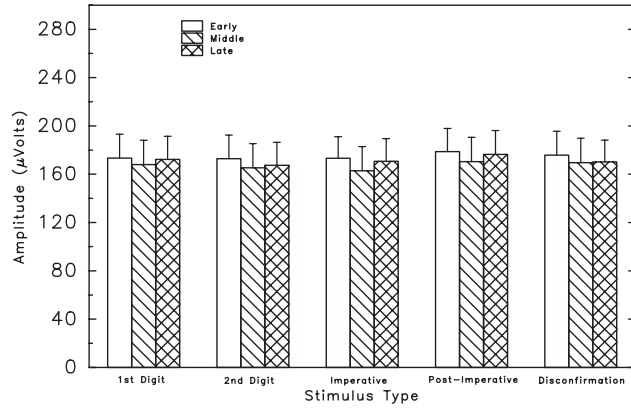
Figure 5. Blink durations as a function of stimulus type and task segment.

Thus, blinks following disconfirmation stimuli (which were at least two digits removed from the next potential imperative stimulus) were longer than the blinks following all other stimuli. Similarly, blinks associated with 1st-digit stimuli were longer than blinks following 2nd-digit and imperative stimuli (which did not, in themselves, differ). Post-imperative stimulus blinks were longer than blinks following imperative stimuli but shorter than disconfirmation-related blinks.

The results of the analysis of blink amplitudes are displayed in Figure 6. Blink amplitude did not vary with TOT [$F(2,30)=1.77; p=0.1980$] but was sensitive to stimulus category [$F(4,60)=5.81; p=0.0015$]. These factors did not interact [$F(8,120)=1.11; p=0.3619$]. The analysis of stimulus effects collapsed across the three task segments revealed that
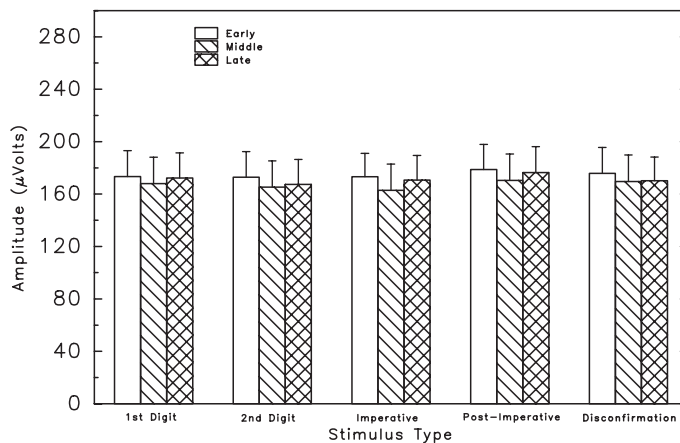

Figure 6. Blink amplitude as a function of stimulus and task segment.

the blinks following post-imperative stimuli were larger than the 1st-digit, 2nd-digit and imperative stimulus blinks, while the disconfirmation-related blinks could be discriminated only from the 2nd-digit blinks. *Taken together, the amplitude and duration results are consistent with the hypothesis that blink morphology adapts in order to minimize processing "down time" due to closure of the eye with increases in the anticipation of a target stimulus.*

Additional analyses were conducted to explore how the timing of blinks may be related to the TOT and stimulus category manipulations. Mean latencies to the first blink following a given stimulus type are displayed in Figure 7. The blink latency measure was sensitive to both TOT [$F(2,30)=12.28;$
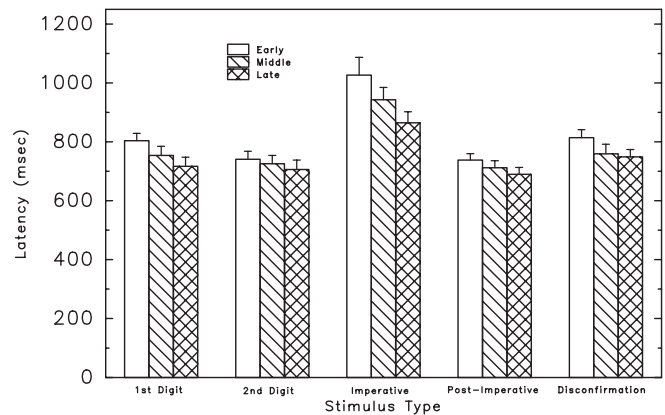

Figure 7. First blink latency as a function of stimulus type and task segment.

$p=0.0008$] and stimulus type [$F(4,60)=23.95; p=0.0001$]. The interaction of these factors was also significant [$F(8,120)=3.78; p=0.0226$]. Follow-up tests confirmed that blink latencies were considerably delayed following the imperative stimulus during all three task segments. Additionally, blink latencies tended to decrease with increased TOT.

The significant interaction effect results because the TOT effects were not equivalent across stimulus categories. Thus, first blink latencies increased from early to middle to late following the imperative and 1st-digit stimuli. For the disconfirmation stimuli, the middle and late segments did not differ from each other but were associated with reduced latencies, in comparison with the early segment. Blink latency following post-imperative stimuli discriminated only between the early and late segments. Finally, no TOT related latency
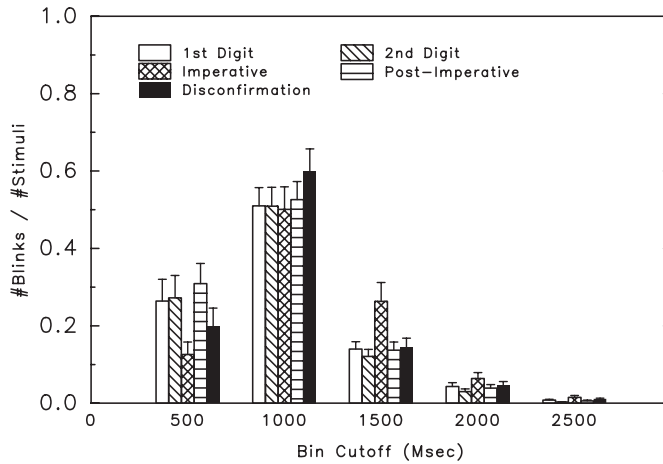
7

Figure 8. Post-stimulus blink distributions as a function of stimulus type and task segment.

differences reached statistical significance for the blinks following 2nd-digit stimuli.

Figure 8 displays the results of a different analysis of blink timing in which the post-stimulus distribution of blinks was calculated for consecutive 500 msec bins. Clearly, blinks were not randomly distributed throughout the interval between stimuli as indicated by a significant main effect of bin [$F(4,60)=58.77$; $p=0.0001$]. The blink distributions also varied as a function of stimulus type giving rise to a significant bin x stimulus type interaction [$F(16,240)=6.12; p=0.0016$]. A post-hoc testing strategy was performed, comparing the distribution of blinks following imperative and disconfirmation stimuli with the blink distributions, collapsed across all the other stimulus categories. These tests were conducted to examine the following hypotheses: a) blinks would be inhibited as a function of the anticipation of a target; b) this inhibition would last longer for the imperative stimuli that actually required a decision to respond.

Both hypotheses were confirmed. Blinks were significantly less likely to occur following both the imperative and disconfirmation stimuli during the first 500 msec post-stimulus. In the 500-1,000 msec interval, the release from inhibition resulting from the recognition that the anticipation of a response had been disconfirmed resulted in an increase in blinking relative to the other stimulus categories. Finally, a delayed release from inhibition following the imperative stimuli produced elevated blink rates during the 1,000-1,500, 1,500-2,000, and the 2,000-2,500 msec post-stimulus intervals.

The relationship between the blink distributions following imperative and disconfirmation stimuli and the latency distribution of post-imperative stimulus

button press responses is displayed in greater detail in Figure 9. In this figure, the data have been binned in 50 msec epochs for each of the early, middle and late task segments. Solid bars indicate bin proportions significantly different from zero.

Blinks were almost completely inhibited during the first 500 msec following stimulus presentation. Blinks were also extremely rare during the 1,000 msec interval preceding the presentation of the next stimulus. The distribution of blinks following disconfirmation stimuli closely parallels the distribution of button press responses following imperative stimuli. The much less sharply peaked distribution of imperative stimulus blinks clearly reflects the delayed release from inhibition associated with the requirement to respond.

The bottom Panel of Figure 9 displays difference scores calculated by subtracting the imperative distribution from the disconfirmation proportions. The bi-phasic nature of the resulting distribution (significantly positive during the 500-750 msec interval and significantly negative during the 750-1,000 msec interval) supports the hypothesis that blink inhibition would be prolonged following imperative stimuli.
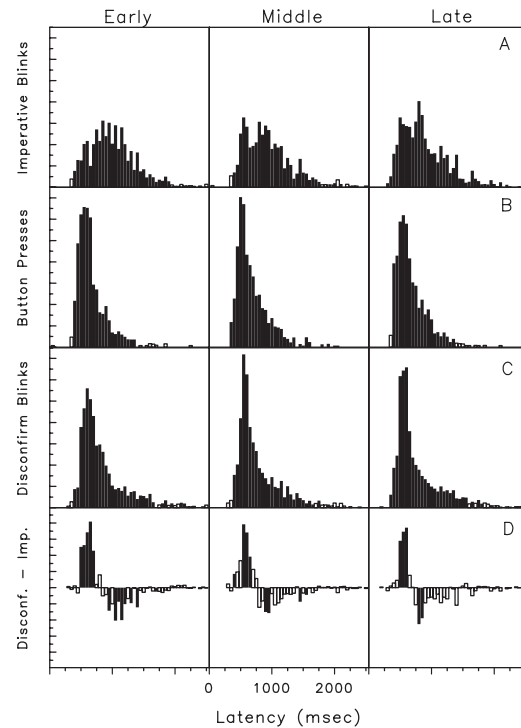

Figure 9. Imperative blink (panel A), button press response (panel B), Disconfirmation blink (panel C), and Disconfirmaiton – Imperative distributions (panel D).

8

Single Trial Relationships Between Blink Parameters and Performance

The sensitivity of the blink closure duration and latency measures to the differing information processing demands associated with the various stimulus categories suggests that variability in these metrics may be systematically related to variance in single trial measures of performance. Several additional analyses were conducted to explore this possibility.

The non-random blink distribution displayed in Figure 8 suggests that blinks are produced during periods when subjects momentarily suspend the active intake and processing of information and are inhibited during other times. Thus, increased processing times on trials where the subject was having trouble making the decision to respond should significantly delay blink initiation latencies. Note that especially long processing delays will begin to encroach upon the interval normally associated with an inhibition of blinking in anticipation of the presentation of the next stimulus and therefore decrease the probability that a blink will be produced. This logic leads to the prediction that imperative trials containing no blinks should be associated with an increase in response time and/or a decrease in response accuracy in comparison with trials containing blinks.

To test this hypothesis, performance measures were sorted as a function of whether or not an associated blink was produced by each subject. This analysis could not be performed on one subject, due to an exceptionally low overall blink rate (6.5 blinks/min). On average, 120 (s.d. = 26.14) of the 147 imperative stimuli contained blinks. The mean RT and percent correct measures associated with the blink present vs. blink absent trials are displayed in Figure 10. The results are clear – RTs on trials containing blinks were approximately 80 msec faster than on trials where blinks were absent [$F(1,14)=6.80$; $p=0.0207$], while performance accuracy remained constant [$F(1,14)=0.02$; $p=0.8946$].

Subsequent analyses were undertaken to determine whether additional relationships between performance and both the timing and morphology of blinks could be demonstrated. As indicated above, blink latency and duration significantly varied as a function of both stimulus type and TOT. To examine whether variability in these measures
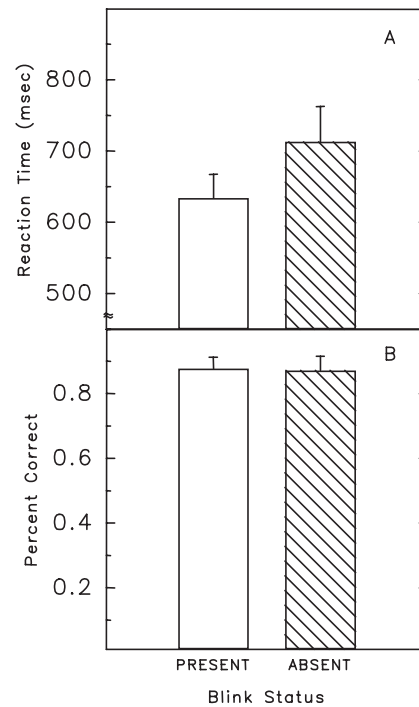

Figure 10. RT (panel A) and performance accuracy (panel B) in the presence and absence of a concurrent blink

could be related to single trial performance variability, imperative stimuli containing blinks were classified as a function of a) blink-duration quartile, and b) blink latency quartile. RTs and percent correct scores, as a function of these blink parameters, were then computed.

The performance measures associated with progressively longer blink durations are presented in Figure 11. RT did not vary as a function of blink duration [$F(1,14)=2.49$; $p=0.1096$]. However, a significant relationship between blink duration and performance accuracy was obtained [$F(1,14)=4.33$; $p=0.0375$]. Follow-up testing indicated that this effect was due to a significant decrease in percent correct scores for the quartile containing the longest duration blinks, as compared with all other quartiles.

Figure 12 displays the results of an identical analysis on the performance data sorted as a function of blink latency quartiles. As with the blink duration analysis, the latency of associated blinks clearly affected performance accuracy [$F(3,42)=9.19$; $p=0.0017$]. The shortest latency blink quartile was associated with poorer accuracy measures than all of
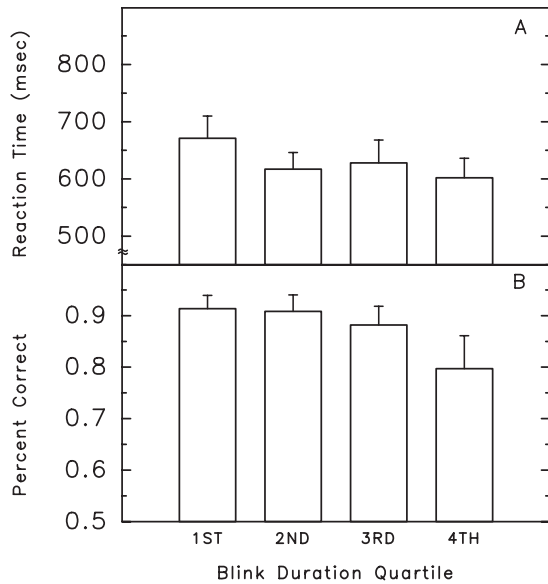
Figure 11. RT (panel A) and performance accuracy as a function of blink duration quartile.

Figure 12. RT (panel A) and performance accuracy (panel B) as a function of blink latency quartile.

the other quartiles. In addition, the second shortest quartile could also be discriminated from the fourth (longest latency) quartile.

The main effect of latency quartile upon RT was not significant [$F(3,42)=2.89$; $p=0.0777$]. A separate test excluding the quartile containing the shortest latency blinks was conducted for two reasons. First, as indicated above, this quartile was associated with significantly poorer performance accuracy; and second, a separate analysis indicated that this quartile was associated with significantly increased RT variability compared with the other quartiles ($t=2.05$; $p=0.0298$). The analysis of latency quartiles 2 through 4 did indicate a significant effect upon RT [$F(2,28)=6.26$; $p=0.0109$] – responses were significantly delayed for the 4th (longest latency) quartile compared with quartiles 2 and 3.

These analyses clearly support the hypothesis that variability in blink parameters could be related to trial-to-trial performance variance. In the absence of a blink, RTs were substantially delayed. On those trials containing blinks, if the blink latency was very short (indicating that the blink occurred during times when the subject should have been engaged in active information processing), errors increased and RTs became more variable. If the blink latency was particularly long, RT was long, consistent with the hypothesis that increased processing times on these trials were associated with increased intervals of blink inhibition. Finally,
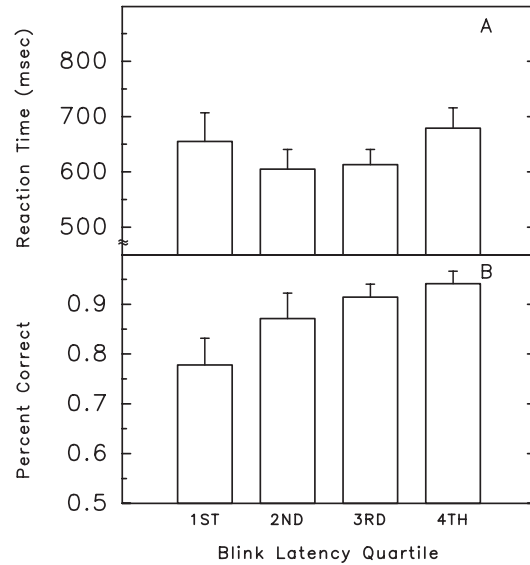
trials containing particularly long duration, presumably fatigue-related blinks, were also associated with decreases in performance accuracy.

Saccade Metrics

Saccadic activity in the horizontal EOG was automatically scored by a computer algorithm described elsewhere (see Oster and Stern, 1980). The performance of the algorithm was monitored by an experimenter who was responsible for rejecting noise artifacts incorrectly identified as saccades, as well as for manually identifying saccades that the computer failed to detect.

Figure 13 displays the post-stimulus saccade proportions (#saccades/#stimuli) associated with the various stimulus categories presented during the three task segments. While horizontal saccade production was not directly affected by TOT [$F(2,30)=0.58$; $p=0.5153$], significant differences as a function of stimulus category [$F(4,60)=24.85$; $p<0.0001$], were mediated by a significant interaction with TOT [$F(8,120)=3.98$; $p=0.0003$].

Separate tests of stimulus category differences were conducted for each of the three task segments to elucidate the nature of this interaction. These tests revealed several relationships that were stable across task segment. Namely, the saccade proportion values obtained following 2nd-Digit stimuli were always larger than the saccadic proportions following the
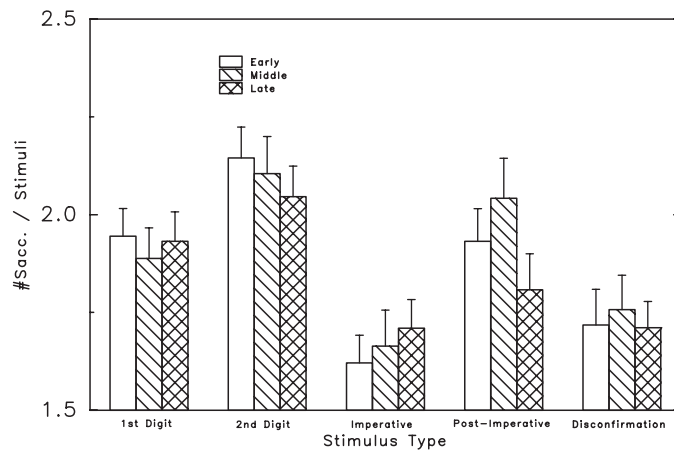
10

Figure 13. Post-stimulus saccade proportions as a function of stimulus type and task segment.

1st-Digit, imperative and disconfirmation stimuli. Thus, the inter-stimulus interval immediately preceding a potential imperative stimulus was associated with the greatest level of saccadic activity. Similarly, the periods immediately following a potential target (i.e., the post-imperative and post-disconfirmation intervals) were associated with a reduction in saccadic activity at each task segment.

The interaction between stimulus category and TOT resulted from variability in saccadic activity following post-imperative stimuli. During the early task segments, the saccadic activity following these stimuli was equivalent to 1st-Digit stimuli. An increase during the middle segment produced a post-imperative level indistinguishable from that, following 2nd-Digit stimuli. Finally, the level of saccades following post-imperative stimuli during the late task segment was comparable to the imperative and disconfirmation stimulus levels.

All of the above results are consistent with the hypothesis that the production of saccades is influenced by the expectancy for a target event. Saccade production was increased in anticipation of a target and was decreased immediately following either a response or a disconfirmation of this expectancy. Finally, saccades following the predictable post-imperative stimuli were associated with the most variable pattern of saccadic activity. This non-random distribution of saccades supports the hypothesis that subjects will adopt different patterns of eye movements as a function of changes in the information processing demands of the task. In particular, the above results indicate that subjects move their eyes to the location of an anticipated stimulus

more frequently when the probability is high that this stimulus will require a response.

The total distribution of saccades was divided into three sub-classes to more precisely determine how patterns of eye movements are related to information processing and performance. The first two categories were labeled reactive and anticipatory saccades. "Reactive" saccades, or saccades triggered by the presentation of a stimulus, serve to center the stimulus on the fovea. A saccade was classified as reactive if it occurred between 100 and 300 msec following stimulus onset. "Anticipatory" saccades were defined as movements which brought the eye towards a stimulus location in advance of the actual presentation of the stimulus.

The third category, labeled "re-check" saccades, consisted of relatively large-amplitude lateral eye-movements that returned after a short pause to a position close to the original point of regard. The rectangular-shaped waves produced by re-check saccades were readily identifiable in the EOG records and presumably reflect the re-acquisition of information concerning the mapping of odd- and even-digit responses to the respective correct response hand. This information was continuously displayed at the extreme lateral locations on the screen, as described in the methods section.

Figures 14-17 display the mean values of the metrics computed for both anticipatory and reactive saccades associated with each stimulus type at all three task segments. Note that the anticipatory saccade measures displayed in these figures refer to activity *before* the associated stimulus, while the reactive metrics, refer to saccades that *followed* stimulus onset. Also note that the imperative and disconfirmation stimulus categories have been collapsed in these analyses because subjects could not distinguish between these two events during the anticipatory (pre-stimulus) and reactive (within 300 msec of stimulus presentation) time intervals.

The production of both anticipatory [$F(3,45)=16.42$; $p<0.0001$] and reactive [$F(3,45)=33.54$; $p<0.0001$] saccades varied as a function of stimulus type — but in an inverse fashion (see Figure 14). Anticipatory saccades were less likely and reactive saccades more frequent for the highly predictable post-imperative stimuli (except that the post-imperative reactive saccade proportions could not be reliably discriminated from the 2nd-Digit stimuli). Conversely,
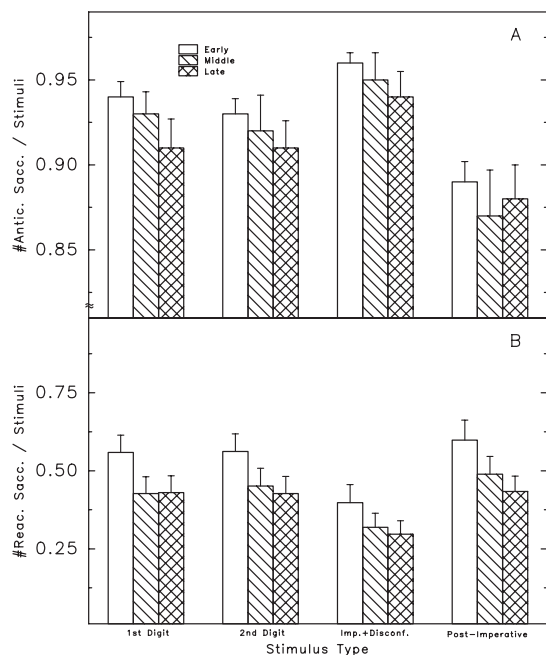
Figure 14. Anticipatory (panel A) and reactive (panel B) saccade proportions as a function of stimulus type and task segment.
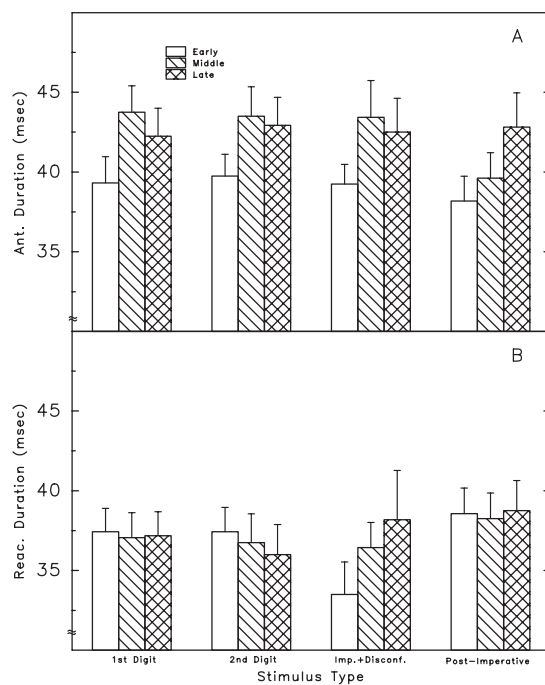

Figure 15. Anticipatory (panel A) and reactive (panel B) saccade durations as a function of stimulus type and task segment.
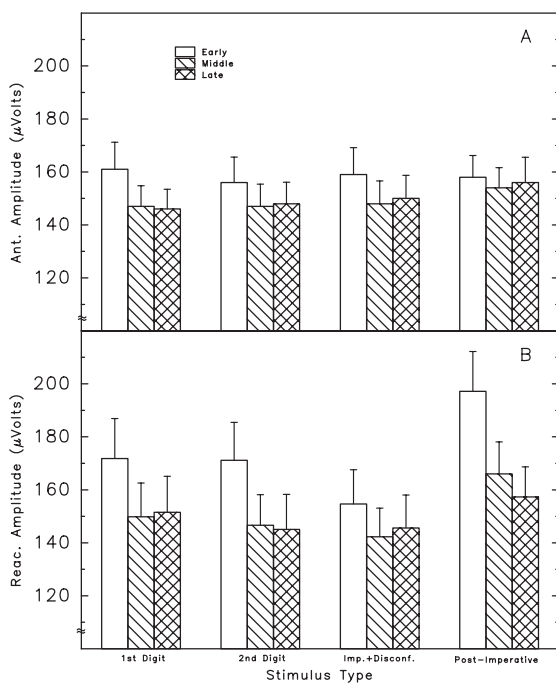

Figure 16. Anticipatory (panel A) and reactive (panel B) saccade amplitudes as a function of stimulus type and task segment.
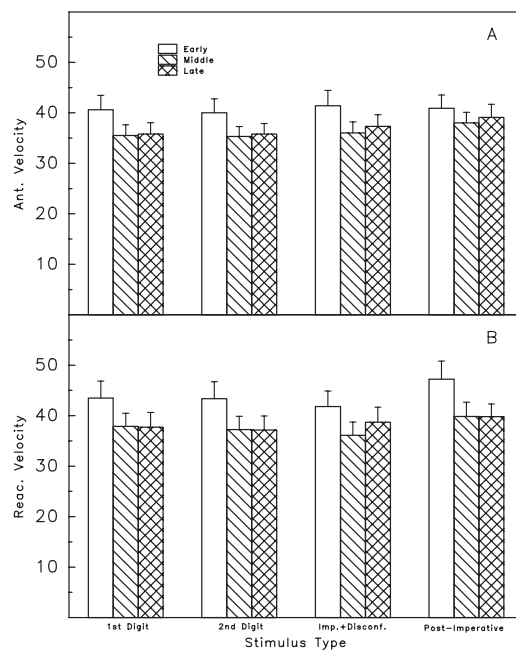

Figure 17. Anticipatory (panel A) and reactive (panel B) saccade velocities as a function of stimulus type and task segment.

anticipatory behavior was the best and reactivity the lowest when an imperative stimulus was expected. For both the anticipatory and reactive sub-types, saccade proportions associated with 1st and 2nd-Digit stimuli were equivalent.

While anticipatory saccade production remained relatively stable [$F(2,30)=0.91; p=0.3827$], reactive saccades were influenced by TOT [$F(2,30)=28.79; p<0.0001$] with fewer reactive saccades occurring during the middle and late segments than during the early portion of the task. The interaction between TOT and stimulus type was not significant for either anticipatory [$F(6,90)=0.44; p=0.6669$] or reactive [$F(6,90)=1.77; p=0.1580$] saccades.

Stimulus type did not reliably affect the duration (see Figure 15) of either anticipatory [$F(3,45)=2.74; p=0.0648$] or reactive [$F(3,45)=3.07; p=0.0572$] saccades. Reactive saccade duration was also unaffected by TOT [$F(2,30)=0.30; p=0.6925$]. Anticipatory saccade duration was influenced by TOT [$F(2,30)=7.35; p=0.0082$] with increased saccade durations associated with the middle and late task segments. This effect did not interact with stimulus type [$F(6,90)=1.80; p=0.1705$].

Differences in saccade amplitude and velocity as a function of TOT were also found (Figures 16 and 17). The significant main effects of TOT upon the amplitude [$F(2,30)=4.32; p=0.0378$] and velocity [$F(2,30)=15.74; p<0.0001$] of anticipatory saccades were due to the generation of smaller and slower saccades during the middle and late task segments. A similar pattern at these task segments also resulted in significant TOT effects upon measures of reactive saccade amplitude [$F(2,30)=17.83; p<0.0001$] and velocity [$F(2,30)=34.04; p<0.0001$]. Anticipatory saccade amplitudes were not influenced by stimulus type [$F(3,45)=1.87; p=0.1716$]. However, this factor did affect anticipatory saccade velocity [$F(3,45)=8.16; p=0.0002$]. Post-imperative and imperative stimuli were both associated with higher saccade velocities than 1st and 2nd-Digit stimuli.

For reactive saccades, both the amplitude [$F(3,45)=9.17; p=0.0007$] and velocity [$F(3,45)=5.98; p=0.0063$] measures were sensitive to stimulus type. Post-imperative reactive saccades were larger and faster than all other stimulus types. Imperative stimulus reactive saccades were smaller and slower than for all other stimulus types (the sole exception being that the 2nd-Digit reactive velocities could not be reliably discriminated). There were no
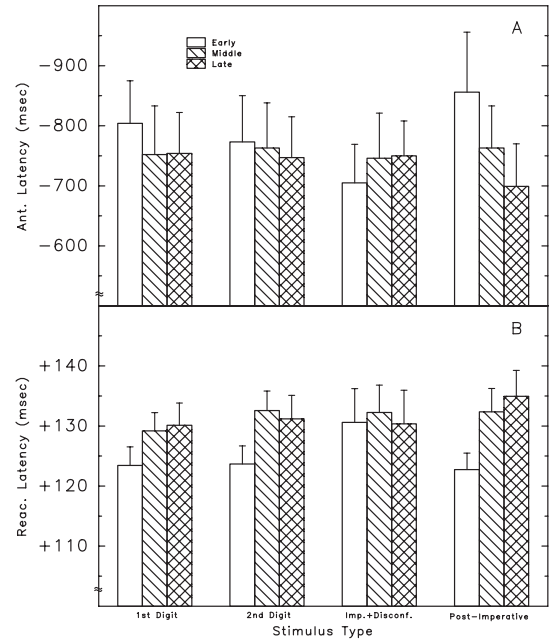

Figure 18. Pre-stimulus anticipatory (panel A) and post-stimulus reactive saccade latencies as a function of stimulus type and task segment.

significant TOT x stimulus type interaction effects upon saccade amplitude or velocity for either the reactive or anticipatory saccade categories (p > 0.05 in all cases).

Reactive saccade latencies (see Figure 18) were sensitive to TOT [$F(2,30)=7.86; p=0.0053$] but not to stimulus type [$F(3,45)=1.31; p=0.2819$], and these factors did not interact [$F(6,90)=2.04; p=0.1329$]. The TOT effect resulted from a latency increase of approximately 7 msec during the middle and late task segments. In contrast, no significant main effects of either TOT or stimulus type upon anticipatory saccade latencies were obtained (p < 0.05). However, a significant interaction [$F(6,90)=4.72; p=0.0049$] resulted from the fact that during the early task segment, anticipatory saccades occurred closer to the time of stimulus presentation when an imperative stimulus was expected than for the post-imperative and 1st-Digit stimuli. No latency differences related to stimulus type were found for the middle and late segments.

Taken as a whole, these data provide compelling support for the claim that saccade timing and morphology are sensitive to information processing demands and TOT. Subjects moved their eyes to an anticipated information source with greater frequency (as evidenced

by increased anticipatory saccade proportions) and with greater accuracy (as indexed by reduced reactive saccade proportions, amplitudes and velocities) when expecting an imperative stimulus. Similarly, in advance of the highly predictable post-imperative stimuli anticipatory behavior was poorer (reduced anticipatory proportions, increased reactive proportions, amplitudes and velocities). These effects were largely impervious to TOT. However, the morphology of saccades did vary with TOT. Both anticipatory and reactive saccades became smaller in amplitude and slower in velocity during middle and late task segments. The duration of anticipatory saccades and the latency of reactive saccades both increased during these segments as well.

Figure 20. Recheck saccade proportions as a function of stimulus type and task segment.

Data concerning the final saccade sub-type – recheck saccades – are presented in Figures 19-21. Recheck saccades could be reliably detected in 14 of the 16 subjects. The proportion of recheck saccades was clearly higher following imperative stimuli (see Figure 19). Sufficient numbers of recheck saccades to enable a factorial analysis of stimulus type (restricted to 1st-Digit, 2nd-Digit and imperative stimuli) by task segment were present for nine of the subjects. The results of such an analysis (see Figure 20) indicated that the proportion of recheck saccades was sensitive to stimulus type [$F(2,16)=7.36; p=0.0208$] but not to TOT [$F(2,16)=1.34; p=0.2860$]. Follow-up testing confirmed that the stimulus type effect was due to a greater proportion of re-check saccades following imperative stimuli (0.839) than following 1st-Digit (0.339) and 2nd-Digit (0.455)
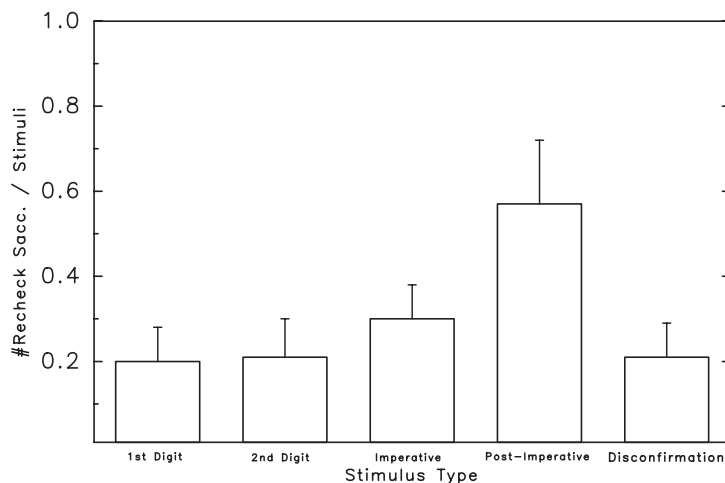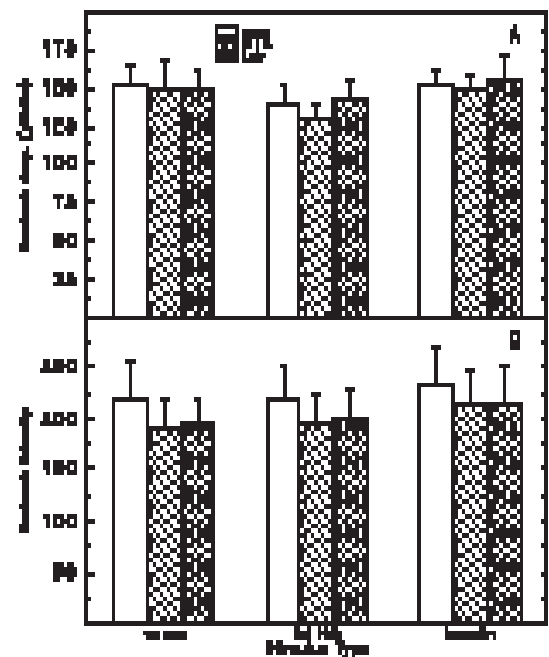

Figure 21. Recheck saccade amplitude (panel A) and duration (panel B) as a function of stimulus type and task segment.

stimuli. This effect did not interact with TOT [$F(4,32)=0.65; p=0.5356$].

Because the anticipatory and reactive saccade analyses had indicated significant TOT effects upon amplitude and velocity measures, recheck saccade morphology was also evaluated along these dimensions. However, since the termination point of these saccades was constrained by the distance required to travel to a continuously displayed information source, it was predicted that the amplitude of re-

Figure 19. Recheck saccade proportions as a function of stimulus type.

check saccades should not be affected by either TOT or stimulus type. In contrast, the speed with which the eye moved to obtain peripherally displayed information was not similarly constrained so that TOT effects upon velocity were expected.

The data confirmed both of these predictions (see Figure 21). Recheck saccade amplitudes remained constant as a function of TOT [$F(2,16)=0.49; p=0.5525$] and stimulus type [$F(2,16)=1.24; p=0.3045$]. Recheck saccade velocities, on the other hand, were influenced by TOT [$F(2,16)=4.50; p=0.0291$] but not by stimulus type [$F(2,16)=1.11; p=0.3278$]. This TOT effect, which did not interact with stimulus type [$F(4,32)=0.26; p=0.7613$], was due to decreased recheck velocities at the middle and late task segments (as was the case for the anticipatory and reactive saccades).

Single Trial Relationships Between Saccade Parameters and Performance

The sensitivity of the anticipatory and reactive saccade measures to the differing information processing demands associated with the various stimulus categories suggests that variability in these metrics may be systematically related to variance in single-trial measures of performance. Therefore, several additional analyses were conducted to explore this possibility.

The non-random saccade distributions detailed above indicate that response expectancies held by subjects are reflected in the proportions of anticipatory and reactive saccades elicited by imperative stimuli. The following analyses were conducted to determine whether imperative trials with accurate anticipatory saccades (indexed by the absence of a reactive saccade) were associated with a decrease in response time and/or response accuracy in comparison to trials containing reactive saccades. One subject was excluded from the analysis due to a virtually complete absence of reactive saccades following imperative stimuli.

On average, 55 (s.d. = 27.24) of the 147 imperative stimuli contained reactive saccades. The mean RT and percent correct measures associated with the anticipatory saccade only vs. reactive trials are displayed in Figure 22. RTs on trials containing reactive saccades were approximately 100 msec longer than on trials where reactive saccades were absent [$F(1,14)=48.06; p<0.0001$]. Performance accuracy was also affected [$F(1,14)=39.96; p<0.0001$] — poorly anticipated trials resulted in increased misses. Furthermore, between subject differences in miss rates were significantly correlated with reactive saccade latency (r=0.44). Thus, subjects
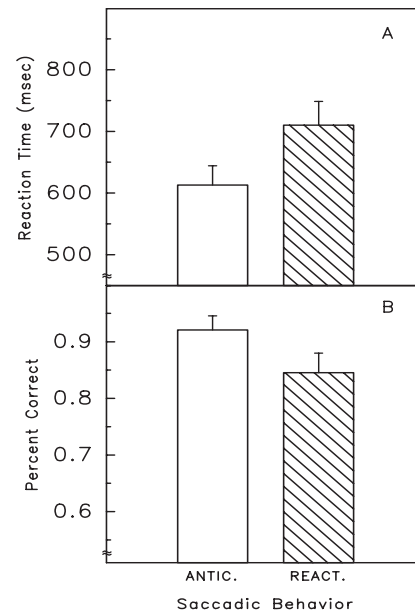

Figure 22. RT (panel A) and performance accuracy (panel B) in the presence and absence of anticipatory saccades.

with long reactive saccade latencies to imperative stimuli were more likely to generate misses.

Subsequent analyses were undertaken to determine whether additional within-subject relationships between performance and both the timing and morphology of reactive saccades could be demonstrated. Imperative stimuli containing reactive saccades were classified as a function of median splits computed on the latency and size of the saccade. RTs based upon these splits were then computed (insufficient misses occurred to reliably estimate median split response accuracies for most subjects).

RT was not systematically related to reactive saccade latency [$F(1,14)=2.10; p=0.1691$]. On the other hand, the amplitude of the reactive saccade did have a significant effect [$F(1,14)=12.10; p= 0.0037$]. RTs following small reactive saccades (mean=696, s.d. = 37.29) were approximately 30 msec faster than if the saccade was large (mean=726, s.d.=40.75).

Ten of the subjects produced sufficient numbers of recheck saccades on imperative stimuli to enable an analysis of how recheck behavior was related to performance variability. A comparison of performance on trials with and without recheck saccades is displayed in Figure 23. On average, 54 (s.d. = 44.76) of the 147 imperative stimuli contained

recheck saccades. RTs on trials containing recheck saccades were approximately 70 msec longer than on trials where recheck saccades did not occur [$F(1,9)=7.60; p<0.0222$]. A tendency towards improved performance accuracy following recheck saccades failed to achieve statistical significance [$F(1,9)=2.92; p<0.1219$].

Imperative stimuli containing recheck saccades were classified as a function of median splits computed on the velocity and size of the saccade. RTs based upon these splits were then computed. Once again, insufficient misses occurred to reliably estimate median split response accuracies for most subjects. As predicted, RT did not vary as a function of recheck saccade amplitude [$F(1,9)=1.71; p=0.2236$]. However, the velocity of the reactive saccade did have a significant effect [$F(1,14)=9.09; p= 0.0146$]. RTs following fast recheck saccades (mean=683, s.d. = 220.04) were approximately 45 msec faster than if the saccade was slow (mean=731, s.d.=221.84).

These analyses clearly support the hypothesis that variability in saccade parameters can be related to trial-to-trial performance variance. In the absence of a reactive saccade, RTs were substantially faster and accuracy was higher. Furthermore, those trials requiring large reactive saccades in order to foveate an imperative stimulus resulted in increased reaction times. The abstraction of peripherally presented information following the presentation of an imperative stimulus also occurred at a cost – trials containing recheck saccades were associated with increased reaction time, and the more slowly the eye moved to the peripheral location, the greater the RT cost.

Electroencephalographic (EEG) Transients

Transient EEG events have been associated with increasing drowsiness (Santamaria and Chiappa, 1987). The occurrence of 25 categories of events were scored by a trained clinical electroencephalographer using impressionistic criteria from the paper record. While many of these events were either infrequent, unevenly distributed across subjects, or both, two of the categories, labelled flat EEG and EEG slowing, occurred with sufficient frequency across subjects to enable an analysis of TOT effects.

The flat EEG metric reflects prolonged periods of high frequency, low-voltage EEG activity characterized by an absence of delta, theta and alpha waves. EEG slowing transients were characterized
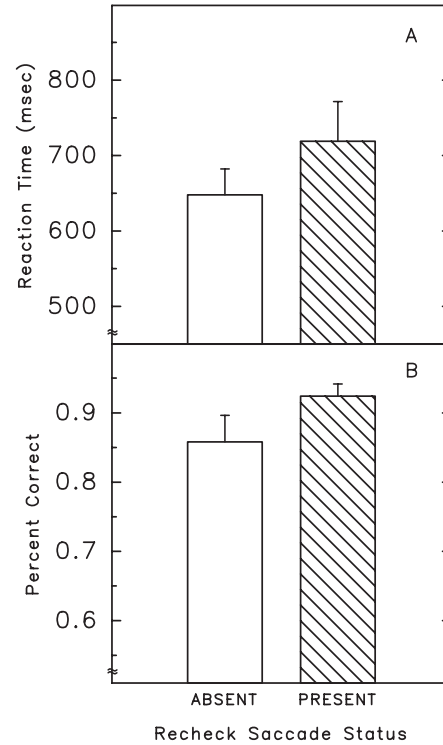

Figure 23. RT (panel A) and performance accuracy (panel B) in the presence and absence of a recheck saccade.

by periods where an initial burst of activity at one frequency (generally alpha) continuously decreased in frequency (generally into the theta range) within a 1 to 2 sec interval.

The combined frequency of the EEG slowing and flatness metrics varied significantly as a function of TOT [$F(2,30)=4.41; p=0.0229$]. Significantly more of these transients occurred during the late task segment (mean = 1.94; s.d. = 0.551) than during the early portion of the task (mean = 0.43; s.d. = 0.223). The frequency of such transients during the middle task segment (mean = 1.31; s.d. = 0.553) could not be discriminated from either the early or the late period.

Event-Related Brain Potential (ERP) Metrics

The ERP is characterized by a transient series of voltage oscillations in the EEG time-locked to the occurrence of discrete events. Thus, the ERP reflects phasic changes in the brain associated with the processing of specific events. The analysis of ERPs is complicated by two factors. First, ERP components typically range in amplitude between

1 to 30 microvolts, while the background EEG may contain voltages in excess of 50 to 100 microvolts — resulting in a poor signal-to-noise ratio. The second complicating factor results from the artifactual contamination of the EEG produced by positional shifts of the polarized eye ball. Thus, both saccades and blinks produce electrical currents that propagate along the scalp, generating substantial voltages at the EEG recording locations.

The following strategy was employed to overcome these difficulties. Periods containing excessive levels of artifact (due to muscle activity for example) were identified from the paper records and data collected during these periods were excluded from all analyses of brain electrical activity. Second, conventional signal averaging techniques, which reduce the contribution of noise by averaging across epochs associated with similar events, were employed.

The issue of eye movement contamination (a particular problem in this eye-movement rich task) was also addressed. Epochs of EEG activity, time-locked to the occurrence of saccades and blinks during the practice session, were extracted and averaged. The contributions of specific EOG events to the EEG recorded at each electrode were estimated via correlational analyses conducted upon these averages. Finally, blink and saccade propagation factors produced by these analyses were used to remove the estimated contribution of the electrical activity generated by blinks and saccades from each time point in the EEG records.

The average ERPs (recorded at Pz) elicited by the events in an imperative stimulus sequence are presented in Figure 24. Two components of the ERP signature are particularly noticeable in this figure. The P300 response (the large positive peak at approximately 500 msec post-stimulus in this Figure) has been shown to be sensitive to a number of factors related to the relevance, salience, and probability of the eliciting stimuli (Johnson, 1986) as well as the cognitive demands of the task. The latter sensitivity has been demonstrated in both laboratory (Sirevaag et al., 1984) as well as simulator (Kramer et al., 1987; Sirevaag et al., 1993) environments. The Contingent Negative Variation (CNV), which can be seen as a slow build-up in negativity preceding stimulus presentation (especially prior to the imperative stimu-

lus in Figure 24) appears to consist of at least two separate processes: an orienting component and a response preparation component (Rohrbaugh and Gaillard, 1983).

Measures of the CNV and P300 were abstracted from averaged epochs recorded at the Fz, Cz and Pz electrode locations. P300 amplitude was computed as the voltage difference between the largest positive deflection occurring in a latency window extending between 300 and 650 msec post-stimulus and the mean voltage of the 100 msec pre-stimulus baseline. P300 latency was defined as the time post-stimulus at which the peak amplitude was detected. Contingent Negative Variation (CNV) amplitude was quantified as the difference between the area under the curve in two 50-msec windows. The first window began 1,000 msec prior to stimulus presentation, the second window was centered around the time of stimulus onset.

The P300 amplitude and latency results for the Pz electrode location are presented in Figures 25 and 26. The amplitude [$F(2,30)=0.99$; $p= 0.3704$] and latency [$F(2,30)=2.25$; $p=0.1333$] of the P300 were unaffected by TOT. However, both the amplitude [$F(4,60)=50.34$; $p<0.0001$] and latency [$F(4,60)=4.16$; $p=0.0261$] measures were influenced by stimulus type. P300s were larger following imperative stimuli than when following all other stimuli, while the disconfirmation-related P300s were larger than the 1st-Digit, 2nd-Digit, and post-imperative stimuli (which did not themselves differ). P300 latencies following the imperative and disconfirmation stimuli were delayed with respect to the latencies following post-imperative stimuli. There were no interactions between the TOT
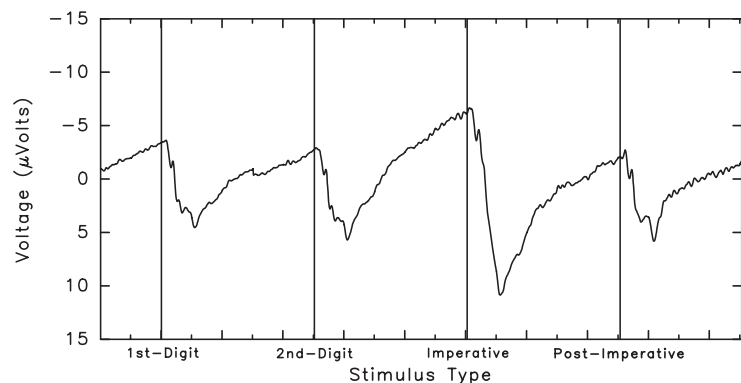

Figure 24. Grand average event-related brain potentials recorded at Pz in response to the stimuli in an imperative response sequence.
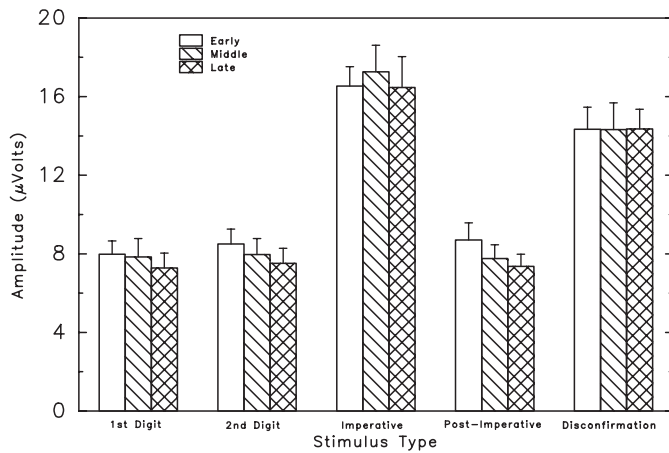
Figure 25. P300 amplittudes at Pz as a function of stimulus type and task segment.
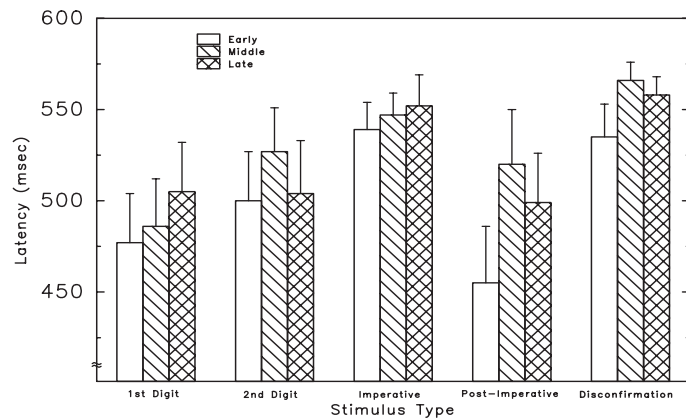

Figure 26. P300 latencies at Pz as a function of stimulus type and task segment.
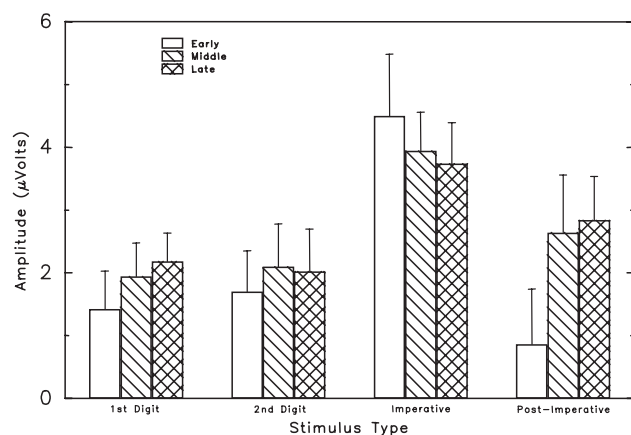

Figure 27. CNV amplitudes at Cz as a function of stimulus type and task segment.

and stimulus type factors for either the amplitude or the latency measures.

The amplitude of the CNV component at Cz is displayed in Figure 27. Note that, because this measure was computed prior to stimulus presentation, the imperative and disconfirmation stimuli have been collapsed into a single category. As was the case for the P300 data, CNV amplitude was not sensitive to TOT [$F(2,30)=1.19; p=0.3098$] but did vary with stimulus type [$F(3,45)=7.92; p=0.0051$] with no interaction. The CNVs generated in anticipation of an imperative stimulus were larger than CNVs for all the other stimulus types.

The amplitude of the P300 elicited by imperative stimuli was significantly correlated with individual differences in RT (r = -0.65). Thus, subjects producing the largest P300s also responded more quickly. CNV amplitude was not systematically related to between subject reaction time differences.

Event-Related Power Metrics

Stimulus locked modulations of power in traditionally defined bands (e.g., alpha, theta, etc.) were evaluated by a technique similar to that proposed by Pfurtscheller (1989): 1) The continuously recorded, eye-movement-corrected EEG was subjected to a Finite Impulse Response (FIR) narrow band-pass filter to exclude all activity other than the band of interest; 2) the voltages recorded at each time point were squared (power is proportional to voltage squared); 3) signal averaging algorithms were employed to compute stimulus-locked event-related modulations in power in a manner identical to that employed for ERPs.

Event-related activity in the theta (passband=4.5-7.5 Hz), gamma (passband=35.0-45.0 Hz), low (passband=7.5-8.5 Hz), medium (passband=9.5-10.5), and high (passband=11.5-12.5) alpha bands recorded at the O2 electrode is displayed in Figure 28. The theta band response (see Panel A) consists of a stimulus-locked increase in power roughly contemporaneous with stimulus presentation. Close inspection of Panel A, reveals stimulus related differences in the interval extending between 600-900 msec post-stimulus. While imperative stimuli are associated with a decrease in theta power during this period, the other stimuli exhibit a small increase in power.

Theta modulations during these two epochs were quantified by first obtaining a baseline
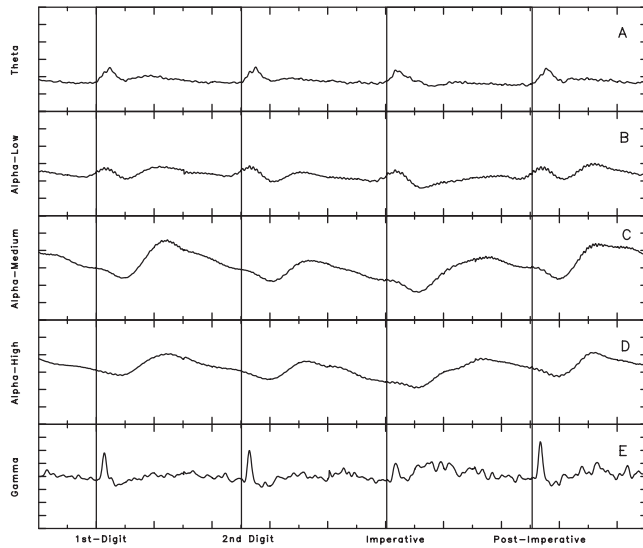
18

Figure 28. Event-related modulations at O2 of activity in the Theta (panel A), low alpha (panel B), medium alpha (panel C), high alpha (panel D) and gamma (panel E) frequency ranges.
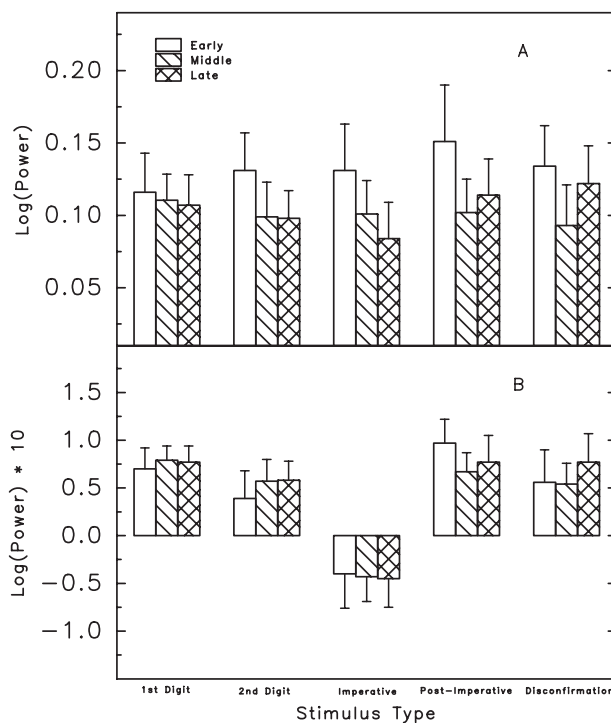

Figure 29. Early (panel A) and late (panel B) theta power modulations as a function of stimulus type and task segment.

estimate of theta activity, computed as the area under the curve in the 300 msec period preceding stimulus onset. This baseline estimate was then subtracted from similar area measures computed during the 0-300 and 600-900 msec post-stimulus intervals, respectively. Mean theta power modulations as a function of TOT and stimulus type are presented in Figure 29.

Panel A displays the results for the early 0-300 msec interval. While the initial increase in theta activity did not vary as a function of stimulus type [$F(4,60)=0.62$; $p=0.5925$], significant TOT effects were obtained [$F(2,30)=7.03$; $p=0.0044$]. The surge in theta power during stimulus presentation decreased during the middle and late task segments. This effect did not interact with stimulus type.

In contrast, theta activity during the 600-900 msec interval (see Panel B of Figure 29) varied significantly as a function of stimulus type [$F(4,60)=13.13$; $p<0.0001$] but remained constant during the different task segments [$F(2,30)=0.13$; $p=0.8703$]. Late theta activity was significantly depressed following imperative stimuli compared with all other stimulus types. Interestingly, the stimuli associated with the next highest level of response expectancy (i.e., 2nd-Digit stimuli) were associated with reduced theta power, compared with the 1st-Digit and post-imperative stimuli. *Thus, while the presentation of a stimulus induced an initial burst of theta activity that decreased with TOT, the requirement to respond resulted in a suppression of subsequent theta activity during all three task segments.*

Inspection of the event-related activity in the three alpha bands displayed in Panels B,C and D of Figure 28 reveals a number of interesting modulations. Five epochs of interest were identified, and a measure of average area under the curve was computed for each. The 1st period (epoch1) reflects the average alpha power in the 300 msec interval beginning 1,000 msec before the presentation of the stimulus. The 2nd period (epoch2) measures alpha activity just prior to and including the period of stimulus onset and was computed as the average area under the curve in the 300 msec interval beginning 200 msec before stimulus presentation. The average area in the 100-300 msec post-stimulus period was defined as epoch3. The 4th period (epoch4) spanned 500 msec beginning 300 msec post-stimulus. The final period that was quantified

(epoch5) subtended the interval from 800-1,900 msec post-stimulus.

The continuous decline in alpha power in anticipation of stimulus presentation, which can be seen most clearly in the waveform reflecting medium band alpha activity (Panel C of Figure 28) will subsequently be referred to as pre-stimulus alpha desynchronization. This activity was quantified by subtracting epoch1 from epoch2. A similar period of reduced alpha power following the stimulus (hereafter labelled post-stimulus alpha desynchronization) was computed as the difference between the epoch2 and epoch4 measures. The difference between epochs 4 and 5 was used to index the large increase in alpha power seen to peak at approximately 1100 msec in Panel C of Figure 28 (termed late alpha hypersynchronization). Finally, the early hypersynchronization most clearly seen in the low alpha band (Panel B of Figure 28) was measured as the difference between epochs 2 and 3.

Figures 30-32 display the mean values of these measures as a function of TOT and stimulus type for the low-, medium-, and high-frequency alpha bands. To facilitate inter-band comparisons, identical scales were used in all three figures. As can be seen in Figure 30, significant pre-stimulus alpha desynchronizations were not present in the low alpha band (reflected in the positive values displayed in Panel A). Similarly, the late hypersynchronization, though detectable, was negligible in this band (see Panel D). Neither of these components of event-related low alpha band modulation varied significantly as a function of either TOT or stimulus type (all p-values > 0.05). However, both components of the bi-phasic response occurring immediately following the presentation of the stimulus were influenced by task demands. The early hypersynchronization (see Panel B) varied as a function of TOT [$F(2,30)=7.18; p=0.0049$], with reductions in the amplitude of this response during the middle- and late-task segments. This effect mirrors the reductions in the early theta response during these task segments.

The measures of post-stimulus low alpha desynchronization reflected in Panel C evidence a significant effect of stimulus type [$F(4,60)=5.45; p=0.0042$]. Greater desynchronization occurred following the presentation of an imperative

Figure 30. Low alpha band pre-stimulus (panel A), and post-stimulus (panel B) desynchronizations, early (panel C) and late (panel D) hypersynchronizations at 02 as a function of stimulus type and task segment.

Figure 31. Medium alpha band pre- (panel A), and post-stimulus (panel B) desynchronizations, early (panel C) and late (panel D) hypersynchronizations at 02 as a function of stimulus type and task segment.
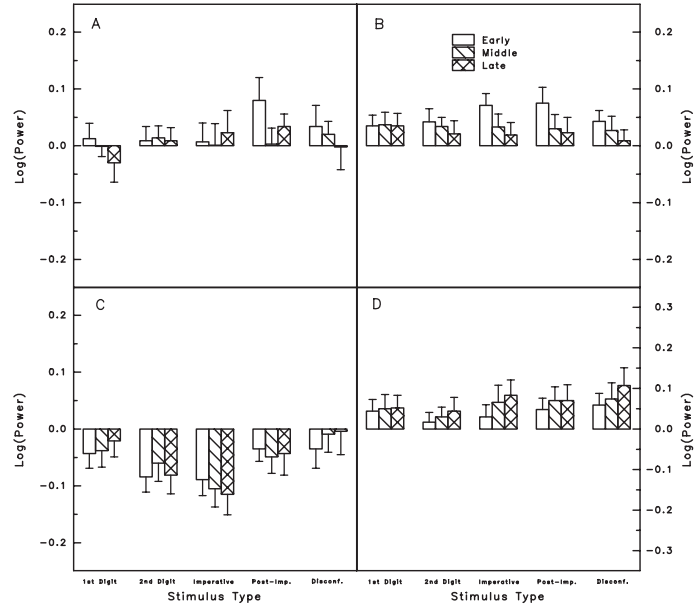
20

Figure 32. High alpha band pre- (panel A), and post-stimulus (panel B) desynchronizations, early (panel C) and late (panel D) hypersynchronizations at 02 as a function of stimulus type and task segment
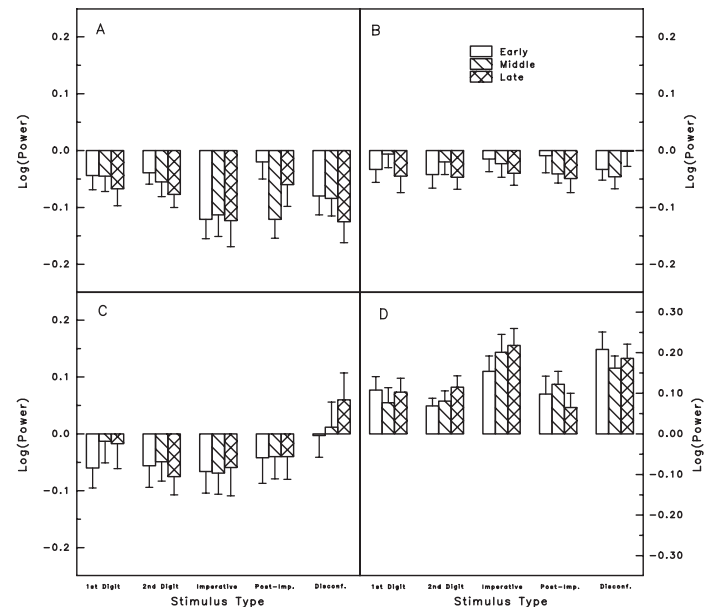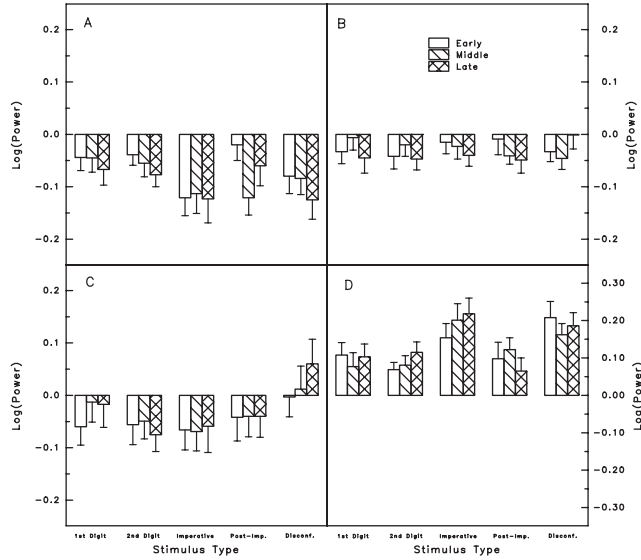
stimulus than for all the other stimulus types (except for the high expectancy level, 2nd-digit stimuli). No other low alpha band, main effects or interactions were obtained.

A different picture emerges with respect to the medium alpha band measures displayed in Figure 31. Now the post-stimulus hypersynchronization is not in evidence (witness the negative values in Panel B). The large pre-stimulus desynchronizations depicted in Panel A did vary as a function of stimulus type [$F(4,60)=3.87$; $p=0.0211$]. Larger desynchronizations occurred prior to the presentation of an imperative or disconfirmation stimulus than in anticipation of either a 1st-Digit or post-imperative (highly predictable) stimulus.

The post-stimulus middle alpha band desynchronization (see Panel C) also varied with stimulus type [$F(4,60)=6.17; p=0.002$]. Note that, in contrast to the pre-stimulus period, subjects can now discriminate between the imperative and the disconfirmation stimuli. Consequently, the post-stimulus desynchronization was smaller following disconfirmations than for all other stimulus types. Increased desynchronizations following imperative stimuli, as compared with 1st-Digit and post-imperative stimuli, were only marginally significant (p-values obtained via the modified Bonferonni procedure < 0.06).

Differences in the amplitude of the late hyper-synchronization in the middle alpha band (see Panel D), as a function of stimulus type, were also present [$F(4,60)=11.69; p<0.0001$]. More alpha power was present during later portions of the inter-stimulus interval following the removal of the expectation to respond (i.e., following imperative and disconfirmation stimuli). None of the components of the middle alpha band response, displayed a sensitivity to TOT effects or stimulus type by TOT interactions.

The responses in the high alpha frequency band are displayed in Figure 32. The overall pattern is quite similar to that obtained for the middle alpha band modulations. Once again, no early hypersynchronization was evident and no main effects of TOT were obtained for any of the components. The magnitude of the pre-stimulus high alpha desynchronization was affected by stimulus type [$F(4,60)=3.50; p=0.0313$]. However, the specific stimulus-related differences, were somewhat different than for the middle band – while the greatest desynchronizations still occurred in anticipation of the imperative and disconfirmation stimuli, these levels differed from those associated with the 1st- and 2nd-Digit stimuli (rather than the 1st-Digit and post-imperative levels discriminated by the middle band).

The magnitude of the post-stimulus desynchronization was also influenced by stimulus type [$F(4,60)=5.84; p=0.0042$] in the high alpha range. Desynchronizations associated with the disconfirmation stimuli were smaller than those following the 2nd-Digit, imperative, and post-imperative stimuli. While the late hypersynchronization was also sensitive to stimulus type [$F(4,60)=13.54; p<0.0001$], these effects were not equivalent during the three task segments, resulting in a significant TOT by stimulus type interaction [$F(8,120)=2.74; p=0.0195$]. Some relationships were constant — the hypersynchronizations following imperative and disconfirmation stimuli were always larger than following 1st- and 2nd-Digit stimuli.

The locus of the interaction effect stems from the fact that, while imperative stimulus hypersynchronizations increased with TOT, disconfirmation hypersynchronizations tended to decrease. Thus, during the early task period, imperative hypersynchronizations were smaller than for disconfirmations; during the middle period they were equivalent, and late in task performance the imperative responses were larger than the disconfirmation hypersynchronizations. Finally, while post-imperative stimuli produced smaller
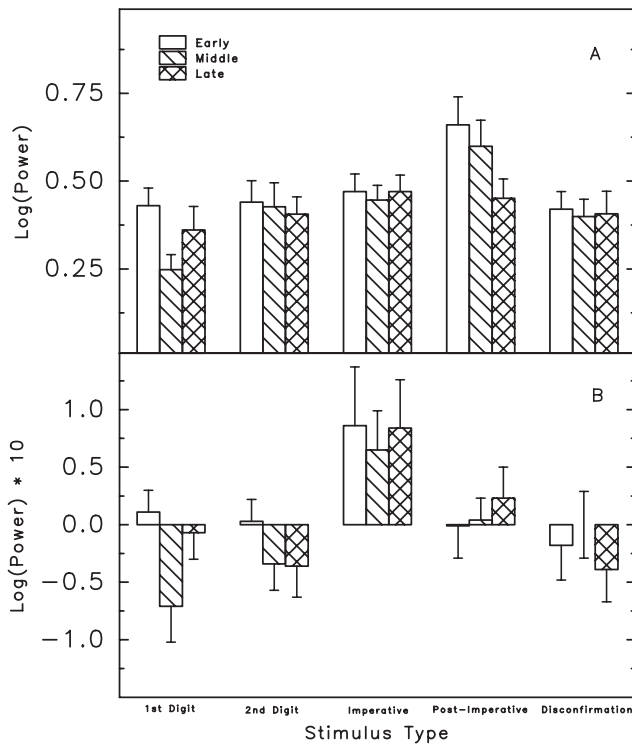
Figure 33. Early (Panel A) and late (panel B) gamma band responses at 02 as a function of stimulus type and task segment.

hypersynchronizations than the imperative stimuli at middle- and late-task segments, these levels could be distinguished from the disconfirmation responses only during the early and late periods. The slightly more complex pattern of results obtained for the high alpha band modulations than for the middle band are most likely because the decreased statistical power, resulting from smaller component amplitudes in the high band, yielded decreased effect sizes and poorer signal-to-noise ratios.

Figure 33 displays the results of analyses exploring the effects of TOT and stimulus type upon evoked gamma band activity. As depicted in Panel D of Figure 28, two distinct periods of increased gamma power can be distinguished – a sharply peaked response during the period of stimulus presentation, and a later, broader response particularly evident following imperative stimuli. These responses were extremely small – and a somewhat different quantification strategy was employed to deal with the poor signal-to-noise ratios, that resulted.

The early response was measured via a base-to-peak procedure similar to that employed for the P300 component of the ERP. Thus, the mean of the 1,000 msec baseline was subtracted from the maximum power value detected

in a window extending from 75 to 200 msec post-stimulus onset. Mean early gamma response power values as a function of stimulus type and TOT are presented in Panel A of Figure 33. Significant TOT [$F(2,30)=4.55; p=0.0192$] and stimulus type [$F(4,60)=11.66; p=0.0178$] effects were obtained. The stimulus type by TOT interaction term was also significant [$F(8,120)=2.87; p=0.0178$].

Follow-up analyses revealed that, during the early and middle task segment, the gamma response to the post-imperative stimulus was enhanced, relative to all other stimulus types. During the middle-task segment, a reduction in 1st-digit elicited gamma power could be detected, in addition to the increase following post-imperative stimuli. Finally, no stimulus-related effects were present during the late task blocks.

Late gamma activity was measured as the difference between the mean of the baseline and the mean power in a window extending from 500 to 1,000 msec post-stimulus (see Panel B). Gamma power late in the stimulus epoch was influenced by stimulus type [$F(4,60)=5.03; p=0.0137$]. Imperative stimuli generated increased late gamma power, as compared with 1st-Digit, 2nd-Digit and disconfirmation stimuli. There were no TOT, or stimulus type by TOT interaction effects upon this measure.

The event-related power analyses present a complex, yet coherent, picture of how modulations in specific frequency bands of the EEG are related to the demands of the task. The interval preceding the presentation of a stimulus was associated with a gradual reduction in middle- and high-frequency alpha power – and the degree of desynchronization increased in the anticipation that a response would be required (i.e., prior to imperative and disconfirmation stimuli).

The presentation of a stimulus elicited a transient increase in power in the theta, low alpha, and gamma bands. The response in all three of these bands diminished during the middle- and late-task segments (however, in the gamma band this reduction was restricted to the burst following post-imperative stimuli). The period immediately following termination of the stimulus was characterized by an additional desynchronization in all three of the alpha bands. In general, though the details differed from band to band, this desynchronization was smallest for those stimuli farthest removed from a potential response.

Finally, the middle portion of the inter-stimulus interval was associated with reduced theta power and increased gamma power following imperative stimuli. While the post-response decision periods following the imperative (go) and disconfirmation (no-go) stimuli contained large middle- and high-alpha band hypersynchronizations.

Electrocardiographic (ECG) Metrics

Individual R-waves in the ECG were detected via an algorithm similar to that employed for the detection of blinks. Thus, a potential R-wave onset was defined as the point at which the slope of ten successive time points exceeded a pre-set criteria. R-wave detection was confirmed if a) a peak was detected within 100 msec of the onset point, b) the amplitude of the R-wave fell within a pre-defined range, and c) the duration of the R-wave fell within predefined parameters. The actual values of the slope, amplitude, and latency criteria were tailored for each subject to maximize detection likelihood. The performance of the automated scoring algorithm was monitored by an experimenter who was responsible for rejecting noise artifacts incorrectly identified as R-waves, as well as for manually inserting R-waves that the computer failed to detect.

Due to the 200 Hz sampling rate employed in this study, the time at which the peak of the R-wave occurred could only be estimated with an accuracy of ± 5 msec. To improve the temporal accuracy of this measure, an algorithm was developed that identified the points around the peak where the slope of the ECG signal transitioned from negative to positive and re-calculated the time of the peak by interpolation.

Data samples consisting of the time between consecutive R-waves, were interpolated to create equidistant time series in the manner prescribed by Mulder (1980). Following the removal of DC offset and linear trend, the data were smoothed using a modified cosine function, zero-padded, and submitted to a Fast Fourier Transform (FFT) analysis. This procedure yielded estimates of the power in spectral bands 0.01 Hz wide, ranging from 0 to 0.50 Hz. The power spectra from 3 consecutive, 5-min periods during the early, middle and late task periods were averaged together to produce stable estimates of ECG variability during the different task segments.

The ECG power spectrum averaged over task periods and subjects is presented in Figure 34. Four distinct peaks in the spectral density function are clearly visible. The physiological mechanisms underlying three of these peaks are relatively well known (see Aasman et al., 1987). Activity in the range from 0.02 to 0.06 Hz is related to vasomotor activity involved in the regulation of body temperature.
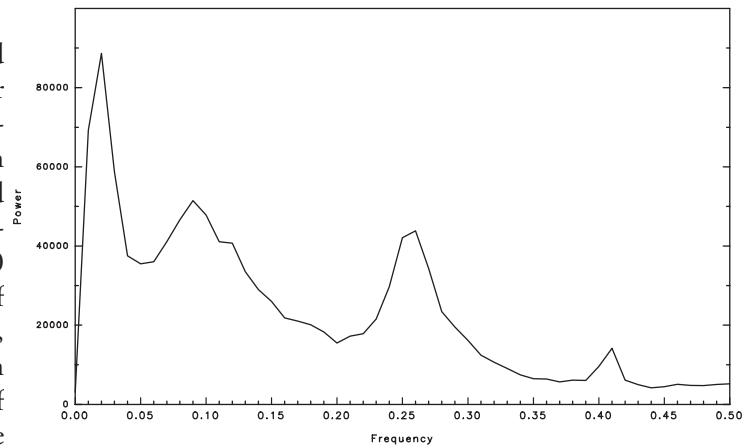

Figure 34. Spectral density (power) function for cardiac IBI variability averaged over subjects and task segments.

Energy in the range from 0.07 to 0.14 Hz (hereafter referred to as the 0.10 Hz component) is related to the short term regulation of arterial blood flow. Previous research has demonstrated the sensitivity of the 0.10 Hz component (which is thought to be primarily influenced by sympathetic autonomic activity) to a variety of variables including (but not limited to) manipulations of cognitive workload (Mulder, 1980; Van Dellen et al. 1985).

Frequencies above 0.14 Hz predominantly reflect the influence of respiration upon the distribution of cardiac inter-beat intervals. Two respiration-related peaks in the spectral density function can be identified. The first, which peaks at approximately 0.25 Hz, corresponds to the traditionally defined Respiratory Sinus Arrhythmia (hereafter termed Spectral RSA). In this study, Spectral RSA was quantified as the energy in the bands from 0.14 to 0.34 Hz. Previous research has suggested that the Spectral RSA component provides a measure of vagal tone in the Autonomic Nervous System (ANS) and can, therefore, be used as a sensitive index of operator state (Porges, 1984).

A smaller peak at approximately 0.40 Hz was found to result from a tendency on the part of some subjects to time-lock respiratory cycles to the rate of stimulus presentation (an ISI of 2511 msec = 0.40 Hz). This component, labelled Stimulus RSA, was computed as the total spectral power in the range from 0.39 to 0.42 Hz. The absolute power in the 0.10 Hz, Spectral and Stimulus RSA bands were submitted to a log transform to diminish the effect of individual differences present in the absolute power scores.

Several measures of cardiovascular function during the early-, middle-, and late-task segments are displayed in Table 2. Mean Interbeat-Interval (IBI) varied considerably across subjects (minimum = 670 msec, maximum = 1,113 msec) but did not vary across task segments (p>0.05). The 0.10 Hz component of IBI variability, on the other hand, was sensitive to TOT (*F(2,30)=10.35; p=0.0015*). Power in this band was greater during the middle- and late-task segments.

In contrast, the two RSA components did not significantly vary with TOT (p-values > 0.05). A 4th spectral measure, presumed to index the relative balance between the sympathetic and parasympathetic branches of the ANS was computed as the ratio of 0.10 Hz power to spectral RSA power (termed S/P ratio). This ratio was sensitive to TOT (*F(2,30)=5.44; p=0.0106*) and, like the 0.10 Hz component, significantly increased during the middle and late task segments.

A major drawback of the spectral technique outlined above is that the measures reflect the activity in 5-min segments. An alternative "peak to valley" technique for estimating RSA (see Grossman, van Beek, & Wientjes, 1990) allows quantification of this component on a minute-to-minute basis. To this end, the RSA2001 program (Uijtdehaage,

1994) was employed to compute Peak-Valley RSA. Across subjects, the Peak-Valley estimates were highly correlated with absolute Spectral RSA power (r=0.91).

Peak-Valley RSA scores for each of the three task segments are also presented in Table 2. As was the case for Spectral RSA, mean levels of Peak-Valley RSA were unaffected by TOT. However, because estimates of RSA and mean heart rate were available for each minute of the task, an analysis similar to that conducted on the performance data could be performed. Thus, for each subject the heart rate and RSA scores were z-transformed and the standard deviations of the z-scores within the early-, middle-, and late-task segments were obtained.

The results of these analyses (termed heart rate variability and RSA variability) are included in Table 2. Heart rate variability was significantly influenced by TOT (*F(2,30)=5.17; p=0.0139*). Increased variability was present during the late task segment in comparison with the early period of task performance. The intermediate middle segment values could not be distinguished from either the early or the late segment. RSA variability was also sensitive to TOT [*F(2,30)=4.72; p=0.0311*]. This measure was lower for both the early and middle segments than late in the task.

Table 2. Mean (and Standard Deviation) of Cardiovascular Metrics by Time Segment

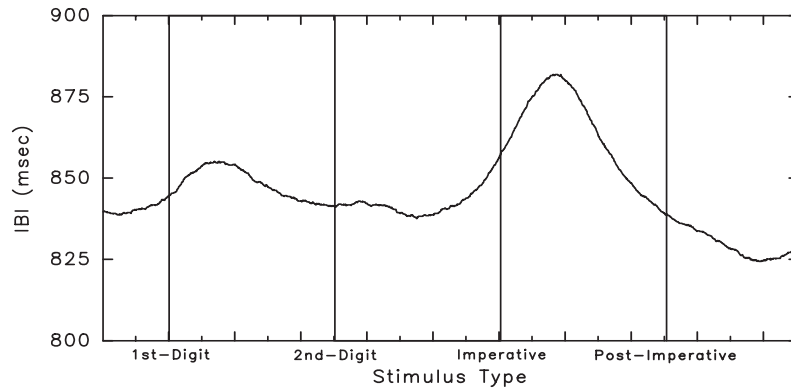| Measures	Task Segment		
	Early	Middle	Late
Heart Rate (beats/min)	71 (11.20)	71 (10.59)	71 (10.39)
0.10 Hz Log(Power)	9.93 (1.215)	10.21 (1.107)	10.49 (0.930)
Spectral RSA Log(Power)	9.40 (1.199)	9.45 (1.103)	9.52 (0.976)
Stimulus RSA Log (Power)	8.71 (1.110)	8.63 (0.860)	8.77 (0.822)
S/P Ratio	1.035 (0.1266)	1.077 (0.1403)	1.099 (0.1160)
Peak-Valley RSA	89.11 (52.20)	89.30 (60.43)	90.53 (47.36)
Heart Rate Variability	0.73 (0.144)	0.81 (0.211)	0.93 (0.169)
RSA Variability	0.85 (0.255)	0.80 (0.202)	1.05 (0.281)

Figure 35. Event-related cardiac IBI response to stimuli in an imperative response sequence.

Event-Related Cardiac Activity

Concurrent with the identification of R-waves, an output file was created containing the instantaneous inter-beat interval every 5 msec. This enabled cardiac responses preceding and following specific stimuli to be evaluated via signal-averaging techniques identical to that employed for ERPs. The average event-related cardiac response to imperative digit sequences (collapsed across subjects) is presented in Figure 35.

A large heart rate deceleration (reflected in an increase in the inter-beat interval) can be seen to precede the presentation of an imperative stimulus, continuing for approximately 1 sec thereafter. The magnitude of the cardiac deceleration associated with each stimulus type was quantified as the difference between the maximum IBI in the post-stimulus epoch between 0 and 1,500 msec and the mean IBI in a 500 msec period beginning 1000 msec prior to stimulus presentation.

Figure 36 displays mean IBI difference scores as a function of stimulus type and task period. While this variable was not directly affected by TOT [$F(2,30)=2.26; p=0.1438$] the absence of decelerative responses associated with 2nd-Digit and post-imperative stimuli generated a significant main effect of stimulus type [$F(4,60)=44.29; p<0.0001$]. The existence of a significant TOT by stimulus type interaction term [$F(8,120)=2.68; p=0.0367$] prompted separate examinations of TOT effects at the three stimuli for which cardiac decelerations were present.
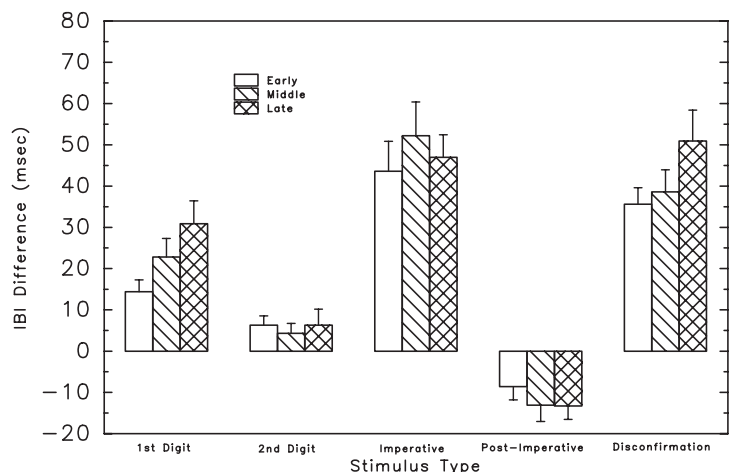
No TOT effects were present for either the imperative or the disconfirmation stimuli (p-values < 0.05). However, the decelerations associated with 1st-Digit stimuli did vary with TOT [$F(2,30)=5.19; p=0.0159$]. Greater decelerations in response to 1st-Digit stimuli occurred during the middle and late portions of the task.

Exploratory Modeling

Multiple regression techniques were employed to explore the extent to which performance variability both between and across subjects could be predicted by measures of oculographic, encephalographic, and cardiovascular function. Models containing up to 15 variables in each of these measurement domains were tested for "optimality" on the basis of the Mallow C(p) statistic. The terms that were subsequently found to contribute significantly to the model were retained.

The results of these analyses are displayed in Table 3. Oculomotor variables accounted for approximately 50% of the variance in both A' and RT. Event-related potentials and frequency band modulations were much better at predicting RT than A', while the reverse was true for the transient EEG measures. Cardiovascular measures in and of themselves, were poor predictors of both RT and A'.

A "Full" psychophysiological model was developed by entering the best parameters from each measurement domain found in the above analyses.

Figure 36. Cardiac decelerations (maximum IBI – baseline IBI) as a function of stimulus type and task segment.

Once again, only parameters significantly contributing to the optimal model were retained. Eleven parameters were identified as predictive of the A' index of performance accuracy. Seven of these parameters were indexes of oculomotor behavior (2nd-Digit and imperative stimulus blink amplitudes and durations, post-imperative stimulus reactive saccade amplitudes and velocities, and 2nd-Digit reactive saccade latencies). The remaining parameters were the amplitude of the late gamma response following imperative stimuli, generalized theta EEG transients, and two cardiovascular parameters related to RSA.

Eight parameters were contained in the optimal model for the prediction of RT – 1st Digit blink latency, P300 latencies following imperative and disconfirmation stimuli, two measures of event-related alpha modulation, the amplitude of the early gamma response following post-imperative stimuli, the S/P cardiovascular ratio, and spectral RSA.

Table 4 displays the actual and predicted A' and RT scores for two high-performance (Subjects 16 and 20) and two low-performance (Subjects 6 and 17) individuals at all three task segments. Note that the "Full" psychophysiological models do a reasonable job of predicting both between as well

Table 3. Performance Modeling

	A Prime		RT	
Measures	# Var.	R-Square	# Var.	R-Square
Oculomotor Activity	5	0.53	4	0.51
Event-Related Brain Activity	4	0.27	7	0.73
Transient Brain Activity	4	0.61	1	0.14
Cardiovascular Activity	3	0.19	4	0.30
"Full" Psychophysiological Model	11	0.90	8	0.70

Table 4. Model Predictions for Low vs. High Performance Subjects

Subject ID	Task Segment	Predicted A'	Actual A'	Predicted RT	Actual RT
16	Early	568	584	0.998	.994
16	Middle	517	523	1.004	1.000
16	Late	493	533	1.001	1.000
20	Early	472	444	1.003	0.999
20	Middle	468	484	0.999	1.000
20	Late	490	461	0.995	1.000
6	Early	710	759	0.992	0.970
6	Middle	799	872	0.926	0.934
6	Late	796	781	0.953	0.946
17	Early	739	826	0.934	0.932
17	Middle	924	877	0.864	0.854
17	Late	749	812	0.907	0.908

as within subject performance variability. Also, note that the predictions are somewhat better for performance accuracy than for RT.

CONCLUSIONS

The data presented here provide compelling evidence that psychophysiological metrics can be used to infer the relative level of alertness and/or attentiveness of human operators during the performance of complex information processing tasks. Figure 37 summarizes the responses across a variety of measurement domains to individual stimuli in the present task. Note that for the sake of simplicity, alpha activity in this figure has been collapsed across the low-, medium-, and high-frequency alpha bands, and that heart rate, rather than mean IBI, is displayed.

The period immediately preceding the presentation of a stimulus was characterized by, an inhibition of blinks, the production of anticipatory saccades and a suppression of alpha band power. Prior to imperative stimuli, a heart rate deceleration coupled with a substantial increase in CNV amplitude was also present. The post-stimulus period also generated a variety of responses. The immediate post-stimulus period was associated with the presence of reactive saccades as well as a continued suppression of blinking. This period also produced bursts of activity in the theta, low alpha, and gamma frequency bands.

Most responses, occurred approximately 500-600 msec post-stimulus. This period also marked the period of maximum likelihood for the production of a blink, as well as the elicitation of a large alpha hypersynchronization. Importantly, the latencies of blinks following imperative stimuli were delayed, and these stimuli also generated larger P300 and late gamma responses.

A number of interesting relationships between the activity in these response domains and performance emerged. While mean performance levels were fairly consistent across early, middle, and late segments of the task, increased TOT did result in increased performance variability. Similar TOT effects were obtained for many physiological measures, including: blink closure durations and latencies; reactive saccade frequencies and latencies; anticipatory and reactive saccade amplitudes and velocities; recheck
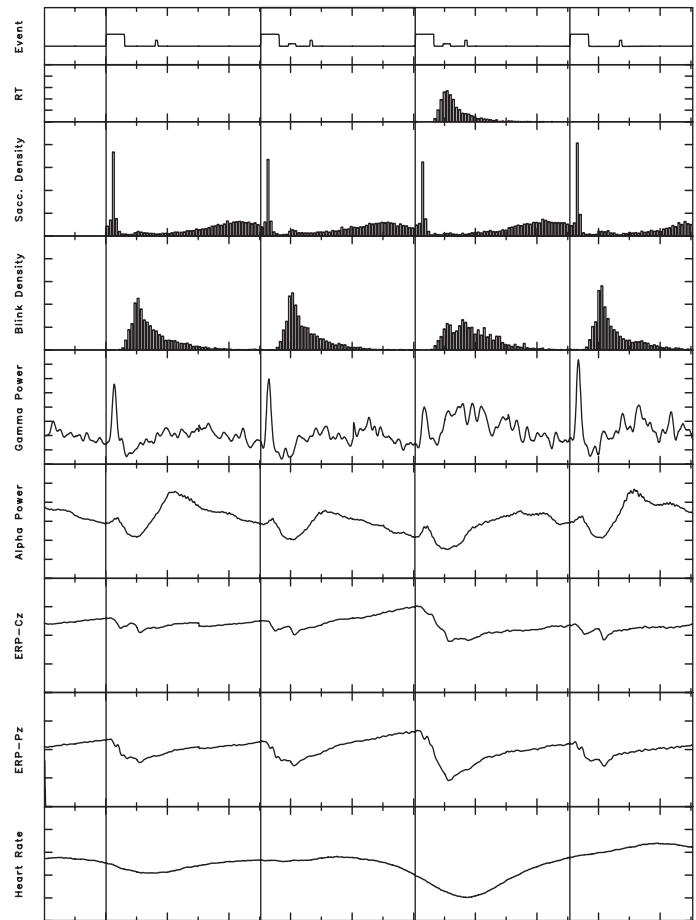

Figure 37. Summary of responses in a variety of systems to the stimuli in an imperative stimulus sequence. Note the increase response lability associated with an imperative stimulus.

saccade velocities; early theta, low-frequency alpha and gamma band hypersynchronizations; and power in the 0.10 Hz component of cardiac IBI variability.

Of greater interest are the many relationships between oculomotor measures and patterns of within-subject performance variability. In the absence of a blink, RTs were substantially delayed — consistent with the hypothesis of an increased period of blink inhibition associated with lengthier and less efficient decision-making processes on these trials. Furthermore, when a blink was present, if the blink latency was too short (and encroached upon intervals associated with stimulus identification and classification processes) errors increased and RTs became more variable. Particularly long blink latencies were associated with increased RTs. Finally, if a long closure duration blink followed the

presentation of an imperative stimulus, the target was more likely to be missed. Saccadic activity was also predictive of within-subject performance variability. Imperative stimuli followed by reactive saccades were responded to with decreased accuracy levels and increased RTs compared with stimuli that were effectively anticipated. Furthermore, the larger the reactive saccade, the greater the increase in RT. Slow-velocity recheck saccades were also associated with increased manual response latencies.

Finally, exploratory multi-variate analyses revealed a number of interesting relationships between overt performance and physiology. Eye-movements alone could account for over 50% of the variance in performance accuracy and response speed. Event-related potentials and modulations in specific frequency bands of the EEG were much better at predicting RT than accuracy. Interestingly, the reverse was true for the measures of more transient EEG events. Cardiovascular measures, as such, were poor predictors of both RT and A'. "Full" psychophysiological models, containing measures from all of the physiological systems, accounted for substantial variance from both within and between subjects, with respect to both the accuracy (R^2 = 0.90) and RT (R^2 = 0.70) measures.

In conclusion, the present study demonstrated a number of intriguing relationships between overt performance and covert psychophysiological variables. These findings will substantially contribute to attempts to develop oculometric, real-time alertness monitoring systems by providing a) an improved understanding of the relationship between the activity of the eyes and ongoing performance; and b) a richer, multi-variate space against which to validate the oculomotor metrics.

REFERENCES

Broughton, R. (1994). Important underemphasized aspects of sleep onset. In R. D. Ogilvie & J. R. Harsh (Eds.), *Sleep Onset: Normal and Abnormal Processes*. Washington, DC: American Psychological Association.

Aasman, J., Mulder, G., & Mulder L. (1987). Operator effort and the measures of heart rate variability. *Human Factors*, *29*, 161-70.

Bills, A.G. (1931). Blocking: A new principle of mental fatigue. *American Journal of Psychology*, *43*, 230-45.

Calderia, J.D. (1980). Parametric assumptions of some "nonparametric" measures of sensory efficiency. *Human Factors*, *22*, 119-30.

Grossman, P., van Beek, J., & Wientjes, C. (1990). A comparison of three quantification methods for estimation of respiratory sinus arrythmia. *Psychophysiology*, *27*, 702-14.

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W.C. (1973). Quantification of sleepiness: A new approach, *Psychophysiology*, *10(4)*, 431-6.

Johnson, R. Jr. (1986). A triarchic model of P300 amplitude. *Psychophysiology*, *23(4)*, 367-84.

Keppel, G. (1991) *Design and Analysis: A Researcher's Handbook* (p. 169-70). New Jersey:Prentice Hall.

Kramer, A., Sirevaag, E., & Braune, R. (1987). A psychophysiological assessment of operator workload during simulated flight missions. *Human Factors*, *29(2)*, 145-60.

Mulder, G. (1980). *The Heart of Mental Effort*. Groningen, The Netherlands:University of Groningen Press.

Oster, Ph. J., & Stern, J.A. (1980). Measurement of eye movements. Electrooculography. In J. Martin and P. H. Venables (Eds.), *Techniques in psychophysiology*. New York:Wiley.

Pfurtscheller, G. (1989). Functional topography during sensorimotor activation studied with event-related desynchronization mapping. *Journal of Clinical Neurophysiology*, *6*, 75-84.

Rohrbaugh, J.W. & Gaillard, A.W.K. (1983). Sensory and motor aspects of the contingent negative variation. In A.W.K. Gaillard & W. Ritter (Eds.), *Tutorials in Event Related Potential Research: Endogenous Components*. Amsterdam:North Holland.

Santamaria, J. & Chiappa K.H. (1987). The EEG of drowsiness in normal adults. *Journal of Clinical Neurophysiology*, *4(4)*, 327-82.

Sirevaag, E., Kramer, A., Coles, M.G.H., & Donchin, E. (1984). Resource reciprocity: An event-related brain potentials analysis. *Acta Psychologica*, *70*, 77-97.

Sirevaag, E., Kramer, A.F., Wickens, C.D., Reisweber, M., Strayer, D., & Grenell, J. (1993). Assessment of pilot performance and mental workload in rotary wing aircraft. *Ergonomics*, *36(9)*, 1121-40.

Uijtdehaage, S. (1994). RSA2001 program published in *The International Journal of Bio-Medical Computing*, *35*, 169-92.

Van Dellen, H., Aasman, J., Mulder, L. & Mulder, G. (1985). Time domain versus frequency domain measures of heart rate variability. In J. Orlebeke, G. Mulder & L. VanDooren (Eds.), *Psychophysiology of Cardiovascular Control: Models, Methods and Data*. New York:Plenum Press.

Wickelgren, W. A. (19xx). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, *41*, 67-85.

Wickens, C.D. (1984). *Engineering Psychology and Human Performance. Columbus:Merrill.*